EVOLUTIONARY AND ECOLOGICAL PROCESSES IN CONSERVATION AND PRESERVATION OF PLANT ADAPTIVE POTENTIAL

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Lionel Nicolas Di Santo

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Environmental and Conservation Sciences

December 2021

Fargo, North Dakota

North Dakota State University
Graduate School

Title

EVOLUTIONARY AND ECOLOGICAL PROCESSES IN CONSERVATION AND PRESERVATION OF PLANT ADAPTIVE POTENTIAL

By

Lionel Nicolas Di Santo

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Jill Hamilton
Chair

Dr. Craig Stockwell

Dr. Changhui Yan

Dr. Jessica Wright

Dr. Phillip McClean

Approved:

01/13/2022
Date

Dr. Craig Stockwell
Department Chair

ABSTRACT

Anthropogenetic disturbances, such as habitat loss and fragmentation, overexploitation, and climate change have diminished population sizes of many species, increasing risks of population extirpation or species extinction. Consequently, conservation of genetic variability, to preserve and maintain rare species' evolutionary potential and avoid within-population inbreeding, is a major goal of conservation biology. For plants, various approaches and guidelines have been developed to preserve species' genetic diversity *ex situ* ("off-site"). However, effective methods to guide conservation and management decisions without relying on the availability of genetic data or knowledge about population size and population genetic structure are lacking. With the first two chapters of my dissertation, I aimed to complement existing *ex situ* strategies by investigating surrogates for estimating genetic variation to optimize conservation of rare species' evolutionary potential when access to genetic data is limited. My results demonstrated that guiding population sampling using environmental and geographic distances, as opposed to randomly selecting source populations, can increase genetic diversity and differentiation captured in simulated *ex situ* collections. Likewise, my research showed that for species with largely heritable seed traits, morphological variation estimated from contemporary seed collections can be used as a proxy for standing genetic variation and help inform sampling efforts aiming to optimize genetic diversity preserved *ex situ.* Although strategies targeted to conserve rare species' evolutionary potential where genetic data may be lacking are needed, the increasing affordability of next-generation sequencing technologies is increasing access to genomic data for rare species. With my third chapter, I investigated whether inferring rare species' evolutionary history from genomic data may help inform conservation practices. My results demonstrated that teasing apart spatial and temporal effects of stochastic

and deterministic processes on population genetic structure may be used to estimate past and contemporary changes in populations' evolutionary potential, as well as to evaluate risks and benefits of genetic rescue as a management strategy. Overall, my PhD research establishes tools and approaches to preserve genetic variation for rare species using different types of data. As world's biodiversity continue to decline, tool development to accommodate species-specific data availability for preservation of genetic variation is crucial.

ACKNOWLEDGMENTS

First and foremost, I would like to express my appreciation to my academic advisor, Dr. Jill Hamilton, for her continuous support and guidance that allowed me to grow as a research scientist. Both your commitment and advice have been invaluable to my research and career. I also would like to thank my committee members, Dr. Phillip McClean, Dr. Changhui Yan, Dr. Craig Stockwell, and Dr. Jessica Wright for seeing the potential in me and accompanying me on this journey. Special thanks to Dr. Craig Stockwell for his mentorship and precious career advice. I also thank past and current members of the Hamilton lab, Zebadiah Yoko, Kate Volk, Jessica Lindstrom, Dr. Pamela Puppo, and Dr. Joseph Braasch who have supported me with insights and friendship. I also would like to express my gratitude to Dr. Sean Hoban for our insightful discussions, and Dr. Thomas Parchman for his invaluable laboratory teachings. Finally, I thank the Department of Biological Sciences at NDSU, NDSU Environmental and Conservation Sciences Program, NDSU College of Science and Mathematics, NDSU Graduate School, and NDSU International Student and Study Abroad Service for support, and acknowledge The Morton Arboretum, ND-EPSCoR (NSF-IIA-1355466), NDSU Environmental and Conservation Sciences Program, the USDA-Forest Service (State and Private Forestry, Gene Conservation Program), the USDA Health Protection Gene Conservation Program, the Western Wildlands Environmental Threat Assessment Center, and the Schatz Center for Tree Molecular Genetics for funding the research presented.

DEDICATION

I would like to dedicate this PhD to my family, if not by birth by heart, and especially to my mother. Your unwavering support, guidance, patience, and love mean the world to me. Completing this work is as much my achievement as it is yours.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance.
CM1	Colonization demographic model 1 (island colonization without subsequent gene flow between populations).
CM2	Colonization demographic model 2 (island colonization with subsequent gene flow between populations).
CM3	Colonization demographic model 3 (mainland colonization without subsequent gene flow between populations).
CM4	Colonization demographic model 4 (mainland colonization with subsequent gene flow between populations).
dbRDA	Distance-based redundancy analysis.
EA	Embryo area (cm^2).
EL	Embryo length (cm).
ELW	Ration between embryo length (cm) and embryo width (cm).
ESA	Endosperm area (cm^2).
EST-SSRs	Expressed sequence tag single-sequence repeat.
EW	Embryo width (cm).
Gen-SNPs	Single-nucleotide polymorphisms within genes.
IBD	[Proportion of genetic differentiation explained by] isolation-by-distance.
IBE	[Proportion of genetic differentiation explained by] isolation-by-environment.
IBD∩IBE	Proportion of genetic differentiation explained by the shared variation between environmental and geographic factors.

IBDUIBE	Proportion of genetic differentiation explained by the combination of geographic and environmental factors.
IUCN	International union for conservation of nature.
maf	Minor allele frequency.
NMDS	Non-metric multidimensional scaling analysis.
PCA	Principal component analysis.
REndS	Relative endosperm size.
RES	Relative embryo size.
RPM1	Remnant populations demographic model 1 (without gene flow between populations).
RPM2	Remnant populations demographic model 2 (with gene flow between populations).
RSCS	Relative seed coat size.
SA	Seed area (cm^2).
SCA	Seed coat area (cm^2).
SCW	Seed coat width (cm).
SEL-SNPs	Single-nucleotide polymorphisms putatively under natural selection.
SL	Seed length (cm).
SLW	Ratio between seed length (cm) and seed width (cm).
SNPs	Single-nucleotide polymorphisms.
SRI	Santa Rosa Island, CA.
SSRs	Single-sequence repeats.
SW	Seed width (cm).
TPSR	Torrey Pine State Reserve, CA.

LIST OF SYMBOLS

θ (theta)	A measurement of the proportion of shared alleles expected between pairs of individuals within a population that are identical by descent, also known as the coancestry or relatedness coefficient.
A_c/A_d	Ratio between the number of alleles captured in simulated *ex situ* collections when *Np* populations are sampled, and the number of alleles present within a dataset (see Figure 1).
A_s/A_p	Ratio between the number of alleles captured when *N* individuals within a population are sampled, and the total number of alleles within the population (see Figure 2).
F_{IS}	A measurement of inbreeding (reproduction among relatives including oneself) within populations, ranging from 0 (panmixia) to 1 (selfing only).
F_{ST}	A measurement of genetic differentiation between populations, ranging from 0 (no differentiation) to 1 (complete differentiation).
H_E	Expected heterozygosity under Hardy-Weinberg equilibrium.
H_O	Observed heterozygosity estimated from empirical genetic/genomic data.
m_{IM}	Migration probability from island population to mainland population.
m_{MI}	Migration probability from mainland population to island population.
$N_{80\%}$	Number of individuals required to capture between 80% - 100% of allelic diversity within a population (see Figure 2).
N_A	Ancestral effective population size.
N_C	Initial effective population size following migration (effective number of migrants).

N_c/N_t	Ratio between the number of unique seed trait values captured in a sample of N_{fam} maternal families, and the total number of unique seed trait values within a seed population (see Figure C5).
Ne	Generic notation for effective population size.
Ne/N	Ratio between effective population size (Ne) and census population size (N), measured as the number of reproductively mature individuals within a population.
N_I	Island effective population size.
N_M	Mainland effective population size.
R^2_c	Proportion of variance explained by fixed and random effects in a linear mixed model.
R^2_m	Proportion of variance explained by fixed effects in a linear mixed model.
r_I	Island (exponential) growth rate.
r_M	Mainland (exponential) growth rate.
T_{Div}	Time of population (island, mainland) divergence.

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

1. GENERAL INTRODUCTION TO CONSERVATION GENETICS

1.1. Consequences of demographic rarity: the extinction vortex

Human-mediated disturbances, including habitat loss and fragmentation (Fahrig 1997), overexploitation (Rosser and Mainka 2002), introduction of alien species (Coblentz 1990), and climate change (Thomas et al. 2004; Keith et al. 2008) contribute to the decline of world's biodiversity. Combined, these factors can have a substantial impact on species demography, increasing populations' susceptibility to stochastic processes. Environmental stochasticity (small and moderate environmental perturbations; Lande 1988; 1993), random catastrophic events (sudden and extreme environmental perturbations; Lande 1993), and demographic stochasticity (variance in individuals' vital rates; Lande 1988; 1993) may considerably reduce species' survival and long-term persistence. Indeed, Lande (1993) showed that under each stochastic scenario, time to extinction increases with increasing population size. Consequently, while impacts of stochastic factors may be negligible in large, widespread populations, they can be devastating in small, isolated populations.

In addition to increased population susceptibility to environmental, demographic, and catastrophic stochasticity, demographic rarity may also exacerbate the loss of genetic diversity within populations due to genetic stochasticity (Ellstrand and Elam 1993). Genetic stochasticity is defined as the stochastic change in allele frequencies within populations due to genetic drift (random sampling of alleles during sexual reproduction or because of environmental and demographic stochasticity). In small, isolated populations, effects of genetic drift and inbreeding (reproduction among genetically related individuals) on genetic variability are often enhanced, resulting in increased allelic fixation and probability of homozygosity. When homozygosity and fixation is increased for deleterious alleles, fitness of individuals within these populations may be

reduced, a phenomenon known as inbreeding depression (Crnokrak and Roff 1999; L. F. Keller and Waller 2002). Ultimately, the combined action of environmental, demographic, and genetic stochasticity could trap rare species in a feedback loop called 'the extinction vortex' (Gilpin and Soulé 1986). For species entering the extinction vortex, environmental and demographic stochasticity reduce the size of populations, exacerbating the effect of both genetic drift and inbreeding. The subsequent loss of genetic variation decreases individuals' fitness (via inbreeding depression) and adaptability, reducing population sizes even more. In the absence of conservation actions to mitigate this loop, populations are likely to keep shrinking until they become extinct (Frankham, Briscoe, and Ballou 2002; Fagan and Holmes 2006).

1.2. The relationship between population size and genetic diversity

Genetic diversity is the raw material for natural selection to act upon, enabling adaptation (Carlson, Cunningham, and Westley 2014; R. D. H. Barrett and Schluter 2008; Gomulkiewicz and Holt 1995; Bell 2013). Genetic diversity is therefore crucial for species long-term persistence under rapidly changing environments. Characterized by small populations sizes (Frankham 1999), rare species may harbor reduced genetic variation relative to their widespread counterparts (Frankham 1995) and could thus be more susceptible to the extinction vortex (Gilpin and Soulé 1986; Fagan and Holmes 2006). In plants, multiple studies have evaluated the relationship between population size and genetic diversity, which often exhibits a positive relationship (reviewed by Leimu et al. 2006; Ilves et al. 2013). The strength and direction of this relationship was identified by Leimu et al. (2006) as depending on certain species features, including species' mating system and, potentially, species rarity. However, these studies have mostly looked at short-lived, insect pollinated plant species, limiting our current understanding of the relationship in long-lived, wind pollinated tree species (but see Tamaki et al. 2018;

Chybicki, Oleksa, and Kowalkowska 2012; Del Castillo et al. 2011). Features associated with such life history traits could mitigate the loss of genetic diversity within rare species. For instance, wind pollination may allow long distance pollen dispersal within and between remnant populations, allowing them to remain genetically connected and alleviating the loss of genetic diversity (Lowe et al. 2005), while a long lifespan may increase the time necessary for the loss of genetic diversity to become perceptible. Therefore, genetic diversity loss associated with reductions in population size may be less predominant in long-lived, wind-pollinated tree species (Victory et al. 2006).

1.3. How can plant genetic diversity be conserved?

For a range of species, plant populations can be conserved in their native ecosystem (termed *in situ* conservation) or samples representing plant populations' genetic diversity can be conserved outside the species native habitat (termed *ex situ* conservation) (Cohen et al. 1991; Potter et al. 2017; Schoen and Brown 2001; D.-Z. Li and Pritchard 2009; Volis and Blecher 2010; Ledig 1988).

1.3.1. *In situ* conservation

Species conserved *in situ* are often preserved on protected lands, including national parks, state parks, research natural areas, wildlife refuges, or private natural areas such as those owned by The Nature Conservancy (Ledig 1988). Conservation of rare species within their native habitat offers the advantage of preserving both natural standing genetic variation and evolutionary processes in response to local abiotic and biotic conditions (Potter et al. 2017; Ledig, Vargas-Hernández, and Johnsen 1998). *In situ* conservation efforts have focused on species of conservation concern. US national parks currently protect multiple plants listed as "endangered" under the Federal Endangered Species Act with, for instance, 8 and 12 endangered

plants preserved at Point Reyes National Seashore, CA and Channel Islands National Park, CA, respectively (https://irma.nps.gov/NPSpecies/Search/SpeciesList). In addition to plants of conservation concern, *in situ* conservation actions have also been undertaken worldwide for economically important species, including crop wild relatives valuable for the improvement of major crops (Harlan 1976; Meilleur and Hodgkin 2004). For example, relatives of wild cereals, wheat, rice, barley, lentil, fruit plants (e.g. apples, peaches, apricots, bananas, mangos), or seed plants (e.g. pistachios, almonds, nuts) are targeted for *in situ* conservation (Meilleur and Hodgkin 2004).

1.3.2. *Ex situ* conservation

Although *in situ* conservation offers many benefits such as the conservation of evolutionary processes, genetic diversity and interactions between groups of species within a protected area (Potter et al. 2017; Ledig 1988; M. B. Hamilton 1994), it also has an important limitation: if natural genetic diversity is lost in the wild, it is lost permanently. Preserving the germplasm of a species outside its native range of occurrence via *ex situ* conservation ensures that at least part of the species' genetic diversity is conserved. Seed, pollen, tissue or whole plants of a targeted species may be stored *ex situ* either in seed banks, clone banks, botanical gardens, arboreta, or through tissue culture (Volis and Blecher 2010; Potter et al. 2017; Ledig 1986; Brown and Briggs 1991; Ledig 1988). Conservation of a species germplasm *ex situ* thus provides insurance against the loss of genetic diversity in nature. Extensive efforts have been undertaken to conserve crops (Westengen, Jeppson, and Guarino 2013), crops wild relatives (Volk et al. 2005; Naredo et al. 2017) and threatened (Q. Li, Xu, and He 2002; Griffith, Lewis, and Francisco-Ortega 2011) species *ex situ*. For example, several millions of seed accessions are stored in 1,750 gene banks worldwide, and approximately 31,199 seed accessions for over 4,800

wild species, mostly representing plant diversity from southwest China, are preserved at the Royal Botanic Gardens Kew and the Chinese Academy of Sciences' Kunming Institute of Botany (Y.-B. Fu, Ahmed, and Diederichsen 2015; D.-Z. Li and Pritchard 2009).

1.4. Optimizing genetic diversity captured *ex situ*

If *in situ* and *ex situ* conservation are conceptually different, they share a common objective: the conservation of plant genetic diversity (Ledig 1988; Griffith et al. 2014; Guerrant Jr, Havens, and Vitt 2014; Hausman et al. 2014; Potter et al. 2017; Ledig, Vargas-Hernández, and Johnsen 1998; Volis and Blecher 2010). Consequently, collection strategies to optimize sampling effort and maximize plant genetic diversity captured *ex situ* are needed. Such guidelines will aid in limiting either over- or under-sampling targeted species. Over-sampling may result in a waste of limited resources (e.g. time, money), whereas under-sampling may result in collections capturing a sub-optimal proportion of a species' genetic diversity (Hoban and Schlarbaum 2014). Over the years, several sampling methods have been proposed to guide threatened species' collection efforts. Commonly used methods include probability-based methods (Brown and Marshall 1995; Lawrence, Marshall, and Davies 1995; Marshall and Brown 1975; Yonezawa 1985) and the resampling method (Caujapé-Castells and Pedrola-Monfort 2004; Gapare, Yanchuk, and Aitken 2008; Namoff et al. 2010). More recently, simulation-based approaches have also been developed to inform *ex situ* collections (Hoban and Schlarbaum 2014; Hoban 2019).

1.4.1. Probability-based strategy

Probability methods rely on probabilistic equations to calculate the minimum number of individuals required to ensure, with high probability, that at least one copy of alleles occurring in a targeted population at specific frequencies is captured. For example, Brown and Marshall

(1995) developed a model to calculate the number of individuals needed to capture 95% of alleles present in a population at frequency greater than 0.05. Using the equation $S \approx -3/\log_e[1-p]$ (1), which calculates the number of random unrelated gametes *S* required to capture, with 95% probability, at least one copy of alleles occurring in a population at a frequency *p*. Based on equation (1), they recommended, as a benchmark criterion, to sample 50 individuals per population and 50 populations per ecoregion. Other probability-based models, in addition to Brown and Marshall's, have been developed (Lawrence, Marshall, and Davies 1995). While they share the same objective, capturing at least one copy of alleles occurring at specific frequencies with high probability, they differ in the model used to calculate allelic capture probability (Lockwood, Richards, and Volk 2007; Hoban and Schlarbaum 2014). For instance, Brown and Marshall (1995) treat populations as independent sampling units (Lockwood, Richards, and Volk 2007), and therefore assume no inter-population gene flow. Contrastingly, Lawrence, Marshall, and Davies (1995) assume all populations are similar. Biologically, this implies populations are genetically similar, and that there is substantial gene flow among populations. Based on their assumptions, Lawrence, Marshall, and Davies (1995) recommended to sample approximately 170 individuals in total, spread over all populations. As they do not require prior information on the ecology, reproductive biology or genetic structure of the targeted species (Lockwood, Richards, and Volk 2007), probability-based methods are easy and fast to implement. This ease and the wide applicability across taxa of these models have pushed governmental agencies, such as the US Forest Service and the Bureau of Land Management, to base their sampling strategies on probabilistic models (Rogers and Montalvo 2004; BLM 2012; Engels et al. 2008).

Nonetheless, probability-based methods make non-generalizable assumptions about the distribution of genetic diversity among populations. For instance, Brown and Marshall (1995) assume no overlap of genetic diversity among populations, while Lawrence, Marshall, and Davies (1995) assume complete overlap of genetic diversity among populations. However, population genetic differentiation results from the interplay between life history and demographic factors (Loveless and Hamrick 1984; Hamrick and Godt 1996; Heuertz et al. 2003), that is most likely to produce a continuum of genetic divergence across populations. Limitations of probabilistic models may thus include overestimation of the amount of total genetic diversity captured in a collection using Lawrence, Marshall, and Davies (1995)'s guidelines, whereas recommendations based on Brown and Marshall (1995)'s model may advise for more sampling effort than may be necessary to capture the desired amount of genetic diversity (Hoban and Schlarbaum 2014).

1.4.2. Resampling-based strategy

Formally, the resampling method uses either informed or stochastic resampling of genetic datasets to estimate genetic diversity that is, or would be captured in an *ex situ* collection (Caujapé-Castells and Pedrola-Monfort 2004; Gapare, Yanchuk, and Aitken 2008; Namoff et al. 2010). As the resampling strategy leverages genetic data, both within- and between-population genetic structure can be evaluated to optimize sampling efforts. For example, using a previously published genetic dataset on core and peripheral populations of Sitka spruce (*Picea stichensis*), Gapare, Yanchuk, and Aitken (2008) designed a sampling strategy aiming at capturing at least 95% of the species' genetic diversity. They found that sampling 150 trees over an area of 144 ha (core populations) and 180 trees over an area of 324 ha (peripheral populations) would be enough to meet their objective. For peripheral populations, sampling of more individuals over an

extended area was necessary to capture equivalent amounts of genetic diversity, as shifts in the mating system, resulting in increased inbreeding, contributed to variance in within-population genetic structure. An important caveat associated with the resampling method is that genetic datasets used must be representative of genetic variation found within natural populations. If not, some alleles (especially rare ones) present within populations may be absent from the dataset. In such a case, expectations of genetic diversity captured in collections would likely be overestimated, resulting in suboptimal sampling recommendations (Hoban and Schlarbaum 2014). Biases in genetic datasets may occur, for example, when individuals collected to generate a dataset are spatially or ecologically biased (Bamberg et al. 2010).

1.4.3. Simulation-based strategy

Hoban and Schlarbaum (2014) developed a simulation-based approach capable of estimating genetic diversity captured in *ex situ* collections when estimates of genetic connectivity (i.e., migration rates) and population sizes are available for modelling. First, genetic variation within and among populations is simulated using SimCoal2 (Laval and Excoffier 2004). Then, genetic diversity captured in a potential collection is calculated by stochastically resampling different numbers of individuals within populations using three distinct sampling strategies: (1) All populations within a region [constrained sampling strategy], (2) one population per region [dispersed sampling strategy], and (3) all populations in all regions [complete sampling strategy]. In summary, Hoban and Schlarbaum's model simulates the distribution of genetic variation within and among populations using species-specific estimates of genetic connectivity and population size, and leverages that information to estimate the number of individuals and populations needed to optimize sampling efforts. When applied to an endangered North American tree, butternut (*Juglans cinerea*), the model estimated that sampling 25 individuals per

population, and one population per region (dispersed sampling strategy) should suffice to optimize genetic diversity preserved.

More recently, another study by Hoban (2019) not only exploited genetic simulations to include realistic population structure in *ex situ* conservation planning using genetic connectivity and population size estimates, but also considered a fundamental aspect of conservation collections maintenance: the loss of plant material. Plant material may be lost via multiple processes following collection, including loss of viability in storage over time and active use for research (Walters, Wheeler, and Grotenhuis 2005; Way 2003). Consequently, population sampling should aim to capture multiple copies of each allele in collections, which would eventually lead to recommending additional sampling efforts. Using the coalescent simulator fastsimcoal, they generated genetic diversity for hypothetical species with various migration rates, population sizes, number of populations, and bottleneck histories. Then, given observed patterns, they determined the number of individuals to sample within each population to capture 95% of alleles occurring at specific frequencies (i.e., all alleles [no frequency threshold], low frequency alleles [0.01-0.10], alleles occurring locally at frequency > 0.05, alleles occurring species-wide at frequency > 0.05). Sampling guidelines are provided for various categories of population sizes (many populations [10-20], few populations [2-7]), targeted allele frequencies (low frequency [0.01-0.10], locally > 0.05, range-wide > 0.05), and desired number of copies for each allele (1, 5, 25). Overall, this study provides "rules of thumb" recommendations without *a priori* data requirements, although general information on population size, genetic connectivity, and demographic history may be used to refine recommendations when available. Alternatively, if estimates for these parameters are known, the approach described by Hoban (2019) may be used to simulate targeted population systems and tailor species-specific recommendations.

1.5. Components of phenotypic variance: implication for plant *ex situ* conservation

Multiple factors underlie trait determination in plants, including environmental conditions experienced by the plant, its genetic background, as well as the interaction between environment and genetic background. Variation in phenotypic traits (V_P) can therefore be split into three different variance components: an environmental component (V_E), a genetic component (V_G) and an interaction component (V_{GxE}). The genetic component (V_G) represents the heritable proportion of phenotypic variance, or the fraction of the variance passed on to the next generation. Consequently, it denotes the raw material upon which natural selection can act to promote plant adaptive evolution (Santiso et al. 2015; Zeng et al. 2017) and may be targeted for conservation of species' evolutionary potential.

Traditionally, heritable genetic variation (V_G) and heritability of quantitative traits (V_G/V_P) have been assessed using common garden experiments (J. A. Hamilton et al. 2017; Toker 2004; Çamaş and Esendal 2006). Common garden experiments establish individuals sampled from different populations in a shared common environment to control for phenotypic variation associated with the environment of origin and, to a lesser extent, genotype by environment interactions so that trait differences observed underlie genetic variability (de Villemereuil et al. 2016). In conservation, such an approach has commonly been used to study local adaptation and fitness consequences of population admixture, as well as to approximate the distribution of heritable genetic variation within and across species' populations (J. A. Hamilton et al. 2017; McKay et al. 2001; Goto et al. 2011; M. Keller, Kollmann, and Edwards 2000; Montalvo and Ellstrand 2001; 2000). Trait variation estimated from common garden-grown individuals may thus provide a valuable means to determine where and how to conserve genetic variation to optimize preservation of rare species' evolutionary potential *ex situ*.

Nonetheless, establishment of common garden experiments is logistically and financially demanding, and may not be well suited to long-lived, slow growing plant species if short-term conservation actions are needed. Where implementation of common garden experiments is not feasible, an alternative approach leveraging natural phenotypic variance in traits previously recognized as largely heritable may still provide a valuable proxy for V_G, and help inform *ex situ* conservation decisions.

1.6. *Ex situ* conservation strategies: synthesis and limitations

To optimize genetic diversity captured in *ex situ* collections, multiple population sampling strategies have been proposed. Each of these strategies require the availability of select types of data or parameters and rely on different sets of assumptions. Determining which approach to use to conserve threatened species *ex situ* necessitate the consideration of several factors. When genetic data or funding to generate genetic data are available, the resampling approach can be used to evaluate the optimal sampling strategy without making any assumptions on population structure (Caujapé-Castells and Pedrola-Monfort 2004; Gapare, Yanchuk, and Aitken 2008; Namoff et al. 2010). When genetic data is unavailable or limited, but connectivity and population size parameters exist or can be estimated, a simulation-based approach would allow simulation and inclusion of population structure into sampling strategies (Hoban and Schlarbaum 2014; Hoban 2019). When no genetic nor connectivity parameters are available, general guidelines established using probability- or simulation-based models provide a means to inform population sampling practices (Lockwood, Richards, and Volk 2007; Hoban 2019). However, these methods come with serious limitations. Probability-based models rely on non-generalizable assumptions about population structure, possibly resulting in under or over-sampling of the targeted species (Hoban and Schlarbaum 2014). Collections prepared in such

conditions may thus either fail to capture the desired level of genetic diversity or waste valued resources. Simulation-based guidelines, although improved relative to those established using probabilistic approaches to include realistic population structure and *ex situ* collection attrition, may only be tailored to specific species if general information on population sizes, migration rates, or demographic history is available. Consequently, risks of under or over-sampling the target species may still exist. Sampling strategies to guide conservation efforts without relying on the availability of genetic data or parameters, for example those that leverage natural variation in heritable quantitative traits provide a potentially valuable metric that remains to be assessed.

1.7. What if conservation of genetic diversity is not enough?

To help species avoid the extinction vortex, conservationists have focused on the conservation and restoration of rare species standing genetic diversity to both preserve their adaptive potential and prevent inbreeding depression. A commonly used approach has been to conserve wild species through habitat protection, habitat restoration, or habitat expansion (Stowell, Pinzone, and Martin 2017). The rationale behind protected areas is providing species with suitable conditions for population size to increase and fitness recovery. Yet, simple protection may sometimes be insufficient. Low standing genetic diversity and low mutation rate may prevent population adaptation and subsequent recovery, a process called 'evolutionary rescue' (Carlson, Cunningham, and Westley 2014).

Where evolutionary rescue seems improbable (i.e., low genetic diversity), actions to augment the size of a population, standing genetic diversity within populations, or both should be considered. A possible action would be to increase the size of the population with genetically similar individuals (demographic rescue). The addition of individuals to populations may help buffer against stochastic processes (e.g. demographic stochasticity) and reduce the Allee effect,

which may provide species with additional time to adapt to environmental changes and increase their fitness (Hufbauer et al. 2015). Alternatively, gene flow among differentiated populations could be artificially increased to promote admixture and increase genetic diversity (genetic rescue). To date, genetic rescue has had multiple designations, including facilitated migration, intentional hybridization or introgression, or admixture rescue (Stowell, Pinzone, and Martin 2017). The main difference between demographic and genetic rescue is that genetic rescue directly contributes to fitness and evolutionary advantages (Whiteley et al. 2015; Hufbauer et al. 2015). Translocating genetically distinct individuals increases genetic variability within rescued populations and may both alleviate inbreeding depression and increase their adaptive potential.

1.7.1. Genetic rescue: a debated conservation strategy

Although genetic rescue may be used to mitigate fitness and evolutionary consequences of severely diminished within-population genetic variation (W. E. Johnson et al. 2010; Westemeier et al. 1998; Madsen et al. 1999; Bossuyt 2007; Willi et al. 2007; Willi and Fischer 2005; Hufbauer et al. 2015), its use remains debated. Current reluctance towards using genetic rescue as a management strategy includes both cultural and biological barriers (Stowell, Pinzone, and Martin 2017; Ralls et al. 2018). Cultural barriers to genetic rescue include the concept of species “integrity” and “purity”. That is, the fear of disrupting species genetic “identity” following admixture. Stowell, Pinzone, and Martin (2017) and Ralls et al. (2017) both identified this fear as stemming from the fact that species are often seen as fixed objects instead of continuously evolving organisms.

The main biological barrier to genetic rescue is the possibility of outbreeding depression. Outbreeding depression is defined as the reduction in mean population fitness following admixture between genetically divergent lineages (Hufford and Mazer 2003). Two mechanisms

may explain this observable loss of fitness in admixed individuals: Dilution of adapted genotypes (also called underdominance) (Montalvo and Ellstrand 2001; Waser and Price 1994) or hybrid breakdown (loss of fitness resulting from the disruption of epistatic interactions following recombination) (M. Keller, Kollmann, and Edwards 2000). Outbreeding depression may pose a serious threat to genetic rescue attempts, and risks associated with inter-population gene flow should be examined. Current guidelines aiming to inform best practices for genetic rescue emphasize the importance of evaluating evolutionary mechanisms underlying populations genetic differentiation (Ralls et al. 2018; Frankham et al. 2011). Genetic rescue applied to endangered populations without this knowledge could be as much beneficial (increasing genetic variability of targeted populations and alleviating inbreeding depression) as deleterious (outbreeding depression). Consequently, understanding evolutionary processes driving population genetic differentiation is crucial when considering genetic rescue as a potential conservation strategy.

1.8. Study system: *Pinus torreyana* Parry ex Carrière

1.8.1. Taxonomy and distribution

Pinus torreyana Parry ex Carrière, commonly named Torrey pine, is a yellow pine (subgenus *Pinus*) belonging to subsection *Sabinianae,* also including *Pinus sabiniana* Douglas (Grey pine) and *Pinus coulteri* D. Don (Coulter pine) (Critchfield and Little 1966). Endemic to south-western California, Torrey pine is distributed in two disjunct populations sometimes referred to as two different subspecies - *Pinus torreyana* subsp. *torreyana* and *Pinus torreyana* subsp. *insularis* (J. R. Haller 1986). One population (*Pinus torreyana* subsp. *torreyana*) located in La Jolla, CA, at the Torrey Pine State Reserve (TPSR) extends over an area of approximately 9.6 km^2 along the mainland coast. Trees typically occur at elevations between 30 m and 125 m

on rapidly eroding slopes and ravines, and are surrounded by a diverse plant life, including coastal dunes, Diegan sage, wide-ranging chaparral, more restricted southern chaparral, desert, and woodland species (J. R. Haller 1986).

The other population (*Pinus torreyana* spp. *insularis*) is located on the north-eastern shore of Santa Rosa Island (SRI), one of the Channel Islands, 57 km southwest of Santa Barbara. Trees typically grow between 30 to 180 m above sea level, in ravines and low ridges of a north-facing slope, and are surrounded by a similar plant life to that found at TPSR, including for examples wide-ranging chaparral and woodland species. However, vegetation flourishing under drier conditions are absent from island Torrey pine stands, and mesic as well as endemic species are more abundant (J. R. Haller 1986).

With approximately 3,000 and 4,000 reproductively mature trees within island (SRI) and mainland (TPSR) populations respectively (J. Franklin and Santos 2011; Hall and Brinkman 2015), Torrey pine is considered one of the rarest pine species in North America (Dusek 1985; Critchfield and Little 1966). In addition, low genetic diversity observed within the species indicates populations may suffer from reduced evolutionary potential, threatening it long-term persistence (Ledig and Conkle 1983; Whittall et al. 2010). Listed as critically endangered based on the IUCN Red List of Threatened Species (IUCN 2021b), extensive conservation efforts have been undertaken to preserve this iconic Californian endemic pine both *in situ* and *ex situ*.

1.8.2. A model to study island-mainland evolutionary dynamics

Island and mainland populations present an ideal system to study impacts complex evolutionary processes have had on phenotypic and genetic structure. First, foundation of island populations by small numbers of immigrants and possible geographic isolation following colonization suggests that loss of genetic diversity and phenotypic differentiation on islands may

be driven by stochastic processes, including founder effects, genetic drift, and inbreeding (S. C. H. Barrett 1996; Frankham 1996). However, ecological differences between island and mainland populations may be considerable, indicating that natural selection and adaptations to local environments may represent important drivers of island-mainland evolution. Various morphological changes in island plants have been recorded and hypothesized to result from differential selective pressures among island and mainland environments (Lens et al. 2013; Burns, Herold, and Wallace 2012; Burns 2016). For instance, larger seed size on islands may have evolved to reduce seed loss via dispersal beyond island limits or increase seedling competitive abilities (Kavanagh and Burns 2014; Burns 2016), and larger leaves may have evolved following released from mainland herbivores (Burns, Herold, and Wallace 2012).

Geological deposits on SRI dating from the Eocene to Miocene epochs suggest the island likely uplifted from ocean floors following tectonic movements during the Tertiary age (Muhs et al. 2014; Schumann et al. 2014). Consequently, it has never been part of the North American continent and establishment of Torrey pine on the island required its colonization. This indicates that neutral evolutionary processes could potentially play a role in driving island and mainland population differentiation. However, data exist suggesting that SRI may have been closer to the coast of San Diego during the mid-Miocene than it currently is (approximately 280 kms) (Ledig and Conkle 1983). Combined with elevated pollen dispersal known in pines (Campbell et al. 1999; Varis et al. 2009; Williams 2010), opportunities for gene flow between populations may have existed and may still exist, hindering population divergence through neutral evolution. Nonetheless, phenotypic differences have repeatedly been observed between island and mainland trees (J. R. Haller 1986; J. A. Hamilton et al. 2017). Natural selection and adaptation to differing island and mainland environments may thus also contribute to population divergence. For

example, differences in biotic selective pressures, including release in seed predators on the island and the presence of the California fivespined ips beetle (*Ips paraconfusus* Lanier) on the mainland (Shea and Neustein 1995; M. Johnson, Vander Wall, and Borchert 2003) may have promoted differentiation.

Overall, Torrey pine represents a valuable model to study the impact stochastic and deterministic process have had on rare species' population structure, and more broadly, evolution across island-mainland systems.

1.9. Research objectives

1.9.1. Objective 1: Investigating environmental and geographic distances as potential surrogates for estimating genetic variation and inform *ex situ* sampling (Chapter 2)

Theory suggests environmental and geographic variation structure genetic diversity among species' populations. Environmental variation by prompting individuals' adaptation to their local environment, and geographic variation by reducing inter-population gene flow and promoting genetic differentiation via genetic drift or founding colonization events (Slatkin 1993; I. J. Wang and Bradburd 2014; Ledig 2000). Consequently, when nothing but the distribution of a rare species is known, environmental and geographic distances may provide the ability to prioritize sampling of genetically divergent populations, optimizing genetic diversity captured in *ex situ* collections. Environmental distance was defined as pairwise Euclidean distances among source populations' climates, and geographic distance as the pairwise Euclidean distance among populations' coordinates (latitude, longitude). Combining a simulation framework with previously published genetic data, I tested the hypothesis that *ex situ* collections prioritized using distance-informed population sampling capture more genetic diversity than *ex situ* collections formed from randomly selected populations, as they would preferably be established from

genetically differentiated populations. Furthermore, I predicted that (i) environmental distance-based sampling would optimize preservation of adaptive genetic diversity by capturing adaptive processes influencing population genetic structure, and (ii) geographic distance-based sampling would optimize preservation of standing genetic variation by capturing neutral processes driving population differentiation.

1.9.2. Objective 2: Evaluating morphological variation in existing seed collections as a tool to provide sampling guidance for future *ex situ* conservation efforts (Chapter 3)

Millions of seed accessions have been stored worldwide (Y.-B. Fu, Ahmed, and Diederichsen 2015; D.-Z. Li and Pritchard 2009), representing a large, yet underexploited, conservation resource. Previous research has shown that seed morphology in conifers is primarily genetically determined (Roy, Thapliyal, and Phartyal 2004; Carles et al. 2009; Zas and Sampedro 2015), as opposed to environmentally controlled. Largely heritable, seed trait variation may thus provide a good proxy for estimating the distribution of genetic variation within and among populations. When existing *ex situ* seed collections are available, morphological variation evaluated from stored seeds may thus potentially be used to tailor species-specific guidelines to optimize genetic diversity captured in future conservation collection. Combining morphological variation estimated from wild-collected seeds of Torrey pine with resampling simulations, this study evaluates the potential use of trait variation within and among seed populations as a tool to establish optimal sampling guidance for future *ex situ* conservation efforts.

1.9.3. Objective 3: Determining whether an understanding of species' evolutionary potential gained from genomic data can be used to help optimize conservation of genetic variation (Chapter 4)

Determining appropriate management strategies to preserve and maintain rare species' genetic diversity is a primary objective in conservation biology. Within this context, knowledge on spatial and temporal changes in effective population sizes is needed to evaluate past and current variation in rare species' adaptive potential, as both inbreeding and loss of genetic diversity negatively correlate with populations' effective sizes. For species suffering from severely low effective population sizes, genetic rescue may represent a potential management strategy to increase both genetic variation and fitness of populations (Whiteley et al. 2015). Nonetheless, outcomes of genetic rescue may vary depending on how different evolutionary processes interact to drive population genetic differentiation. While this strategy may restore gene flow between small, isolated populations, mitigating inbreeding depression and increasing populations' performance, it may also have the opposite effect by introducing maladapted variation within targeted populations (Hufford and Mazer 2003). With this study, I teased apart the respective roles neutral and adaptive processes have had on population genetic variation in Torrey pine and evaluated whether rare species' evolutionary history inferred from genomics data can be used to inform conservation management decisions.

1.9.4. Objective 4: Providing conclusive remarks and future directions (Chapter 5)

With the last chapter of this dissertation, I synthesized my most significant results and discussed future work that could complement and build upon past and present research aiming to conserve species' evolutionary potential.

2. USING ENVIRONMENTAL AND GEOGRAPHIC DATA TO OPTIMIZE *EX SITU* COLLECTIONS AND PRESERVE EVOLUTIONARY POTENTIAL[1]

2.1. Abstract

Maintenance of biodiversity through seed banks and botanical gardens, where the wealth of species' genetic variation may be preserved *ex situ,* is a major goal of conservation. However, challenges can persist in optimizing *ex situ* collections where trade-offs exist between expense, effort, and conserving species evolutionary potential, particularly when genetic data is not available. Within this context, we evaluate the genetic consequences of guiding population preservation using geographic (isolation-by-distance, IBD) and/or environmental (isolation-by-environment, IBE) distance for *ex situ* collections where population provenance is available. We use 19 genetic and genomic datasets from 15 plant species to (i) assess the proportion of population genetic differentiation explained by geographic, environmental or both factors, and (ii) simulate *ex situ* collections prioritizing source populations using pairwise geographic distance, environmental distance, or both. Specifically, we test the impact prioritizing sampling based on these distances may have on capturing neutral, functional, or putatively adaptive genetic diversity and differentiation. Individually, IBD and IBE alone explain limited population genetic differences across all three genetic marker classes (IBD: 10%-16%, IBE: 1%-5.5%). However, together, they explain a substantial proportion of population genetic differences for functional (45%) and adaptive (71%) variation. Consistent with these results, simulated *ex situ*

[1] The material presented in this chapter was co-authored by Lionel N. Di Santo and Jill A. Hamilton. Lionel Di Santo had primary responsibility for collecting the data, performing statistical analysis, and developing conclusions advanced here. Lionel Di Santo also drafted and revised all versions of this chapter. Jill Hamilton served as proofreader and contributed to writing and improving all versions of this chapter. A modified version of this chapter is published in *Conservation Biology*: Di Santo, Lionel N, and Jill A Hamilton. 2020. "Using Environmental and Geographic Data to Optimize Ex Situ Collections and Preserve Evolutionary Potential." *Conservation Biology* 35 (2): 733–44. https://doi.org/10.1111/cobi.13568.

collections reveal that inclusion of IBD and/or IBE increases both allelic diversity and genetic differentiation captured among populations, particularly for loci that may be important for adaptation. Thus, prioritizing population collections using environmental and/or geographic distance data can optimize genetic variation captured *ex situ*. For the vast majority of plant species for which we have no genetic information, these data are invaluable to conservation, guiding preservation of genetic variation needed to maintain evolutionary potential within collections.

2.2. Introduction

Genetic variation is fundamentally a prerequisite for adaptive evolution (Carlson, Cunningham, and Westley 2014). Consequently, to maintain species' evolutionary potential, conservation often focuses on the preservation and maintenance of genetic variation. *Ex situ* collections provide one approach to preserve genetic diversity outside species' native ranges. This includes extensive efforts to collect, preserve, and maintain variation across the range of different crop species, wild relatives, and rare or threatened species (Q. Li, Xu, and He 2002; Westengen, Jeppson, and Guarino 2013; Naredo et al. 2017). The Global Strategy for Plant Conservation (GSPC) aims to have at least 75% of endangered plant species preserved *ex situ* by 2020 and available for use in recovery or restoration (Target 8; https://plants2020.net/). While significant progress has been made, major gaps remain in the maintenance of genetic variation within collections (Sharrock, Hoft, and Dias 2018). Consequently, *ex situ* programs designed to maintain genetic diversity are yet needed.

Traditionally, *ex situ* methods rely on either probabilistic equations (Brown and Marshall 1995; Lawrence, Marshall, and Davies 1995), or stochastic resampling using pre-existing genetic datasets to optimize sampling efforts (Caujapé-Castells and Pedrola-Monfort 2004; Gapare,

Yanchuk, and Aitken 2008). However, these approaches have limitations as they require either the availability of genetic data (population resampling strategy) or make ungeneralizable assumptions of within species population structure (probability-based strategy; Lockwood, Richards, and Volk 2007). More recently, simulation-based strategies have been developed and tested to guide sampling practices (Hoban and Schlarbaum 2014; Hoban 2019). Simulation-based approaches do not require previously published genetic datasets but enable realistic simulations of population structure using available estimates of population size and genetic connectivity. To overcome challenges associated with *a priori* data requirements, the use of surrogate data, such as environmental or spatial data to estimate neutral and nonneutral genetic variation has received considerable attention (Guerrant Jr, Havens, and Vitt 2014; R. Whitlock et al. 2016; Hanson et al. 2017). Empirical work has focused mainly on testing these data surrogates in preserving genetic diversity *in situ* or in wild populations (R. Whitlock et al. 2016; Hanson et al. 2017). However, using environmental and geographic data to optimize *ex situ* sampling could have substantial value to conservation.

Evolutionary processes have predictable impacts on the distribution of standing genetic variation, which may be used to guide *ex situ* collections. IBD or "isolation-by-distance" (Wright 1943) arises when gene flow between geographically distant populations is not enough to counteract the accumulation of genetic differences via genetic drift or following successive founder events during colonization (Slatkin 1993; Ledig 2000). In this way, IBD is a proxy for the relationship between pairwise population geographic and genetic distances associated with spatial structure and serial colonization across a landscape. Likewise, IBE or "isolation-by-environment" (I. J. Wang and Summers 2010) describes the accumulation of genetic differences between environmentally distinct populations. IBE predicts that environmental differences are

correlated with genetic differences (I. J. Wang and Bradburd 2014), as selection differs across environments (M. Keller, Kollmann, and Edwards 2000; McBride and Singer 2010), providing a proxy for the relationship between genetic and environmental distance. The influence of geographic and environmental variation in structuring patterns of genetic variation, either independently or collectively, has received extensive support across taxa (Sexton, Hangartner, and Hoffmann 2014). Given these observations, spatial and environmental data may provide valuable proxies in designing *ex situ* conservation collections that optimize the preservation of neutral and nonneutral evolutionary processes.

The impact of IBD and IBE on population genetic structure is expected to differ for neutral and adaptive genetic variation (Table 1). This includes the prediction that IBD will have a greater influence at neutral loci relative to IBE. IBD reflects past and current demographic history, as well as the interplay between drift and gene flow in structuring genetic variation, whereas IBE is influenced by natural selection, largely reflecting adaptive genetic variation. Cumulatively, we predict that IBD and IBE will explain the greatest proportion of genetic differences among populations for nonneutral loci. Finally, for those genetic markers underlying functional genetic diversity, including polymorphisms within genes or expressed sequences, we predict patterns of IBE and IBD will be intermediate as they may reflect a combination of adaptive and neutrally evolving loci.

The explosion of genetic and genomic datasets publicly available provides a timely opportunity to compare the contribution of IBD and IBE to genetic structure. In the present study, we compare the influence of genetic marker type on IBD and IBE. We classify single-sequence repeats (SSRs) and genome-wide single-nucleotide polymorphisms (SNPs) as neutral genetic variation (neutral class), SNPs identified previously as candidate loci for selection using

statistical or empirical methods as underlying adaptive genetic diversity (adaptive class), and genetic markers within known genes or expressed sequences (genic SNPs or expressed sequence tag SSRs) as a functional class. We distinguish functional polymorphisms from neutral and adaptive classes as these markers estimate quantitative genetic variation and likely represent a combination of neutral and adaptive processes.

Table 1. Evolutionary processes [a] contributing to genetic structure across neutral and adaptive genetic markers and their predicted weight [b] on expected patterns of among-population genetic differentiation (Random, IBD and IBE).

Neutral genetic markers	Random	IBD	IBE
Stochastic processes (e.g., genetic drift, inbreeding)	++	-	-
Demographic history (e.g., founder events)	++	+	-
Genetic drift combined with gene flow	-	+++	-
Natural selection	-	-	+
Adaptive genetic markers	**Random**	**IBD**	**IBE**
Stochastic processes (e.g., genetic drift, inbreeding)	- (+)	-	-
Demographic history (e.g., founder events)	-	-	-
Genetic drift combined with gene flow	-	+	-
Natural selection	-	-	+++

[a] Here genetic drift alone is a stochastic evolutionary force and genetic drift combined with gene flow is a process leading to a pattern of IBD.
[b] Influence of the evolutionary forces on the specified pattern; -: no, +: small, ++: intermediate, and +++: important.

To optimize sampling of genetic variation and differentiation *ex situ*, we have re-analyzed existing genetic and genomic datasets to (i) quantify the impact of IBD and IBE have on population genetic structure across neutral, functional, and putatively adaptive genetic datasets, and (ii) to evaluate whether inclusion of IBD and IBE during population sampling influences genetic diversity captured at neutral, functional, and adaptive loci using simulated *ex situ* collections. We use variation partitioning to disentangle the effect of IBD, IBE, their intersection, and union on population genetic structure and then simulate *ex situ* collections using geographic and environmental distance metrics to optimize genetic variation and differentiation

conserved. This study seeks to advance understanding of the role non-genetic factors play in the distribution of genetic variation across natural populations, and to provide new parameters to optimize *ex situ* sampling designs where genomic data may be limited or non-existent.

2.3. Materials and methods

2.3.1. Source of genetic and geographic data

We searched the Dryad Digital Repository (https://datadryad.org/) to identify genetic or genomics datasets for plant species using three discrete search categories: "Population structure plant", "SSR population structure" and "SNP population structure". Following this, for inclusion in our study, a dataset or a subset of a dataset had to meet the following criteria:

- Populations were collected range-wide or were sampled across an isolated fraction of a species' distribution.
- Geographic coordinates (latitude, longitude) were available for each population sampled.
- Genetic data, categorized as SSRs (single-sequence repeats), EST-SSRs (expressed sequence tag SSRs) or SNPs (single-nucleotide polymorphism), were available.

Range-wide sampling or sampling of populations spanning a large, isolated fraction of a species' distribution were required to ensure the majority of a species' ecological niche space was captured. In addition, sampling a broad range of environmental and geographic distances can reduce the likelihood of covariance between environmental and geographic factors (I. J. Wang and Bradburd 2014). Using publicly available databases, population-specific latitude and longitude were used to model climatic variation associated with geographic provenance. These data were used in variation partitioning analyses and to calculate pairwise population

environmental and geographic distances for each species. To calculate genetic distances, we included studies using SSRs, SNPs or EST-SSRs. SNP genotyping varied across studies, therefore we divided SNP datasets into two categories: SNPs assessed genome-wide (SNPs) and SNPs assessed within genes (Gen-SNPs). If specific SNPs were identified as being under selection based on previous work, we included a fifth category, SEL-SNPs. Finally, genetic markers were broadly classified as either putatively neutral (neutral class: SSRs, SNPs), underlying functional variation (functional class: EST-SSRs, Gen-SNPs) or putatively adaptive (adaptive class: SEL-SNPs).

Overall, we gathered 17 genetic or genomic datasets, in addition to two genomic datasets received directly from Holliday et al. (2010) (Table 2; Table A1). To meet the above criteria, datasets associated with seven of the 15 studied species were sub-sampled and individual geographic coordinates for one study were averaged to create population-scale coordinates (Table 2; Appendix B).

2.3.2. Environmental data

We used latitude, longitude and elevation associated with population provenance to extract annual, seasonal, and monthly climate variables using ClimateNA (North America), ClimateSA (South America), ClimateEU (Europe) or ClimateAP (Asia Pacific) (https://sites.ualberta.ca/~ahamann/data.html) (Table A2). Where elevation was not provided, GPS Visualizer (http://www.gpsvisualizer.com/elevation) was used to assign population elevation values. In total, 80 environmental variables were assigned to each population, including 79 climate-related variables and elevation. For each of the species, all environmental variables associated with population origin were filtered, standardized, and transformed to

Table 2. Proportion of genetic differentiation explained by environmental and geographic variables [a] obtained using variation partitioning analyses and correlation coefficients estimated between pairwise geographic and environmental Euclidean distances for all 19 genetic and genomic datasets.

Study system		Data		Results				
Species	Distribution	Number of Populations	Genetic Marker [b]	IBD (Adj. R2)	IBE (Adj. R2)	IBD∩IBE (Adj. R2)	IBD∪IBE (Adj. R2)	Corr. (r)
Betula maximowicziana	Japan	48	EST-SSRs	0.02	0.02	0.42	0.46[e]	0.48[e]
Centaurea solstitialis [c]	Eurasia	25	SNPs	0.14[e]	0.33[e]	0	0.47[e]	-0.02
Helianthus annuus	North America	15	SNPs	0.1[e]	0.08[f]	0.02	0.2[e]	0.93[e]
Helianthus argophyllus [c]	Texas	51	Gen-SNPs	0.02	0.04[e]	0.32	0.38[e]	0.9[e]
Mimulus guttatus [c]	United Kingdom	14	SNPs	0.14	0.09	0	0.23	0.56[e]
Mimulus lacinatus [c]	California	23	SSRs	0.01	0.03	0.04	0.08[f]	0.35[e]
Narcissus papyraceus [c]	Spain and Morocco	26	SSRs	0.12[f]	0.03	0.02	0.17[f]	0.08
Nothofagus alpina	Chile	12	SSRs	0	0	0.18	0.18	0.49[e]
Nothofagus glauca	Chile	8	SSRs	0.75[e]	0.05	0.06	0.86[e]	0.2
Nothofagus obliqua	Chile	20	SSRs	0.17[e]	0.06	0.39	0.62[e]	0.31[e]
Picea sitchensis [c]	North America	10	Gen-SNPs	0.07	0	0.37	0.44	0.44[e]
		10	SEL-SNPs	0.15	0	0.56	0.71[f]	0.44[e]
Populus balsamifera [c]	North America	31	Gen-SNPs	0.35[e]	0.01	0.3	0.66[e]	0.42[e]
		31	SEL-SNPs	0.32[e]	0.01	0.42	0.75[e]	0.42[e]

Table 2. Proportion of genetic differentiation explained by environmental and geographic variables [a] obtained using variation partitioning analyses and correlation coefficients estimated between pairwise geographic and environmental Euclidean distances for all 19 genetic and genomic datasets. (continued)

Study system		Data		Results				
Species	Distribution	Number of Populations	Genetic Marker [b]	IBD (Adj. R2)	IBE (Adj. R2)	IBD∩IBE (Adj. R2)	IBD∪IBE (Adj. R2)	Corr. (r)
Populus tremula [d]	Sweden	12	Gen-SNPs [control set]	0.02	0	0.02	0.04	0.71[e]
		12	Gen-SNPs [defense set]	0.15	0.05	0.33	0.53[e]	0.71[e]
		12	SEL-SNPs	0.16	0.07	0.25	0.48[e]	0.71[e]
Rhododendron oldhamii	Taiwan	18	EST-SSRs	0.13[e]	0.05	0.24	0.42[e]	0.29[e]
Shorea leprosula	South-East Asia	24	EST-SSRs	0.24[e]	0.03	0.25	0.52[e]	0.27[e]

[a] Proportion of population genetic differentiation explained by pure geographic factors (IBD), pure environmental factors (IBE), the shared variation between environmental and geographic factors (IBD∩IBE), and both environmental and geographic factors combined (IBD∪IBE). Note that negative adjusted-R^2 values are listed and interpreted as zeros.

[b] SSR: single-sequence repeat (neutral class), EST-SSR: expressed sequence tag single-sequence repeat (functional class), SNPs: genome-wide single-nucleotide polymorphism (neutral class), Gen-SNPs: genic single-nucleotide polymorphism (functional class), and SEL-SNPs: single-nucleotide polymorphism identified as potentially under selection (adaptive class).

[c] Subsampled genetic or genomic datasets.

[d] Adjusted geographical coordinates.

[e] Fractions of variation explained and correlation coefficients are significant at α=0.05.

[f] Fractions of variation explained and correlation coefficients are significant at α=0.1.

summarize environmental differences among populations. First, dataset-specific environmental variables exhibiting no population-level variation were excluded from analyses. Environmental variables were then standardized and used to conduct a principal component analysis (PCA). PCA was used to reduce the overall number of environmental variables by summarizing environmental differences across two major axes of differentiation, which together explain more than 70% (range = 73.9-93%) of the environmental variation observed among populations (Table A3). These two major PC axes were considered as predictor variables for variation partitioning and used to calculate population pairwise environmental distances in simulations. Although two environmental PC axes are used here to summarize major environmental differences between populations, the inclusion of additional axes may capture subtler environmental differences. Therefore, if the objective of a collection is to preserve fine-scale adaptive genetic variation, increasing the number of PC axes may capture subtler environmental variation important for adaptation.

2.3.3. Variation partitioning analysis

To quantify the contribution of IBD and IBE to genetic divergence within each of the 19 datasets, we conducted a variation partitioning analysis in R (R core Team 2018) using the "vegan" package (Oksanen et al. 2007). We used standard estimates of population genetic differentiation re-calculated for all population pairs within each dataset as our response variable. To account for variation in genetic markers, we used Nei's F_{ST} (Nei 1987), as this metric can provide comparable estimates of population genetic differentiation for both biallelic (e.g. SNPs) and multi-allelic (e.g. SSRs) loci. For each dataset, population divergence was partitioned between two sets of predictor variables, including the geographic coordinates (latitude, longitude) and the two major environmental PC axes (PC1, PC2) associated with each population

within a dataset. Following variation partitioning, we conducted a partial distance-based redundancy analysis (dbrda) on each dataset to test the significance of (i) variance explained by each set of predictor variables alone (IBD, IBE; Table 2), and (ii) the variance explained by the union of predictor variables – that is the total variance explained by geographic and environmental factors - (IBD∪IBE; Table 2). We did not evaluate the significance of the variance explained by the intersection of predictor variables – or the variance explained by the shared variation between geographic and environmental factors – (IBD∩IBE; Table 2), as this variance fraction is not testable using dbrda.

2.3.4. Quantifying the correlation between genetic, environmental, and geographic distances

Geographic and environmental distance between population pairs was measured as the Euclidean distance between populations' geographic coordinates (latitude, longitude) or between populations' two major environmental PC axes (PC1, PC2), respectively. To visualize and evaluate the covariance structure between genetic, environmental, and geographic distance matrices, we graphed and estimated the correlation between all distance metrics (Table 2; Figure A1). Correlation coefficients were estimated using the nonparametric mantel test implemented in the R package "adegenet" (Jombart 2008) for each dataset separately.

2.3.5. Simulating an *ex situ* collection: an idealized framework

We simulated an idealized *ex situ* conservation collection for each dataset using a customized R script relying on R packages "adegenet" (Jombart 2008), "hierfstat" (Goudet 2005) and "data.table" (Dowle and Srinivasan 2019). This simulation measured the amount of genetic differentiation and the proportion of allelic diversity captured in *ex situ* collections that prioritize population sampling based on environmental and geographic distances. We simulated *ex situ* collections using four different population sampling strategies. This included random sampling,

as well as sampling prioritized based on distances between populations' two major environmental PC axes (Euclidean environmental distance), sampling based on distances between populations' geographic coordinates (Euclidean geographic distance) or both (Figure 1).

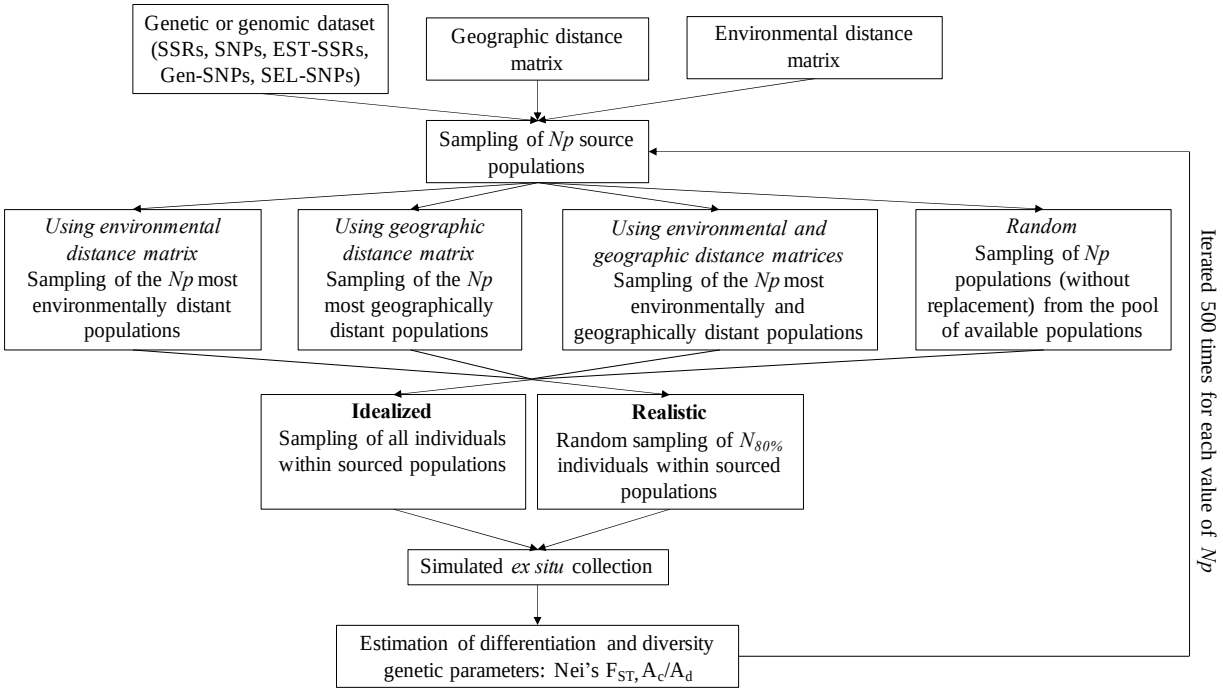


Figure 1. Simulation framework used to estimate genetic variation and differentiation parameters in *ex situ* collections simulated under two different within-population sampling scenarios (realistic and idealized) and four distinct population prioritization strategies (random, based on environmental distance, based on geographic distance, and based on both environmental and geographic distance combined). Simulations using this framework were conducted on each dataset independently. Computation proceeds from top to bottom.

Ex situ collections were simulated using between two and the total number of populations available for each dataset (*Np*, Figure 1). Randomized sampling sampled populations without replacement from the pool of available populations. Environmentally or geographically prioritized simulations sampled population pairs with the greatest pairwise distances in decreasing order. Collections simulated using the combination of environmental and geographic distances sampled population pairs that exhibited the greatest sum of environmental and

geographic distances following standardization, prioritized in decreasing order. All individuals within each population were sampled as part of the idealized simulation.

To compare genetic diversity captured across simulated collections, we estimated two genetic parameters: Nei's F_{ST} and allelic diversity captured (A_c/A_d). These indices were chosen as they quantify different aspects of population genetic diversity. Nei's F_{ST} provides an estimate of genetic differentiation across sampled populations and A_c/A_d provides an estimate of the number of alleles captured in collections (A_c) relative to the total number of alleles present within a dataset (A_d). All genetic parameters were estimated in R using the "hierfstat" package.

Population sampling and associated genetic summary statistics were simulated 500 times for each dataset to account for the variance introduced through randomly sampling across populations. Summary statistics were estimated based on average values across all 500 simulations. No replication was used for environmental and/or geographic distance-based population sampling, as neither provenance of source populations nor genetic summary statistics would have changed with repeated iterations.

For these idealized simulations, all individuals were sampled within each target population (equivalent to protecting the entire population), regardless of collection strategy, assuming 100% of the standing genetic variation was captured. However, monetary, or logistical constraints usually impact the number of individuals that could be sampled within a target population. Given this, we predict that genetic diversity captured within source populations will vary. To assess whether insights gained from idealized simulations were maintained under more realistic conditions, we conducted additional simulations, introducing differences in the amount of genetic diversity captured between populations (hereafter referred to as realistic simulations, see below).

2.3.6. Simulating an *ex situ* collection: a realistic framework

To simulate a realistic *ex situ* collection, a subset of individuals was sampled within each population. This provides the opportunity to evaluate the impact varying genetic diversity captured within populations may have on total genetic diversity and differentiation captured across populations collected. We assume that *ex situ* collections aim to preserve as much genetic variation as possible within each population. Within this framework, we postulated that at least 80% of within-population allelic diversity would be captured *ex situ*. Therefore, for each dataset, we assessed the number of individuals ($N_{80\%}$) that when sampled captures between 80%-100% of allelic diversity within each population, stochastically introducing variation in genetic diversity captured across populations.

An additional simulation was used to determine the value of $N_{80\%}$ for each dataset (Figure 2). For every population, *N* individuals (ranging from one up to the size of the smallest population within the assessed dataset) were randomly sampled without replacement. Following this, the number of alleles captured for *N* individuals (A_s) divided by the total number of alleles in the population (A_p) was quantified for each population. Sampling of individuals and quantification of allelic diversity captured was replicated 500 times for each population and value of *N* to calculate confidence intervals around A_s/A_p ratios. The number of individuals required to capture 80% or more ($A_s/A_p \geq 0.8$) of allelic diversity in every population ($N_{80\%}$) was visually assessed for each dataset independently (Figure A2) and used to parametrize realistic simulations (Figure 1). *Ex situ* collections were simulated 500 times using the realistic scenario to estimate genetic summary statistics regardless of the population sampling strategy used (Figure 1). For these simulations, $N_{80\%}$ were often much lower than the existing size of most

populations and performing repeated iterations accounted for the variation in genetic summary statistics introduced by small values of $N_{80\%}$.

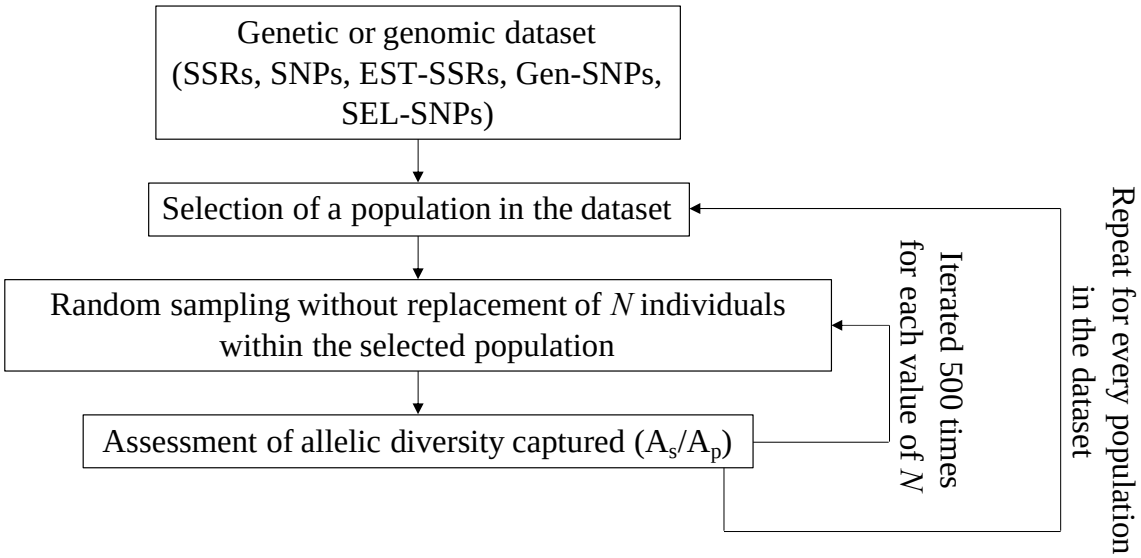


Figure 2. Simulation framework used to estimate the number of individuals required to capture between 80-100% of allelic diversity in each population of a dataset ($N_{80\%}$). Simulations using this framework were conducted on each dataset independently. Computation proceeds from top to bottom.

Maintaining the range of A_s/A_p ratios across datasets is crucial as unbalanced variance may confound the influence of prioritization strategies in downstream analyses. Four of the 19 datasets (*H. argophyllus* (Gen-SNPs), *M. lacinatus* (SSRs), *R. oldhamii* (EST-SSRs), and *S. leprosula* (EST-SSRs)) were discarded from realistic simulations, as $N_{80\%}$ values were not reached for these datasets (Figure A2). These same datasets were also removed from idealized simulations to ensure that differences in summary statistics between idealized and realistic simulations originated solely from variation in allelic diversity captured across populations introduced in the latter. See Table A4 for a complete list of parameters tested and used for simulations.

2.3.7. Analysis of simulated data

We tested whether prioritizing source population collection using environmental and/or geographic distance data influences genetic variation and differentiation captured *ex situ*. For every number of populations sampled (*Np*), genetic summary statistics simulated using random sampling were subtracted from values based on prioritization strategies using environmental distances, geographic distances, or both. Summary statistics were averaged for each dataset following repeated iterations, grouped by distance-based strategies, genetic marker class, and simulation framework (idealized or realistic) (Figure 4). Differences in genetic summary statistics are provided based on the proportion of populations sampled as the number of populations sampled for analysis varied across studies. For each dataset, we selected four numbers of populations sampled (*Np*) representing between 30-40%, 50-60%, 70-80%, and 90-100% of populations present in a dataset (Table A4).

Finally, we fitted a linear model between proportions of populations sampled and differences in genetic summary statistics for every combination of genetic marker class, distance-based prioritization strategy, and simulation framework (Figure 4). A negative relationship indicates that a given distance-informed sampling generally increases the genetic summary statistics relative to random sampling while a positive relationship would suggest the opposite. In addition, it is important to note that a significant relationship (positive or negative) will always be approaching zero as the proportion of populations sampled increases. This is because with additional populations sourced, the probability that identical populations are sampled randomly or via distance-based strategies increases and will reach one when all populations are sampled. As the number of shared populations between sampling strategies increases, the difference in genetic summary statistics decreases.

2.4. Results

2.4.1. Relative contributions of IBD and IBE to population genetic differentiation

Variation partitioning revealed that IBD explained significantly more among-population genetic differences (13%) than IBE alone (5.5%) or IBD∩IBE (3%) for neutral genetic datasets (Figure 3A). This contrasts with functional and adaptive datasets, where a significantly higher proportion of among-population genetic differences was explained by geographically structured environmental variables relative to environmental or geographic factors alone (Figure 3A). Overall, 31% and 42% of population genetic differences were explained by IBD∩IBE for functional and adaptive datasets, respectively, while only a small proportion was explained by IBD (functional: 10%, adaptive: 16%) and IBE alone (functional: 2.5%, adaptive:1%).

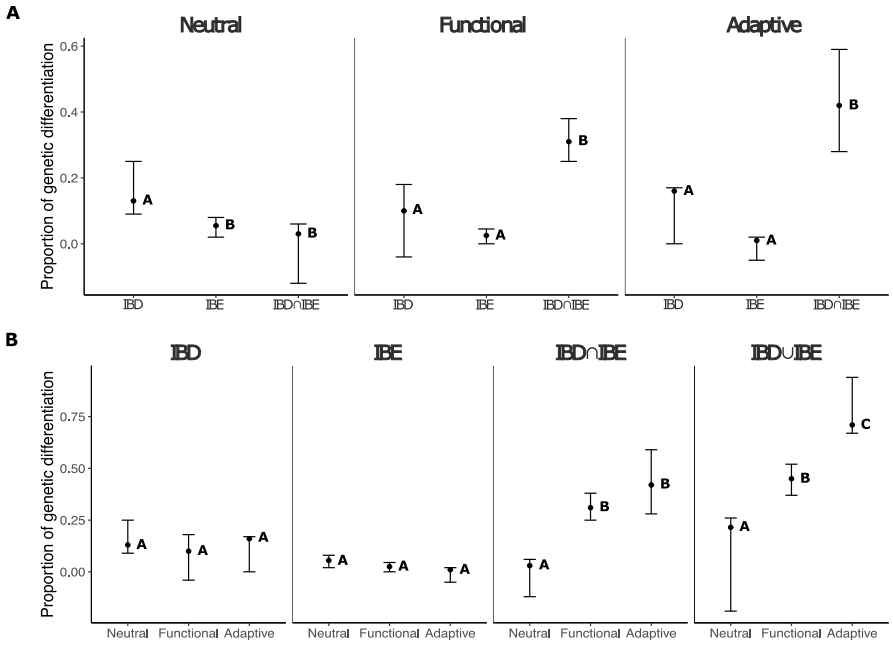


Figure 3. Median proportion of population genetic differentiation explained by (A) purely geographic factors (IBD), purely environmental factors (IBE), and the shared variation between environmental and geographic factors (IBD∩IBE) across genetic marker classes (neutral, functional, adaptive) and (B) IBD, IBE, IBD∩IBE, and both environmental and geographic factors combined (IBD∪IBE) across genetic marker classes. Error bars represent 95% confidence intervals obtained by bootstrapping (2,000 bootstrap replicates). Two medians are significantly different (α=0.05) if their confidence intervals do not overlap. Significant difference between medians is notified by nonmatching capital letters (A, B, C).

While significant differences in the proportion of genetic differentiation explained were observed across genetic marker classes for IBD∩IBE and IBD∪IBE, no significant differences were observed in the individual contribution of IBD and IBE (Figure 3B). IBD∪IBE explained the greatest proportion of genetic differences for adaptive genetic markers (71%), followed by functional (45%) and neutral (21.5%) genetic markers, respectively. Interestingly, IBD∩IBE explained substantial among-population genetic differences for both functional and adaptive datasets but explained limited variation for neutral datasets (Figure 3B). The contribution of IBD∩IBE to population genetic differentiation for adaptive and functional datasets likely reflect high correlations observed between environmental and geographic distance matrices (Table 2; Figure A1). Therefore, the relative contribution of geography and environment should be interpreted with caution for these genetic marker classes, as population genetic differentiation could not be partitioned solely by IBD or IBE.

2.4.2. Genetic diversity and differentiation captured in simulated *ex situ* collections

2.4.2.1. Genetic differentiation (Nei's F_{ST})

Significant negative relationships were observed between proportions of populations sampled and changes in genetic differences (F_{ST}) captured for collections simulated using both adaptive and functional datasets, but not neutral genetic datasets (Figure 4A). This suggests that using environmental and/or geographic distance to prioritize population sampling may potentially increase adaptive and functional genetic variation captured in *ex situ* collections. Simulations revealed that using all three distance-based population sampling strategies increased genetic differentiation captured among adaptive loci (Figure 4A). This contrasts with the results obtained for functional datasets, where sampling prioritizing source populations using environmental distance, or the combination of both environmental and geographic distances

increased genetic differences captured. For both adaptive and functional genetic makers classes, simulations based on realistic and idealized within-population sampling scenarios led to similar slopes, regardless of the distance-based population sampling strategy used (Figure 4A; Table A5). This indicates that the ability of distance-based population sampling strategies to increase F_{ST} among functional and adaptive loci was not impacted by the within-population sampling scenarios simulated.

2.4.2.2. Proportion of allelic diversity captured (A_c/A_d)

Both realistic and idealized *ex situ* collection simulations using functional and adaptive genetic datasets indicated allelic diversity captured (A_c/A_d) is likely sensitive to within-population sampling. Prioritizing population sampling using environmental distances increased allelic diversity captured at functional loci under realistic within-population sampling conditions but had no impact using idealized within-population sampling scenario (Figure 4B). This contrasts with results obtained for adaptive datasets, where the opposite pattern was observed. Prioritizing population sampling using environmental or the combination of environmental and geographic distances increased A_c/A_d under idealized within-population sampling conditions (Figure 4B). For neutral genetic datasets no consistent change in allelic diversity was observed in response to varying proportions of population sampled, regardless of population prioritization strategy tested or within-population sampling scenario simulated (Figure 4B). Together, these results suggest that incorporating environmental and/or geographic distances to prioritize collections may increase allelic diversity captured at both functional and adaptive loci. Nonetheless, simulations also indicate that increasing allelic diversity captured in *ex situ* collections is dependent on within-population sampling scenarios and may thus only be achieved under specific sampling conditions.

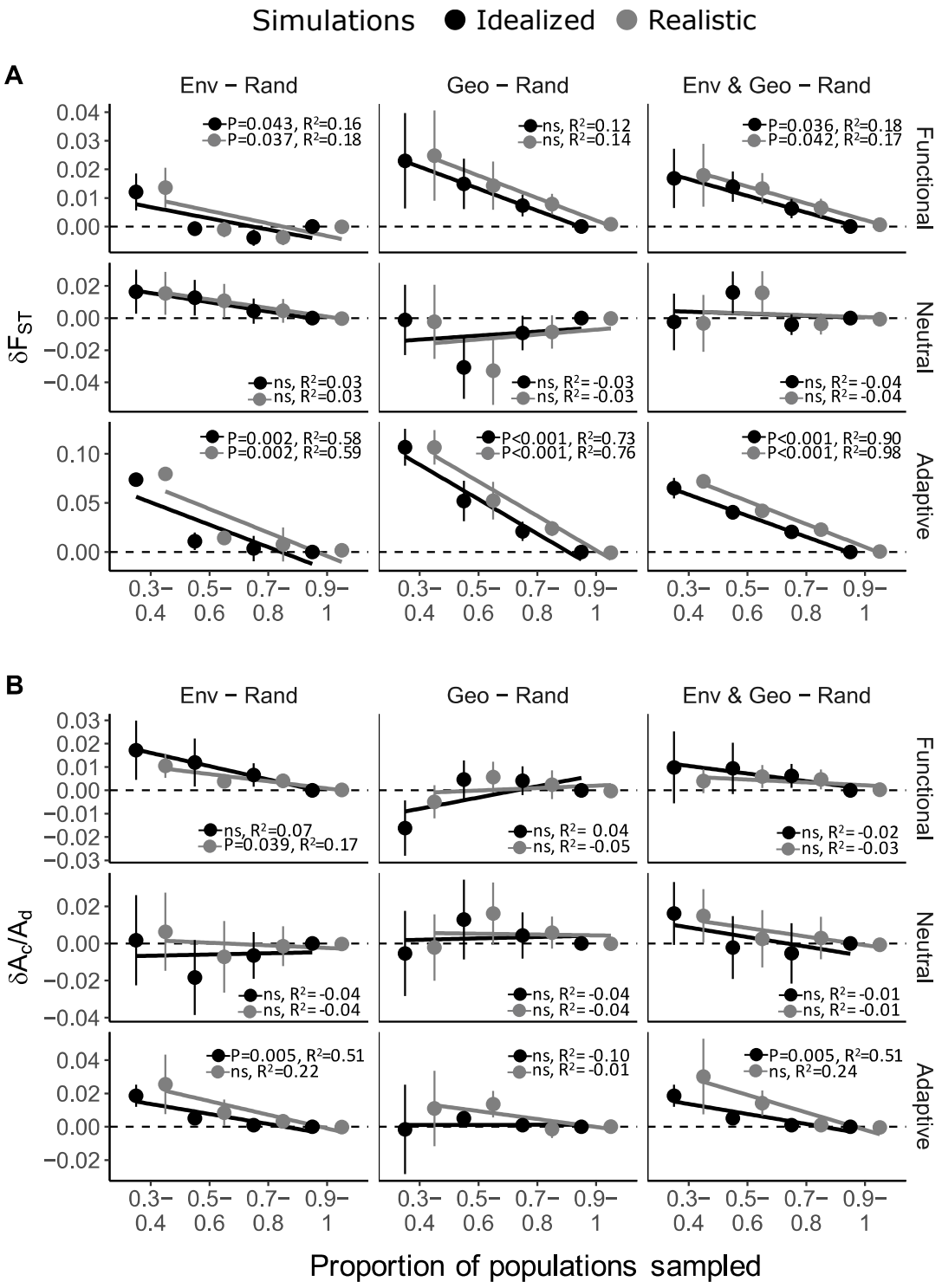


Figure 4. Average differences (±SE) across datasets in (A) population genetic differentiation (Nei's F_{ST}) and (B) allelic diversity (A_c/A_d) estimated from *ex situ* collections simulated using distance-informed (Env: environmental, Geo: geographic, and Env & Geo: environmental and geographic) and random (Rand) population sampling strategies separated by genetic marker classes (functional, neutral, and adaptive). Differences in genetic summary statistics were estimated for different proportions of populations sampled. ns: non-significant.

2.5. Discussion

Optimizing efforts to conserve genetic variation relies upon an understanding for how non-genetic factors, geographic and environmental variation, contribute to population genetic structure. Here, we leverage population provenance and environmental data to optimize genetic differences captured in simulated conservation collections. Environmental and geographic factors explain genetic differences observed among populations, although the extent differs by genetic marker class. The proportion of genetic differentiation explained by IBD∪IBE was significantly higher for adaptive and functional datasets relative to neutral datasets (Figure 3B). This suggests that geographic and environmental data may provide a useful guide when designing *ex situ* population sampling, particularly where the goal is to conserve adaptive and functional genetic variation. We simulated *ex situ* sampling and found that, as predicted, strategies that included environmental, geographic, or both distances to prioritize population sampling increased genetic differences and diversity captured at functional and adaptive loci. Overall, we suggest inclusion of IBD and IBE in guiding *ex situ* sampling can ensure adaptive and functional genetic variation is conserved, crucial for long-term preservation and maintenance of species' evolutionary potential. Although optimizing ex situ collections likely requires a comprehensive cost-effectiveness analysis (Griffith and Husby 2010; Griffith, Lewis, and Francisco-Ortega 2011), leveraging these existing data to prioritize collections provides an economical and timely means to achieve conservation goals.

Consistent with previous studies, our results demonstrate that genetic differentiation across neutral, functional, and adaptive loci can, at least partly, be explained by environmental and geographic factors (Sexton, Hangartner, and Hoffmann 2014; Nadeau et al. 2016; Xia et al. 2018) (Table 2). Interestingly, limited genetic differentiation was explained by IBD or IBE alone

across all three genetic marker classes. For functional and adaptive datasets, this is likely due to the fact that substantial genetic structure is explained by their intersection (Figure 3A). Indeed, IBD∩IBE reflects covariance between geographic and environmental factors that cannot be teased apart. Additional empirical work minimizing this covariance would be required to completely disentangle these factors. Nonetheless, when combined, environmental and geographic factors explained a substantial proportion of population genetic differentiation for functional and adaptive datasets (IBD∪IBE; Figure 3B). This suggests that geographic and environmental differences contribute largely to genetic divergence at nonneutral loci (Huang et al. 2016; Xia et al. 2018). Consequently, the inclusion of IBD and IBE may provide a means to capture adaptive and functional genetic variation *ex situ*. For neutral datasets, geographic and environmental factors, either individually (IBD, IBE) or cumulatively (IBD∪IBE), explained very small proportions of among-population genetic differences (Figure 3). This indicates that stochastic processes, such as genetic drift, bottlenecks or founding events (Stern and Orgogozo 2009; Maruyama and Fuerst 1985; Gavrilets and Hastings 1996), likely influence neutral genetic structure. Overall, our findings suggest that environmental and geographic distance metrics can be used to target genetic differences that likely reflect adaptive or functional genetic variation over neutral genetic variation.

Ex situ strategies have previously optimized variation captured in collections relying on existing genetic datasets (Caujapé-Castells and Pedrola-Monfort 2004; Gapare, Yanchuk, and Aitken 2008) or genetic simulations (Hoban and Schlarbaum 2014; Hoban 2019). These approaches require substantial *a priori* information and target neutral genetic variation. Where knowledge of population location is available, pairwise geographic and environmental distances may be leveraged to extend previous sampling to conserve adaptive and functional genetic

variation. Our simulations demonstrate that *ex situ* collections prioritized using environmental or the combination of environmental and geographic distances increase both Nei's F_{ST} and A_c/A_d captured for adaptive and functional datasets relative to random sampling (Figure 4). This suggests that divergent selection contributes to genetic differentiation at nonneutral loci (Hancock et al. 2011; T. Wang et al. 2016). Interestingly, IBE-based prioritization strategies indicate that part of additional functional and adaptive genetic differences captured in collections consist of environmentally or spatially restricted alleles (Figure 4B). However, simulations also revealed that increasing allelic diversity captured in collections using distance-based prioritization strategies depends on within-population sampling conditions (realistic or idealized). These results have important applications to applied conservation efforts. While a realistic sampling scenario was sufficient to increase Nei's F_{ST} at adaptive and functional loci, only an idealized sampling scenario increased A_c/A_d captured at adaptive loci (Figure 4). This suggests that IBD- and IBE-based sampling strategies can likely increase genetic differences captured among populations by sampling only a subset of their individuals, whereas extensive within-population sampling may be needed to increase adaptive allelic diversity captured in collections. Overall, simulations demonstrate that inclusion of IBD and IBE in population prioritization provides a means to target genetic variation that may be needed to maintain adaptive potential within collections, without the need for prior genetic data.

Despite the fact conservation has long valued environmental and geographic data (Brown and Marshall 1995; Guerrant, Havens-Young, and Maunder 2004; Guerrant Jr, Havens, and Vitt 2014), use of these data for conservation planning have only emerged during the past decade (Vinceti et al. 2013; Hanson et al. 2017; R. Whitlock et al. 2016). Consistent with previous work, we observe inconsistent benefits of leveraging geography for the preservation of neutral genetic

diversity (Figure 4). This could be due to the fact that gene flow between populations may be disturbed by landscape characteristics (Dudaniec et al. 2016), or some species may exhibit greater gene flow between geographically distant populations (O'Connell, Mosseler, and Rajora 2007). Our results, however, do provide further empirical support for inclusion of environmental and geographic data in conservation planning to target and increase adaptive variation conserved (Hanson et al. 2017) (Figure 4). In addition, this study is the first to provide evidence that distance-based population prioritization strategies may increase genetic differentiation and diversity captured at functional loci. This indicates that using environmental and geographic surrogates may not only preserve current adaptive genetic variation but may also secure variation crucial for future evolution.

Although simulations are a powerful inferential tool, they can include a number of assumptions. Here, we assumed that maternal plants used in realistic and idealized simulations were collected for storage *ex situ*. However, the progeny of these plants more accurately reflects those likely to be included in collections (FAO 2010). Future studies will need to consider empirical or simulated progeny data to evaluate whether strategies leveraging environmental or geographic data capture genetic variation across generations. In addition, by estimating environmental differentiation between populations using principal component analysis, we made two assumptions: linearity between environmental variables and low dimensionality of species' environmental space. To test the susceptibility of our results to these assumptions, we re-ran idealized and realistic simulations using non-metric multidimensional scaling (NMDS) instead of PCA, and three instead of two environmental PC axes to estimate environmental distances (Figure A3; Figure A4). Despite minor quantitative changes in the outcome of individual simulations, these re-analyses lead to the same broad conclusions. Finally, we grouped different

genetic markers into genetic diversity classes to test the effect of prioritizing population sampling using environmental data, geographic data, or both at a broader scale. However, allelic distributions and mutation models largely differ between these genetic markers. Thus, future work should evaluate marker-specific patterns associated with IBD- and IBE-based prioritization strategies.

Anthropogenic changes have had substantial impacts on biodiversity, resulting in a global call for its preservation. While the cost of genotyping and sequencing continues to drop, genetic testing of *ex situ* collections still remains largely prohibitive due to cost and effort required. This research expands existing *ex situ* population sampling strategies, leveraging geographic provenance and environmental distance to increase functional and adaptive genetic differences conserved in collections. Incorporating our understanding of evolutionary and ecological processes influencing population structure for collections provides a low cost and timely method to meet conservation objectives, particularly critical for species at risk where limited genetic information may be available.

2.6. Acknowledgements

We thank the Hamilton Lab, T. L. Parchman, S. M. Hoban, and anonymous reviewers for their valuable comments on early versions of this article. This work was supported by a new faculty award from the office of the North Dakota Experimental Program to Stimulate Competitive Research (ND-EPSCoR NSF-IIA-1355466) and funding from the NDSU Environmental and Conservation Sciences Program to J. A. Hamilton.

3. SEED MORPHOLOGICAL TRAITS AS A TOOL TO QUANTIFY VARIATION MAINTAINED IN *EX SITU* COLLECTIONS: A CASE STUDY IN *PINUS TORREYANA*[2] (PARRY)

3.1. Abstract

Understanding the within- and among-population distribution of trait variation within seed collections may provide a means to approximate standing genetic variation and inform plant conservation. This study aimed to estimate population- and family-level seed trait variability for existing seed collections of Torrey pine (*Pinus torreyana*), and to use these data to guide sampling of future collections. We quantified variation in 14 seed morphological traits and seedling emergence within and among Torrey pine populations. Using a simulation-based approach, we used estimates of within-population variance to assess the number of maternal families required to capture 95% of trait variation within each existing seed collection. Substantial structure was observed both within and among Torrey pine populations, with island and mainland seeds varying in seed size and seed coat thickness. Despite morphological differences, seedling emergence was similar across populations. Simulations revealed that 83% and 71% of all maternal families within island and mainland seed collections respectively needed to be resampled to capture 95% of seed trait variation within existing collections. From a conservation perspective, our results indicate that to optimize genetic diversity captured in

[2] The material presented in this chapter was co-authored by Lionel N. Di Santo, Monica Polgar, Storm Nies, Paul Hodgkiss, Courtney A. Canning, Jessica W. Wright, and Jill A. Hamilton. Lionel Di Santo had primary responsibility for collecting the data, performing statistical analysis, and developing conclusions advanced here. Lionel Di Santo also drafted and revised all versions of this chapter. Monica Polgar contributed to collecting the data and conducting statistical analyses. Storm Nies, Paul Hodgkiss, and Courtney Canning contributed to collecting the data. Jessica Wright contributed to collecting the data, designing the research, and editing final versions of this chapter. Jill Hamilton contributed to collecting the data, designing the research, as well as writing and improving all versions of this chapter. A modified version of this chapter is published in *AoB PLANTS*: Di Santo, Lionel N, Monica Polgar, Storm Nies, Paul Hodgskiss, Courtney A Canning, Jessica W Wright, and Jill A Hamilton. 2021. "Seed Morphological Traits as a Tool to Quantify Variation Maintained in Ex Situ Collections: A Case Study in Pinus Torreyana (Parry)." *AoB PLANTS* 13 (5): plab058. https://doi.org/10.1093/aobpla/plab058.

Torrey pine seed collections, maximizing the number of maternal families sampled within each population will be necessary.

3.2. Introduction

Ex situ seed collections preserve species genetic diversity outside of their native range, providing the raw material for species reintroductions and germplasm to augment restoration (Potter et al. 2017; Guerrant Jr, Havens, and Vitt 2014). Ensuring *ex situ* collections represent genetic variation found in natural populations is critical to both contemporary conservation and potential future restoration efforts (Schaal and Leverich 2004; Basey, Fant, and Kramer 2015). An invaluable conservation resource, particularly for rare species, *ex situ* collections protect against biodiversity loss in the wild, while preserving species' evolutionary potential. However, the cost and logistical constraints associated with seed collection pose a significant challenge. Given this challenge, means are needed to optimize *ex situ* sampling efforts (Di Santo and Hamilton 2020; Hoban and Schlarbaum 2014).

One approach may be to use the distribution of trait variation existing within contemporary *ex situ* seed collections as a proxy for quantifying standing genetic variation within and among populations. Although multiple factors influence plant phenotypes (Monty et al. 2013; Villellas et al. 2014), seed morphological variation is often considered highly heritable. For example, seed length, seed width, and seed mass have a heritability (or repeatability) estimated between 0.33 and 0.98 in conifers, including maritime pine (*Pinus pinaster*), chir pine (*Pinus roxburghii*), and white spruce (*Picea glauca*) (Roy, Thapliyal, and Phartyal 2004; Carles et al. 2009; Zas and Sampedro 2015). In addition, traits such as seed shape, seed coat thickness, or embryo length also exhibit high heritability, with values estimated between 0.59 and 0.96 for agronomic species, including soybean (*Glycine max*), narrow-leafed lupin (*Lupinus*

angustifolius) and rice (*Oryza sativa*) (Cober, Voldeng, and Fregeau-Reid 1997; Hakim and Suyamto 2017; Pandey, Seshu, and Akbar 1994; Mera et al. 2004). Given these observations, variation in seed morphological traits likely has a genetic basis and may reflect standing genetic variation within and among populations. In addition, morphological variation of seeds stored *ex situ* may reflect variation attributable to the maternal environment (Platenkamp and Shaw 1993; Singh et al. 2017). However, for rare species where existing genetic data are limited, quantifying within and between population variation for traits largely considered heritable within existing seed collections may be invaluable to optimizing future collections, even if estimates of genetic variation do not control for maternal environment.

The distribution of heritable genetic variation estimated via common garden experiments – experimental approaches used to understand the genetic contribution to phenotypic variation under common environmental conditions – (Weber and Kolb 2014; J. A. Hamilton et al. 2017; Yoko et al. 2020) or molecular genetic data (Tamaki et al. 2018; Hausman et al. 2014; Zhang and Zhou 2013) can be used to quantify the distribution of standing genetic variation. However, when common garden experiments or molecular genetic data are unavailable, quantifying trait variation within and among *ex situ* seed population collections may provide a reasonable proxy for the distribution of genetic variation. Millions of seed accessions have been stored in gene banks internationally (FAO 2010), representing a large conservation and research resource. Although common garden experiments are preferred when available, heritability of seed morphological traits and ease of access to seeds through existing *ex situ* collections suggests that quantifying seed morphological variation may provide a timely approach to estimating variation preserved in collections. In addition, where the goal is to limit the loss of biodiversity and

preserve evolutionary potential for rare species, existing seed morphological data may be leveraged to optimize supplemental conservation collections.

Pinus torreyana Parry (Torrey pine), is one of the rarest pines in the world (Dusek 1985; Critchfield and Little 1966), endemic to two discrete natural populations in California. Torrey pine occupies one mainland population (*Pinus torreyana* subsp. *torreyana*) of approximately 6,000 trees at the Torrey Pine State Reserve in La Jolla, CA, and an island population (*Pinus torreyana* subsp. *insularis*) of approximately 3,000 reproductively mature trees on Santa Rosa Island, CA, one of the Channel Islands (J. Franklin and Santos 2011; J. R. Haller 1986; Hall and Brinkman 2015) (Figure 5). Listed as critically endangered by the IUCN (2020), Torrey pine is of critical conservation concern due to multiple factors, including low population size (J. Franklin and Santos 2011; Hall and Brinkman 2015), low genetic diversity (Ledig and Conkle 1983; Whittall et al. 2010), climate change, and environmental and human-mediated disturbances (J. Franklin and Santos 2011; J. A. Hamilton et al. 2017). While *in situ* conservation has preserved the whole of the species' range, with fewer than 10,000 reproductively mature individuals in native populations, there are substantial risks for population-level extirpation events. To mitigate potential losses in the wild, conservation efforts have focused on preservation of seeds *ex situ*. While *ex situ* seed collections provide an invaluable conservation resource, they may also be used to quantify species' trait variation needed to inform future conservation efforts.

In this study, we evaluate morphological trait variation in a large *ex situ* conservation collection of Torrey pine seed sourced from the two native extant populations. Specifically, we quantify the distribution of variation for 14 seed morphology traits and assess differences in emergence between island and mainland seedlings. In addition, we use existing *ex situ* collection

data to provide supplemental population sampling guidance for future Torrey pine collections. For this latter objective, we use simulations to estimate the number of maternal families required to capture 95% of seed morphological variation existing in contemporary *ex situ* collections, for both island and mainland population independently. This study evaluates the distribution of seed morphological variation in *ex situ* collections as a proxy for standing genetic diversity, quantifying variation attributable to within and between population differences. These data are then used to inform population sampling necessary to meet conservation objectives in future seed collections. Although presented here using Torrey pine, our approach is broadly applicable for *ex situ* collections within species with largely heritable seed trait variation.

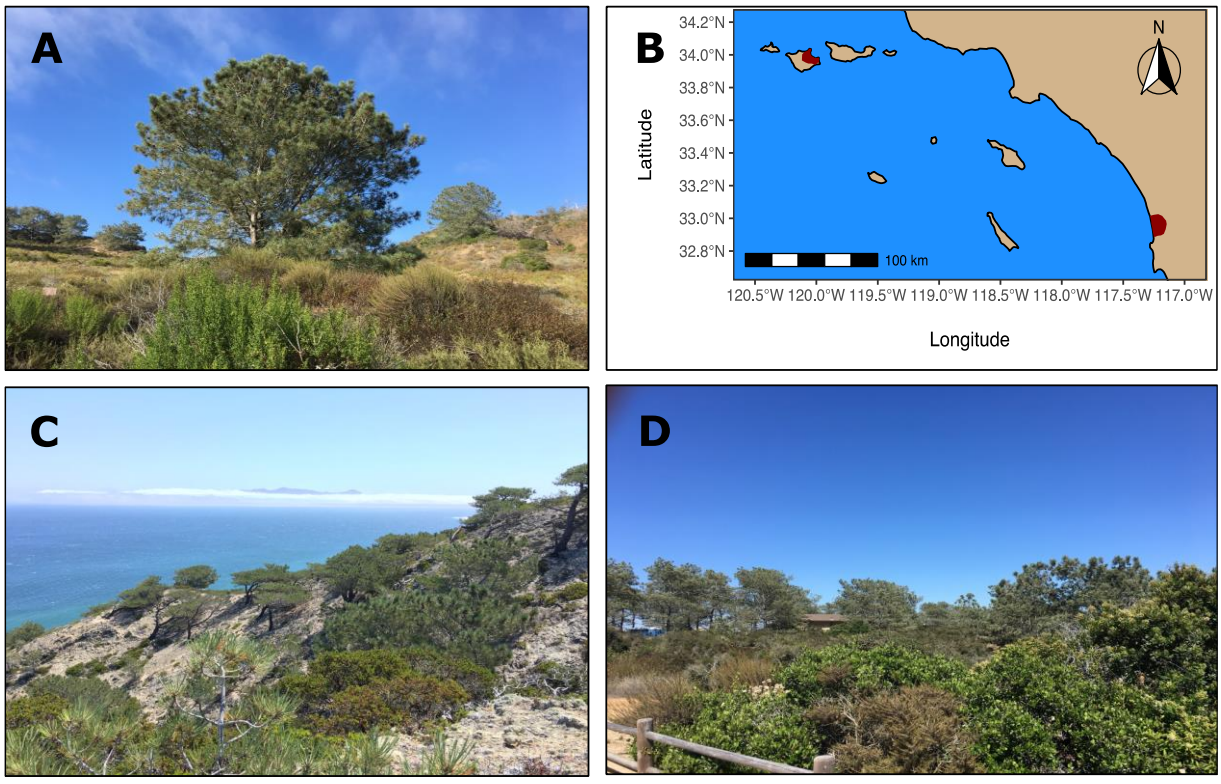


Figure 5. (A) *Pinus torreyana* individual. (B) *Pinus torreyana* distribution map, including Torrey pine distribution on Santa Rosa Island, CA (*Pinus torreyana* subsp. *Insularis*, top left red-shaded area) and at the Torrey Pine State Reserve, CA (*Pinus torreyana* subsp. *torreyana*, bottom right red-shaded area). (C) Torrey pine stand on Santa Rosa Island, CA. (D) Torrey pine stand at the Torrey Pine State Reserve, CA.

3.3. Materials and methods

3.3.1. Cone collection and seed processing

Mature, open-pollinated Torrey pine (*Pinus torreyana* Parry) cones were collected from native extant populations as part of a large *ex situ* conservation collection between June and July of 2017. Cones were collected from 157 trees on Santa Rosa Island (Channel Islands National Park), CA (island population) and 201 trees at the Torrey Pine State Reserve in La Jolla, CA (mainland population), representing the species' entire natural distribution (Figure 5; Figure C1). Where possible, we collected approximately five cones per maternal tree at each location. Sampling of reproductive maternal trees was evenly spaced; however, adjacent maternal trees were occasionally sampled to ensure enough cones were collected. The average pairwise distance between all maternal trees sampled was approximately 714 (range = 1 – 2,092) and 397 (range = 1 – 1,131) meters for the mainland and island population, respectively. Seeds, organized by individual maternal tree were then extracted from cones using a combination of mallet and pliers and processed for inclusion in a long-term *ex situ* conservation collection (see below).

3.3.2. Seed viability tests

Estimating viability of seeds preserved *ex situ* is necessary given their potential role in restoration, reforestation, or reintroduction. Given this, the potential viability of Torrey pine seeds was tested using two complementary approaches prior to inclusion in the final *ex situ* collection. A float test was first used as a rapid, low-cost approach to assess seed viability. Floating seeds were presumed to lack an endosperm or embryo, while seeds that sunk were presumed filled. Seeds were dropped into water for approximately 15 seconds to differentiate presumed non-viable, floating seeds from presumed viable, sinking seeds (Gribko and Jones 1995; Morina et al. 2017). Those seeds classed as likely viable were organized by maternal tree

using paper bags, and then placed in a Blue M drying oven (Thermal Product Solutions, White Deer, Pennsylvania, USA) maintained at 37°C for 24 hours to remove potential surface moisture. Following this, seeds from a haphazard sample of maternal families were x-rayed at the Placerville Nursery, CA. In addition to visualizing seed morphological variation, x-ray radiographs were used to verify viability based on float tests. Acrylic seed trays [20.3 cm x 25.4 cm x 0.48 cm], with a 9 x 11 array of wells, were used to separate and position each Torrey pine seed over the x-ray film. Kodak x-OMAT HBT film (20.3 cm x 25.4 cm) was placed in a lightproof x-ray film cassette which was positioned in the x-ray machine with the seed tray centered on top of the film, with a shelf height of 55.9 cm. The x-ray was taken using a 17 kVP exposure for a total of two minutes, based on standardized conditions established previously for *Pinus coulteri* (Sara Wilson, USDA Forest Service, pers. comm.). X-ray radiographs were digitized using a Nikon D40 digital camera mounted on a tripod over a light box.

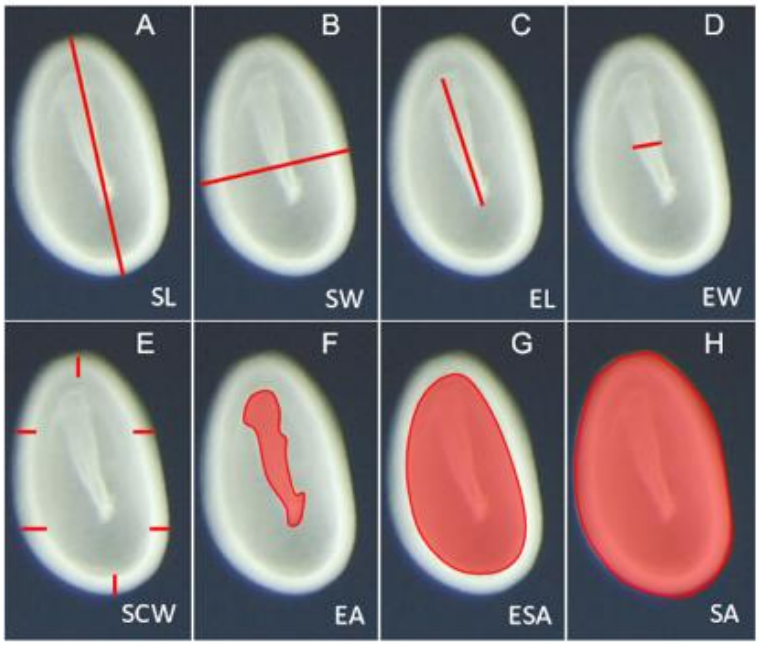


Figure 6. Visual of morphological measurements taken using ImageJ for seeds collected on Santa Rosa Island and at the Torrey pine State Reserve. (A) Seed length [cm]. (B) Seed width [cm]. (C) Embryo length [cm]. (D) Embryo width [cm]. (E) Seed coat width [cm]. (F) Embryo area [cm^2]. (G) Endosperm area [cm^2]. (H) Seed area [cm^2].

3.3.3. Morphological measurement of seed traits

Using ImageJ (Abràmoff, Magalhães, and Ram 2004), eight seed morphological traits were measured across 80 mainland maternal families and 30 island maternal families, representing a haphazard subset of the complete collection (Figure 6; Table 3). Although a balanced design is ideal, logistical constraints limited the number of island maternal families assessed. To evaluate the influence of an unbalanced design we compared analyses using all 80 mainland and 30 island maternal families with 30 island and a random subset of 30 mainland maternal families. Overall, we found that while equalizing sample sizes across populations impacted quantitative results, qualitative trends and patterns generally remained similar regardless of the number of maternal families used (Table C1, Table C2, Figure C2, Figure C3, and Figure C4). Given these observations, a sample size of 80 mainland and 30 island trees were retained for subsequent analyses, as it is most likely to provide a better estimation of morphological trait variation within populations, ultimately improving conservation conclusions for the species. Each x-ray radiograph was scaled using the diameter of a seed tray well (1.87 cm) to express pixels as trait values in centimeters. Directly measured seed traits included seed length (SL, cm), seed width (SW, cm), embryo length (EL, cm), embryo width (EW, cm), seed coat width (SCW, cm), seed area (SA, cm^2), endosperm area (ESA, cm^2), and embryo area (EA, cm^2). We selected these traits as they can readily be measured from x-ray radiographs of seeds and provide a ubiquitous means to evaluate morphological variation for plants preserved *ex situ*. Using measured seed traits, six additional traits were derived (Table 3), including seed length/width ratio (SLW), embryo length/width ratio (ELW), relative embryo size (RES), relative endosperm size (REndS), seed coat area (SCA, cm^2), and relative seed coat size (RSCS).

Table 3. Morphological measurements of Torrey pine seeds sourced from Santa Rosa Island (Island, n=30) and Torrey Pine State Reserve (Mainland, n=80), CA. Listed are population mean estimates (±SE) of measured (A) and derived (B) seed traits summarized by maternal families. Measurable traits: seed length (SL), seed width (SW), embryo length (EL), embryo width (EW), seed coat width (SCW), seed area (SA), endosperm area (ESA), and embryo area (EA). Derived traits: seed length/width ratio (SLW), embryo length/width ratio (ELW), relative embryo size (RES), relative endosperm size (REndS), seed coat area (SCA), and relative seed coat size (RSCS). Differences in seed morphology between mainland and island populations were significant (α=0.05) for all 14 seed traits.

A – Measured seed traits

Population	SL (cm)	SW (cm)	SCW (cm)	EL (cm)	EW (cm)	SA (cm^2)	ESA (cm^2)	EA (cm^2)
Mainland	1.415 ±0.016	0.768 ±0.009	0.089 ±0.002	1.082 ±0.012	0.155 ±0.003	0.887 ±0.023	0.548 ±0.014	0.169 ±0.004
Island	1.674 ±0.020	0.966 ±0.015	0.117 ±0.002	1.198 ±0.020	0.186 ±0.004	1.410 ±0.031	0.807 ±0.021	0.232 ±0.007

B – Derived seed traits

Population	SLW (SL/SW)	ELW (EL/EW)	SCA (cm^2) (SA-ESA)	RES (EA/ESA)	REndS (ESA/SA)	RSCS (SCA/SA)
Mainland	1.857 ±0.013	7.275 ±0.094	0.339 ±0.010	0.313 ±0.005	0.619 ±0.003	0.381 ±0.003
Island	1.769 ±0.037	6.640 ±0.136	0.603 ±0.013	0.287 ±0.004	0.571 ±0.004	0.429 ±0.004

These traits were derived as they provide a means to relate different morphological traits to each other and can provide a fine-scale estimate of the relative contribution of growth and size traits within individual seeds. We measured five randomly selected seeds per maternal tree, including three technical replicates per seed for each trait (the same seed was measured three times for any given morphological trait). Measurements were averaged across technical replicates to summarize the mean trait value per seed. In total, 550 seeds were measured from across 110 maternal trees spanning the two Torrey pine populations.

3.3.4. Seedling emergence test

Within a restoration or reintroduction context, concurrent seedling emergence is often preferred for nursery plantings. To evaluate the timing and probability of seedling emergence within Torrey pine populations, as well as test the viability of seeds in population collections, a trial was conducted in January 2018 using a random subset of seeds preserved *ex situ*, including seeds sourced from Torrey Pine State Reserve and Santa Rosa Island, CA. Following x-ray, seeds were stored at 4°C in sealed mylar bags (USA emergency supply, Beaverton, Oregon, USA) placed in plastic boxes; each box contained desiccant crystals to decrease ambient moisture and reduce likelihood of mold. Seeds from eight maternal families per population were selected for the emergence trial. Between eight to ten seeds per maternal tree were weighed and then stratified under cold, moist conditions for 30 days (placed in plastic boxes on a moist paper towel at 4°C). Seeds were sown directly into a 164 mL Ray Leach "Cone-tainer" ™ (Stuewe & Sons, Tangent, Oregon, USA) filled with Sunshine® Mix #4 (Sungro horticulture, Agawam, Massachusetts, USA), pressed halfway into the soil, and then covered with a thin layer of gravel. For approximately one month following planting, seeds were misted for one minute at hourly intervals over a daily eight-hour period (9am – 4pm). Following emergence, seedlings were hand

watered to saturation weekly to biweekly as needed. Emergence was quantified across three separate timepoints as the proportion of seeds per maternal family that successfully developed into living seedlings from the total initially planted. Timepoints included Feb 06/2018 (32 days since sowing), Feb 16/2018 (42 days since sowing), and Mar 07/2018 (61 days since sowing).

3.3.5. Evaluating the distribution of seed trait variation

We conducted a principal component analysis (PCA) using all 14 measured and derived seed traits averaged by maternal family to evaluate population-specific differentiation in seed morphology. Prior to performing the PCA, to account for differences in measurement units, all seed traits were standardized using the *scale()* function in R implementing the z-score standardization: $z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$ (1), where x_{ij} is the non-transformed trait value, μ_j is the mean of a given seed trait across populations, and σ_j the standard deviation of the same seed trait across populations. Subsequently, to test for seed trait differences between population means, we used either Student's two-sample test or its non-parametric equivalent when normality was not met, Wilcoxon's two-sample test, within the R package "exactRankTests" (Hothorn and Hornik 2021). Normality was assessed using Shapiro-Wilk's test of normality within each population. In total, four of the fourteen traits were distributed normally in both Torrey pine populations, including seed width (mainland: W = 0.97, P = 0.06; island: W = 0.97, P = 0.52), embryo length (mainland: W = 0.98, P = 0.29; island: W = 0.95, P = 0.21), embryo width (mainland: W = 0.97, P = 0.09; island: W = 0.93, P = 0.05), and embryo area (mainland: W = 0.98, P = 0.45; island: W = 0.96, P = 0.26).

To evaluate the distribution of morphological trait variation within and between Torrey pine populations, we quantified the proportion of variation attributed to population and maternal tree families using measured and derived morphological traits summarized by seeds. For each

trait, we fit a linear mixed model using the R package "lme4" (Bates et al. 2015) with population considered a fixed effect and maternal families within populations considered a nested random effect: $Y_{ij} = \mu + \pi_i + r_{i/j} + e_{ij}$, where Y_{ij} is the observed seed trait value, μ is the seed trait overall mean, π_i is the effect of population origin on the seed trait mean, $r_{i/j}$ is the effect of maternal family within populations on the observed seed trait value, and e_{ij} are the effects on the seed trait value of any other variables unaccounted for in the model (residual error). For each model, normality of residual errors was visually assessed and significance of fixed- and random-effect terms was tested using the functions *anova()* and *ranova()* respectively, implemented in the R package "stats" (R Core Team 2020) and "lmerTest" (Kuznetsova, Brockhoff, and Christensen 2017). Proportions of seed morphological variance explained by populations (marginal R^2, R^2_m), both populations and maternal families (conditional R^2, R^2_c), and maternal families alone ($R^2_c - R^2_m$) were quantified for each model independently using the function *r.squaredGLMM()* implemented in the R package "MuMIn" (Bartoń 2020).

3.3.6. Assessing differences in seedling emergence across populations

To test for differences in the probability and the timing of seedling emergence in Torrey pine, we evaluated the proportion of seeds that produced seedlings both within and between populations across timepoints. First, we used Friedman's rank sum test (non-parametric repeated measures ANOVA) followed by Wilcoxon' paired two-sample test, both implemented in the R package "rstatix" (Kassambara 2020), to assess differences in the proportion of emerged seedlings between timepoints within populations. We used a non-parametric approach for both Torrey pine populations because normality could not be assumed at select timepoints due to high frequency of zero values. We accounted for multiple testing using Benjamini and Hochberg (1995)'s False Discovery Rate (FDR) correction implemented in the *wilcox_test()* function.

Following this, we evaluated timepoint-specific population differences in seedling emergence. We used Shapiro-Wilk's test to assess populations' deviation from normality at each timepoint and either Student's (for timepoints passing the normality test) or Wilcoxon's two-sample test (for timepoints failing the normality test) to evaluate differences in population emergence. Timepoints Feb 16/2018 – 42 days since sowing – (mainland: W = 0.94, P = 0.57; island: W = 0.9, P = 0.28) and Mar 07/2018 – 61 days since sowing – (mainland: W = 0.96, P = 0.85; island: W = 0.88, P = 0.21) passed the normality test, while timepoint Feb 06/2018 – 32 days since sowing – (mainland: W = 0.52, P < 0.001; island: W = 0.73, P = 0.004) failed the normality test. All statistical analyses were performed using R version 4.0.2 and 4.0.5 (R Core Team 2020; 2021).

3.3.7. Simulating variation captured in the *ex situ* collection using seed morphological traits

For each of the 14 measured and derived seed traits, we conducted a separate simulation quantifying morphological variation captured when increasing the number of maternal families sampled from contemporary Torrey pine seed collections. Simulations were conducted in R version 3.6.3 (R Core Team 2020) using a customized script (Figure C5). Resampling of *ex situ* collections were performed for island and mainland Torrey pine populations independently, using between one and the total number of maternal families available within each *ex situ* population collection (mainland: 80 maternal families, island: 30 maternal families) (N_{fam}). Maternal trees were sampled randomly without replacement from the pool of available families. All seeds within each selected maternal family were sampled as part of this simulation, except those with missing values for the trait simulated. Overall, between two to five seeds per maternal family were sampled within each population.

To evaluate the number of maternal families needed to capture 95% of seed trait variation in both island and mainland populations, we estimated the number of unique seed trait values captured in a sample of N_{fam} maternal families (N_c) relative to the total number of unique seed trait values present in a seed population (N_t). Here, we define "unique seed trait values" as the number of non-redundant standardized measurements for the seed trait simulated rounded to the first digit. Seed morphological measurements were rounded to the first digits as we believe that seed trait variation estimated using additional digits is more likely to fail to capture meaningful biological variation. Standardization of the data was performed so that all seed traits share the same unit (the number of standard deviations a value is from the overall trait mean across populations, see equation (1) above) and become comparable. Sampling of maternal families and estimation of the summary statistic, defined as the proportion of total seed trait variance captured (N_c/N_t), were repeated 500 times for each seed morphological trait and Torrey pine population. In this way, N_c/N_t accounts for potential variation in number of seeds sampled per maternal family or variation in maternal families included.

Finally, for each number of maternal trees sampled (N_{fam}), we averaged the summary statistic across all 500 replicates. This process was repeated for each of the 14 seed morphological traits and performed for each Torrey pine population separately. Following this, the summary statistic was averaged across all seed traits and separated by populations (see Results below). Proportions of total seed trait variance captured (N_c/N_t) are provided based on proportions of maternal families sampled (instead of the number of maternal families sampled) as sample sizes varied across Torrey pine populations.

3.4. Results

3.4.1. Island-mainland differentiation in seed morphology

A principal component analysis (PCA) using all 14 measured and derived seed traits averaged by maternal family revealed substantial differences in seed morphology between island and mainland populations of Torrey pine (Figure 7). The first PC axis explained 57.8% of variation in seed morphological traits, primarily separating the island from the mainland population. Seed length, seed width, seed area, endosperm area, and seed coat area exhibited the five highest loadings (absolute values) on PC1 (Table C3), indicating that seed size and seed coat thickness can largely discriminate island from mainland individuals. On average, seeds collected on island trees were longer, wider, larger, and thicker than seeds collected on mainland trees (Table 3). The second PC axis explained 15.9% of seed trait variation and summarizes within population variability in seed morphology (Figure 7). Relative seed coat size, relative endosperm size, and relative embryo size had the three highest loadings (absolute values) on PC2 (Table C3). This suggests that once corrected for seed size, seed coat thickness, endosperm size, and embryo size are traits contributing to within-population variation.

3.4.2. Contribution of population origin and maternal family to seed trait variation

Consistent with our principal component analysis, linear mixed models constructed for each of the 14 measured and derived seed traits demonstrated that considerable variation in seed morphology in Torrey pine is explained by population origin (Figure 8). On average, population origin explained 23% (range = 0.02–0.57) of variation across the species' distribution (Table C4). Traits associated with seed size and seed coat thickness exhibited the highest proportion of variance explained by population origin. These include seed coat area (0.57; $F_{1,107.60} = 221.91$, $P < 0.001$), seed area (0.49; $F_{1,107.56} = 156.45$, $P < 0.001$), endosperm area (0.37; $F_{1,108.07} = 100.58$,

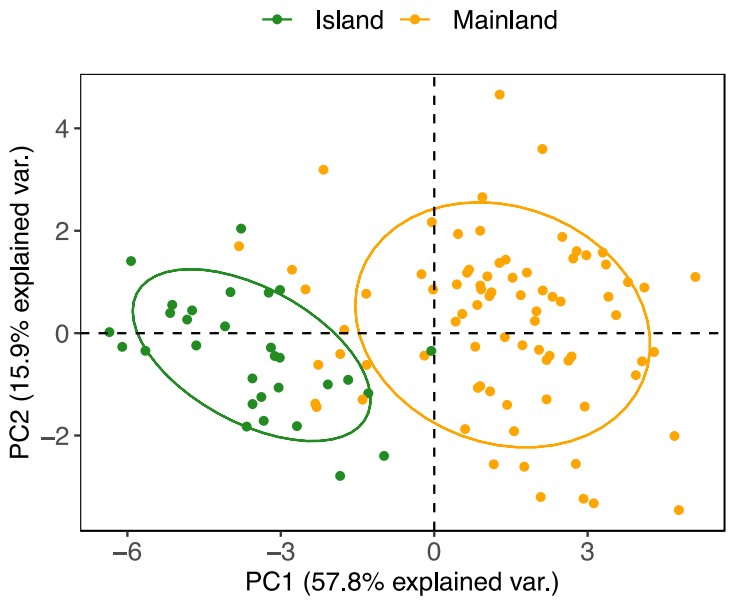


Figure 7. Principal components analysis (PCA) using all 14 seed morphological traits measured and derived from maternal plants collected on Santa Rosa Island (green) and at the Torrey Pine State Reserve (orange).

P < 0.001), seed width (0.36; $F_{1,108}$ = 126.04, P < 0.001), seed coat width (0.32; $F_{1,108}$ = 96.04, P < 0.001), and seed length (0.30; $F_{1,108.50}$ = 78.92, P < 0.001). Overall, this suggests seed size and seed coat thickness are major discriminants of island and mainland Torrey pine seeds.

While population origin explained substantial variation across populations, assessment of maternal seed families within populations indicated substantial family structure to seed trait variation (Figure 8). On average, maternal seed family explained 24% (range = 0.07–0.37) of variation within populations (Table C4). Embryo length (0.37; χ^2 = 124.82, df = 1, P < 0.001), seed length (0.35; χ^2 = 180.72, df = 1, P < 0.001), endosperm area (0.34; χ^2 = 211.87, df = 1, P < 0.001), seed area (0.31; χ^2 = 256.71, df = 1, P < 0.001), embryo area (0.29; χ^2 = 100.16, df = 1, P < 0.001), and seed coat width (0.28; χ^2 = 126.93, df = 1, P < 0.001) exhibited the highest proportion of seed trait variation explained by within-population maternal families. This suggests that there is substantial family-level structure to seed size, endosperm size, embryo size, and seed coat thickness within Torrey pine populations.

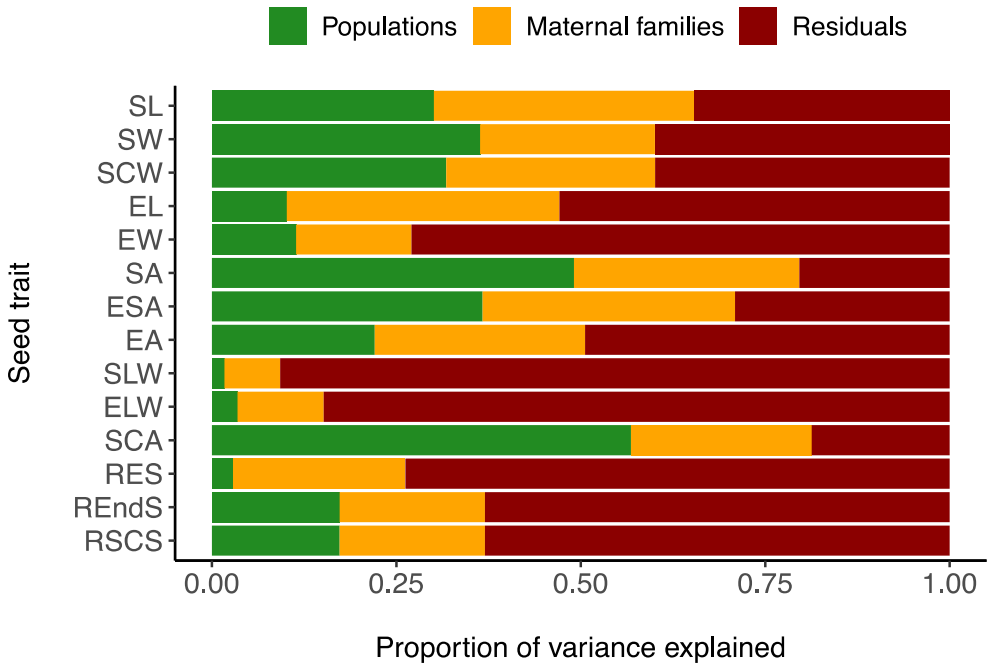


Figure 8. Proportion of variance in seed morphology explained by populations (green), maternal families within populations (orange), and other variables not accounted for in the model (residuals; dark red) for each of the 14 measured and derived seed traits. SL, seed length [cm]; SW, seed width [cm]; SCW, seed coat width [cm]; EL, embryo length [cm]; EW, embryo width [cm]; SA, seed area [cm^2]; ESA, endosperm area [cm^2]; EA, embryo area [cm^2]; SLW, seed length/width ratio; ELW, embryo length/width ratio; SCA, seed coat area [cm^2]; RES, relative embryo size; REndS, relative endosperm size; RSCS, relative seed coat size. See Table C4 for numerical estimates.

3.4.3. Impact of population seed trait differentiation on seedling emergence

The proportion of emerged seedlings increased over time for both island ($Q = 15.5$, $df = 2$, $P < 0.001$) and mainland ($Q = 15.2$, $df = 2$, $P < 0.001$) populations (Figure 9). However, we found no significant differences in the proportion of individuals emerging between populations across observed timepoints. On average, 7% and 9% of mainland and island seedlings had emerged 32 days after sowing (Feb 06/2018; $W = 28$, $P = 0.64$), 63% and 53% of mainland and island seedlings had emerged 42 days after sowing (Feb 16/2018; $t = 0.81$, $df = 14$, $P = 0.43$), and 78% of mainland and island seedlings had emerged 61 days after sowing (Mar 07/2018; $t = -0.06$, $df = 14$, $P = 0.95$). Overall, this indicates that under controlled conditions, timing and

probability of emergence may not be impacted by population differences in seed morphology for Torrey pine seedlings.

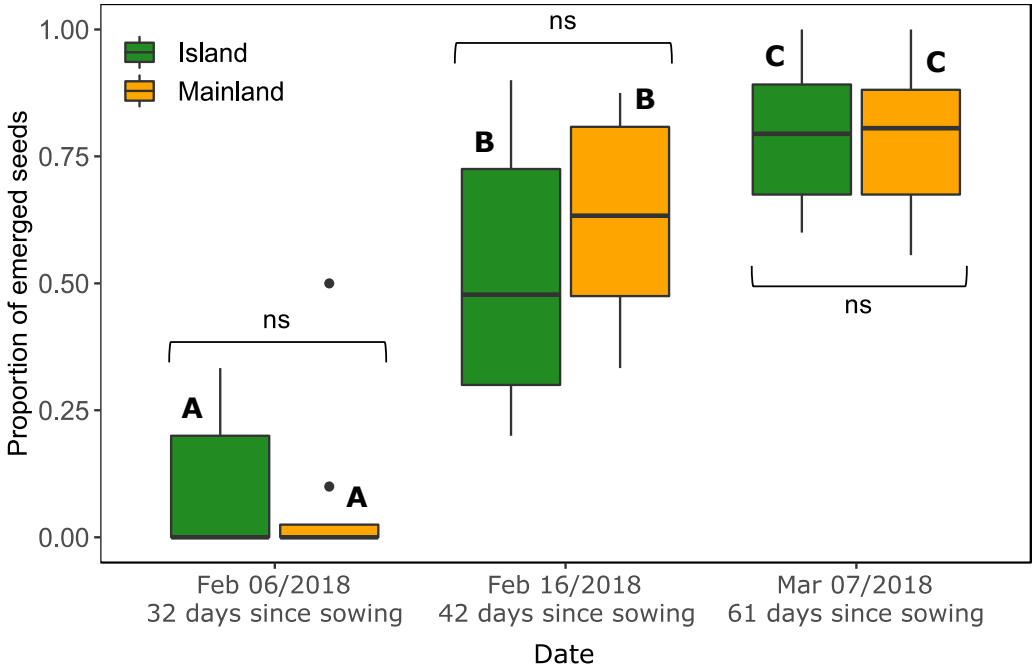


Figure 9. Proportion of emerged seedlings recorded at three different timepoints for seeds sampled on Santa Rosa Island (green) and at the Torrey Pine State Reserve (orange). Significant differences in emergence time across timepoints within populations are indicated with different capital letters [A, B, C]. Comparisons between populations at each timepoint is indicated with square brackets. ns, non-significant difference (α=0.05).

3.4.4. Morphological variation captured in simulated seed collections

Simulations revealed that to capture 95% of seed trait variation present in our existing e*x situ* collections, on average 83% (25 out of 30) and 71% (57 out of 80) of all island and mainland families would need to be resampled, respectively (Figure 10). This indicates that both island and mainland populations harbor considerable within-population structure for seed morphological traits. Interestingly, capturing equal morphological variation across seed collections always required a higher proportion of island maternal families to be collected relative to the mainland population.

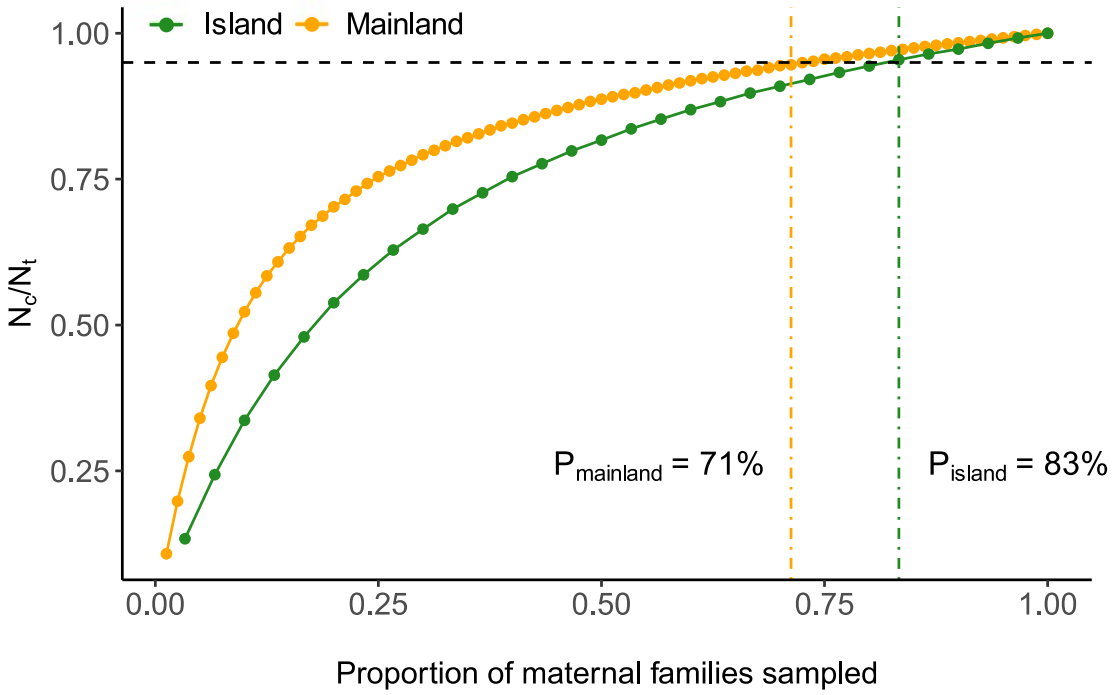


Figure 10. Phenotypic variation captured across seed traits in simulated collections (N_c) relative to total phenotypic variation present in seed populations (N_t). Average proportion of phenotypic variation captured (N_c/N_t) was estimated for various proportions of maternal families sampled. P_{island} and $P_{mainland}$ represent the proportion of maternal families required to capture 95% of morphological variation (horizontal dashed line) present in island (green) and mainland (orange) *ex situ* seed populations, respectively.

3.5. Discussion

Evaluating between- and within-population morphological variation in contemporary *ex situ* seed collections may aid in understanding the distribution of variation needed to guide future conservation efforts. Here, we quantified the distribution of trait variation within an *ex situ* collection of Torrey pine seeds, with an aim to optimize future supplemental collections. Morphologically, island and mainland seeds were significantly different from each other. Island seeds were larger on average with thicker seed coats relative to their mainland counterparts. These morphological differences may be explained by a combination of stochastic and deterministic factors associated with population origin, including genetic bottlenecks following colonization, genetic drift, and selection associated with unique biotic pressures. Interestingly,

despite substantial morphological differentiation, seedling emergence did not vary among populations, suggesting that either the probability and timing of emergence under controlled conditions is not impacted by differences in seed morphology, or that island and mainland seeds respond similarly to an artificial germination protocol. In addition to population origin, a considerable proportion of seed trait variation within each population was explained by maternal family. This suggests that there is substantial within-population variation that will be important to conserve and maintain populations' evolutionary potential. Finally, simulations demonstrated that 83% and 71% of all maternal families within island and mainland seed collection, respectively, were necessary to capture 95% of morphological variation within the existing collections. To ensure we maintain representative levels of trait variation in future seed collections, our results suggest that the number of maternal families sampled within natural populations should be maximized, with the island population potentially requiring additional sampling efforts.

Species distributed between island and mainland origins often exhibit marked among-population phenotypic differentiation, including differentiation in seed morphology (Burns, Herold, and Wallace 2012; Lens et al. 2013; Kavanagh and Burns 2014; Burns 2016). Our results revealed considerable seed morphological differences between Torrey pine populations (Figure 7), primarily in seed size and seed coat thickness (Figure 8; Table C3). On average, islands seeds were larger and had thicker seed coats than seeds collected on the mainland (Table 3). These results are consistent with previous studies of island-mainland systems that noted island populations exhibited larger seeds relative to mainland populations (Kavanagh and Burns 2014; Burns 2016; Biddick, Hendriks, and Burns 2019). A combination of different factors could

contribute to morphological variation among seed populations, including both stochastic and deterministic forces.

On islands, seed traits associated with long-distance dispersal may be selected against as they can increase the probability an individual would disperse beyond an island's limits (Cody and Overton 1996; Kavanagh and Burns 2014; Ottaviani et al. 2020, but see Burns 2018). For Torrey pine, increased seed size on the island may have evolved to limit potential seed losses via wind-dispersal, as seed mass negatively correlates with dispersal distance in pines (Greene and Johnson 1993; Debain, Curt, and Lepart 2003, but see Wyse and Hulme 2021). Nonetheless, Torrey pine seeds possess degenerated wings (Ledig and Conkle 1983), suggesting that other mechanisms likely contribute to seed dispersal in this species. Rodents and birds both feed on Torrey pine, suggesting that seeds may undergo animal-mediated dispersal (M. Johnson, Vander Wall, and Borchert 2003). Thus, seed predation may contribute to differences in seed size observed between populations. On the island, *Peromyscus maniculatus* (Deer mouse) is the only rodent present to predate on Torrey pine seeds (M. Johnson, Vander Wall, and Borchert 2003). This contrasts with the mainland, where multiple seed predators have been documented, including *Peromyscus boylei* (Brush mice), *Peromyscus maniculatus* (Deer mice), *Peromyscus eremicus* (Cactus mice), *Chaetodipus californicus* (California pocket mice), *Spermophilus beecheyi* (California ground squirrels), or *Aphelocoma californica* (Scrub jays) (M. Johnson, Vander Wall, and Borchert 2003). If large seeds are preferentially targeted by seed predators (Reader 1993; Gómez 2004), reduced seed size on the mainland may have evolved as a consequence of the trade-off between attracting predators to promote seed dispersal and mitigating fitness loss due to seed consumption.

While selection may contribute to population differences, differentiation in seed morphology may result from stochastic evolutionary forces. Founder effects associated with the colonization of Santa Rosa island by mainland individuals, and genetic drift in the face of limited gene flow, may have led to morphological differentiation between Torrey pine populations (Ledig and Conkle 1983). Alternatively, more complex demographic histories of the two populations, including colonization, extinction, and recolonization events, or divergence of island and mainland populations from an ancestral population following tectonic movement may have led to the differences observed between populations (J. R. Haller 1986; Di Santo, Hoban, et al. 2021). While both stochastic and deterministic factors may contribute to population differences in seed morphology, additional experiments are required to test mechanistic hypotheses. Seeds evaluated in this manuscript were collected from natural populations. To tease apart the contribution of environment and genetics to seed trait differences observed among populations, a common garden experiment is required. Furthermore, a reciprocal transplant experiment would be the most effective test of the action of natural selection in shaping morphological differences between island and mainland seeds.

Despite significant differences in seed morphology between populations, timing and probability of emergence was similar across populations (Figure 9). Emergence rates were high throughout the trial, with 78% of island and mainland seedlings emerging 61 days after sowing. The absence of differences in seedling emergence between populations was surprising, as seed size often negatively correlates with time to germination (Daws, Garwood, and Pritchard 2005; Tanveer et al. 2013). However, seed coat thickness can also influence rates of emergence. Seeds with thick seed coats relative to their mass often germinate later than seeds with thinner seed coats (Daws, Garwood, and Pritchard 2005; K. N. Hamilton et al. 2013). For Torrey pine, J. A.

Hamilton et al. (2017) found that island seeds germinate on average two days after mainland seeds. Interestingly, island seeds were not only larger, but also had thicker seed coats relative to mainland seeds (Table 3). Even after correcting for differences in seed size, seed coat thickness (relative seed coat size; RSCS) remained moderately higher in island seeds. Together, these results predict that island seedlings should emerge at similar or later timepoints relative to mainland seedlings, which is consistent with current and previous observations.

Similar emergence rates may also result from our experimental design. Abe and Matsunaga (2011), in a mainland-island comparison study, observed that cold stratification attenuates differences in germination rates between populations of *Rhaphiolepis umbellata*. Additionally, complete and rapid germination of pine seeds is generally observed when pretreated under cold and moist conditions (Krugman and Jenkinson 2008). Overall, this suggests that cold stratification may mask population-specific differences in seedling emergence. Concurrent seedling emergence from both Torrey pine populations coupled with high emergence success suggests a cold stratification protocol is valuable for Torrey pine, particularly where simultaneous emergence for nursery-grown seedlings is desired. Note, however, that variation in the proportion of emerged seedlings within populations across timepoints may have concealed population-specific differences in emergence rates. Consequently, weak differences in the timing and probability of seedling emergence observed between island and mainland populations may be an artifact of small numbers of seeds and maternal families used during emergence trials.

Although population origin explained a substantial proportion of seed trait variation, linear mixed models demonstrated that maternal seed families within populations explained as much variation (Figure 8; Table C4). Given generally high heritability for seed morphological traits and the half-sib design of our collection (Cober, Voldeng, and Fregeau-Reid 1997; Hakim

and Suyamto 2017; Mera et al. 2004; Pandey, Seshu, and Akbar 1994; Roy, Thapliyal, and Phartyal 2004; Carles et al. 2009; Zas and Sampedro 2015), family-level seed trait variation likely provides a useful proxy for assessing within-population genetic diversity. With nearly 25% of variation explained on average by maternal families (Table C4), this suggests there is substantial genetic structure within Torrey pine populations. These results were notable as previous studies using allozymes and chloroplast DNA indicated that the species exhibits little to no within-population genetic variability (Ledig and Conkle 1983; Waters and Schaal 1991; Whittall et al. 2010). However, the common garden experiment initiated by Thomas Ledig indicated substantial family-level variation in tree height within both island and mainland populations (J. A. Hamilton et al. 2017). Overall, these results indicate that Torrey pine populations may possess within-population genetic variation necessary for natural selection to act upon. From a conservation perspective, these findings suggest that a strategy maximizing the number of maternal families sampled would optimize genetic diversity preserved in future *ex situ* seed collections and increased distance among individuals may limit relatedness among maternal trees.

Generally, *ex situ* seed collections aim to capture 95% of genetic diversity present throughout a species' distribution (Brown and Marshall 1995; Marshall and Brown 1975; Gapare, Yanchuk, and Aitken 2008; Q. Li, Xu, and He 2002). Simulations revealed that, to capture 95% of morphological variation currently maintained *ex situ*, 25 (83% of island collection) and 57 (71% of mainland collection) maternal families within each seed collection would need to be sampled (Figure 10). These data indicate that sampling more maternal families from the island population may be necessary to achieve the same level of representation of morphological variation. Assuming increased phenotypic variation observed on the island

results from higher allelic diversity, capturing 95% of genetic variation within the island population will always require more maternal families relative to the mainland population. For these simulations, we assumed that contemporary *ex situ* collections captured all morphological variation both within and between populations, including seed phenotype frequencies. However, if this is not the case, these conclusions may result in suboptimal sampling of standing variation within targeted populations. This caveat is important because the number of x-rayed maternal families differed between island (30 maternal families) and mainland (80 maternal families) seed collections. To address this caveat, it will be important to have a general understanding of the fraction of natural morphological variation captured across *ex situ* seed populations and adapt sampling efforts accordingly.

Practical and cost-effective, long-term storage of seeds *ex situ* is widely used to capture and maintain rare species genetic diversity. These seed collections represent an invaluable resource to quantify within and between population trait variation that may be used to guide future *ex situ* sampling efforts. Using Torrey pine as a model, we demonstrate that incorporating existing information from *ex situ* collections offers a unique opportunity to monitor and optimize conservation objectives, particularly important for rare species. While our results and conclusions may be specific to Torrey pine, the empirical, statistical, and simulation-based approaches presented here are broadly applicable to heritable traits across *ex situ* seed collections. Nonetheless, although our approach offers a powerful tool to guide and inform conservation decisions across a wide variety of plant taxa, its applicability to other species needs to be tested.

3.6. Acknowledgements

The authors thank Annette Delfino Mix, Valerie Gallup, Andrew Bower, Drew Peterson, Emma Ordemann, Conner Harrington, Forest Swaciak, Jill Wulf and Stephen Johnson for their help collecting the Torrey pine cones. Sara Wilson provided training and assistance for the use of the x-ray machine. We also thank the USDA Forest Service (Pacific Southwest Research Station) for their support of this project, as well as the Michumash and the Kumeyaay people as the traditional caretakers of the Torrey pine ecosystems sampled for this study. Finally, we thank Sean Hoban from the Morton Arboretum Center for Tree Science for helpful discussions during the writing of this manuscript. The collections were funded by the USDA Forest Service, State and Private Forestry, Gene Conservation program to J.W.W. and J.A.H. This work was also supported by a new faculty award from the office of the North Dakota Experimental Program to Stimulate Competitive Research (ND-EPSCoR NSF-IIA-1355466), the NDSU Environmental and Conservation Sciences Program, and the Schatz Center for Tree Molecular Genetics to J.A.H. Any use of product names is for information purposes only and does not imply endorsement by the US Government.

4. REDUCED REPRESENTATION SEQUENCING TO UNDERSTAND THE EVOLUTIONARY HISTORY OF TORREY PINE (*PINUS TORREYANA* PARRY) WITH IMPLICATIONS FOR RARE SPECIES CONSERVATION[3]

4.1. Abstract

Understanding the contribution of neutral and adaptive evolutionary processes to population differences is often necessary for better informed management and conservation of rare species. In this study, we focused on *Pinus torreyana* Parry (Torrey pine), one of the world's rarest pines, endemic to one island and one mainland population in California. Small population size, low genetic diversity, and susceptibility to abiotic and biotic stresses suggest Torrey pine may benefit from inter-population genetic rescue to preserve the species' evolutionary potential. We leveraged reduced representation sequencing to tease apart the respective contributions of stochastic and deterministic evolutionary processes to population differentiation. We applied these data to model spatial and temporal demographic changes in effective population sizes and genetic connectivity, to assess loci possibly under selection, and evaluate genetic rescue as a potential conservation strategy. Overall, we observed exceedingly low standing variation reflecting consistently low effective population sizes across time and limited genetic differentiation suggesting maintenance of gene flow following divergence. However, genome scans identified more than 2000 SNPs candidates for divergent selection. Combined with

[3] The material presented in this chapter was co-authored by Lionel N. Di Santo, Sean Hoban, Thomas L. Parchman, Jessica W. Wright, and Jill A. Hamilton. Lionel Di Santo had primary responsibility for collecting the data, designing the study, performing statistical analysis, and developing conclusions advanced here. Lionel Di Santo also drafted and revised all versions of this chapter. Sean Hoban and Thomas Parchman contributed to analyzing the data and editing final versions of this chapter. Jessica Wright contributed to collecting the data, designing the study, and editing final versions of this chapter. Jill Hamilton contributed to collecting the data, designing the study, as well as writing and improving all versions of this chapter. A modified version of this chapter is available on BioRxiv: Di Santo, Lionel N, Sean Hoban, Thomas L Parchman, Jessica W Wright, and Jill A Hamilton. 2021. "Reduced Representation Sequencing to Understand the Evolutionary History of Torrey Pine (Pinus Torreyana Parry) with Implications for Rare Species Conservation." *BioRxiv*. https://doi.org/10.1101/2021.07.02.450939.

previous observations indicating population phenotypic differentiation, this indicates that natural selection has likely contributed to population genetic differences. Thus, while reduced genetic diversity, small effective population size, and genetic connectivity between populations suggest genetic rescue could mitigate the adverse effect of rarity, divergent selection between populations indicates that genetic mixing could disrupt adaptation. Further work evaluating the fitness consequences of inter-population admixture is necessary to empirically evaluate the trade-offs associated with genetic rescue in Torrey pine.

4.2. Introduction

Conservation biology aims to preserve rare species and determine the appropriate management strategies necessary for long-term persistence and maintenance of evolutionary potential (Young, Brown, and Zich 1999; Swarts et al. 2009; Ralls et al. 2018; Di Santo and Hamilton 2020). Rare species may have reduced effective population sizes (Ne), impeding populations' ability to adapt to change (e.g., Ne < 1000) or increasing probability of inbreeding (e.g., Ne < 100) (Frankham, Bradshaw, and Brook 2014), ultimately increasing the risk of local extirpation. Combined, rarity and isolation are often associated with stochastic loss of genetic variation (Young, Boyle, and Brown 1996; Aguilar et al. 2008; Hague and Routman 2016). Genetic rescue is one conservation strategy that could be used in both animals and plants to mitigate consequences of severely reduced genetic diversity (Westemeier et al. 1998; Madsen et al. 1999; Bossuyt 2007; Willi et al. 2007; W. E. Johnson et al. 2010; Hedrick et al. 2014). Genetic rescue introduces or restores gene flow between populations to alleviate the fitness consequences of inbreeding through the introduction of genetic variation. However, while rare species may exhibit small effective population sizes and reduced adaptive potential, disruption of local adaptation may lead to outbreeding depression, or reduced fitness of progeny following

admixture between genetically differentiated lineages (Hufford and Mazer 2003). Thus, the contribution of natural selection to the evolution of population genetic differences is a consideration for genetic rescue, as it may ultimately lead to migrants or translocated individuals being maladapted (Nosil, Vines, and Funk 2005; Lowry, Rockwood, and Willis 2008). For rare species conservation, an understanding of contemporary effective population size is therefore required to assess immediate genetic threats to population persistence. Following these threats, however, understanding the history of population connectivity, the distribution of genetic variation within and among populations, and the role of selection in shaping population differences will be critical to informed management decisions.

In addition to guiding conservation management strategies, understanding rare species' demographic and evolutionary history may prove valuable to optimizing strategies necessary to preserve neutral and nonneutral genetic diversity *ex situ*. *Ex situ* conservation collections, or the preservation of species outside their natural range of occurrence, can complement *in situ* conservation strategies (Pritchard et al. 2012; Cavender et al. 2015; Potter et al. 2017), providing a critical resource for the preservation of genetic variation, restoration or reintroduction (Guerrant Jr, Havens, and Vitt 2014; Potter et al. 2017). *Ex situ* sampling designs traditionally rely on neutral population genetic structure to guide sampling decisions (Caujapé-Castells and Pedrola-Monfort 2004; Gapare, Yanchuk, and Aitken 2008; Hoban and Schlarbaum 2014; Hoban 2019). However, concerns exist regarding the sole use of neutral genetic variability for species conservation, as variation at neutral loci is unlikely to reflect adaptive genetic diversity (McKay and Latta 2002; Holderegger, Kamm, and Gugerli 2006; Bonin et al. 2007; Teixeira and Huber 2021). *Ex situ* population sampling may need to evaluate the impact different evolutionary processes have had on population genetic structure to optimize neutral and adaptive variation

collected. Thus, an understanding of population connectivity and the impact of selection across populations can inform *ex situ* collection design. If empirical or simulated data suggest populations are genetically connected and genetic differentiation is low, then most neutral genetic variation may be captured within one or a few populations. However, if selection overcomes the homogenizing effects of gene flow, ensuring adaptive genetic differences are preserved for all populations will require the inclusion of diverse population origins, separation of such populations *ex situ*, and consideration of population origin in breeding programs.

With the advent of next-generation sequencing (NGS) and the ever-decreasing costs associated with these technologies, genome-wide estimates of genetic diversity can be readily assessed and used to guide conservation management strategies. Combined with statistical and simulation-based tools, these data provide a powerful and timely means to evaluate aspects of population genetic variation and spatial genetic structure critical to informing genetic rescue and *ex situ* conservation plans, including both populations' demographic and adaptive history (Abebe, Naz, and Léon 2015; Xia et al. 2018; Liu et al. 2020; X. Wang, Bernhardsson, and Ingvarsson 2020). However, in conifers, despite extensive use in characterizing genomes as well as neutral and adaptive variation (Namroud et al. 2008; Eckert et al. 2010; Nystedt et al. 2013; Stevens et al. 2016; Tyrmi et al. 2020; X. Wang, Bernhardsson, and Ingvarsson 2020), these data have only rarely been used to inform conservation management decisions.

Torrey pine (*Pinus torreyana* Parry) is a critically endangered pine (IUCN 2021a), endemic to California. One of the rarest pine species in the world (Critchfield and Little 1966; Dusek 1985), Torrey pine's distribution spans one island population (*Pinus torreyana* subsp. *insularis*) of approximately 3,000 reproductive individuals (Santa Rosa Island, CA), and one mainland population (*Pinus torreyana* subsp. *torreyana*) of approximately 4,000 reproductive

individuals (Torrey Pine State Reserve in La Jolla, CA) (J. Franklin and Santos 2011; Hall and Brinkman 2015). In addition to low population size, and despite current *in situ* and *ex situ* conservation efforts, Torrey pine suffers from exceedingly low genetic variation and faces both anthropogenic and environmental disturbances (Ledig and Conkle 1983; Waters and Schaal 1991; Whittall et al. 2010; J. Franklin and Santos 2011; J. A. Hamilton et al. 2017). For these reasons, the species may be at imminent risk for population-scale extirpation events and thus a potential candidate for genetic rescue. Inter-population admixture may increase population genetic diversity, alleviating potential fitness consequences associated with Torrey pine's low genetic diversity (J. A. Hamilton et al. 2017), and increase evolutionary potential necessary to respond to current and future ecological challenges (Carlson, Cunningham, and Westley 2014). Previous research observed heterosis following one generation of admixture between island and mainland individuals, suggesting that genetic rescue may alleviate fitness consequences associated with reduce genetic variation (J. A. Hamilton et al. 2017). However, if adaptive genetic differences have evolved between island and mainland populations, fitness consequences following the disruption of co-adapted gene complexes may not be observed in the first generation cross. Thus, although the combination of exceedingly low genetic diversity and conservation status suggest Torrey pine may be a candidate for genetic rescue, evaluation of the species' demographic and adaptive evolutionary history will be necessary prior to inform conservation management decisions.

With this study, we use genomic data to quantify and model aspects of populations' demographic and evolutionary history necessary to preserve rare species' evolutionary potential. Specifically, we delineate the contribution of stochastic and deterministic processes to genomic differentiation in *Pinus torreyana*, asking three questions: (i) what are current effective

population sizes and how have they changed over time, (ii) have populations remained genetically connected following isolation, and (iii) is there evidence of adaptive divergence between populations that may indicate distinct evolutionary trajectories? We identify the most probable demographic scenario for Torrey pine using Approximate Bayesian Computing (ABC) and coalescent simulations. To evaluate the role of selection, we identify loci that may be important to adaptation using various F_{ST} outlier methods and assess the functional significance of these loci by annotating candidates with the gene ontology (GO) resource. Using this knowledge, we discuss conservation strategies for Torrey pine, including *ex situ* sampling to preserve neutral and nonneutral processes within species collections, and possible risks associated with genetic rescue. This study demonstrates the benefits and necessity of understanding the demographic and adaptive history of rare species to guide conservation.

4.3. Materials and methods

4.3.1. Population sampling and DNA extraction

Between June and July 2017, needle tissue was collected from individuals spanning the entire natural distribution of *Pinus torreyana* (Torrey pine; Figure 11). A total of 286 individuals were sampled, including 146 individuals from the mainland population at the Torrey Pine State Reserve (TPSR) near La Jolla, CA and 140 individuals from the island population on Santa Rosa Island, CA (SRI), one of the Channel Islands. Needles were dried in silica gel following which genomic DNA was extracted using between 25-30 mg of dry needle tissue and a modified CTAB protocol (Doyle and Doyle 1987). To reduce DNA shearing, slow manual shaking of tubes was used. Following extraction, the concentration and purity of DNA extracted was quantified for each sample using a NanoDrop 1000 Spectrophotometer (Thermo scientific) to ensure all

samples had a concentration of at least 85 ng/µl and purity ratios of 1.4 and above (average across samples; 260/280 = 1.85, 260/230 = 1.96).

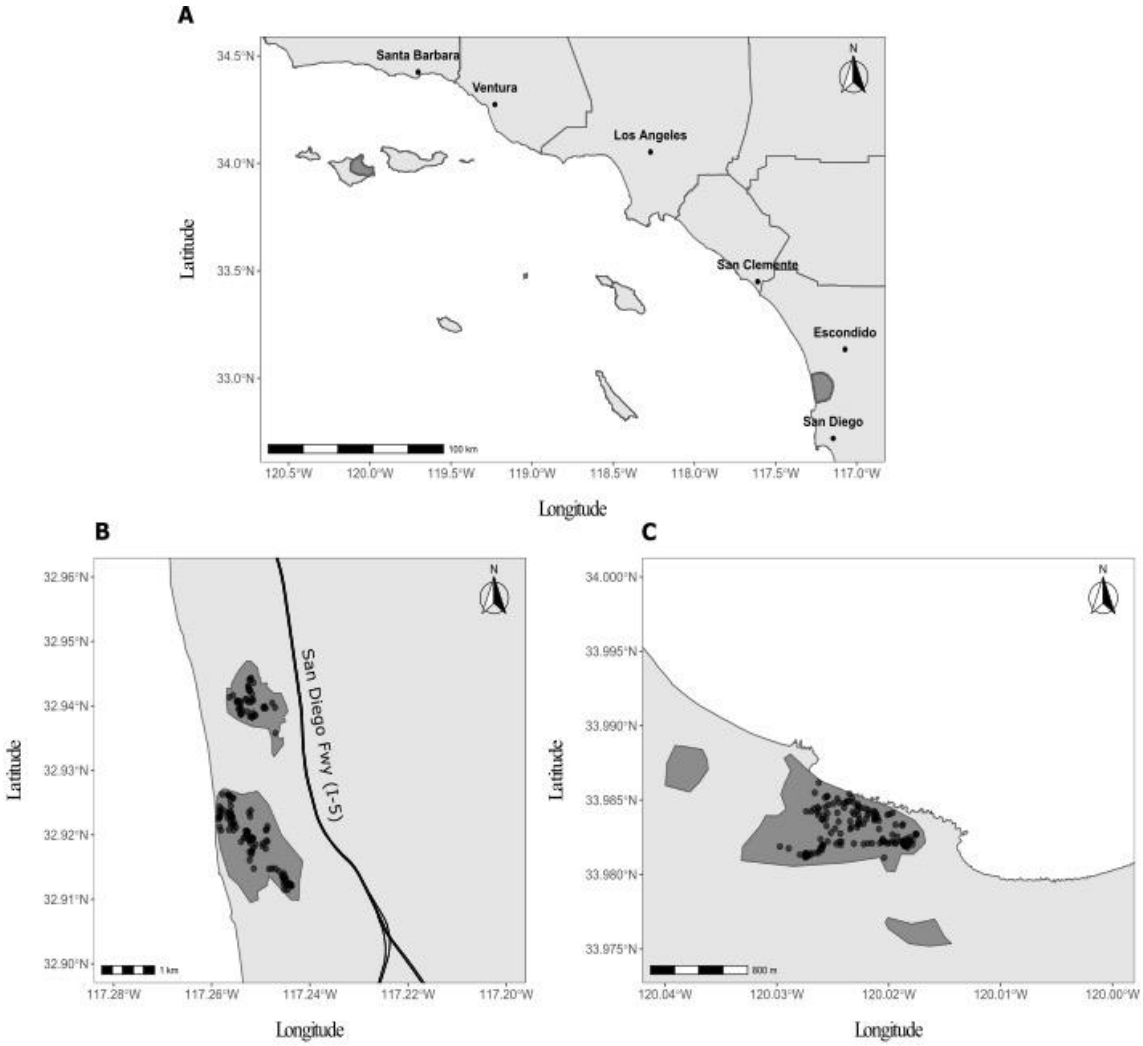


Figure 11. (A) Distribution of the last two remanent populations of Torrey pine (dark grey shades). Top left: Santa Rosa Island, Channel Islands, CA. Bottom right: Torrey Pine State Reserve, La Jolla, CA. (B-C) Population-specific distribution of Torrey pine (dark grey shades) and trees sampled for needle tissue (circles) at the Torrey Pine State Reserve (TPSR, B) and on Santa Rosa Island (SRI, C).

4.3.2. Genomic library preparation and ddRAD sequencing

Genomic libraries were prepared for all 286 individuals following the protocol of Parchman et al. (2012). Briefly, 510 ng (6 µl at 85 ng/µl) of DNA was digested using endonucleases *Eco*RI and *Mse*I (New England BioLabs, Inc.) after which barcoded (*Eco*RI cut

site) and non-barcoded (*Mse*I cut site) adapters compatible with Illumina sequencing were ligated to each end of DNA fragments using T4 ligase (New England BioLabs, Inc.). A different barcode sequence was used for each of the 286 samples. Due to the large, highly repetitive nature of pines' genome (Stevens et al. 2016), we used the methylation-sensitive enzyme *Eco*RI, as it effectively reduces the presence of repetitive and non-coding DNA sequences in genomic libraries (Parchman et al. 2012). Restriction-ligation products were amplified using two successive PCR reactions to produce genomic libraries with concentrations necessary for sequencing (> 2nM). All PCR-amplified genomic libraries were subsequently pooled and sent to the Genomic Sequencing and Analysis Facility (GSAF; Austin, TX) for size selection of fragments within the range of 450-500 bp and sequenced on 5 lanes of an Illumina HiSeq 2500 using the 100 bp single-end sequencing protocol (1 x 100 bp).

4.3.3. De novo assembly and SNP calling

Demultiplexing of sequence files was performed using IPYRAD v0.9.12 (Eaton and Overcast 2020) allowing one mismatch in the barcode sequence. Reads were filtered, assembled *de novo*, and used to call SNPs within the *dDocent* v2.7.8 pipeline (Puritz, Hollenbeck, and Gold 2014; Puritz et al. 2014). Reads were filtered by removing low-quality bases at the beginning and end of reads (PHRED score < 20), Illumina adapters, and trimmed when the average PHRED score fell below 10 within a 5 bp window using the program TRIMMOMATIC (Bolger, Lohse, and Usadel 2014). As a contiguous genome assembly for *P. torreyana* is not available, we first generated a reference of genomic regions sampled with our sequencing design using a *de novo* approach. Reads were clustered based on sequence similarity and assembled into a reference assembly using the program CD-HIT (W. Li and Godzik 2006; L. Fu et al. 2012). To be included in *de novo* assembly, reads had to have a minimum of 3x within-individual coverage and be

present in at least 5 individuals. To form a cluster (locus), reads had to have a minimum of 86% sequence similarity, a cutoff previously used in pines (Menon et al. 2018). This threshold was chosen as a tradeoff to avoid the clustering of paralogous loci while still accounting for the presence of missing bases, errors, or polymorphisms between true homologous sequences. Finally, reads were mapped onto *de novo* assembled loci using BWA MEM (H. Li 2013) and genetic variants were called using the software FREEBAYES (Garrison and Marth 2012). Read mapping was performed using BWA default parameters, including a match value of 1, a mismatch penalty of 4 and a gap penalty of 6. This yielded a set of 652,492 genetic variants that were subjected to downstream filtering. Variants with genotype quality (GQ) < 20 and genotype depth (DP) < 3 were first marked as missing. Then, variants with PHRED scores (QUAL) ≤ 30, minor allele counts < 3, minor allele frequencies < 0.01, call rate across all individuals < 0.95, mean depth across samples > 57 (based on the equation: $d + 4\sqrt{d}$, where d is the average read depth across variants, H. Li (2014)), F_{IS} estimates < -0.5 or > 0.5, and linkage score (r^2) > 0.5 within a 95 bp window were removed from the raw dataset. Additionally, insertion/deletion polymorphisms (INDELs) and SNPs with more than two alleles were discarded. Following filtering, a total of 93,085 biallelic SNPs were kept and used for analysis (hereafter referred to as the full dataset). Note that 16 individuals with > 40% missing data were also discarded, leaving 270 genotyped individuals (SRI: 130 individuals, TPSR: 140 individuals) for inclusion in analyses.

4.3.4. Population structure and genetic diversity analyses

To describe and quantify contemporary genetic differences between Torrey pine populations, we first assessed genetic structure of populations using principal component analysis (PCA) implemented in the R package “adegenet” (Jombart 2008; Jombart and Ahmed

2011). Unless otherwise stated, analyses were performed in R version 3.6.2 and 4.0.2 (R Core Team 2019; 2020). To quantify genetic differences between island and mainland populations, we calculated Nei's F_{ST} statistic (Nei 1987) for each SNP using the "hierfstat" R package (Goudet and Jombart 2020) and averaged estimates across loci. A 95% confidence interval around the mean was constructed by bootstrapping the empirical F_{ST} distribution 10,000 times using R packages "boot" (Canty and Ripley 2021; Davison and Hinkley 1997) and "simpleboot" (Peng 2019).

To determine the extent of contemporary within-population genetic structure, the most likely number of genetic clusters was independently evaluated for SRI and TPSR using the function *find.clusters()* implemented in the R package "adegenet". This function transforms genomic data using principal component analysis and performs successive K-means clustering with an increasing number of clusters (*k*). For each successive value of *k*, the Bayesian Information Criterion (BIC) is computed and was used to assess the optimal number of clusters (*k*). For each population, we assessed between 1 to 10 clusters while retaining principal components necessary to explain 90% of the variation after ordination (SRI: 114, TPSR: 122). For TPSR, two individuals (TPSR5107, TPSR3189) clustered distantly from the population, which may mask subtle within-population genetic structure. Consequently, we re-ran the analysis excluding the two individuals while maintaining the same range for *k* (1 to 10) and retaining 121 principal components (90.38% of variation explained after ordination).

To evaluate Torrey pine evolutionary potential, contemporary standing genetic variation within the species was estimated by calculating expected heterozygosity (H_E), inbreeding coefficients (F_{IS}), and coancestry coefficients ($\boldsymbol{\theta}$) for both island and mainland populations independently. Values of H_E and F_{IS} were calculated for each SNP separately using the R

package “adegenet” and averaged across loci to provide population-level estimates. To evaluate $\boldsymbol{\theta}$, we used the R package “related” (Pew et al. 2015) that estimates genetic relatedness between all possible pairs of individuals within a population. Specifically, we used the triadic likelihood (TrioML) estimate of relatedness assuming no inbreeding within populations. Averaging pairwise $\boldsymbol{\theta}$ values across all individuals within TPSR and SRI provided population estimates of genetic relatedness. To build 95% confidence intervals around H_E, F_{IS} and $\boldsymbol{\theta}$ averages, the empirical distribution for each parameter within each population was bootstrapped 10,000 times in R using the “boot” and “simpleboot” packages.

4.3.5. ABC demographic modeling

4.3.5.1. Demographic models

To evaluate the impact genetic drift and gene flow have had to patterns of neutral genetic variation in Torrey pine, we quantified changes in effective population size over time, interpopulation migration rate, and time since population divergence. To do so, we tested six distinct demographic models that were classified into two broad categories: (1) isolation with/without migration (Figure 12A), and (2) two-population demic expansion (Figure 12B, C).

Models of isolation with or without migration (RPM1, RPM2) were developed to test a hypothesis formulated by Ledig and Conkle (1983). This hypothesis predicts that there was once a single ancestral population of Torrey pine that diverged to form one island and one mainland population following tectonic movement (Figure 12A). These models assume that an ancestral population with an effective size N_A diverged T_{Div} generations before present to form two populations with current effective sizes N_I (island, SRI) and N_M (mainland, TPSR). Following divergence, to assess whether gene flow has occurred between populations, bidirectional

migration was either prevented (RPM1, $m_{IM}=m_{MI}=0$) or permitted (RPM2, $m_{IM}=m_{MI}>0$). Both models assume constant island and mainland effective population size following divergence.

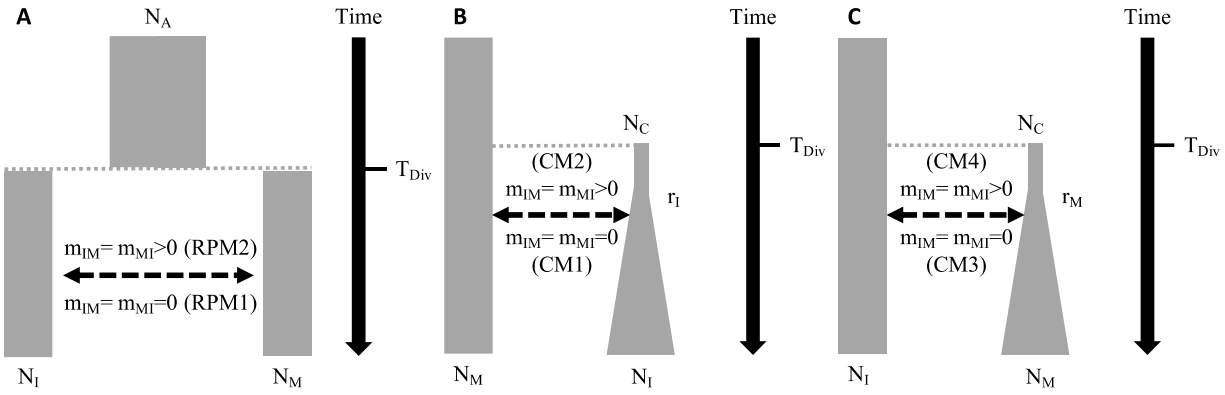


Figure 12. Schematics of demographic scenarios simulated. Rectangles represent current or ancestral populations, dashed arrows represent migration between population, and solid arrows represent time. (A) Scenarios of isolation without (RPM1) or with (RPM2) migration between populations. (B) Island colonization scenarios without (CM1) or with (CM2) subsequent migration between populations. (C) Mainland colonization scenarios without (CM3) or with (CM4) subsequent migration between populations. N_A, ancestral effective population size; N_I, island effective population size; N_M, mainland effective population size; N_C initial effective population size following migration (number of migrants); m_{IM}, migration probability from island to mainland; m_{MI}, migration probability from mainland to island; T_{Div}, time of population divergence; r_I, island (exponential) growth rate; r_M, mainland (exponential) growth rate.

The remaining four models (CM1, CM2, CM3, CM4) tested two different hypotheses of land colonization where one population was founded by the other (Figure 12B, C). Models CM1 and CM2 specifically test the hypothesis that Santa Rosa Island was colonized by a subset of mainland individuals (Ledig and Conkle 1983). In this scenario, SRI was founded by N_C effective migrants from TPSR T_{Div} generations ago and grew exponentially at a rate r_I ($r_I = log\left(\frac{N_C}{N_I}\right)/T_{Div}$) to form a population with an effective size N_I. TPSR effective population size (N_M) was assumed constant. As above, bidirectional migration between populations was either prevented (CM1, $m_{IM}=m_{MI}=0$) or permitted (CM2, $m_{IM}=m_{MI}>0$) to evaluate whether gene flow has occurred between populations following colonization. Models CM3 and CM4 test the hypothesis that the mainland population was founded by a subset of island individuals (J. R.

Haller 1986). This scenario assumes TPSR was founded by N_C effective migrants from SRI T_{Div} generations before present. SRI effective population size (N_I) was assumed constant while TPSR effective population size (N_M) grew exponentially at rate r_M ($r_M = log\left(\frac{N_C}{N_M}\right)/T_{Div}$). Once again, to test whether gene flow has occurred between populations since colonization, exchange of migrants between population was either prevented (CM3, $m_{IM}=m_{MI}=0$) or permitted (CM4, $m_{IM}=m_{MI}>0$).

For all six models, uniform priors were used except for T_{Div}, N_C, and N_A for which we used log-uniform priors. Priors on a logarithmic scale increases the weight given to small values and is recommended when parameters' ranges span several orders of magnitude (Wegmann, Leuenberger, and Excoffier 2009). Details on demographic parameters and their prior distributions are provided in Table 4.

4.3.5.2. Additional filtering and down-sampling of genetic variants

Accurate estimation of demographic parameters using coalescence simulations requires the use of neutrally evolving genetic markers. To ensure neutrality of SNPs, we filtered the full dataset for SNPs that did not deviate significantly from Hardy-Weinberg equilibrium (HWE). Since population structure may create departures from HWE, we applied this filter to island and mainland populations independently and removed SNPs that deviated significantly from HWE ($P < 0.05$) in both populations. This was performed using a customized R function relying on R packages "adegenet" and "pegas" (Paradis 2010). In total, 73,928 SNPs were retained following filtering for use in demographic modelling (hereafter referred to as the HWE-filtered dataset).

For computational efficiency, we down sampled the HWE-filtered dataset from 73,928 to 9,795 variants first by generating bivariate bins based on observed heterozygosity and Nei's F_{ST} (0.05-interval bins), and then by subsampling each bin proportionally to the number of SNPs

they contained. In this way, each bin is subsampled to reflect its contribution to the HWE-filtered dataset (Figure D1). Following subsampling, we conducted a principal component analysis using the down-sampled dataset to ensure patterns of genetic diversity and population structure were maintained between datasets (Figure D2).

Table 4. Demographic parameters with their prior distributions and occurrence in each of the six models simulated.

Present in models	Parameter	Symbol	Prior distribution	Unit
RPM1, RPM2	Ancestral effective population size	N_A [a]	logU(100:100,000)	ind.
All	Island effective population size	N_I [b]	U(100:6,000)	ind.
All	Mainland effective population size	N_M [b]	U(100:6,000)	ind.
CM1, CM2, CM3, CM4	Initial effective population size following colonization	N_C [c]	logU(2:300)	ind.
All	Time of population divergence	T_{Div} [d]	logU(400:50,000)	gen.
RPM2, CM2, CM4	Migration from island to mainland and vice-versa.	m_{IM}, m_{MI} [e]	U(0.001:0.01)	-
All	Minor allele frequency	maf [f]	0.01	-

[a] A wide prior range for Torrey pine ancestral effective population size was used to enable the simulation of complex demographic histories, including population size expansion, population size reduction or a combination of both.
[b] Census size of reproductively mature trees is of approximately 3,000 to 4,000 individuals in island and mainland population respectively (Hall and Brinkman 2015; J. Franklin and Santos 2011). We selected a wide prior range around those estimates to account for the fact that effective and not census population sizes were simulated.
[c] Number of potential effective migrants were chosen based on seed capacity of a Torrey pine cone, which can hold up to several hundred seeds (unpublished data).
[d] Time since populations' isolation was estimated between 4,300 and 430,000 years ago (Ledig and Conkle 1983). Generation time for Torrey pine being approximately 10 years, this translates into a time since divergence between 430 and 43,000 generations.
[e] Low genome-wide differentiation observed between populations in this study (F_{ST} = 0.013) suggest that gene flow may have occurred between populations of Torrey pine. We believe a probability of migration between 0.001 and 0.01 should appropriately reflect low to moderate gene exchange expected between geographically distant populations (approximately 280 km) of this wind-pollinated, long-lived, and primarily outcrossing pine species (F_{IS} = -0.127 and -0.124 for island and mainland population respectively).
[f] Minor allele frequency of 0.01 was chosen to reflect empirical genetic data, filtered for variants with frequencies of or above 0.01 (see Material and methods).

4.3.5.3. Generating coalescent simulations and estimating summary statistics

To evaluate and compare demographic scenarios, a set of 200,000 simulations was generated using ABCSAMPLER for each model, a wrapper program included in the package ABCTOOLBOX (Wegmann et al. 2010). For each simulation, ABCSAMPLER samples prior ranges of demographic parameters and uses these values as inputs for coalescence simulations within a user-defined simulation program. We used FASTSIMCOAL version 2.6.0.3 (Excoffier and Foll 2011; Excoffier et al. 2013) to simulate 9,795 unlinked SNPs with a minor allele frequency of 0.01, reflecting the composition of the down-sampled genetic dataset. For each model, simulated data were output as genotypes and fed to a user-defined program by ABCSAMPLER to estimate population genetic summary statistics. ARLEQUIN version 3.5.2.2 (Excoffier and Lischer 2010) was used to calculate ten distinct population genetic summary statistics, specifically aiming at quantifying genetic variation and divergence within and between Torrey pine populations. These included genetic diversity (i.e., population-specific heterozygosity and number of alleles, average heterozygosity and number of alleles over loci and populations, and mean total heterozygosity), genetic differentiation (i.e., pairwise F_{ST}), and variance (i.e., standard deviation over populations of the average heterozygosity and number of alleles) statistics. Finally, to obtain observed population genetic summary statistics, ARLEQUIN version 3.5.2.2 was rerun using the down-sampled dataset. Characterizing and summarizing the amount and distribution of genetic variation within each dataset, these statistics were used to calculate posterior probabilities and identify the demographic model with greatest support, as well as estimate demographic parameters associated with that model (see below).

4.3.5.4. ABC parameter estimation

Demographic parameters were estimated in R using the "abc" package (Csilléry, François, and Blum 2012). Cross-validation simulations were conducted first to evaluate the ability of summary statistics to distinguish between demographic models (Figure D3). We performed leave-one-out cross-validations, consisting of selecting one simulation of a demographic scenario and then assigning it to one of the six models using posterior probabilities estimated from all remaining simulations. This was repeated one hundred times for each demographic model, generating a confusion matrix. If misclassification rates (proportions of simulations incorrectly assigned to a model) are low, then computed summary statistics can distinguish between our different demographic scenarios. The posterior probability of each demographic model was approximated as the proportion of accepted simulations and used to select the best model following cross-validation. To ensure the best model provided a good fit to the data, we performed a goodness-of-fit test as implemented in the function *gfit()*. Finally, demographic parameters associated with the best model were estimated as the weighted medians of posterior distributions using the *weighted.median()* function implemented in the R package "spatstat" (Baddeley, Rubak, and Turner 2015). Posterior distributions were created from the set of accepted simulations using a non-linear postsampling regression adjustment conducted on log-transformed data (Blum and François 2010). Ninety-five percent confidence intervals around weighted medians were estimated using 10,000 bootstrap replicates of posterior distributions in R, using "boot", "simpleboot" and "spatstat" packages. The validity and accuracy of each estimated parameter were tested using additional, yet distinct, 100-fold leave-one-out cross-validation simulations (Figure D4). The cross-validation process begins with the random selection of one simulation generated by the best demographic model. Summary statistics

associated with that simulation are considered as pseudo-observed data and its demographic parameters are estimated using remaining simulations for the model. If pseudo-observed parameters can accurately be predicted, then inferred demographic parameters from true observed data can be considered valid and accurate. Cross-validation simulations, model selection, model validation, and parameters estimation were conducted using a tolerance threshold of 0.01, a tradeoff between retaining a reasonable number of simulations to estimate posterior distributions and keeping the tolerance value as low as possible (S. Li and Jakobsson 2012). Finally, as census sizes for Torrey pine populations are available (J. Franklin and Santos 2011; Hall and Brinkman 2015), we calculated the proportion of the census size (N) to effective population size (Ne) for each population separately. Of all reproductive trees present within a population (census size), this ratio estimated the proportion contributing to the next generation (effective size).

4.3.6. Simulating the null F_{ST} distribution

To evaluate the influence natural selection may have had on the genomic structure of Torrey pine populations, we compared the distribution of Nei's F_{ST} estimated from the full SNP dataset with a simulated distribution based on 93,085 independent SNPs from 270 individuals generated using SIMCOAL2 version 12.09.07 (Laval and Excoffier 2004). We used weighted medians estimated from posterior distributions borrowed from the best demographic model (see Results) as input parameters for neutral simulations. For each simulated SNP, the minimum allele frequency was set to 0.01 to reflect filters applied to the full dataset. Locus-specific Nei's F_{ST} values were estimated for both full and simulated datasets in R using the "hierfstat" package (Figure D5).

4.3.7. Outlier detection analyses

To estimate the potential contribution of local adaptation to genomic differentiation between Torrey pine populations, we used the full dataset of 93,085 SNPs to identify loci putatively under selection using three distinct methods: the software BAYESCAN version 2.1 (Foll and Gaggiotti 2008), and R packages "OutFLANK" (M. C. Whitlock and Lotterhos 2014) and "pcadapt" (Privé et al. 2020). These three approaches were selected for their ability to account for neutral population structure (BAYESCAN, OutFLANK) and to handle genetic admixture between individuals (pcadapt). For both BAYESCAN and OutFLANK, we grouped all 270 individuals by populations (mainland: 140 individuals, island: 130 individuals). Below is a brief description of all three methods used to identify candidate SNPs.

The first approach we used was BAYESCAN. For each locus, BAYESCAN uses a Bayesian approach to decompose F_{ST} coefficients into population- and locus-specific components using a logistic regression. Loci are identified as putatively under selection if the locus-specific component is needed to explain the observed distribution of genetic diversity. Our analysis was conducted using BAYESCAN default parameters. Results were visualized and analyzed in R. Only SNPs with a false discovery rate of 1% or below were retained and considered as candidate loci. The second approach we used was OutFLANK, an R package which identifies outliers by inferring a null F_{ST} distribution approximated from empirical F_{ST} values. This distribution was produced by removing loci with an expected heterozygosity below 0.1 (Hmin = 0.1), trimming 5% of lowest and highest empirical F_{ST} values (RightTrimFraction = LeftTrimFraction = 0.05), and retaining only loci passing a built-in neutrality test with a false discovery rate of 0.1% (qthreshold = 0.001). Using the latter threshold provided a conservative estimate for neutral F_{ST}. Outlier loci were identified by comparing the empirical F_{ST} distribution

to the inferred null F_{ST} distribution using a built-in chi-squared test with a false discovery rate of 1% (qthreshold = 0.01). Note that loci with $H_E < 0.1$ were excluded while conducting the chi-squared test and therefore could not be identified as potential outliers (Hmin = 0.1). The third and last approach we used was pcadapt. Also implemented in R, pcadapt is a package that assesses population structure using PCA. Consequently, this approach does not require individuals to be grouped into populations. Following PCA, candidate loci are identified as those substantially correlating with population structure. We ran pcadapt retaining the first axis of differentiation (K = 1) and SNPs with a minor allele frequency of 0.01 (min.maf = 0.01). We only considered the first principal component to calculate the test statistic, as additional axes did not ascertain population structure (Figure D6). Candidate loci were identified as the set of SNPs with a false discovery rate of 1% or below. To minimize the presence of false positives within our dataset, only outlier SNPs common to all three approaches were considered as candidate loci and included in subsequent analyses. Finally, to both visualize and quantify genetic structure at putatively adaptive loci, we conducted a principal component analysis based only on candidate SNPs shared by all three methods using the R package "adegenet".

4.3.8. Functional categorization of candidate loci

To identify biological processes or molecular functions that may play a role in adaptation to mainland or island environments, *de novo* assembled sequences containing candidate SNPs common to all genome scans were first extracted and then annotated using BLASTx version 2.6.0+. Sequences were blasted against the UniProt protein database filtered for sequences from species within the *Pinaceae* family (Taxon identifier 3318). Sequence similarity was assessed using BLASTx default parameters, an e-value hit filter of 10^{-3} (-evalue 0.001), and a number of

database hits to retain of 1 (-max_target_seqs 1). Gene ontology terms were mapped onto annotated sequences in R using the "UniprotR" package (Soudy et al. 2020).

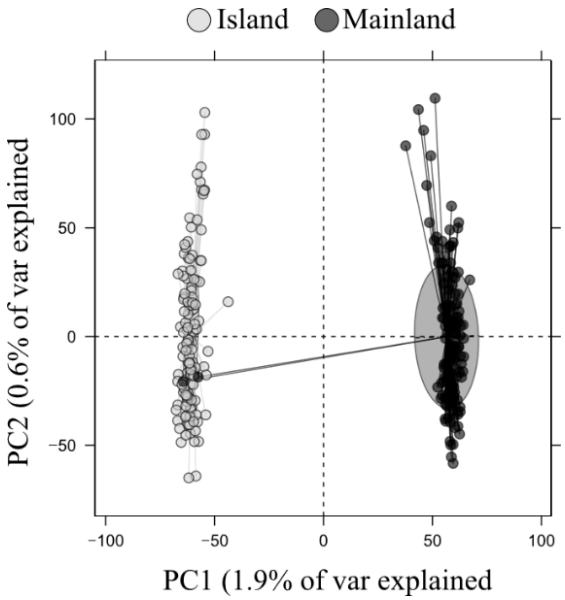


Figure 13. Principal component analysis using 93,085 SNPs for 270 Torrey pine individuals, including individuals from both mainland (black) and island (grey) populations. Variation explained by the first two principal components is provided in parentheses.

4.4. Results

4.4.1. Population genetic structure and variation

Using all 93,085 SNP markers, the principal component analysis revealed little genome-wide differentiation in Torrey pine with the first two principal components (PC1, PC2) explaining only 1.9 and 0.6% of observed genetic differences, respectively (Figure 13). Nonetheless, PC1 unambiguously separated island from mainland individuals, suggesting some level of genetic structure exists between populations. These results were further supported by the low average coefficient of genetic differentiation found across loci (Nei's F_{ST}), estimated at 0.013 (95% CI: 0.012 - 0.013). In addition to population-scale genetic differentiation, within-population genetic differentiation was estimated to evaluate whether local inbreeding, possibly increased in small populations, could have contributed to fine-scale genetic structure in Torrey

pine. For both island and mainland populations, we found no evidence of within-population genetic structure. Population-specific principal component analyses identified no clear genetic clusters and revealed that, combined, the first two axes of differentiation (PC1, PC2) only explained approximately 1.9% and 2.2% of mainland and island within-population genetic differences, respectively (Figure 14). In addition, evaluating the likelihood between 1 to 10 genetic clusters (*k*) within each population using BIC indicated that populations appear largely homogeneous (most likely $k = 1$) (Figure D7). Interestingly, average estimates of inbreeding and coancestry coefficients across loci were low for both the island and the mainland population (Table 5). While low inbreeding coefficients support the lack of observable within-population genetic structure, this also indicates that reproduction among relatives or unequal reproductive success among individuals is unlikely to have contributed to low expected heterozygosity observed within populations (Table 5). Combined, our results indicate that Torrey pine exhibits exceedingly low genetic diversity, with the majority of variation distributed within genetically unstructured populations.

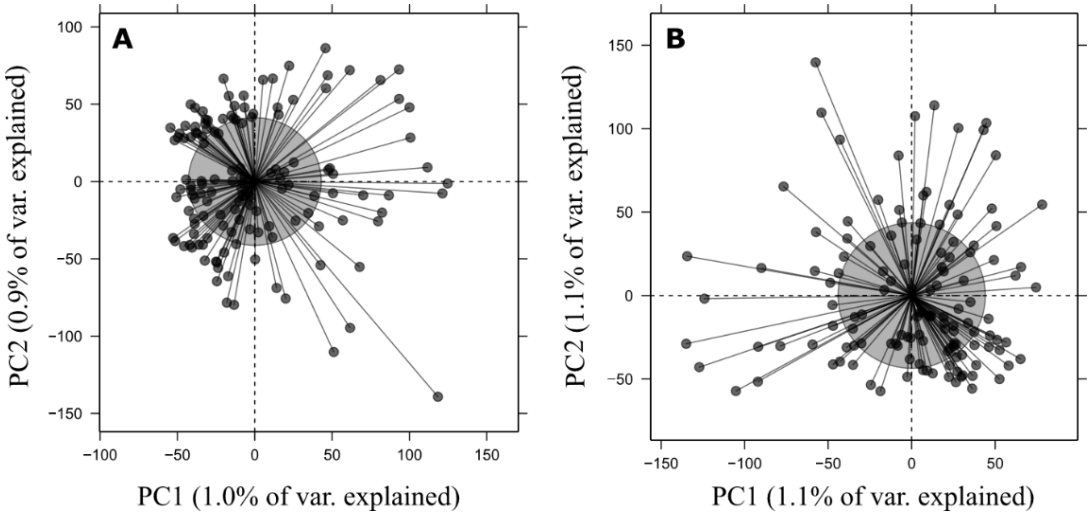


Figure 14. Population-specific principal component analysis based on 93,085 SNP variants. (A) Mainland population (TPSR, n = 138). Note that two individuals were removed from analysis to visualize within-population structure on a finer scale (see Material and methods). (B) Island population (SRI, n = 130).

Table 5. Average genetic summary statistics across loci, including expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), and coancestry (relatedness) coefficient (θ) for island (SRI) and mainland (TPSR) Torrey pine populations. Ninety-five percent confidence interval (95% CI) around mean estimates were obtained by bootstrapping.

Population	H_E (95% CI)	F_{IS} (95% CI)	θ (95% CI)
Island	0.185 (0.184, 0.186)	-0.127 (-0.128, -0.126)	0.023 (0.023, 0.024)
Mainland	0.184 (0.183, 0.184)	-0.124 (-0.125, -0.124)	0.022 (0.022, 0.023)

4.4.2. Demographic history of Torrey pine

Of the six demographic models evaluated (Figure 12), the isolation with migration model (RPM2) received the most support with a posterior probability of 92.18%. The remaining five models exhibited lower posterior probabilities ranging from 0% to 3.98%. With low misclassification rates, cross-validations indicated that simulated summary statistics were able to confidently distinguish between different demographic scenarios (Figure D3). The goodness-of-fit test revealed that simulated summary statistics did not significantly differ from observed ones ($P = 0.76$), providing a good fit to the data.

Based on RPM2, an ancestral Torrey pine population with an effective size of approximately 1,124 individuals (95% CI: 939 - 1,213) diverged during the early Pleistocene approximately 1.2 million YBP (95% CI: 1,195,367 - 1,296,186, assuming a generation time of ten years) to form one island and one mainland population with effective sizes of approximately 2,305 (95% CI: 2,166 - 2,338) and 1,715 (95% CI: 1,616 - 1,759) individuals, respectively (Figure 15). This resulted in a 0.75 ($N_I/N = 2,305/3,063$) proportion of the census to effective population size on the island, and a 0.45 ($N_M/N = 1,715/3,806$) proportion of the census to effective population size on the mainland. Following divergence, some gene flow was maintained between populations with an estimated migration rate of 8.34×10^{-3} (95% CI: 8.15 x

10^{-3} - 8.76 x 10^{-3}) per generation. In general, cross-validation simulations indicated low prediction errors (Figure D4), suggesting high accuracy of inferred parameters. Nonetheless, note that for T_{Div}, the associated prediction error was higher than for other parameters.

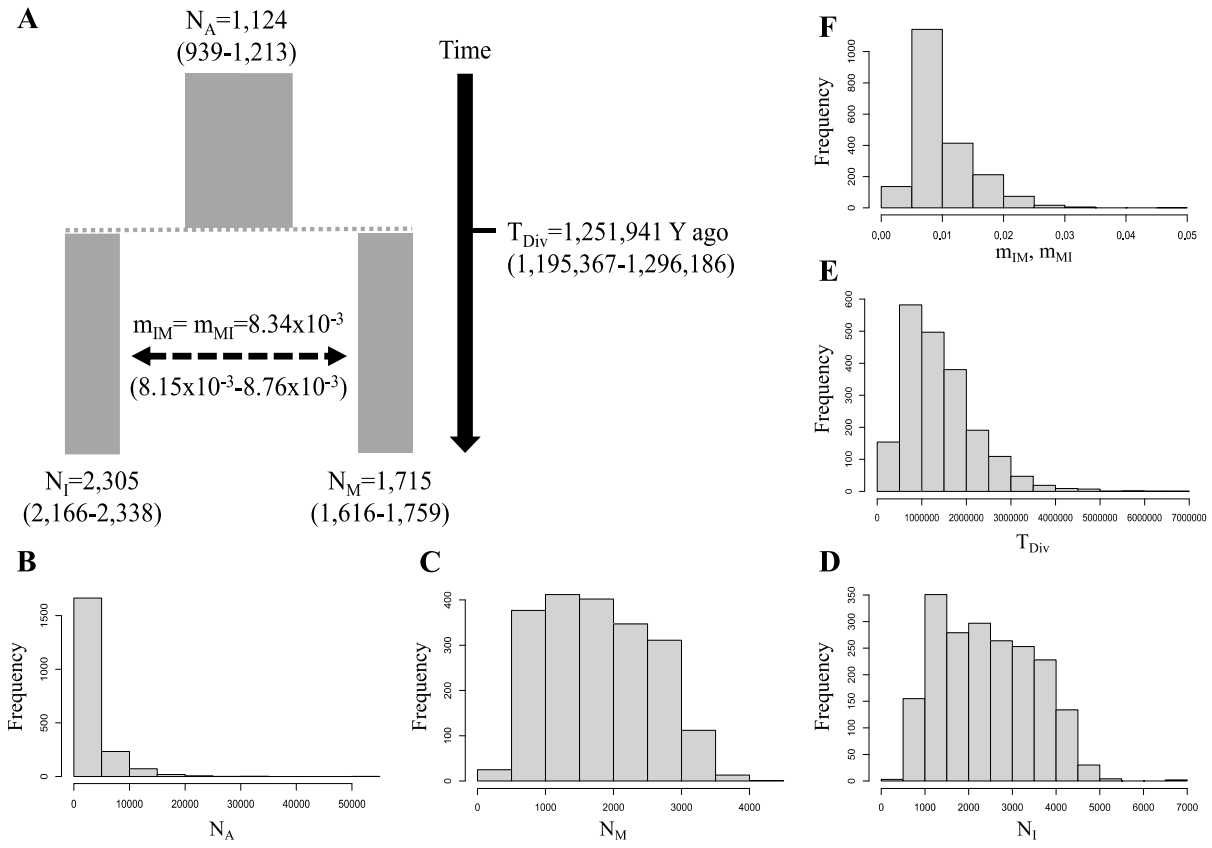


Figure 15. (A) Demographic parameters (weighted medians) and 95% confidence intervals (in parentheses) estimated using RPM2. Each rectangle represents a population, either contemporary or ancestral. The solid arrow represents time, while the dashed arrow indicates gene flow between populations. (B-F) Posterior distribution of each demographic parameter inferred using a tolerance rate of 0.01. N_A, ancestral effective population size; N_I, island effective population size; N_M, mainland effective population size; m_{IM}, migration probability from island to mainland; m_{MI}, migration probability from mainland to island; T_{Div}, time of population divergence.

4.4.3. Divergent selection between island and mainland populations

Despite limited genetic variation within populations, we found some evidence for the evolution of genetic differences among populations at a subset of loci. We compared the neutral F_{ST} distribution generated from the simulated RPM2 demographic model with the empirical

distribution based on our full dataset of SNP variants (Figure D5). For select SNPs, moderate to high empirical F_{ST} values (from approximately 0.2 to 0.65) could not be generated through neutral simulations, suggesting they may be candidate for selection. Pcadapt, BAYESCAN, and OutFLANK identified 3,138 (3.37%), 2,163 (2.32%), and 3,942 (4.23%) outlier SNPs, respectively (Figure 16A). Of these outlier loci, 2,053 (2.21%) were common to all three methods and contribute to genomic structure between Torrey pine populations (Figure 16B). Indeed, the principal component analysis revealed that the first axis of differentiation (PC1) unambiguously differentiated island from mainland populations based on common outlier SNPs, explaining over 20% of observed genetic differences. Consistent with these results, F_{ST} values for putatively adaptive loci ranged between 0.1 and 0.63 and either could not be generated or could only be generated at low frequency through neutral simulations, representing only 0.078% of all simulated F_{ST} values.

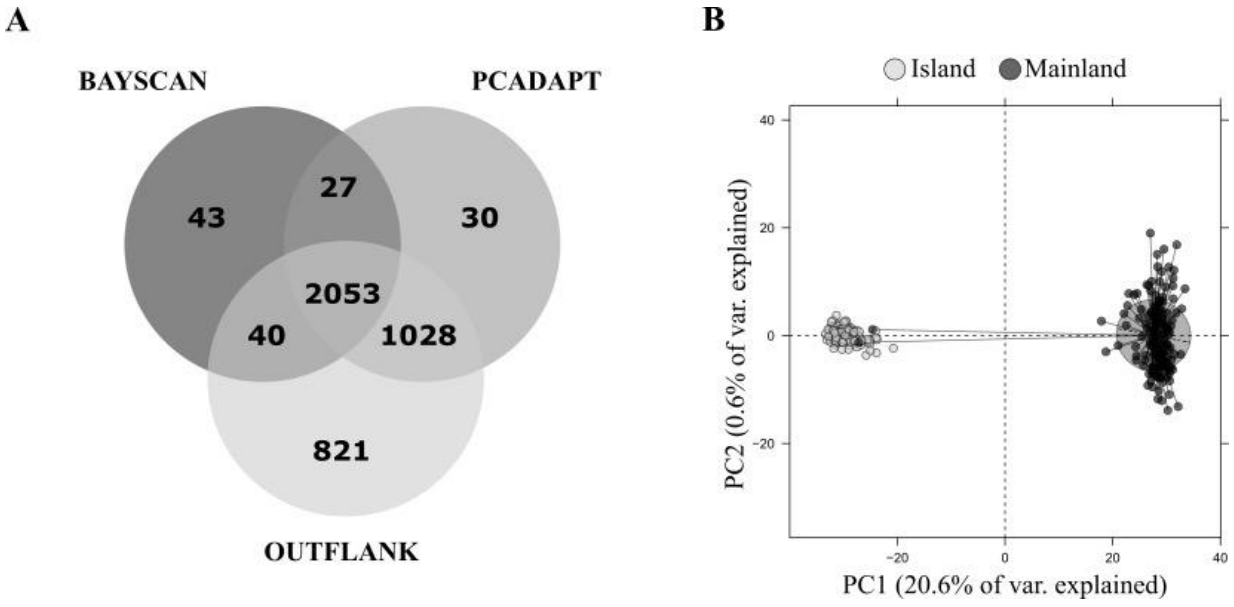


Figure 16. (A)Venn diagram showing the number of putatively unique and shared adaptive SNPs detected by BAYESCAN, pcadapt, and OutFLANK. (B) Principal component analysis based on all 2,053 shared candidate SNPs for Torrey pine individuals, including individuals from both mainland (black, n = 140) and island (grey, n = 130) populations. Variation explained by the first two principal components is provided in parentheses.

Functional categorization of common outlier loci was performed by blasting *de novo* assembled contigs carrying outlier SNPs against the *Pinaceae* UniProt protein database and retrieving each hit's Gene Ontology terms. Overall, 110 (7.51%) contigs were annotated. After accounting for redundancy in the data (i.e., different contigs aligning to the same locus), we identified a total of 80 putative adaptive genes with homologous sequences in *Larix*, *Picea*, or *Pinus* species (Table D1) that may be targets of selection. Functionally, these genes are primarily encoded in mitochondria, involved in the process of DNA integration, or with processes associated with molecular functions such as RNA-directed DNA polymerase activity or nucleic acid binding (Table 6).

4.5. Discussion

An understanding of demographic and adaptive evolutionary history is invaluable for rare species of conservation concern, particularly when management decisions impact populations at risk of extinction. Teasing apart the contribution of both stochastic and deterministic evolutionary processes to population genomic differentiation over time and space can be used to inform species conservation decisions, including potential consequences of genetic rescue (Hufford and Mazer 2003; Frankham et al. 2011; Ralls et al. 2018). Here, we evaluated the genomics of Torrey pine, a critically endangered endemic isolated to two populations in California. We modeled demographic change and connectivity over time and tested the influence of neutral and selective processes have had on contemporary population genomic structure. We observed that Torrey pine populations exhibited exceedingly low genetic variation, particularly for a conifer (see below), with little within- or among-population structure. Although some connectivity has been maintained between island and mainland populations, demographic modeling indicates that Torrey pine has consistently suffered from low effective population size.

Table 6. Functional categorization of 80 putatively adaptive genes between Torrey pine populations. Listed are the ten most frequent GO terms within each of the three GO classes (Biological Process, Molecular Function, Cellular Component).

GO class	GO term	Frequency (%)
Biological Process	DNA integration	6.2
	Methylation	2.5
	Actin filament polymerization	1.2
	Arp2/3 complex-mediated actin nucleation	1.2
	Carbohydrate metabolic process	1.2
	Carbohydrate transport	1.2
	Cell differentiation	1.2
	Defense response to bacterium	1.2
	Defense response to fungus	1.2
	Gene silencing by RNA	1.2
Molecular Function	RNA-directed DNA polymerase activity	11.2
	Nucleic acid binding	7.5
	ATP binding	3.8
	ADP binding	2.5
	Magnesium ion binding	2.5
	O-methyltransferase activity	2.5
	Oxidoreductase activity	2.5
	Protein kinase activity	2.5
	4-lactone oxidase activity	1.2
	Actin binding	1.2
Cellular Component	Mitochondrion	38.8
	Integral component of membrane	7.5
	Arp2/3 protein complex	1.2
	Cell wall	1.2
	Chloroplast	1.2
	Chloroplast thylakoid membrane	1.2
	Cytoplasm	1.2
	Extracellular region	1.2
	Membrane	1.2
	Nucleus	1.2

Genome scans revealed over 2000 loci that were candidates for selection. Consistent with previous observations indicating phenotypic differences between island and mainland populations, these data suggest adaptive genetic differences have evolved among populations (J. R. Haller 1986; J. A. Hamilton et al. 2017; Di Santo, Polgar, et al. 2021). From a conservation standpoint, these results lead to contrasting recommendations with respect to genetic rescue. A history of reduced effective population size and low genome-wide differentiation at neutral loci indicate little genetic differentiation among populations that may impact a genetic rescue program. However, previous observations of phenotypic differences paired with loci associated with divergent selection point towards the importance of adaptive evolution among Torrey pine populations. These results suggest increased genetic variation via inter-population crosses may be needed in this species, but admixture should be evaluated first to quantify its fitness impact.

Torrey pine populations exhibited extremely low genetic variation in comparison with other conifers which often exhibit an expected heterozygosity around 0.3 within populations (Namroud et al. 2008; Tsumura et al. 2012). Average expected heterozygosity (H_E) and contemporary effective population size (Ne) were estimated at 0.185 and 2,305 individuals for the island population, and 0.184 and 1,715 individuals for the mainland population, respectively (Table 5; Figure 15). Although low, this study is the first to find genetic variability within island and mainland populations of Torrey pine. While previous genetic analyses using allozymes and chloroplast DNA markers identified fixed genetic differences between populations, they failed to observe genetic variation within populations (Ledig and Conkle 1983; Whittall et al. 2010). Ledig and Conkle (1983) hypothesized reduced genetic diversity was attributable to drastically reduced mainland and island populations in the recent geological past. They suggested that Torrey pine populations declined to fewer than 50 individuals, following which a recovery led to

approximately 3,000 to 4,000 reproductively mature trees (J. Franklin and Santos 2011; Hall and Brinkman 2015). Demographic models herein, however, suggest that effective population size has always been low for Torrey pine (Figure 15). The best fit demographic model predicted an ancestral effective population size (N_A) of 1,124 and only limited change following population divergence. Long-term reduced effective population size has thus likely exacerbated the consequences of genetic drift within Torrey pine populations, eventually leading to an extreme lack of genetic variability.

Despite extremely low genetic variation and small effective population sizes, there was no evidence for inbreeding (F_{IS}) or excessive relatedness ($\boldsymbol{\theta}$) within populations (Table 5). These findings support high ratios between effective and census population size (Ne/N) found in both Torrey pine populations (island = 0.75, mainland = 0.45) and may, at least partly, explain why these estimates were higher than ratio averages of 0.1 to 0.2 typically recommended for conservation management (Frankham, Bradshaw, and Brook 2014). Indeed, it is not uncommon for plants to exhibit Ne/N ratios greater than 0.2 (Waples et al. 2013; Hoban et al. 2020). Combined with a lack of within population genetic structure (Figure 14, Figure D7), these results also indicate that neither reproduction among relatives nor unequal reproductive success has likely contributed to reduced genetic variation within populations. Wind pollination and zoochorous seed dispersal have likely contributed to homogenizing the gene pool within populations (Loveless and Hamrick 1984; M. Johnson, Vander Wall, and Borchert 2003). Pines also possess mechanisms that can reduce the probability of self-fertilization, including the embryo lethal system (Williams, Zhou, and Hall 2001; Williams 2009). This self-incompatibility system selectively induces death in embryos resulting from self-fertilization (Bramlett and Popham 1971; Williams, Barnes, and Nyoka 1999). Consequently, increased dispersal potential

paired with post-zygotic barriers limiting the probability of mating among relatives have likely reduced within-population genetic structure in Torrey pine.

Demographic modelling using neutral genomic variation supports the maintenance of some genetic connectivity following population divergence approximately 1.2 MYA (Figure 15), estimating the probability of gene exchange at $8.34x10^{-3}$ per generation. Despite geographic isolation among populations and reduced potential for inter-population gene flow, contemporary estimates of F_{ST}=0.013 indicate only subtle genome-wide differentiation between island and mainland populations. Reduced genetic differentiation is typical of many conifers (Tyrmi et al. 2020; Eckert et al. 2010; Namroud et al. 2008), as pollen may maintain connectivity over very long distances (Campbell et al. 1999; Varis et al. 2009; Williams 2010). Gene flow between populations may also have been maintained via seed dispersal. Birds represent potential seed dispersers for Torrey pine and may play a prominent role in long-distance seed dispersal (M. Johnson, Vander Wall, and Borchert 2003; Pesendorfer et al. 2016; Viana et al. 2016). Interestingly, higher estimates of contemporary effective population size (N_I, N_M) relative to the ancestral population size (N_A) suggest that island and mainland populations have experienced genetic bottlenecks following one or multiple moderate population expansion events (Figure 15). Overall, our findings indicate that despite the increased probability of genetic drift due to genetic bottlenecks and low population sizes, gene flow maintained between island and mainland populations may have been sufficient to prevent extensive genomic differentiation at neutral loci following population isolation. Note, however, that coalescent simulations assume non-overlapping generations, which may limit their ability to accurately estimate demographic parameters in long-lived species, including conifers. Consequently, gene flow estimated between island and mainland populations may have been overestimated or may possibly be an artefact

resulting from an attempt of the demographic model to account for shared ancestral genetic variation among populations.

Phenotypic monitoring using common garden experiments or *in situ* morphological observations for cone, seed, and needle morphology have previously suggested genetically-based phenotypic divergence among Torrey pine populations (J. R. Haller 1986; J. A. Hamilton et al. 2017; Di Santo, Polgar, et al. 2021). To test for the role of selection across loci, we simulated a null F_{ST} distribution to compare with our empirical F_{ST} distribution, which indicated that a few thousand loci may be under divergent selection (Figure D5). Genome scans further supported this observation, identifying 2,053 (2.21%) candidate SNPs with accentuated divergence between island and mainland populations (Figure 16). Annotation of sequences containing these outliers SNPs suggested that adaptive evolution in Torrey pine may not only result from genetic differentiation at the nuclear level, but also at the mitochondrial level (Table 6). This could be consistent with previous observations of the importance of cytoplasmic genetic differences as a factor contributing to local adaptation in plants (Leinonen, Remington, and Savolainen 2011; Leinonen et al. 2013; J. A. Hamilton and Aitken 2013; for a review see Bock, Andrew, and Rieseberg 2014). For example, Leinonen, Remington, and Savolainen (2011) found using a reciprocal transplant experiment that individuals of *Arabidopsis lyrata* harboring the local cytoplasmic genome had higher fitness than individuals harboring the non-local cytoplasmic genome, suggesting that cytoplasmic genetic variation may contribute to local adaptation.

Overall, GO annotation indicated that genes important for mechanisms such as DNA integration, methylation, gene silencing, carbohydrate transport and metabolic processes, and defense against pathogens (bacteria and fungi) were candidates for selection (Table 6). This suggests that between the island and mainland environments, modification of genetic

composition and architecture following DNA integration, changes in gene expression or protein function following methylation and gene silencing, and direct or indirect selection against pathogens may have played an important role in population divergence following isolation. For example, a candidate gene associated with defense against bacteria and fungi (UniProt accession: B8LLJ5, GO terms: GO:0042742, GO:0050832, GO:0031640) suggests that phenotypic differentiation may have evolved in response to pests or pathogens. Indeed, the mainland population of Torrey pine may have faced substantial selection associated with the recent outbreak of the California five-spined ips (*Ips paraconfusus* Lanier) (Shea and Neustein 1995; J. Franklin and Santos 2011), whereas the island population may not have been exposed to that selective pressure. Noteworthy with these results is the fact that pine genomes are enormous (Stevens et al. 2016; Grotkopp et al. 2004), and our sequencing approach (reduced representation sequencing; see Material and methods) represents only a fraction of the Torrey pine genome. This suggests that, despite differences observed, some variation playing a critical role in local adaptation for this species has most likely been overlooked.

While identification of the appropriate effective population size necessary to protect adaptive evolutionary potential for rare species is still debated, recommendations generally range between 500 to 5000 individuals (I. R. Franklin and Frankham 1998; Lynch and Lande 1998; Frankham, Bradshaw, and Brook 2014). Torrey pine, critically endangered and endemic to just two native populations, suffers from extremely low effective population size (N_I = 2,305, N_M = 1,715) relative to other pines (Menon et al. 2018; Xia et al. 2018). Given historical and contemporary estimates of effective population size as well as contemporary estimates of expected heterozygosity, our results indicate that Torrey pine may not retain the genetic variation within populations needed to adapt to change. Current monitoring within the mainland

population suggests that a lack of recruitment (personal observation), infestation by *Ips* beetles (personal observation), and climate warming (Diffenbaugh, Swain, and Touma 2015) may increase extinction risk. Thus, for Torrey pine, increased Ne and greater genetic diversity may be required for long-term persistence.

As genetic variation is extremely low within populations, one conservation strategy that may facilitate the maintenance of genetic variation within populations at risk is a genetic rescue program. A genetic rescue program would facilitate inter-population breeding as a means to increase heterozygosity, increasing rates of inter-population gene flow. Indeed, demographic modeling indicates that following population isolation some gene flow has been maintained, resulting in low genome-wide differentiation among populations (Nei's $F_{ST} = 0.013$). However, the combination of observed phenotypic differences and large number of genes that appear targets of selection suggest island and mainland populations of Torrey pine have undergone distinct evolutionary trajectories necessary for adaptation following isolation. Thus, a genetic rescue program should be considered with caution as gene flow between populations may disrupt local adaptation and further reduce population performance (Montalvo and Ellstrand 2001; Hufford and Mazer 2003; Goto et al. 2011). Despite this word of caution, preliminary data comparing mainland, island, and F1 individuals from a common garden experiment planted outside the species natural distribution indicate that F1s exhibit increased fitness relative to mainland and island populations (J. A. Hamilton et al. 2017). Consequently, future monitoring is needed to empirically quantify fitness consequences of advanced-generation admixture (F2, Backcross-Island (BC-I), Backcross-Mainland (BC-M)) following early-generation heterosis.

Given the challenge to conserve and manage rare species in a rapidly changing environment, the use of genomic data to model evolutionary history, assess demographic change,

and tease apart the contributions of neutral and adaptive processes will be critical. For Torrey pine, the fact that there is low genome-wide differentiation among populations, a consistent history of low effective population size, and indications that some gene flow is maintained among populations may suggest that one population (island or mainland) could be targeted to effectively preserve neutral genetic variation. However, the combination of outlier loci and previously observed phenotypic differences suggest if the goal is to preserve adaptive genetic variation, a strategy that favors conservation efforts across both mainland and island populations will be needed. If a conservation strategy such as genetic rescue is considered, assessment of multiple admixed generations within a common environment will provide the necessary empirical test to evaluate the consequences of enhancing genetic exchange among populations.

4.6. Acknowledgements

This work was funded by USDA Health Protection Gene Conservation program and Western Wildlands Environmental Threat Assessment Center to JAH and JWW, Morton Arboretum Center for Tree Science Fellowship to JAH, and a new faculty award from the office of North Dakota Experimental Program to Stimulate Competitive Research (ND-EPSCoR NSF-IIA-1355466) and the NDSU Environmental and Conservation Sciences Graduate Program to JAH and LNDS. For their help collecting needle tissue, the authors would like to thank Stephen Johnson, Jill Wulf, Forest Swaciak, Conner Harrington, Emma Ordemann, Drew Peterson, Andrew Bower, Gary Man, Valerie Gallup, Annette Delfino Mix, Stephanie Steel, and tree climbers from San Diego Zoo Wildlife Alliance. For their help coordinating sampling on Santa Rosa Island and at the Torrey Pine State Reserve, the authors also would like to thank Paula Power (NPS), Rocky Rudolph (NPS), Kathryn McEachern (USGS), Robyn Shea (California State University – Channel Islands), and Darren Smith (California State Parks). Finally, the

authors would like to acknowledge the Michumash and the Kumeyaay people as the traditional caretakers of the Torrey pine ecosystems sampled for this study. Any use of product names is for informational purposes only and does not imply endorsement by the US Government.

5. GENERAL CONCLUSIONS

5.1. Introduction

The extent of current global biodiversity loss is unparalleled: 18% of vertebrates, 22% of invertebrates, 40% of plants, and 28% of fungi and protists are considered threatened (IUCN 2021b). This rapid loss of species diversity has led to discussions about what may arguably be the sixth global extinction event, through which a large proportion of the world's biodiversity is predicted to be lost (Cannon 2017; Barnosky et al. 2011; Butchart et al. 2010). In response to this alarming trend, extensive conservation efforts have been undertaken to conserve species' genetic diversity. Important for both individuals' fitness and populations' evolutionary potential (L. F. Keller and Waller 2002; Carlson, Cunningham, and Westley 2014), preservation of genetic variation is critical for the long-term persistence of rare species.

Ex situ conservation preserves genetic variability outside species' natural distribution, providing an important resource for restoration and reforestation should anything happen to natural populations. Within this context, *ex situ* collections have traditionally aimed to capture 95% of genetic variation present within species. Various approaches have been developed over the years to meet this objective (Gapare, Yanchuk, and Aitken 2008; Brown and Marshall 1995; Hoban and Schlarbaum 2014; Hoban 2019; Caujapé-Castells and Pedrola-Monfort 2004). However, effective approaches and guidelines to optimize population sampling when genetic data may be lacking remained limited. With the first two chapters of my dissertation, I aimed to fill this gap, examining potential surrogates for estimating genetic variation that may be used to optimize sampling efforts in *ex situ* conservation planning.

Nonetheless, the advent of next-generation sequencing, and the ever-increasing affordability of these technologies, has provided scientists with greater access to genomic data.

As the availability of genome-wide markers (i.e., single-nucleotide polymorphisms) via reduced representation sequencing (e.g., restriction-site-associated DNA sequencing, genotyping by sequencing) and whole genome sequencing increased, new methods were developed to analyze these data, including coalescent simulators (Gutenkunst et al. 2010; Excoffier and Foll 2011) and genome scans (Foll and Gaggiotti 2008; M. C. Whitlock and Lotterhos 2014; Privé et al. 2020). Combined, these new analytical methods provide a means to infer species' demographic and adaptive histories (X. Wang, Bernhardsson, and Ingvarsson 2020; Menon et al. 2018; Xia et al. 2018). Within a conservation context, access to such information may aid in establishing optimal conservation decisions, especially when preservation of standing and adaptive genetic diversity or genetic rescue are considered as potential management strategies. The fourth research chapter of my dissertation focused on inferring the evolutionary history of Torrey pine and assessing whether an understanding of species' evolutionary history inferred from genomic data can be leveraged to guide conservation of rare species' genetic variation.

In a world where ecological and anthropogenic disturbances predominate, these approaches provide timely and necessary tools to informing species' conservation. Specifically, this study demonstrated that where preservation of species' evolutionary potential is desired, strategies leveraging environmental, geographic, and genomic variation can aid in establishing optimal conservation recommendations to preserving populations' genetic diversity. Overall, methods presented throughout this dissertation provide a set of new and complementary approaches to be included within the conservation genetics toolbox.

5.2. Guiding *ex situ* sampling: potential surrogates to estimate genetic variation

Tapping into ecological, evolutionary, and quantitative genetic theories, I evaluated and proposed two potential proxies to estimate genetic variation within species that may be used to guide and optimize conservation efforts.

5.2.1. Population provenance and environmental variation

As genetic differences may arise among populations from both adaptive (natural selection and local adaptation) and neutral (genetic drift) evolutionary processes, I first hypothesized that a sampling strategy prioritizing source populations based on environmental and geographic distances would capture increased genetic diversity and differences relative to a sampling strategy prioritizing populations randomly. Consistent with my expectations, results from chapter 2 demonstrated that population provenance and environmental variation can be used to target and increase genetic diversity and differences captured in *ex situ* collections, particularly variation and differentiation at adaptive and functional loci. Although previous studies already demonstrated the benefits of leveraging environmental and geographic data to optimize adaptive genetic variation preserved (Hanson et al. 2017), this study is the first to show that distance-informed population sampling can also increase variation and differentiation captured at functional loci. This result has important conservation implications, as it indicates that population sampling based on environmental and geographic distances may not only capture contemporary adaptation, but may also preserve genetic diversity critical for future adaptive evolution. Note, however, that I assumed in this study that genetic material of maternal plants was sampled for inclusion in *ex situ* collections, while plant germplasm, such as seeds, are generally more likely to be targeted (D.-Z. Li and Pritchard 2009; Y.-B. Fu, Ahmed, and Diederichsen 2015). Consequently, future work evaluating whether geographically- and

environmentally-informed sampling can optimize genetic diversity captured *ex situ* across generations, for example using observed or simulated offspring genetic data, is needed. In addition, benefits of using environmental and geographic variation to optimize genetic variability captured across broad genetic diversity categories (neutral, functional, adaptive) was tested by pooling distinct genetic markers together (i.e., microsatellites, single-nucleotide polymorphisms). Future work may thus consider assessing marker-specific patterns associated with environmental and geographic distances to evaluate the ability of these distances to estimate genetic diversity at markers with contrasting allelic and mutation models.

5.2.2. Seed morphological variation

Phenotypic variation in plants is influenced by the environment the plant is experiencing (V_E), its genetic composition (V_G), and the interaction between its genetic composition and environment (V_{GxE}). Traits for which genetic variability explains a considerable proportion of phenotypic variation may be of particular interest to conservation, as they could potentially be used as surrogates for estimating within- and among-population genetic diversity. Interestingly, various traits underlying seed morphology in agronomically important species and coniferous trees have been shown to be highly heritable (genetically determined) (Roy, Thapliyal, and Phartyal 2004; Carles et al. 2009; Zas and Sampedro 2015; Pandey, Seshu, and Akbar 1994; Cober, Voldeng, and Fregeau-Reid 1997; Mera et al. 2004; Hakim and Suyamto 2017). For plant species conserved in seed banks, seed morphological variation could thus possibly be used to evaluate genetic variation maintained in *ex situ* collections. With chapter 3, I evaluated and described an empirical approach leveraging seed trait variation estimated from existing seed collections to inform future *ex situ* sampling efforts for rare species, using Torrey pine as a model. Torrey pine represented an ideal study system as large *ex situ* seed collections were

available for morphological assessment across the range of the species and seed traits in conifers have been found to be largely heritable. The results of this study indicated that both population origin (island, mainland) and maternal family within populations explained a significant portion of observed variation in seed morphology, suggesting within- and between-population genetic structure exist within the species. From an *ex situ* conservation perspective, these findings have important implications. First, population morphological differentiation indicated that optimizing trait variation preserved *ex situ* would require sampling of both island and mainland populations. Second, within-population differentiation in seed morphology suggested that to optimize morphological variation preserved in future seed collections, a strategy maximizing the number of maternal families sampled within populations should be favored. Third, greater morphological variation observed within the island population indicated that, if one aim to capture equal amount of morphological variation across populations, additional sampling efforts on the island would be needed. Specifically, resampling simulations demonstrated that to capture 95% of trait variation present within existing seed collections, 71% (57 individuals) and 83% (25 individuals) of maternal families within mainland and island seed populations should be sampled, respectively. Although providing a powerful tool to optimize seed morphological variation, and thus possibly genetic variation, captured in future *ex situ* conservation collections, this approach comes with a few caveats. Theoretically applicable to any plant species with largely heritable seed traits, this approach has only been tested in Torrey pine, which possesses rather large seeds. Future work evaluating this approach in species with smaller seeds would provide valuable insights into the breadth of its applicability. In addition, as discussed in chapter 3, seed traits may often be targets for natural selection. Depending on the strength with which natural selection shape seed morphology, potentially reducing additive genetic variance in these traits, genetic diversity

approximated from seed trait variation may be underestimated. Future work jointly modeling phenotypic and genetic diversity through simulation, for example using SLiM3 (B. C. Haller and Messer 2019), would help quantify the impact selection for fix phenotypes may have on estimating genetic variation using selectable quantitative traits. Finally, recommendations made using this method will always be relative to seed morphological variation captured within existing seed collections. To extrapolate to wild populations, one would need to assume that contemporary seed collections captured all natural seed trait variation occurring within populations, a condition that is not likely met.

5.3. Preserving rare species' evolutionary potential in the genomics era

Chapter 4 demonstrated that where genomic data are available, they may not only be used to evaluate the distribution of genetic variability within species, but may also be leveraged to infer species' evolutionary history. Understanding a species' evolutionary history can help inform conservation decisions by tracking temporal changes in populations' effective sizes (Ne), and thus variation in populations' evolutionary potential and probability of inbreeding across time. Additionally, insights into populations' migration and adaptive history gained from such knowledge can help evaluate the contribution of neutral and nonneutral evolutionary processes to population genetic differentiation. Inference regarding the demographic and adaptive history of the critically endangered Torrey pine indicated that populations have consistently suffered from low effective sizes across time, suggesting that the species may suffer from reduced evolutionary potential. Consequently, a conservation strategy to increase genetic diversity within population, such as genetic rescue, may be desired. However, demographic simulations and genome scans also revealed that, despite evidence for limited gene flow between populations, a number of loci may be targets for selection. Potential disruption of local adaptation following admixture thus

advise consideration of risks associated with increasing inter-population gene flow within the species. In addition, while low genome-wide differentiation and history of gene-flow between populations suggested most of genetic variation within Torrey pine may be captured by conserving only one population (island or mainland), patterns of divergent selection suggest that to preserve adaptive genetic diversity, conservation of both populations is needed. Note, however, that limitations associated with methods warrant caution when interpreting these results. First, coalescent simulators often assume discrete, non-overlapping generations. For long-lived species, such as Torrey pine, shared ancestral genetic variation among populations unaccounted for in simulations may lead to biased demographic parameters, including overestimated effective population sizes and migrations rate, as well as underestimated time of population divergence. In addition, pines are well known for their substantial genomes, varying between approximately 20-40 GB in size (Grotkopp et al. 2004; Stevens et al. 2016). Consequently, the reduced representation sequencing approach (ddRAD) used in this study is likely to have sampled only a fraction of Torrey pine's genome. This may have reduced the accuracy with which genetic summary statistics were estimated, as well as limited our ability to uncover genetic variation involved in adaptations to island and mainland environments. Future work using sequencing technologies enabling greater genomic representation, such as PacBio or nanopore, would help evaluate the impact short-read sequencing may have on estimating genetic summary statistics in species with large genomes, and improve our understanding of local adaptation in Torrey pine. Finally, future collaborative efforts towards sequencing and annotating Torrey pine's genome will be invaluable to study the distribution of genetic variants and their physical linkage across the genome, providing a powerful means to decipher the genomic basis of adaptive evolution within the species.

Although associated with limitations deserving consideration, this research revealed that understanding species' evolutionary history may help assess risks and benefits of inter-population gene flow and inform conservation actions needed to preserve rare species' evolutionary potential. Nonetheless, while this approach can caution against potential risks associated with genetic rescue, only an advanced-generation common garden can empirically evaluate fitness outcomes following population crossing and predict the success of genetic rescue as a conservation strategy. Indeed, studying fitness-related traits in first- and later-generation hybrids will help evaluate whether population crossing increases or reduces fitness in progenies, and whether these beneficial or adverse fitness effects are maintained across generations.

5.4. Future direction

As species threatened with extinction continue to multiply, tools to effectively guide and optimize conservation decisions are needed. This research focused on developing new and complementary approaches to guide conservation of species' evolutionary potential, particularly when limited genetic information is available. I demonstrated that environmental and geographic distances, as well as variation in seed morphology estimated from contemporary seed collections can provide useful proxies for genetic diversity and serve as a basis for *ex situ* conservation decisions making. Additionally, I showed that an understanding of rare species' evolutionary history gained from genomic data can be used to tease apart stochastic and deterministic forces driving population genetic differentiation and inform conservation management strategies, including genetic rescue. Here, I focused on approaches leveraging species' geographic distributions (Chapter 2), phenotypic variation (Chapter 3), and genomic variation (Chapter 4) for conservation decision making. However, as next-generation sequencing costs continue to decrease, availability of genomic datasets will increase, facilitating integration of genomic data

into phenotypic studies and opening up new perspectives for the preservation of species' adaptive potential.

Species exhibiting exceedingly low genetic variation may benefit from a conservation strategy aiming to increase genetic diversity within populations, such as genetic rescue. However, if reproductive barriers have evolved among populations following environmental or geographic isolation, implementation of such a strategy may reduce the fitness of admixed populations rather than promoting their recovery. In this study, we used demographic modeling and genome scans as tools to infer species' evolutionary history. However, while genome scans can suggest the action of divergent selection in population genetic differentiation and caution against potential risks associated with augmenting or restoring gene flow between isolated populations, they do not allow the assessment of reproductive isolation. Although only advanced-generation common garden experiments leveraging parent and hybrid populations can empirically test for the evolution of reproductive barriers, combining fitness-related trait variation with genomic ancestry from earlier-generation common gardens may provide a means to unravel the genetic basis of reproductive isolation among species' populations. Indeed, future studies pairing genetic simulations with genome-wide association analyses to detect low-admixture genomic regions associated with reduced fitness in heterozygous F1 hybrids may help identify areas of species' genome putatively important for reproductive isolation. Ultimately, these data could be used to inform species' conservation, including, for example, the delineation of conservation units or the viability of genetic rescue as a management strategy.

5.5. Conclusion

We are currently witnessing a revolution in the availability of next-generation sequencing technologies for non-model systems and the continuing development of new and more

sophisticated analytical and simulation tools for application to evolutionary and ecological questions. From a conservation perspective, combining these different approaches will advance our understanding of the factors influencing the distribution of species' genomic variation, will tease apart the genomic complexity of phenotypic trait determination, and address outstanding questions regarding the evolution of inter-specific genetic boundaries. Broadly, these advances are providing the scientific community critical tools essential to the preservation, maintenance, and promotion of biological diversity.

REFERENCES

Abe, Tetsuto, and Michio Matsunaga. 2011. "Geographic Variation in Germination Traits in Melia Azedarach and Rhaphiolepis Umbellata." *American Journal of Plant Sciences* 2 (1): 52–55. https://doi.org/10.4236/ajps.2011.21007.

Abebe, Tiegist D, Ali A Naz, and Jens Léon. 2015. "Landscape Genomics Reveal Signatures of Local Adaptation in Barley (Hordeum Vulgare L.)." *Frontiers in Plant Science* 6: 813. https://doi.org/10.3389/fpls.2015.00813.

Abràmoff, Michael D, Paulo J Magalhães, and Sunanda J Ram. 2004. "Image Processing with ImageJ." *Biophotonics International*. Laurin Publishing.

Aguilar, Ramiro, Mauricio Quesada, Lorena Ashworth, Yvonne Herrerias-Diego, and Jorge Lobo. 2008. "Genetic Consequences of Habitat Fragmentation in Plant Populations: Susceptible Signals in Plant Traits and Methodological Approaches." *Molecular Ecology* 17 (24): 5177–88. https://doi.org/10.1111/j.1365-294X.2008.03971.x.

Baddeley, Adrian, Ege Rubak, and Rolf Turner. 2015. *Spatial Point Patterns: Methodology and Applications with R*. New York: CRC press.

Bamberg, John, Alfonso Del Rio, Charles Fernandez, Alberto Salas, Sandra Vega, Cinthya Zorrilla, Willy Roca, and David Tay. 2010. "Comparison of 'Remote' versus 'Easy' in Situ Collection Locations for USA Wild Solanum (Potato) Germplasm." *American Journal of Potato Research* 87 (3): 277–84. https://doi.org/10.1007/s12230-010-9133-8.

Barker, Brittany S, Krikor Andonian, Sarah M Swope, Douglas G Luster, and Katrina M Dlugosch. 2017. "Population Genomic Analyses Reveal a History of Range Expansion and Trait Evolution across the Native and Invaded Range of Yellow Starthistle (Centaurea Solstitialis)." *Molecular Ecology* 26 (4): 1131–47. https://doi.org/10.1111/mec.13998.

Barnosky, Anthony D, Nicholas Matzke, Susumu Tomiya, Guinevere O U Wogan, Brian Swartz, Tiago B Quental, Charles Marshall, Jenny L McGuire, Emily L Lindsey, and Kaitlin C Maguire. 2011. "Has the Earth's Sixth Mass Extinction Already Arrived?" *Nature* 471 (7336): 51. https://doi.org/10.1038/nature09678.

Barrett, Rowan D H, and Dolph Schluter. 2008. "Adaptation from Standing Genetic Variation." *Trends in Ecology & Evolution* 23 (1): 38–44. https://doi.org/10.1016/j.tree.2007.09.008.

Barrett, Spencer C H. 1996. "The Reproductive Biology and Genetics of Island Plants." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351 (1341): 725–33. https://doi.org/10.1098/rstb.1996.0067.

Bartoń, Kamil. 2020. "MuMIn: Multi-Model Inference." https://cran.r-project.org/package=MuMIn.

Basey, Adrienne C, Jeremie B Fant, and Andrea T Kramer. 2015. “Producing Native Plant Materials for Restoration: 10 Rules to Collect and Maintain Genetic Diversity.” *Native Plants Journal* 16 (1): 37–53. https://doi.org/10.3368/npj.16.1.37.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bell, Graham. 2013. “Evolutionary Rescue and the Limits of Adaptation.” *Phil. Trans. R. Soc. B* 368 (1610): 20120080. https://doi.org/10.1098/rstb.2012.0080.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bernhardsson, Carolina, and Park Ingvarsson. 2012. “Geographical Structure and Adaptive Population Differentiation in Herbivore Defence Genes in European Aspen (Populus Tremula L., Salicaceae).” *Molecular Ecology* 21 (9): 2197–2207. https://doi.org/10.1111/j.1365-294X.2012.05524.x.

Biddick, Matthew, Annemieke L H Hendriks, and Kevin C Burns. 2019. “Plants Obey (and Disobey) the Island Rule.” *Proceedings of the National Academy of Sciences* 116 (36): 17632–34. https://doi.org/10.1073/pnas.1907424116.

BLM. 2012. “Bureau of Land Management Technical Protocol for the Collection, Study, and Conservation of Seeds from Native Plant Species for Seeds of Success.”

Blum, Michael G B, and Olivier François. 2010. “Non-Linear Regression Models for Approximate Bayesian Computation.” *Statistics and Computing* 20 (1): 63–73. https://doi.org/10.1007/s11222-009-9116-0.

Bock, Dan G, Rose L Andrew, and Loren H Rieseberg. 2014. “On the Adaptive Value of Cytoplasmic Genomes in Plants.” *Molecular Ecology* 23 (20): 4899–4911. https://doi.org/10.1111/mec.12920.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bonin, Aurelie, Florence Nicole, Francois Pompanon, Claude Miaud, and Pierre Taberlet. 2007. “Population Adaptive Index: A New Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for Conservation.” *Conservation Biology* 21 (3): 697–708. https://doi.org/10.1111/j.1523-1739.2007.00685.x.

Bossuyt, Beatrijs. 2007. “Genetic Rescue in an Isolated Metapopulation of a Naturally Fragmented Plant Species, Parnassia Palustris.” *Conservation Biology* 21 (3): 832–41. https://doi.org/10.1111/j.1523-1739.2006.00645.x.

Bramlett, David L, and Thomas W Popham. 1971. “Model Relating Unsound Seed and Embryonic Lethal Alleles in Self-Pollinated Pines.” *Silvae Genet* 20: 192–93.

Brown, Anthony H D, and J D Briggs. 1991. “Sampling Strategies for Genetic Variation in Ex Situ Collections of Endangered Plant Species.” *Genetics and Conservation of Rare Plants. Oxford University Press, New York*, 99–119.

Brown, Anthony H D, and D R Marshall. 1995. “A Basic Sampling Strategy: Theory and Practice.” In *Collecting Plant Genetic Diversity: Technical Guidelines*, edited by Luigi Guarino, V Ramanatha Rao, and Robert Reid, 75–91. Wallingford: CAB International.

Burns, Kevin C. 2016. “Size Changes in Island Plants: Independent Trait Evolution in Alyxia Ruscifolia (Apocynaceae) on Lord Howe Island.” *Biological Journal of the Linnean Society* 119 (4): 847–55. https://doi.org/10.1111/bij.12851.

———. 2018. “Time to Abandon the Loss of Dispersal Ability Hypothesis in Island Plants: A Comment on García-Verdugo, Mairal, Monroy, Sajeva and Caujapé-Castells (2017).” *Journal of Biogeography* 45 (6): 1219–22. https://doi.org/10.1111/jbi.13223.

Burns, Kevin C, Nadine Herold, and Ben Wallace. 2012. “Evolutionary Size Changes in Plants of the South-west Pacific.” *Global Ecology and Biogeography* 21 (8): 819–28. https://doi.org/10.1111/j.1466-8238.2011.00730.x.

Butchart, Stuart H M, Matt Walpole, Ben Collen, Arco Van Strien, Jörn P W Scharlemann, Rosamunde E A Almond, Jonathan E M Baillie, Bastian Bomhard, Claire Brown, and John Bruno. 2010. “Global Biodiversity: Indicators of Recent Declines.” *Science*, 1187512. https://doi.org/10.1126/science.1187512.

Çamaş, Necdet, and Enver Esendal. 2006. “Estimates of Broad-sense Heritability for Seed Yield and Yield Components of Safflower (Carthamus Tinctorius L.).” *Hereditas* 143 (2006): 55–57. https://doi.org/10.1111/j.2006.0018-0661.01914.x.

Campbell, Ian D, Karen McDonald, Michael D Flannigan, and Joanni Kringayark. 1999. “Long-Distance Transport of Pollen into the Arctic.” *Nature* 399 (6731): 29–30. https://doi.org/10.1038/19891.

Cannon, Charles H. 2017. “Assisted Diversification for an Era of Habitat Extinction.” In *Proceedings of Workshop on Gene Conservation of Tree Species–Banking on the Future*, 71. Portland: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. https://www.fs.usda.gov/treesearch/pubs/55174.

Canty, Angelo, and B D Ripley. 2021. “Boot: Bootstrap R (S-Plus) Functions.” https://cran.r-project.org/package=boot.

Carles, Sylvie, Mohammed S Lamhamedi, Jean Beaulieu, Debra C Stowe, Fabienne Colas, and Hank A Margolis. 2009. “Genetic Variation in Seed Size and Germination Patterns and Their Effect on White Spruce Seedling Characteristics.” *Silvae Genetica* 58 (1–6): 152–61. https://doi.org/10.1515/sg-2009-0020.

Carlson, Stephanie M, Curry J Cunningham, and Peter A H Westley. 2014. “Evolutionary Rescue in a Changing World.” *Trends in Ecology and Evolution* 29 (9): 521–30. https://doi.org/10.1016/j.tree.2014.06.005.

Castillo, Rafael F Del, Sonia Trujillo-Argueta, Nahúm Sánchez-Vargas, and Adrian C Newton. 2011. “Genetic Factors Associated with Population Size May Increase Extinction Risks and Decrease Colonization Potential in a Keystone Tropical Pine.” *Evolutionary Applications* 4 (4): 574–88. https://doi.org/https://doi.org/10.1111/j.1752-4571.2010.00177.x.

Caujapé-Castells, Juli, and Joan Pedrola-Monfort. 2004. “Designing Ex-Situ Conservation Strategies through the Assessment of Neutral Genetic Markers: Application to the Endangered Androcymbium Gramineum.” *Conservation Genetics* 5 (2): 131–44. https://doi.org/10.1023/B:COGE.0000029997.59502.88.

Cavender, Nicole, Murphy Westwood, Catherine Bechtoldt, Gerard Donnelly, Sara Oldfield, Martin Gardner, David Rae, and William McNamara. 2015. “Strengthening the Conservation Value of Ex Situ Tree Collections.” *Oryx* 49 (3): 416–24. https://doi.org/10.1017/S0030605314000866.

Chybicki, Igor J, Andrzej Oleksa, and Katarzyna Kowalkowska. 2012. “Variable Rates of Random Genetic Drift in Protected Populations of English Yew: Implications for Gene Pool Conservation.” *Conservation Genetics* 13 (4): 899–911. https://doi.org/10.1007/s10592-012-0339-9.

Cober, Elroy R, Harvey D Voldeng, and Judith A Fregeau-Reid. 1997. “Heritability of Seed Shape and Seed Size in Soybean.” *Crop Science* 37 (6): 1767–69. https://doi.org/10.2135/cropsci1997.0011183X003700060017x.

Coblentz, Bruce E. 1990. “Exotic Organisms: A Dilemma for Conservation Biology.” *Conservation Biology* 4 (3): 261–65. https://doi.org/10.1111/j.1523-1739.1990.tb00287.x.

Cody, Martin L, and Jacob McC Overton. 1996. “Short-Term Evolution of Reduced Dispersal in Island Plant Populations.” *Journal of Ecology* 84 (1): 53–61. https://doi.org/10.2307/2261699.

Cohen, Joel I, J Trevor Williams, Donald L Plucknett, and Henry Shands. 1991. “Ex Situ Conservation of Plant Genetic Resources: Global Development and Environmental Concerns.” *Science(Washington)* 253 (5022): 866–72. https://doi.org/10.1126/science.253.5022.866.

Critchfield, William B, and Elbert L Little. 1966. *Geographic Distribution of the Pines of the World*. Washington, DC: US Department of Agriculture, Forest Service.

Crnokrak, Peter, and Derek A Roff. 1999. “Inbreeding Depression in the Wild.” *Heredity* 83 (3): 260–70. https://doi.org/10.1038/sj.hdy.6885530.

Csilléry, Katalin, Olivier François, and Michael G B Blum. 2012. “Abc: An R Package for Approximate Bayesian Computation (ABC).” *Methods in Ecology and Evolution* 3 (3): 475–79. https://doi.org/10.1111/j.2041-210X.2011.00179.x.

Davison, Anthony C, and David V Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Daws, Matthew I, Nancy C Garwood, and Hugh W Pritchard. 2005. “Traits of Recalcitrant Seeds in a Semi-deciduous Tropical Forest in Panamá: Some Ecological Implications.” *Functional Ecology* 19 (5): 874–85. https://doi.org/10.1111/j.1365-2435.2005.01034.x.

Debain, Sandrine, Thomas Curt, and Jacques Lepart. 2003. “Seed Mass, Seed Dispersal Capacity, and Seedling Performance in a Pinus Sylvestris Population.” *Ecoscience* 10 (2): 168–75. https://doi.org/10.1080/11956860.2003.11682764.

Diffenbaugh, Noah S, Daniel L Swain, and Danielle Touma. 2015. “Anthropogenic Warming Has Increased Drought Risk in California.” *Proceedings of the National Academy of Sciences* 112 (13): 3931–36. https://doi.org/10.1073/pnas.1422385112.

Dowle, Matt, and Arun Srinivasan. 2019. “Data. Table: Extension of ‘Data. Frame.’” https://cran.r-project.org/package=data.table.

Doyle, Jeff J, and Jane L Doyle. 1987. “A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue.” *Phytochemical Bulletin* 19 (1): 11–15. http://worldveg.tind.io/record/33886.

Dudaniec, Rachael Y, Jessica Worthington Wilmer, Jeffrey O Hanson, Matthew Warren, Sarah Bell, and Jonathan R Rhodes. 2016. “Dealing with Uncertainty in Landscape Genetic Resistance Models: A Case of Three Co-occurring Marsupials.” *Molecular Ecology* 25 (2): 470–86. https://doi.org/10.1111/mec.13482.

Dusek, Karen H. 1985. “Update on Our Rarest Pine.” *American Forests* 91: 26–29, 61, 63.

Eaton, Deren A R, and Isaac Overcast. 2020. “Ipyrad: Interactive Assembly and Analysis of RADseq Datasets.” *Bioinformatics* 36 (8): 2592–94. https://doi.org/10.1093/bioinformatics/btz966.

Eckert, Andrew J, Joost van Heerwaarden, Jill L Wegrzyn, Dana C Nelson, Jeffrey Ross-Ibarra, Santíago C González-Martínez, and David B Neale. 2010. “Patterns of Population Structure and Environmental Associations to Aridity across the Range of Loblolly Pine (Pinus Taeda L., Pinaceae).” *Genetics* 185 (3): 969–82. https://doi.org/10.1534/genetics.110.115543.

Ellstrand, Norman C, and Diane R Elam. 1993. “Population Genetic Consequences of Small Population Size: Implications for Plant Conservation.” *Annual Review of Ecology and Systematics* 24 (1): 217–42. https://doi.org/10.1146/annurev.es.24.110193.001245.

Engels, J M M, L Maggioni, N Maxted, and M E Dulloo. 2008. “Complementing in Situ Conservation with Ex Situ Measures.” In *Conserving Plant Genetic Diversity in Protected Areas: Population Management of Crop Wild Relatives.*, edited by J M Iriondo, N Maxted, and M E Dulloo, 169. Wallingford: CAB International. https://doi.org/10.1079/9781845932824.0169.

Excoffier, Laurent, Isabelle Dupanloup, Emilia Huerta-Sánchez, Vitor C Sousa, and Matthieu Foll. 2013. “Robust Demographic Inference from Genomic and SNP Data.” *PLoS Genetics* 9 (10): e1003905. https://doi.org/10.1371/journal.pgen.1003905.

Excoffier, Laurent, and Matthieu Foll. 2011. “Fastsimcoal: A Continuous-Time Coalescent Simulator of Genomic Diversity under Arbitrarily Complex Evolutionary Scenarios.” *Bioinformatics* 27 (9): 1332–34. https://doi.org/10.1093/bioinformatics/btr124.

Excoffier, Laurent, and Heidi E L Lischer. 2010. “Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows.” *Molecular Ecology Resources* 10 (3): 564–67. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Fagan, William F, and Elizabeth E Holmes. 2006. “Quantifying the Extinction Vortex.” *Ecology Letters* 9 (1): 51–60. https://doi.org/10.1111/j.1461-0248.2005.00845.x.

Fahrig, Lenore. 1997. “Relative Effects of Habitat Loss and Fragmentation on Population Extinction.” *The Journal of Wildlife Management* 61 (3): 603–10. https://doi.org/10.2307/3802168.

FAO. 2010. “The Second Report on the State of the World’s Plant Genetic Resources for Food and Agriculture.” Rome: Food & Agriculture Org.

Foll, Matthieu, and Oscar Gaggiotti. 2008. “A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective.” *Genetics* 180 (2): 977–93. https://doi.org/10.1534/genetics.108.092221.

Frankham, Richard. 1995. “Conservation Genetics.” *Annual Review of Genetics* 29 (1): 305–27. https://doi.org/10.1146/annurev.ge.29.120195.001513.

———. 1996. “Relationship of Genetic Variation to Population Size in Wildlife.” *Conservation Biology* 10 (6): 1500–1508. https://doi.org/10.1046/j.1523-1739.1996.10061500.x.

———. 1999. “Quantitative Genetics in Conservation Biology.” *Genetics Research* 74 (3): 237–44. https://doi.org/10.1017/S001667239900405X.

Frankham, Richard, Jonathan D Ballou, Mark D B Eldridge, Robert C Lacy, Katherine Ralls, Michele R Dudash, and Charles B Fenster. 2011. “Predicting the Probability of Outbreeding Depression.” *Conservation Biology* 25 (3): 465–75. https://doi.org/10.1111/j.1523-1739.2011.01662.x.

Frankham, Richard, Corey J A Bradshaw, and Barry W Brook. 2014. “Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses.” *Biological Conservation* 170: 56–63. https://doi.org/10.1016/j.biocon.2013.12.036.

Frankham, Richard, David A Briscoe, and Jonathan D Ballou. 2002. *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808999.

Franklin, Ian R, and Richard Frankham. 1998. “How Large Must Populations Be to Retain Evolutionary Potential?” *Animal Conservation* 1 (1): 69–70. https://doi.org/10.1111/j.1469-1795.1998.tb00228.x.

Franklin, Janet, and Elizabeth V Santos. 2011. “A Spatially Explicit Census Reveals Population Structure and Recruitment Patterns for Narrowly Endemic Pine, Pinus Torreyana.” *Plant Ecology* 212 (2): 293–306. https://doi.org/10.1007/s11258-010-9822-x.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. “CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data.” *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

Fu, Yong-Bi, Zaheer Ahmed, and Axel Diederichsen. 2015. “Towards a Better Monitoring of Seed Ageing under Ex Situ Seed Conservation.” *Conservation Physiology* 3 (1): cov026. https://doi.org/10.1093/conphys/cov026.

Gapare, Washington J, Alvin D Yanchuk, and Sally N Aitken. 2008. “Optimal Sampling Strategies for Capture of Genetic Diversity Differ between Core and Peripheral Populations of Picea Sitchensis (Bong.) Carr.” *Conservation Genetics* 9 (2): 411–18. https://doi.org/10.1007/s10592-007-9353-8.

Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read Sequencing.” *ArXiv*. https://arxiv.org/abs/1207.3907.

Gavrilets, Sergey, and Alan Hastings. 1996. “Founder Effect Speciation: A Theoretical Reassessment.” *The American Naturalist* 147 (3): 466–91. https://doi.org/10.1086/285861.

Gilpin, Michael E, and Michael E Soulé. 1986. “Minimum Viable Populations: Processes of Species Extinction.” In *Conservation Biology: An Evolutionary-Ecological Perspective*, edited by Michael E Soulé and Bruce A Wilcox, 19–34. Sunderland: Sinauer associates.

Gómez, José M. 2004. “Bigger Is Not Always Better: Conflicting Selective Pressures on Seed Size in Quercus Ilex.” *Evolution* 58 (1): 71–80. https://doi.org/10.1111/j.0014-3820.2004.tb01574.x.

Gomulkiewicz, Richard, and Robert D Holt. 1995. “When Does Evolution by Natural Selection Prevent Extinction?” *Evolution* 49 (1): 201–7. https://doi.org/10.2307/2410305.

Goto, Susumu, Hayato Iijima, Hitomi Ogawa, and Kazumi Ohya. 2011. "Outbreeding Depression Caused by Intraspecific Hybridization between Local and Nonlocal Genotypes in Abies Sachalinensis." *Restoration Ecology* 19 (2): 243–50. https://doi.org/10.1111/j.1526-100X.2009.00568.x.

Goudet, Jérôme. 2005. "Hierfstat, a Package for R to Compute and Test Hierarchical F-statistics." *Molecular Ecology Resources* 5 (1): 184–86. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Goudet, Jérôme, and Thibaut Jombart. 2020. "Hierfstat: Estimation and Tests of Hierarchical F-Statistics." https://cran.r-project.org/package=hierfstat.

Greene, David F, and Edward A Johnson. 1993. "Seed Mass and Dispersal Capacity in Wind-Dispersed Diaspores." *Oikos* 67 (1): 69–74. https://doi.org/10.2307/3545096.

Gribko, Linda S, and William E Jones. 1995. "Test of the Float Method of Assessing Northern Red Oak Acorn Condition." *Tree Planters' Notes* 46 (4): 143–47.

Griffith, Patrick, Michael Calonje, Alan W Meerow, Freddy Tut, Andrea T Kramer, Abby Hird, Tracy M Magellan, and Chad E Husby. 2014. "Can a Botanic Garden Cycad Collection Capture the Genetic Diversity in a Wild Population?" *International Journal of Plant Sciences* 176 (1): 1–10. https://doi.org/10.1086/678466.

Griffith, Patrick, and Chad Husby. 2010. "The Price of Conservation: Measuring the Mission and Its Cost." *BGJournal* 7 (1): 12–14. https://www.jstor.org/stable/24811072.

Griffith, Patrick, Carl Lewis, and Javier Francisco-Ortega. 2011. "Palm Conservation at a Botanic Garden: A Case Study of the Keys Thatch Palm." *Palms* 55 (2): 93–101.

Grotkopp, Eva, Marcel Rejmánek, Michael J Sanderson, and Thomas L Rost. 2004. "Evolution of Genome Size in Pines (Pinus) and Its Life-History Correlates: Supertree Analyses." *Evolution* 58 (8): 1705–29. https://doi.org/10.1111/j.0014-3820.2004.tb00456.x.

Guerrant, Edward O, Kayri Havens-Young, and Mike Maunder. 2004. *Ex Situ Plant Conservation: Supporting Species Survival in the Wild*. Edited by Edward O Guerrant, Kayri Havens-Young, Mike Maunder, and Michael Maunder. Illustrate. Vol. 3. Washington, DC: Island Press.

Guerrant Jr, Edward O, Kayri Havens, and Pati Vitt. 2014. "Sampling for Effective Ex Situ Plant Conservation." *International Journal of Plant Sciences* 175 (1): 11–20. https://doi.org/10.1086/674131.

Gutenkunst, Ryan, Ryan Hernandez, Scott Williamson, and Carlos Bustamante. 2010. "Diffusion Approximations for Demographic Inference: DaDi." *Nature Precedings*, 1–1. https://doi.org/10.1038/npre.2010.4594.1.

Hague, Michael T J, and Eric J Routman. 2016. “Does Population Size Affect Genetic Diversity? A Test with Sympatric Lizard Species.” *Heredity* 116 (1): 92–98. https://doi.org/10.1038/hdy.2015.76.

Hakim, Lukman, and Suyamto Suyamto. 2017. “Gene Action and Heritability Estimates of Quantitative Characters among Lines Derived from Varietal Crosses of Soybean.” *Indonesian Journal of Agricultural Science* 18 (1): 25–32. https://doi.org/10.21082/ijas.v18n1.2017.p25-32.

Hall, Travis, and Andrew Brinkman. 2015. “Population Dynamics of the Island Torrey Pine (Pinus Torreyana Ssp. Insularis) on Santa Rosa Island, CA.” California State University Channel Islands.

Haller, Benjamin C, and Philipp W Messer. 2019. “SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model.” *Molecular Biology and Evolution* 36 (3): 632–37. https://doi.org/10.1093/molbev/msy228.

Haller, John R. 1986. “Taxonomy and Relationships of the Mainland and Island Populations of Pinus Torreyana (Pinaceae).” *Systematic Botany* 11 (1): 39–50. https://doi.org/10.2307/2418944.

Hamilton, Jill A, and Sally N Aitken. 2013. “Genetic and Morphological Structure of a Spruce Hybrid (Picea Sitchensis× P. Glauca) Zone along a Climatic Gradient.” *American Journal of Botany* 100 (8): 1651–62. https://doi.org/10.3732/ajb.1200654.

Hamilton, Jill A, Raphaël Royauté, Jessica W Wright, Paul Hodgskiss, and Thomas F Ledig. 2017. “Genetic Conservation and Management of the California Endemic, Torrey Pine (Pinus Torreyana Parry): Implications of Genetic Rescue in a Genetically Depauperate Species.” *Ecology and Evolution* 7 (18): 7370–81. https://doi.org/10.1002/ece3.3306.

Hamilton, Kim N, Catherine A Offord, Peter Cuneo, and Myrna A Deseo. 2013. “A Comparative Study of Seed Morphology in Relation to Desiccation Tolerance and Other Physiological Responses in 71 Eastern Australian Rainforest Species.” *Plant Species Biology* 28 (1): 51–62. https://doi.org/10.1111/j.1442-1984.2011.00353.x.

Hamilton, Matthew B. 1994. “Ex Situ Conservation of Wild Plant Species: Time to Reassess the Genetic Assumptions and Implications of Seed Banks.” *Conservation Biology* 8 (1): 39–49. https://doi.org/10.1046/j.1523-1739.1994.08010039.x.

Hamrick, J L, and M J W Godt. 1996. “Effects of Life History Traits on Genetic Diversity in Plant Species.” *Philosophical Transactions: Biological Sciences*, 1291–98. https://doi.org/10.1098/rstb.1996.0112.

Hancock, Angela M, Benjamin Brachi, Nathalie Faure, Matthew W Horton, Lucien B Jarymowycz, Gianluca F Sperone, Chris Toomajian, Fabrice Roux, and Joy Bergelson. 2011. “Adaptation to Climate Across the Arabidopsis Thaliana Genome.” *Science* 334 (6052): 83–86. https://doi.org/10.1126/science.1209244.

Hanson, Jeffrey O, Jonathan R Rhodes, Cynthia Riginos, and Richard A Fuller. 2017. "Environmental and Geographic Variables Are Effective Surrogates for Genetic Variation in Conservation Planning." *Proceedings of the National Academy of Sciences* 114 (48): 12755–60. https://doi.org/10.1073/pnas.1711009114.

Harlan, Jack R. 1976. "Genetic Resources in Wild Relatives of Crops." *Crop Science* 16 (3): 329–33. https://doi.org/10.2135/cropsci1976.0011183X001600030004x.

Hausman, Constance E, Michelle M Bertke, John F Jaeger, and Oscar J Rocha. 2014. "Genetic Structure of Green Ash (Fraxinus Pennsylvanica): Implications for the Establishment of Ex Situ Conservation Protocols in Light of the Invasion of the Emerald Ash Borer." *Plant Genetic Resources* 12 (3): 286–97. https://doi.org/10.1017/S1479262114000033.

Hedrick, Philip W, Rolf O Peterson, Leah M Vucetich, Jennifer R Adams, and John A Vucetich. 2014. "Genetic Rescue in Isle Royale Wolves: Genetic Analysis and the Collapse of the Population." *Conservation Genetics* 15 (5): 1111–21. https://doi.org/10.1007/s10592-014-0604-1.

Heuertz, Myriam, Xavier Vekemans, J-F Hausman, M Palada, and Olivier J Hardy. 2003. "Estimating Seed vs. Pollen Dispersal from Spatial Genetic Structure in the Common Ash." *Molecular Ecology* 12 (9): 2483–95. https://doi.org/10.1046/j.1365-294X.2003.01923.x.

Hoban, Sean. 2019. "New Guidance for Ex Situ Gene Conservation: Sampling Realistic Population Systems and Accounting for Collection Attrition." *Biological Conservation* 235: 199–208. https://doi.org/10.1016/j.biocon.2019.04.013.

Hoban, Sean, Michael Bruford, Josephine D'Urban Jackson, Margarida Lopes-Fernandes, Myriam Heuertz, Paul A Hohenlohe, Ivan Paz-Vinas, et al. 2020. "Genetic Diversity Targets and Indicators in the CBD Post-2020 Global Biodiversity Framework Must Be Improved." *Biological Conservation* 248: 108654. https://doi.org/10.1016/j.biocon.2020.108654.

Hoban, Sean, and Scott Schlarbaum. 2014. "Optimal Sampling of Seeds from Plant Populations for Ex-Situ Conservation of Genetic Biodiversity, Considering Realistic Population Structure." *Biological Conservation* 177: 90–99. https://doi.org/10.1016/j.biocon.2014.06.014.

Holderegger, Rolf, Urs Kamm, and Felix Gugerli. 2006. "Adaptive vs. Neutral Genetic Diversity: Implications for Landscape Genetics." *Landscape Ecology* 21 (6): 797–807. https://doi.org/10.1007/s10980-005-5245-9.

Holliday, Jason A, Kermit Ritland, and Sally N Aitken. 2010. "Widespread, Ecologically Relevant Genetic Markers Developed from Association Mapping of Climate-related Traits in Sitka Spruce (Picea Sitchensis)." *New Phytologist* 188 (2): 501–14. https://doi.org/10.1111/j.1469-8137.2010.03380.x.

Hothorn, Torsten, and Kurt Hornik. 2021. "ExactRankTests: Exact Distributions for Rank and Permutation Tests." https://cran.r-project.org/package=exactRankTests.

Hsieh, Y-C, J-D Chung, Chun-Neng Wang, Chung-Te Chang, C-Y Chen, and Shih-Ying Hwang. 2013. “Historical Connectivity, Contemporary Isolation and Local Adaptation in a Widespread but Discontinuously Distributed Species Endemic to Taiwan, Rhododendron Oldhamii (Ericaceae).” *Heredity* 111 (2): 147–56. https://doi.org/10.1038/hdy.2013.31.

Huang, Chun-Lin, Jui-Hung Chen, Chung-Te Chang, Jeng-Der Chung, Pei-Chun Liao, Jenn-Che Wang, and Shih-Ying Hwang. 2016. “Disentangling the Effects of Isolation-by-Distance and Isolation-by-Environment on Genetic Differentiation among Rhododendron Lineages in the Subgenus Tsutsusi.” *Tree Genetics & Genomes* 12 (3): 53. https://doi.org/10.1007/s11295-016-1010-2.

Hufbauer, Ruth A, Marianna Szűcs, Emily Kasyon, Courtney Youngberg, Michael J Koontz, Christopher Richards, Ty Tuff, and Brett A Melbourne. 2015. “Three Types of Rescue Can Avert Extinction in a Changing Environment.” *Proceedings of the National Academy of Sciences* 112 (33): 10557–62. https://doi.org/10.1073/pnas.1504732112.

Hufford, Kristina M, and Susan J Mazer. 2003. “Plant Ecotypes: Genetic Differentiation in the Age of Ecological Restoration.” *Trends in Ecology & Evolution* 18 (3): 147–55. https://doi.org/10.1016/S0169-5347(03)00002-8.

Ilves, Aigi, Kaire Lanno, Marek Sammul, and Kadri Tali. 2013. “Genetic Variability, Population Size and Reproduction Potential in Ligularia Sibirica (L.) Populations in Estonia.” *Conservation Genetics* 14 (3): 661–69. https://doi.org/10.1007/s10592-013-0459-x.

IUCN. 2020. “The IUCN Red List of Threatened Species. Version 2020-2.” 2020. https://www.iucnredlist.org.

———. 2021a. “The IUCN Red List of Threatened Species. Version 2021-1.” 2021. https://www.iucnredlist.org.

———. 2021b. “The IUCN Red List of Threatened Species. Version 2021-2.” 2021. https://www.iucnredlist.org.

Johnson, Matthew, Stephen B Vander Wall, and Mark Borchert. 2003. “A Comparative Analysis of Seed and Cone Characteristics and Seed-Dispersal Strategies of Three Pines in the Subsection Sabinianae.” *Plant Ecology* 168 (1): 69–84. https://doi.org/10.1023/A:1024470224134.

Johnson, Warren E, David P Onorato, Melody E Roelke, Darrell E Land, Mark Cunningham, Robert C Belden, Roy McBride, Deborah Jansen, Mark Lotz, and David Shindle. 2010. “Genetic Restoration of the Florida Panther.” *Science* 329 (5999): 1641–45. https://doi.org/10.1126/science.1192891.

Jombart, Thibaut. 2008. “Adegenet: A R Package for the Multivariate Analysis of Genetic Markers.” *Bioinformatics* 24 (11): 1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Jombart, Thibaut, and Ismaïl Ahmed. 2011. “Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data.” *Bioinformatics* 27 (21): 3070–71. https://doi.org/10.1093/bioinformatics/btr521.

Kassambara, Alboukadel. 2020. “Rstatix: Pipe-Friendly Framework for Basic Statistical Tests.” https://cran.r-project.org/package=rstatix.

Kavanagh, Patrick H, and Kevin C Burns. 2014. “The Repeated Evolution of Large Seeds on Islands.” *Proceedings of the Royal Society B: Biological Sciences* 281 (1786): 20140675. https://doi.org/10.1098/rspb.2014.0675.

Keith, David A, H Resit Akçakaya, Wilfried Thuiller, Guy F Midgley, Richard G Pearson, Steven J Phillips, Helen M Regan, Miguel B Araújo, and Tony G Rebelo. 2008. “Predicting Extinction Risks under Climate Change: Coupling Stochastic Population Models with Dynamic Bioclimatic Habitat Models.” *Biology Letters* 4 (5): 560–63. https://doi.org/10.1098/rsbl.2008.0049.

Keller, Lukas F, and Donald M Waller. 2002. “Inbreeding Effects in Wild Populations.” *Trends in Ecology & Evolution* 17 (5): 230–41. https://doi.org/10.1016/S0169-5347(02)02489-8.

Keller, M, Johannes Kollmann, and Peter J Edwards. 2000. “Genetic Introgression from Distant Provenances Reduces Fitness in Local Weed Populations.” *Journal of Applied Ecology* 37 (4): 647–59. https://doi.org/10.1046/j.1365-2664.2000.00517.x.

Keller, Stephen R, Vikram E Chhatre, and Matthew C Fitzpatrick. 2017. “Influence of Range Position on Locally Adaptive Gene–Environment Associations in Populus Flowering Time Genes.” *Journal of Heredity* 109 (1): 47–58. https://doi.org/10.1093/jhered/esx098.

Keller, Stephen R, Nicholas Levsen, Matthew S Olson, and Peter Tiffin. 2012. “Local Adaptation in the Flowering-Time Gene Network of Balsam Poplar, Populus Balsamifera L.” *Molecular Biology and Evolution* 29 (10): 3143–52. https://doi.org/10.1093/molbev/mss121.

Krugman, Stanley L, and James L Jenkinson. 2008. “Pinus L. Pine.” In *The Woody Plant Seed Manual*, edited by Franklin T Bonner and Robert P Karrfalt, 727:809–47. Washington, DC: US Department of Agriculture, Forest Service.

Kuznetsova, Alexandra, Per B Brockhoff, and Rune H B Christensen. 2017. “LmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software* 82 (13): 1–26. https://doi.org/10.18637/jss.v082.i13.

Lande, Russell. 1988. “Genetics and Demography in Biological Conservation.” *Science* 241 (4872): 1455–60. https://doi.org/10.1126/science.3420403.

———. 1993. “Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes.” *The American Naturalist* 142 (6): 911–27. https://doi.org/10.1086/285580.

Laval, Guillaume, and Laurent Excoffier. 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): 2485–87. https://doi.org/10.1093/bioinformatics/bth264.

Lawrence, M J, David F Marshall, and P Davies. 1995. “Genetics of Genetic Conservation. I. Sample Size When Collecting Germplasm.” *Euphytica* 84 (2): 89–99. https://doi.org/10.1007/BF01677945.

Ledig, Thomas F. 1986. “Conservation Strategies for Forest Gene Resources.” *Forest Ecology and Management* 14 (2): 77–90. https://doi.org/10.1016/0378-1127(86)90093-9.

———. 1988. “The Conservation of Diversity in Forest Trees: Why and How Should Genes Be Conserved?” *BioScience* 38 (7): 471–79. https://doi.org/10.2307/1310951.

———. 2000. “Founder Effects and the Genetic Structure of Coulter Pine.” *Journal of Heredity* 91 (4): 307–15. https://doi.org/10.1093/jhered/91.4.307.

Ledig, Thomas F, and Thompson M Conkle. 1983. “Gene Diversity and Genetic Structure in a Narrow Endemic, Torrey Pine (Pinus Torreyana Parry Ex Carr.).” *Evolution* 37 (1): 79–85. https://doi.org/10.2307/2408176.

Ledig, Thomas F, J Jesús Vargas-Hernández, and Kurt H Johnsen. 1998. “The Conservation of Forest Genetic Resources: Case Histories from Canada, Mexico, and the United States.” *Journal of Forestry* 96 (1): 32–41. https://doi.org/10.1093/jof/96.1.32.

Leimu, Roosa, P I A Mutikainen, Julia Koricheva, and Markus Fischer. 2006. “How General Are Positive Relationships between Plant Population Size, Fitness and Genetic Variation?” *Journal of Ecology* 94 (5): 942–52. https://doi.org/10.1111/j.1365-2745.2006.01150.x.

Leinonen, Päivi H, David L Remington, Johanna Leppälä, and Outi Savolainen. 2013. “Genetic Basis of Local Adaptation and Flowering Time Variation in Arabidopsis Lyrata.” *Molecular Ecology* 22 (3): 709–23. https://doi.org/10.1111/j.1365-294X.2012.05678.x.

Leinonen, Päivi H, David L Remington, and Outi Savolainen. 2011. “Local Adaptation, Phenotypic Differentiation, and Hybrid Fitness in Diverged Natural Populations of Arabidopsis Lyrata.” *Evolution: International Journal of Organic Evolution* 65 (1): 90–107. https://doi.org/10.1111/j.1558-5646.2010.01119.x.

Lens, Frederic, Nicolas Davin, Erik Smets, and Marcelino del Arco. 2013. “Insular Woodiness on the Canary Islands: A Remarkable Case of Convergent Evolution.” *International Journal of Plant Sciences* 174 (7): 992–1013. https://doi.org/10.1086/670259.

Li, De-Zhu, and Hugh W Pritchard. 2009. “The Science and Economics of Ex Situ Plant Conservation.” *Trends in Plant Science* 14 (11): 614–21. https://doi.org/10.1016/j.tplants.2009.09.005.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *ArXiv*. https://arxiv.org/abs/1303.3997.

———. 2014. “Toward Better Understanding of Artifacts in Variant Calling from High-Coverage Samples.” *Bioinformatics* 30 (20): 2843–51. https://doi.org/10.1093/bioinformatics/btu356.

Li, Qiaoming, Zaifu Xu, and Tianhua He. 2002. “Ex Situ Genetic Conservation of Endangered Vatica Guangxiensis (Dipterocarpaceae) in China.” *Biological Conservation* 106 (2): 151–56. https://doi.org/10.1016/S0006-3207(01)00240-3.

Li, Sen, and Mattias Jakobsson. 2012. “Estimating Demographic Parameters from Large-Scale Population Genomic Data Using Approximate Bayesian Computation.” *BMC Genetics* 13 (1): 1–15. https://doi.org/10.1186/1471-2156-13-22.

Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics* 22 (13): 1658–59. https://doi.org/10.1093/bioinformatics/btl158.

Liu, Detuan, Lu Zhang, Jihua Wang, and Yongpeng Ma. 2020. “Conservation Genomics of a Threatened Rhododendron: Contrasting Patterns of Population Structure Revealed from Neutral and Selected SNPs.” *Frontiers in Genetics* 11: 757. https://doi.org/10.3389/fgene.2020.00757.

Lockwood, Dale R, Christopher M Richards, and Gayle M Volk. 2007. “Probabilistic Models for Collecting Genetic Diversity: Comparisons, Caveats, and Limitations.” *Crop Science* 47 (2): 861–66. https://doi.org/10.2135/cropsci2006.04.0262.

Loveless, Marilyn D, and James L Hamrick. 1984. “Ecological Determinants of Genetic Structure in Plant Populations.” *Annual Review of Ecology and Systematics* 15 (1): 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433.

Lowe, Andrew J, David Boshier, M Ward, Cecile F E Bacles, and Carlos M Navarro. 2005. “Genetic Resource Impacts of Habitat Loss and Degradation; Reconciling Empirical Evidence and Predicted Theory for Neotropical Trees.” *Heredity* 95 (4): 255–73. https://doi.org/10.1038/sj.hdy.6800725.

Lowry, David B, Cotton R Rockwood, and John H Willis. 2008. “Ecological Reproductive Isolation of Coast and Inland Races of Mimulus Guttatus.” *Evolution: International Journal of Organic Evolution* 62 (9): 2196–2214. https://doi.org/10.1111/j.1558-5646.2008.00457.x.

Lynch, Michael, and Russell Lande. 1998. “The Critical Effective Size for a Genetically Secure Population.” *Animal Conservation* 1 (1): 70–72. https://doi.org/10.1111/j.1469-1795.1998.tb00229.x.

Madsen, Thomas, Richard Shine, Mats Olsson, and Håkan Wittzell. 1999. “Restoration of an Inbred Adder Population.” *Nature* 402 (6757): 34–35. https://doi.org/10.1038/46941.

Marshall, D R, and Anthony H D Brown. 1975. “Optimum Sampling Strategies in Genetic Conservation.” In *Crop Genetic Resources for Today and Tomorrow*, edited by O H Frankel and J G Hawks, 53–80. London: Cambridge University Press, Cambridge.

Maruyama, Takeo, and Paul A Fuerst. 1985. “Population Bottlenecks and Nonequilibrium Models in Population Genetics. II. Number of Alleles in a Small Population That Was Formed by a Recent Bottleneck.” *Genetics* 111 (3): 675–89. http://www.genetics.org/content/111/3/675.abstract.

McAssey, Edward V, Jonathan Corbi, and John M Burke. 2016. “Range-Wide Phenotypic and Genetic Differentiation in Wild Sunflower.” *BMC Plant Biology* 16 (1): 249. https://doi.org/10.1186/s12870-016-0937-7.

McBride, Carolyn S, and Michael C Singer. 2010. “Field Studies Reveal Strong Postmating Isolation between Ecologically Divergent Butterfly Populations.” *PLoS Biology* 8 (10): e1000529. https://doi.org/10.1371/journal.pbio.1000529.

McKay, John K, John G Bishop, Jing-Zhong Lin, James H Richards, Anna Sala, and Thomas Mitchell-Olds. 2001. “Local Adaptation across a Climatic Gradient despite Small Effective Population Size in the Rare Sapphire Rockcress.” *Proceedings of the Royal Society of London B: Biological Sciences* 268 (1477): 1715–21. https://doi.org/10.1098/rspb.2001.1715.

McKay, John K, and Robert G Latta. 2002. “Adaptive Population Divergence: Markers, QTL and Traits.” *Trends in Ecology & Evolution* 17 (6): 285–91. https://doi.org/10.1016/S0169-5347(02)02478-3.

Meilleur, Brien A, and Toby Hodgkin. 2004. “In Situ Conservation of Crop Wild Relatives: Status and Trends.” *Biodiversity & Conservation* 13 (4): 663–84. https://doi.org/10.1023/B:BIOC.0000011719.03230.17.

Menon, Mitra, Justin C Bagley, Christopher J Friedline, Amy V Whipple, Anna W Schoettle, Alejandro Leal-Sàenz, Christian Wehenkel, Francisco Molina-Freaner, Lluvia Flores-Rentería, and Socorro M Gonzalez-Elizondo. 2018. “The Role of Hybridization during Ecological Divergence of Southwestern White Pine (Pinus Strobiformis) and Limber Pine (P. Flexilis).” *Molecular Ecology* 27 (5): 1245–60. https://doi.org/10.1111/mec.14505.

Mera, Mario, R Jerez, H Miranda, and J L Rouanet. 2004. “Seed Coat Specific Weight in Lupinus Angustifolius: Influence of Genotype and Environment and Relationship with Seed Coat Proportion.” *Australian Journal of Agricultural Research* 55 (11): 1189–95. https://doi.org/10.1071/AR04095.

Montalvo, Arlee M, and Norman C Ellstrand. 2000. “Transplantation of the Subshrub Lotus Scoparius: Testing the Home-site Advantage Hypothesis.” *Conservation Biology* 14 (4): 1034–45. https://doi.org/10.1046/j.1523-1739.2000.99250.x.

———. 2001. “Nonlocal Transplantation and Outbreeding Depression in the Subshrub Lotus Scoparius (Fabaceae).” *American Journal of Botany* 88 (2): 258–69. https://doi.org/10.2307/2657017.

Monty, Arnaud, Jean-Philippe Bizoux, José Escarré, and Grégory Mahy. 2013. “Rapid Plant Invasion in Distinct Climates Involves Different Sources of Phenotypic Variation.” *PLoS One* 8 (1): e55627. https://doi.org/10.1371/journal.pone.0055627.

Morina, Daniel L, Marcus A Lashley, M Colter Chitwood, Christopher E Moorman, and Christopher S DePerno. 2017. “Should We Use the Float Test to Quantify Acorn Viability?” *Wildlife Society Bulletin* 41 (4): 776–79. https://doi.org/10.1002/wsb.826.

Moyers, Brook T, and Loren H Rieseberg. 2016. “Remarkable Life History Polymorphism May Be Evolving under Divergent Selection in the Silverleaf Sunflower.” *Molecular Ecology* 25 (16): 3817–30. https://doi.org/10.1111/mec.13723.

Muhs, Daniel R, Kathleen R Simmons, R Randall Schumann, Lindsey T Groves, Stephen B DeVogel, Scott A Minor, and DeAnna Laurel. 2014. “Coastal Tectonics on the Eastern Margin of the Pacific Rim: Late Quaternary Sea-Level History and Uplift Rates, Channel Islands National Park, California, USA.” *Quaternary Science Reviews* 105: 209–38. https://doi.org/10.1016/j.quascirev.2014.09.017.

Nadeau, Simon, Patrick G Meirmans, Sally N Aitken, Kermit Ritland, and Nathalie Isabel. 2016. “The Challenge of Separating Signatures of Local Adaptation from Those of Isolation by Distance and Colonization History: The Case of Two White Pines.” *Ecology and Evolution* 6 (24): 8649–64. https://doi.org/10.1002/ece3.2550.

Namoff, Sandra, Chad E Husby, Javier Francisco-Ortega, Larry R Noblick, Carl E Lewis, and M Patrick Griffith. 2010. “How Well Does a Botanical Garden Collection of a Rare Palm Capture the Genetic Variation in a Wild Population?” *Biological Conservation* 143 (5): 1110–17. https://doi.org/10.1016/j.biocon.2010.02.004.

Namroud, Marie-Claire, Jean Beaulieu, Nicolas Juge, Jerome Laroche, and Jean Bousquet. 2008. “Scanning the Genome for Gene Single Nucleotide Polymorphisms Involved in Adaptive Population Differentiation in White Spruce.” *Molecular Ecology* 17 (16): 3599–3613. https://doi.org/10.1111/j.1365-294X.2008.03840.x.

Naredo, Maria Elizabeth Boncodin, Sheila Mae Quilloy Mercado, Maria Celeste Naynes Banaticla-Hilario, Myrricar Loren Berdos, Mario Alcantara Rodriguez, Kenneth Lee McNally, and Ruaraidh Sackville Hamilton. 2017. “Genetic Diversity Patterns in Ex Situ Collections of Oryza Officinalis Wall. Ex G. Watt Revealed by Morphological and Microsatellite Markers.” *Genetic Resources and Crop Evolution* 64 (4): 733–44. https://doi.org/10.1007/s10722-016-0396-x.

Nei, Masatoshi. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nosil, Patrik, Timothy H Vines, and Daniel J Funk. 2005. “Reproductive Isolation Caused by Natural Selection against Immigrants from Divergent Habitats.” *Evolution* 59 (4): 705–19. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x.

Nystedt, Björn, Nathaniel R Street, Anna Wetterbom, Andrea Zuccolo, Yao-Cheng Lin, Douglas G Scofield, Francesco Vezzi, Nicolas Delhomme, Stefania Giacomello, and Andrey Alexeyenko. 2013. “The Norway Spruce Genome Sequence and Conifer Genome Evolution.” *Nature* 497 (7451): 579–84. https://doi.org/10.1038/nature12211.

O’Connell, Lisa M, Alex Mosseler, and Om P Rajora. 2007. “Extensive Long-Distance Pollen Dispersal in a Fragmented Landscape Maintains Genetic Diversity in White Spruce.” *Journal of Heredity* 98 (7): 640–45. https://doi.org/10.1093/jhered/esm089.

Ohtani, Masato, Toshiaki Kondo, Naoki Tani, Saneyoshi Ueno, Leong S Lee, Kevin K S Ng, Norwati Muhammad, Reiner Finkeldey, Mohamad Na’iem, and Sapto Indrioko. 2013. “Nuclear and Chloroplast DNA Phylogeography Reveals P Leistocene Divergence and Subsequent Secondary Contact of Two Genetic Lineages of the Tropical Rainforest Tree Species S Horea Leprosula (D Ipterocarpaceae) in S Outh-E Ast A Sia.” *Molecular Ecology* 22 (8): 2264–79. https://doi.org/10.1111/mec.12243.

Oksanen, Jari, Roeland Kindt, Pierre Legendre, Bob O’Hara, and Henry M H Stevens. 2007. “The Vegan Package.” *Community Ecology Package* 10 (631–637): 719. https://cran.r-project.org/package=vegan.

Ottaviani, Gianluigi, Gunnar Keppel, Lars Götzenberger, Susan Harrison, Øystein H Opedal, Luisa Conti, Pierre Liancourt, Jitka Klimešová, Fernando A O Silveira, and Borja Jiménez-Alfaro. 2020. “Linking Plant Functional Ecology to Island Biogeography.” *Trends in Plant Science* 25 (4): 329–39. https://doi.org/10.1016/j.tplants.2019.12.022.

Pandey, M P, D V Seshu, and M Akbar. 1994. “Genetics of Embryo Size and Its Relationship with Seed and Seedling Vigour in Rice (Oryza Sativa L.).” *Indian Journal of Genetics and Plant Breeding* 54 (3): 258–68.

Pantoja, Pauline O, Violeta I Simón-Porcar, Joshua R Puzey, and Mario Vallejo-Marín. 2017. “Genetic Variation and Clonal Diversity in Introduced Populations of Mimulus Guttatus Assessed by Genotyping at 62 Single Nucleotide Polymorphism Loci.” *Plant Ecology & Diversity* 10 (1): 5–15. https://doi.org/10.1080/17550874.2017.1287785.

Paradis, Emmanuel. 2010. “Pegas: An R Package for Population Genetics with an Integrated–Modular Approach.” *Bioinformatics* 26 (3): 419–20. https://doi.org/10.1093/bioinformatics/btp696.

Parchman, Thomas L, Zachariah Gompert, Joann Mudge, Faye D Schilkey, Craig W Benkman, and Alex C Buerkle. 2012. “Genome-wide Association Genetics of an Adaptive Trait in Lodgepole Pine.” *Molecular Ecology* 21 (12): 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x.

Peng, Roger D. 2019. “Simpleboot: Simple Bootstrap Routines.” https://cran.r-project.org/package=simpleboot.

Pesendorfer, Mario B, Scott T Sillett, Walter D Koenig, and Scott A Morrison. 2016. “Scatter-Hoarding Corvids as Seed Dispersers for Oaks and Pines: A Review of a Widely Distributed Mutualism and Its Utility to Habitat Restoration.” *The Condor: Ornithological Applications* 118 (2): 215–37. https://doi.org/10.1650/CONDOR-15-125.1.

Pew, Jack, Jinliang Wang, Paul Muir, and Tim Frasier. 2015. “Related: An R Package for Analyzing Pairwise Relatedness Data Based on Codominant Molecular Markers.” https://github.com/timothyfrasier/related.

Platenkamp, Gerrit A J, and Ruth G Shaw. 1993. “Environmental and Genetic Maternal Effects on Seed Characters in Nemophila Menziesii.” *Evolution* 47 (2): 540–55. https://doi.org/10.2307/2410070.

Potter, Kevin M, Robert M Jetton, Andrew Bower, Douglass F Jacobs, Gary Man, Valerie D Hipkins, and Murphy Westwood. 2017. “Banking on the Future: Progress, Challenges and Opportunities for the Genetic Conservation of Forest Trees.” *New Forests* 48: 153–80. https://doi.org/10.1007/s11056-017-9582-8.

Pritchard, Diana J, John E Fa, Sara Oldfield, and Stuart R Harrop. 2012. “Bring the Captive Closer to the Wild: Redefining the Role of Ex Situ Conservation.” *Oryx* 46 (1): 18–23. https://doi.org/10.1017/S0030605310001766.

Privé, Florian, Keurcien Luu, Bjarni J Vilhjálmsson, and Michael G B Blum. 2020. “Performing Highly Efficient Genome Scans for Local Adaptation with R Package Pcadapt Version 4.” *Molecular Biology and Evolution* 37 (7): 2153–54. https://doi.org/10.1093/molbev/msaa053.

Puritz, Jonathan B, Christopher M Hollenbeck, and John R Gold. 2014. “DDocent: A RADseq, Variant-Calling Pipeline Designed for Population Genomics of Non-Model Organisms.” *PeerJ* 2: e431. https://doi.org/10.7717/peerj.431.

Puritz, Jonathan B, Mikhail V Matz, Robert J Toonen, Jesse N Weber, Daniel I Bolnick, and Christopher E Bird. 2014. “Demystifying the RAD Fad.” *Molecular Ecology* 23 (24): 5937–42. https://doi.org/10.1111/mec.12965.

R Core Team. 2019. “R: A Language and Environment for Statistical Computing.” Vienna: R Foundation for Statistical Computing, Austria. https://www.r-project.org/.

———. 2020. “R: A Language and Environment for Statistical Computing.” Vienna: R Foundation for Statistical Computing, Austria. https://www.r-project.org/.

———. 2021. “R: A Language and Environment for Statistical Computing.” Vienna: R Foundation for Statistical Computing, Austria. https://www.r-project.org/.

Ralls, Katherine, Jonathan D Ballou, Michele R Dudash, Mark D B Eldridge, Charles B Fenster, Robert C Lacy, Paul Sunnucks, and Richard Frankham. 2018. “Call for a Paradigm Shift in the Genetic Management of Fragmented Populations.” *Conservation Letters* 11 (2): e12412. https://doi.org/10.1111/conl.12412.

Reader, R J. 1993. “Control of Seedling Emergence by Ground Cover and Seed Predation in Relation to Seed Size for Some Old-Field Species.” *Journal of Ecology* 81 (1): 169–75. https://doi.org/10.2307/2261232.

Rogers, Deborah L, and Arlee M Montalvo. 2004. “Genetically Appropriate Choices for Plant Materials to Maintain Biological Diversity.” *Report to the USDA Forest Service*. University of California, Lakewood, CO.

Rosser, Alison M, and Sue A Mainka. 2002. “Overexploitation and Species Extinctions.” *Conservation Biology* 16 (3): 584–86. http://www.jstor.org/stable/3061203.

Roy, S Mukherjee, R C Thapliyal, and S S Phartyal. 2004. “Seed Source Variation in Cone, Seed and Seedling Characteristic across the Natural Distribution of Himalayan Low Level Pine Pinus Roxburghii Sarg.” *Silvae Genetica* 53 (1–6): 116–23. https://doi.org/10.1515/sg-2004-0021.

Santiso, Xabier, Lúa López, Kimberly J Gilbert, Rodolfo Barreiro, Michael C Whitlock, and Rubén Retuerto. 2015. “Patterns of Genetic Variation within and among Populations in Arbutus Unedo and Its Relation with Selection and Evolvability.” *Perspectives in Plant Ecology, Evolution and Systematics* 17 (3): 185–92. https://doi.org/10.1016/j.ppees.2015.02.006.

Di Santo, Lionel N, and Jill A Hamilton. 2020. “Using Environmental and Geographic Data to Optimize Ex Situ Collections and Preserve Evolutionary Potential.” *Conservation Biology* 35 (2): 733–44. https://doi.org/10.1111/cobi.13568.

Di Santo, Lionel N, Sean Hoban, Thomas L Parchman, Jessica W Wright, and Jill A Hamilton. 2021. “Reduced Representation Sequencing to Understand the Evolutionary History of Torrey Pine (Pinus Torreyana Parry) with Implications for Rare Species Conservation.” *BioRxiv*. https://doi.org/10.1101/2021.07.02.450939.

Di Santo, Lionel N, Monica Polgar, Storm Nies, Paul Hodgkiss, Courtney A Canning, Jessica W Wright, and Jill A Hamilton. 2021. “Seed Morphological Traits as a Tool to Quantify Variation Maintained in Ex Situ Collections: A Case Study in Pinus Torreyana.” *AoB PLANTS* 13 (5): plab058. https://doi.org/10.1093/aobpla/plab058.

Schaal, Barbara, and W Leverich. 2004. “Population Genetic Issues in Ex Situ Plant Conservation.” In *Ex Situ Plant Conservation: Supporting Species Survival In The Wild*, edited by Edward O Guerrant, Kayri Havens-Young, and Michael Maunder, 267–84. Washington, DC: Island Press.

Schoen, Daniel J, and Anthony H D Brown. 2001. “The Conservation of Wild Plant Species in Seed Banks: Attention to Both Taxonomic Coverage and Population Biology Will Improve the Role of Seed Banks as Conservation Tools.” *BioScience* 51 (11): 960–66. https://doi.org/10.1641/0006-3568(2001)051[0960:TCOWPS]2.0.CO;2.

Schumann, R Randall, Scott A Minor, Daniel R Muhs, and Jeffrey S Pigati. 2014. “Landscapes of Santa Rosa Island, Channel Islands National Park, California.” *Monographs of the Western North American Naturalist* 7 (1): 48–67. https://doi.org/10.3398/042.007.0108.

Sexton, Jason P, Sandra B Hangartner, and Ary A Hoffmann. 2014. “Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common?” *Evolution* 68 (1): 1–15. https://doi.org/10.1111/evo.12258.

Sexton, Jason P, Matthew B Hufford, Ashley C Bateman, David B Lowry, Harald Meimberg, Sharon Y Strauss, and Kevin J Rice. 2016. “Climate Structures Genetic Variation across a Species’ Elevation Range: A Test of Range Limits Hypotheses.” *Molecular Ecology* 25 (4): 911–28. https://doi.org/10.1111/mec.13528.

Sharrock, Suzanne, Robert Hoft, and Braulio Ferreira de Souza Dias. 2018. “An Overview of Recent Progress in the Implementation of the Global Strategy for Plant Conservation-a Global Perspective.” *Rodriguésia* 69 (4): 1489–1511. https://doi.org/10.1590/2175-7860201869401.

Shea, Patrick J, and Michael Neustein. 1995. “Protection of a Rare Stand of Torrey Pine from Ips Paraconfusus.” In *Application of Semiochemicals for Management of Bark Beetle Infestations - Proceedings of and Informal Conference*, 39–43. USDA, Forest Service, General Technical Report INT: GTR-318.

Simón-Porcar, Violeta I, Xavier F Picó, and Juan Arroyo. 2015. “Range-wide Population Genetics and Variation in Morph Ratio in Style-dimorphic Narcissus Papyraceus.” *American Journal of Botany* 102 (3): 449–56. https://doi.org/10.3732/ajb.1400209.

Singh, Jugpreet, Jose A Clavijo Michelangeli, Salvador A Gezan, Hyungwon Lee, and C Eduardo Vallejos. 2017. “Maternal Effects on Seed and Seedling Phenotypes in Reciprocal F1 Hybrids of the Common Bean (Phaseolus Vulgaris L.).” *Frontiers in Plant Science* 8: 42. https://doi.org/10.3389/fpls.2017.00042.

Slatkin, Montgomery. 1993. “Isolation by Distance in Equilibrium and Non-equilibrium Populations.” *Evolution* 47 (1): 264–79. https://doi.org/10.1111/j.1558-5646.1993.tb01215.x.

Soudy, Mohamed, Ali Mostafa Anwar, Eman Ali Ahmed, Aya Osama, Shahd Ezzeldin, Sebaey Mahgoub, and Sameh Magdeldin. 2020. “UniprotR: Retrieving and Visualizing Protein Sequence and Functional Information from Universal Protein Resource (UniProt Knowledgebase).” *Journal of Proteomics* 213: 103613. https://doi.org/10.1016/j.jprot.2019.103613.

Stern, David L, and Virginie Orgogozo. 2009. “Is Genetic Evolution Predictable?” *Science* 323 (5915): 746–51. https://doi.org/10.1126/science.1158997.

Stevens, Kristian A, Jill L Wegrzyn, Aleksey Zimin, Daniela Puiu, Marc Crepeau, Charis Cardeno, Robin Paul, Daniel Gonzalez-Ibeaz, Maxim Koriabine, and Ann Holtz-Morris. 2016. “Sequence of the Sugar Pine Megagenome.” *Genetics* 204 (4): 1613–26. https://doi.org/10.1093/genetics/204.4.NP.

Stowell, Sierra M Love, Cheryl A Pinzone, and Andrew P Martin. 2017. “Overcoming Barriers to Active Interventions for Genetic Diversity.” *Biodiversity and Conservation* 26 (8): 1753–65. https://doi.org/10.1007/s10531-017-1330-z.

Swarts, Nigel D, Elizabeth A Sinclair, Siegfried L Krauss, and Kingsley W Dixon. 2009. “Genetic Diversity in Fragmented Populations of the Critically Endangered Spider Orchid Caladenia Huegelii: Implications for Conservation.” *Conservation Genetics* 10 (5): 1199–1208. https://doi.org/10.1007/s10592-008-9651-9.

Tamaki, Satoshi, Keiya Isoda, Makoto Takahashi, Hiroo Yamada, and Yumiko Yamashita. 2018. “Genetic Structure and Diversity in Relation to the Recently Reduced Population Size of the Rare Conifer, Pseudotsuga Japonica, Endemic to Japan.” *Conservation Genetics* 19 (5): 1243–55. https://doi.org/10.1007/s10592-018-1092-5.

Tanveer, A, M Tasneem, A Khaliq, M M Javaid, and M N Chaudhry. 2013. “Influence of Seed Size and Ecological Factors on the Germination and Emergence of Field Bindweed (Convolvulus Arvensis).” *Planta Daninha* 31 (1): 39–51. https://doi.org/10.1590/S0100-83582013000100005.

Teixeira, João C, and Christian D Huber. 2021. “The Inflated Significance of Neutral Genetic Diversity in Conservation Genetics.” *Proceedings of the National Academy of Sciences* 118 (10): e2015096118. https://doi.org/10.1073/pnas.2015096118.

Thomas, Chris D, Alison Cameron, Rhys E Green, Michel Bakkenes, Linda J Beaumont, Yvonne C Collingham, Barend F N Erasmus, Marinez Ferreira De Siqueira, Alan Grainger, and Lee Hannah. 2004. “Extinction Risk from Climate Change.” *Nature* 427 (6970): 145–48. https://doi.org/10.1038/nature02121.

Toker, Cengiz. 2004. “Estimates of Broad-sense Heritability for Seed Yield and Yield Criteria in Faba Bean (Vicia Faba L.).” *Hereditas* 140 (3): 222–25. https://doi.org/10.1111/j.1601-5223.2004.01780.x.

Tsuda, Yoshiaki, K Nakao, Y Ide, and Y Tsumura. 2015. “The Population Demography of Betula Maximowicziana, a Cool-temperate Tree Species in Japan, in Relation to the Last Glacial Period: Its Admixture-like Genetic Structure Is the Result of Simple Population Splitting Not Admixing.” *Molecular Ecology* 24 (7): 1403–18. https://doi.org/10.1111/mec.13123.

Tsumura, Yoshihiko, K Uchiyama, Y Moriguchi, S Ueno, and T Ihara-Ujino. 2012. “Genome Scanning for Detecting Adaptive Genes along Environmental Gradients in the Japanese Conifer, Cryptomeria Japonica.” *Heredity* 109 (6): 349–60. https://doi.org/10.1038/hdy.2012.50.

Tyrmi, Jaakko S, Jaana Vuosku, Juan J Acosta, Zhen Li, Lieven Sterck, Maria T Cervera, Outi Savolainen, and Tanja Pyhäjärvi. 2020. “Genomics of Clinal Local Adaptation in Pinus Sylvestris under Continuous Environmental and Spatial Genetic Setting.” *G3: Genes, Genomes, Genetics* 10 (8): 2683–96. https://doi.org/10.1534/g3.120.401285.

Varis, Saila, Anne Pakkanen, Aina Galofré, and Pertti Pulkkinen. 2009. “The Extent of South-North Pollen Transfer in Finnish Scots Pine.” *Silva Fennica* 43 (5): 717–26. https://doi.org/10.14214/sf.168.

Vergara, Rodrigo, Matthew A Gitzendanner, Douglas E Soltis, and Pamela S Soltis. 2014. “Population Genetic Structure, Genetic Diversity, and Natural History of the South American Species of Nothofagus Subgenus Lophozonia (Nothofagaceae) Inferred from Nuclear Microsatellite Data.” *Ecology and Evolution* 4 (12): 2450–71. https://doi.org/10.1002/ece3.1108.

Viana, Duarte S, Laura Gangoso, Willem Bouten, and Jordi Figuerola. 2016. “Overseas Seed Dispersal by Migratory Birds.” *Proceedings of the Royal Society B: Biological Sciences* 283 (1822): 20152406. https://doi.org/10.1098/rspb.2015.2406.

Victory, Erin R, Jeffrey C Glaubitz, Olin E Rhodes Jr, and Keith E Woeste. 2006. “Genetic Homogeneity in Juglans Nigra (Juglandaceae) at Nuclear Microsatellites.” *American Journal of Botany* 93 (1): 118–26. https://doi.org/10.3732/ajb.93.1.118.

Villellas, Jesús, Regina Berjano, Anass Terrab, and María B García. 2014. “Divergence between Phenotypic and Genetic Variation within Populations of a Common Herb across Europe.” *Ecosphere* 5 (5): 1–14. https://doi.org/10.1890/ES13-00291.1.

Villemereuil, Pierre de, Oscar E Gaggiotti, Médéric Mouterde, and Irene Till-Bottraud. 2016. “Common Garden Experiments in the Genomic Era: New Perspectives and Opportunities.” *Heredity* 116 (3): 249.

Vinceti, Barbara, Judy Loo, Hannes Gaisberger, Maarten J van Zonneveld, Silvio Schueler, Heino Konrad, Caroline A C Kadu, and Thomas Geburek. 2013. “Conservation Priorities for Prunus Africana Defined with the Aid of Spatial Analysis of Genetic Data and Climatic Variables.” *PloS One* 8 (3): e59987. https://doi.org/10.1371/journal.pone.0059987.

Volis, Sergei, and Michael Blecher. 2010. “Quasi in Situ: A Bridge between Ex Situ and in Situ Conservation of Plants.” *Biodiversity and Conservation* 19 (9): 2441–54. https://doi.org/10.1007/s10531-010-9849-2.

Volk, Gayle M, Christopher M Richards, Ann A Reilley, Adam D Henk, Philip L Forsline, and Herb S Aldwinckle. 2005. "Ex Situ Conservation of Vegetatively Propagated Species: Development of a Seed-Based Core Collection for Malus Sieversii." *Journal of the American Society for Horticultural Science* 130 (2): 203–10. https://doi.org/10.21273/JASHS.130.2.203.

Walters, Christina, Lana M Wheeler, and Judith M Grotenhuis. 2005. "Longevity of Seeds Stored in a Genebank: Species Characteristics." *Seed Science Research* 15 (1): 1–20. https://doi.org/10.1079/SSR2004195.

Wang, Ian J, and Gideon S Bradburd. 2014. "Isolation by Environment." *Molecular Ecology* 23 (23): 5649–62. https://doi.org/10.1111/mec.12938.

Wang, Ian J, and Kyle Summers. 2010. "Genetic Structure Is Correlated with Phenotypic Divergence Rather than Geographic Isolation in the Highly Polymorphic Strawberry Poison-dart Frog." *Molecular Ecology* 19 (3): 447–58. https://doi.org/10.1111/j.1365-294X.2009.04465.x.

Wang, Ting, Zhen Wang, Fan Xia, and Yingjuan Su. 2016. "Local Adaptation to Temperature and Precipitation in Naturally Fragmented Populations of Cephalotaxus Oliveri, an Endangered Conifer Endemic to China." *Scientific Reports* 6 (1): 25031. https://doi.org/10.1038/srep25031.

Wang, Xi, Carolina Bernhardsson, and Pär K Ingvarsson. 2020. "Demography and Natural Selection Have Shaped Genetic Variation in the Widely Distributed Conifer Norway Spruce (Picea Abies)." *Genome Biology and Evolution* 12 (2): 3803–17. https://doi.org/10.1093/gbe/evaa005.

Waples, Robin S, Gordon Luikart, James R Faulkner, and David A Tallmon. 2013. "Simple Life-History Traits Explain Key Effective Population Size Ratios across Diverse Taxa." *Proceedings of the Royal Society B: Biological Sciences* 280 (1768): 20131339. https://doi.org/10.1098/rspb.2013.1339.

Waser, Nickolas M, and Mary V Price. 1994. "Crossing-distance Effects in Delphinium Nelsonii: Outbreeding and Inbreeding Depression in Progeny Fitness." *Evolution* 48 (3): 842–52. https://doi.org/10.1111/j.1558-5646.1994.tb01366.x.

Waters, Elizabeth R, and Barbara A Schaal. 1991. "No Variation Is Detected in the Chloroplast Genome of Pinus Torreyana." *Canadian Journal of Forest Research* 21 (12): 1832–35. https://doi.org/10.1139/x91-253.

Way, Michael J. 2003. "Collecting Seed from Non-Domesticated Plants for Long-Term Conservation." In *Seed Conservation: Turning Science into Practice*, edited by Roger D Smith, John B Dickie, and Hugh W Pritchard, 165–201. Royal Botanic Garden, Kew.

Weber, Anne, and Annette Kolb. 2014. "Differences in Heritable Trait Variation among Populations of Varying Size in the Perennial Herb Phyteuma Spicatum." *Conservation Genetics* 15 (6): 1329–37. https://doi.org/10.1007/s10592-014-0620-1.

Wegmann, Daniel, Christoph Leuenberger, and Laurent Excoffier. 2009. “Using ABCtoolbox.” http://cmpg.unibe.ch/software/ABCtoolbox/.

Wegmann, Daniel, Christoph Leuenberger, Samuel Neuenschwander, and Laurent Excoffier. 2010. “ABCtoolbox: A Versatile Toolkit for Approximate Bayesian Computations.” *BMC Bioinformatics* 11 (1): 1–7. https://doi.org/10.1186/1471-2105-11-116.

Westemeier, Ronald L, Jeffrey D Brawn, Scott A Simpson, Terry L Esker, Roger W Jansen, Jeffery W Walk, Eric L Kershner, Juan L Bouzat, and Ken N Paige. 1998. “Tracking the Long-Term Decline and Recovery of an Isolated Population.” *Science* 282 (5394): 1695–98. https://doi.org/10.1126/science.282.5394.1695.

Westengen, Ola T, Simon Jeppson, and Luigi Guarino. 2013. “Global Ex-Situ Crop Diversity Conservation and the Svalbard Global Seed Vault: Assessing the Current Status.” *PloS One* 8 (5): e64146. https://doi.org/10.1371/journal.pone.0064146.

Whiteley, Andrew R, Sarah W Fitzpatrick, W Chris Funk, and David A Tallmon. 2015. “Genetic Rescue to the Rescue.” *Trends in Ecology & Evolution* 30 (1): 42–49. https://doi.org/10.1016/j.tree.2014.10.009.

Whitlock, Michael C, and Katie Lotterhos. 2014. “OutFLANK: Fst Outliers with Trimming.” https://github.com/whitlock/OutFLANK.

Whitlock, Raj, Helen Hipperson, Des B A Thompson, Roger K Butlin, and Terry Burke. 2016. “Consequences of In-Situ Strategies for the Conservation of Plant Genetic Diversity.” *Biological Conservation* 203: 134–42. https://doi.org/10.1016/j.biocon.2016.08.006.

Whittall, J B, J Syring, M Parks, J Buenrostro, C Dick, A Liston, and R Cronn. 2010. “Finding a (Pine) Needle in a Haystack: Chloroplast Genome Sequence Divergence in Rare Widespread Pines.” *Molecular Ecology* 19 (s1): 100–114. https://doi.org/10.1111/j.1365-294X.2009.04474.x.

Willi, Yvonne, and M Fischer. 2005. “Genetic Rescue in Interconnected Populations of Small and Large Size of the Self-Incompatible Ranunculus Reptans.” *Heredity* 95 (6): 437–43. https://doi.org/10.1038/sj.hdy.6800732.

Willi, Yvonne, Mark Van Kleunen, Stefan Dietrich, and Markus Fischer. 2007. “Genetic Rescue Persists beyond First-Generation Outbreeding in Small Populations of a Rare Plant.” *Proceedings of the Royal Society of London B: Biological Sciences* 274 (1623): 2357–64. https://doi.org/10.1098/rspb.2007.0768.

Williams, Claire G. 2009. *Conifer Reproductive Biology*. Dordrecht: Springer.

———. 2010. “Long-distance Pine Pollen Still Germinates after Meso-scale Dispersal.” *American Journal of Botany* 97 (5): 846–55. https://doi.org/10.3732/ajb.0900255.

Williams, Claire G, R D Barnes, and I Nyoka. 1999. “Embryonic Genetic Load for a Neotropical Conifer, Pinus Patula Schiede et Deppe.” *Journal of Heredity* 90 (3): 394–98. https://doi.org/10.1093/jhered/90.3.394.

Williams, Claire G, Yi Zhou, and Sarah E Hall. 2001. “A Chromosomal Region Promoting Outcrossing in a Conifer.” *Genetics* 159 (3): 1283–89. https://doi.org/10.1093/genetics/159.3.1283.

Wright, Sewall. 1943. “Isolation by Distance.” *Genetics* 28 (2): 114–38.

Wyse, Sarah V, and Philip E Hulme. 2021. “Limited Evidence for a Consistent Seed Mass-dispersal Trade-off in Wind-dispersed Pines.” *Journal of Ecology* 109 (1): 284–93. https://doi.org/10.1111/1365-2745.13477.

Xia, Hanhan, Baosheng Wang, Wei Zhao, Jin Pan, Jian Feng Mao, and Xiao Ru Wang. 2018. “Combining Mitochondrial and Nuclear Genome Analyses to Dissect the Effects of Colonization, Environment, and Geography on Population Structure in Pinus Tabuliformis.” *Evolutionary Applications* 11 (10): 1931–45. https://doi.org/10.1111/eva.12697.

Yatabe, Yoko, Nolan C Kane, Caroline Scotti-Saintagne, and Loren H Rieseberg. 2007. “Rampant Gene Exchange across a Strong Reproductive Barrier between the Annual Sunflowers, Helianthus Annuus and H. Petiolaris.” *Genetics* 175 (4): 1883–93. https://doi.org/10.1534/genetics.106.064469.

Yoko, Zebadiah G, Kate L Volk, Ned A Dochtermann, and Jill A Hamilton. 2020. “The Importance of Quantitative Trait Differentiation in Restoration: Landscape Heterogeneity and Functional Traits Inform Seed Transfer Guidelines.” *AoB Plants* 12 (2): plaa009. https://doi.org/10.1093/aobpla/plaa009.

Yonezawa, K. 1985. “A Definition of the Optimal Allocation of Effort in Conservation of Plant Genetic Resources-with Application to Sample Size Determination for Field Collection.” *Euphytica* 34 (2): 345–54. https://doi.org/10.1007/BF00022927.

Young, Andrew G, Tim Boyle, and Tony Brown. 1996. “The Population Genetic Consequences of Habitat Fragmentation for Plants.” *Trends in Ecology & Evolution* 11 (10): 413–18. https://doi.org/10.1016/0169-5347(96)10045-8.

Young, Andrew G, A H D Brown, and F A Zich. 1999. “Genetic Structure of Fragmented Populations of the Endangered Daisy Rutidosis Leptorrhynchoides.” *Conservation Biology* 13 (2): 256–65. https://doi.org/10.1046/j.1523-1739.1999.013002256.x.

Zas, Rafael, and Luis Sampedro. 2015. “Heritability of Seed Weight in Maritime Pine, a Relevant Trait in the Transmission of Environmental Maternal Effects.” *Heredity* 114 (1): 116–24. https://doi.org/10.1038/hdy.2014.76.

Zeng, Xueqin, Walter Durka, Erik Welk, and Markus Fischer. 2017. “Heritability of Early Growth Traits and Their Plasticity in 14 Woody Species of Chinese Subtropical Forest.” *Journal of Plant Ecology* 10 (1): 222–31. https://doi.org/10.1093/jpe/rtw086.

Zhang, De-Quan, and Nong Zhou. 2013. “Genetic Diversity and Population Structure of the Endangered Conifer Taxus Wallichiana Var. Mairei (Taxaceae) Revealed by Simple Sequence Repeat (SSR) Markers.” *Biochemical Systematics and Ecology* 49: 107–14. https://doi.org/10.1016/j.bse.2013.03.030.

APPENDIX A. SUPPLEMENTARY MATERIAL (CHAPTER 2)

Table A1. Reference and availability information associated with genetic and genomic datasets of all 15 plant species under study.

Species	Literature reference	Dryad link
Betula maximowicziana	(Tsuda et al. 2015)	https://doi.org/10.5061/dryad.dj17c
Centaurea solstitialis	(Barker et al. 2017)	https://doi.org/10.5061/dryad.pf550
Helianthus annuus	(McAssey, Corbi, and Burke 2016)	https://doi.org/10.5061/dryad.6p1c4
Helianthus argophyllus	(Moyers and Rieseberg 2016)	https://doi.org/10.5061/dryad.3c769
Mimulus guttatus	(Pantoja et al. 2017)	https://doi.org/10.5061/dryad.91v3n
Mimulus lacinatus	(Sexton et al. 2016)	https://doi.org/10.5061/dryad.8qc40
Narcissus papyraceus	(Simón-Porcar, Picó, and Arroyo 2015)	https://doi.org/10.5061/dryad.jh21r
Nothofagus alpina	(Vergara et al. 2014)	https://doi.org/10.5061/dryad.h3d26
Nothofagus glauca	(Vergara et al. 2014)	https://doi.org/10.5061/dryad.h3d26
Nothofagus obliqua	(Vergara et al. 2014)	https://doi.org/10.5061/dryad.h3d26
Picea sitchensis	(Holliday, Ritland, and Aitken 2010)	See note [a]
Populus balsamifera	(S. R. Keller, Chhatre, and Fitzpatrick 2017)	https://doi.org/10.5061/dryad.gp78p
Populus tremula	(Bernhardsson and Ingvarsson 2012)	https://doi.org/10.5061/dryad.0vr6m66t
Rhododendron oldhamii	(Hsieh et al. 2013)	https://doi.org/10.5061/dryad.nc221
Shorea leprosula	(Ohtani et al. 2013)	https://doi.org/10.5061/dryad.1mt1h

[a] Datasets received directly from the authors.

Table A2. Raw set of climatic variables assigned to populations of every genetic and genomic dataset using ClimateNA, SA, EU, or AP (https://sites.ualberta.ca/~ahamann/data.html).

Annual variables	Seasonal variables	Monthly variables
Mean Annual Temperature (˚C)	Winter mean maximum Temperature (˚C)	January – December mean Temperatures (˚C)
Mean Warmest Month Temperature (˚C)	Spring mean maximum Temperature (˚C)	January – December maximum Temperatures (˚C)
Mean Coldest Month Temperature (˚C)	Summer mean maximum Temperature (˚C)	January – December minimum Temperatures (˚C)
Continentality (˚C)	Autumn mean maximum Temperature (˚C)	January – December Precipitation (mm)
Mean Annual Precipitation (mm)	Winter mean minimum Temperature (˚C)	
Annual Heat-Moisture index	Spring mean minimum Temperature (˚C)	
Degree-Days below 0˚C	Summer mean minimum Temperature (˚C)	
Degree-Days above 5˚C	Autumn mean minimum Temperature (˚C)	
Degree-Days below 18˚C	Winter mean Temperature (˚C)	
Degree-Days above 18˚C	Spring mean Temperature (˚C)	
Number of frost-free days	Summer mean Temperature (˚C)	
Precipitation as snow (mm)	Autumn mean Temperature (˚C)	
Extreme Minimum Temperature over past 30 years (˚C)	Winter Precipitation (mm)	
Hargreaves reference evaporation (mm)	Spring Precipitation (mm)	
Hargreaves climatic moisture deficit (mm)	Summer Precipitation (mm)	
	Autumn Precipitation (mm)	

Table A3. Proportion of environmental differences among populations explained by the first PC axis (PC1), the second PC axis (PC2), and the combination of both PC axes (PC1 and PC2) for all 15 plant species.

Species	Variance explained by PC1 (%)	Variance explained by PC2 (%)	Variance explained by PC1 and PC2 (%)
Betula maximowicziana	62.5	20.7	83.2
Centaurea solstitialis	52.9	21.8	74.7
Helianthus annuus	77.7	12.8	90.5
Helianthus argophyllus	53.6	38.6	92.2
Mimulus guttatus	65.7	21.7	87.4
Mimulus lacinatus	80.2	11.6	91.8
Narcissuspapyraceus	53	28.7	81.7
Nothofagus alpina	65	24.1	89.1
Nothofagus glauca	54.7	27.4	82.1
Nothofagus obliqua	59.8	28.7	88.5
Picea sitchensis	75	18	93
Populus balsamifera	61.8	20.2	82
Populus tremula	75.7	16.1	91.8
Rhododendron oldhamii	66.1	19.4	85.5
Shorea leprosula	54.7	19.2	73.9

Table A4. List of parameters tested and used for realistic and idealized simulations given per dataset.

Species	Genetic marker (Number of loci)	Number of populations	Number of individuals per population			$N_{80\%}$	Np			
			Min	Max	Mean (SD)		30-40%	50-60%	70-80%	90-100%
Betula maximowicziana	EST-SSRs (12)	48	21	49	29.5 (4.8)	11	15	24	36	48
Centaurea solstitialis	SNPs (747)	25	2	19	9 (6.2)	2	8	13	19	25
Helianthus annuus	SNPs (246)	15	12	20	19.1 (2.3)	3	5	8	11	15
Helianthus argophyllus	Gen-SNPs (68)	51	1	29	10.9 (9.7)	NA	NA	NA	NA	NA
Mimulus guttatus	SNPs (62)	14	10	21	18.6 (2.6)	3	5	7	11	14
Mimulus lacinatus	SSRs (8)	23	30	49	41.3 (5.5)	NA	NA	NA	NA	NA
Narcissus papyraceus	SSRs (8)	26	13	20	16.2 (2.2)	13	8	13	20	26
Nothofagus alpina	SSRs (7)	12	16	16	16 (0)	10	4	6	9	12
Nothofagus glauca	SSRs (7)	8	16	16	16 (0)	10	3	4	6	8
Nothofagus obliqua	SSRs (6)	20	16	16	16 (0)	10	6	10	15	20
Picea sitchensis	Gen-SNPs (339)	10	12	46	28.6 (13.2)	4	3	5	8	10
	SEL-SNPs (34)	10	12	46	28.6 (13.2)	5	3	5	8	10

Table A4. List of parameters tested and used for realistic and idealized simulations given per dataset. (continued)

Species	Genetic marker (Number of loci)	Number of populations	Number of individuals per population			$N_{80\%}$	Np			
			Min	Max	Mean (SD)		30-40%	50-60%	70-80%	90-100%
Populus balsamifera	Gen-SNPs (284)	31	10	18	14.3 (1.4)	3	10	16	23	31
	SEL-SNPs (33)	31	10	18	14.3 (1.4)	4	10	16	23	31
Populus tremula	Gen-SNPs (93) [control set]	12	6	10	9.6 (1.2)	2	4	6	9	12
	Gen-SNPs (71) [defense set]	12	6	10	9.2 (1.1)	3	4	6	9	12
	SEL-SNPs (10)	12	6	10	9.2 (1.1)	3	4	6	9	12
Rhododendron oldhamii	EST-SSRs (26)	18	8	31	18.7 (6.9)	NA	NA	NA	NA	NA
Shorea leprosula	EST-SSRs (27)	24	10	59	30.25 (16.1)	NA	NA	NA	NA	NA

Table A5. Regression statistics of genetic parameters assessed (F_{ST} and A_c/A_d ; see Figure 4 for details) separated first by comparisons (Env-Rand, Geo-Rand, Env & Geo-Rand) and then by within-population sampling scenarios (realistic vs idealized). For each genetic parameter, slopes and SEs are given per genetic marker class.

Genetic marker class	Env-Rand		Geo-Rand		Env & Geo-Rand	
	realistic slope (SE)	idealized slope (SE)	realistic slope (SE)	idealized slope (SE)	realistic slope (SE)	idealized slope (SE)
F_{ST}						
Neutral	-0.005(0.004)	-0.006(0.004)	0.003(0.007)	0.002(0.007)	-0.001(0.005)	-0.001(0.005)
Functional	-0.004(0.002)[a]	-0.004(0.002)[a]	-0.008(0.004)	-0.008(0.004)	-0.006(0.003)[a]	-0.006(0.003)[a]
Adaptive	-0.024(0.006)[a]	-0.023(0.006)[a]	-0.035(0.006)[a]	-0.035(0.006)[a]	-0.023(0.001)[a]	-0.022(0.002)[a]
A_c/A_d						
Neutral	-0.001(0.007)	0.001(0.007)	0(0.006)	0.001(0.007)	-0.005(0.005)	-0.005(0.006)
Functional	-0.003(0.001)[a]	-0.006(0.004)	0.001(0.003)	0.005(0.004)	-0.001(0.002)	-0.003(0.004)
Adaptive	-0.008(0.004)	-0.006(0.002)[a]	-0.005(0.005)	0(0.005)	-0.01(0.005)	-0.006(0.002)[a]

[a] Indicates that observed slopes are significantly different from zero (α=0.05).

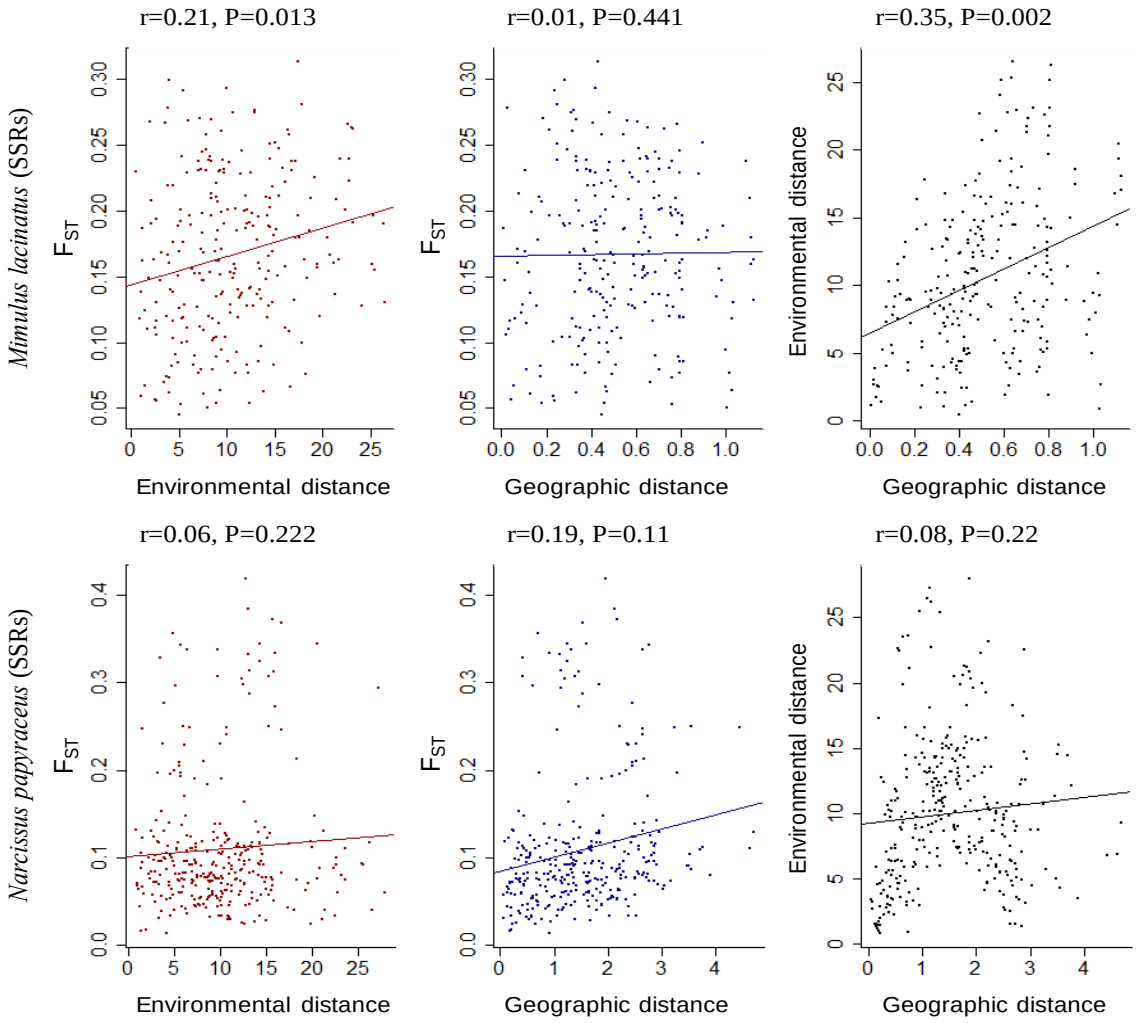


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study. Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

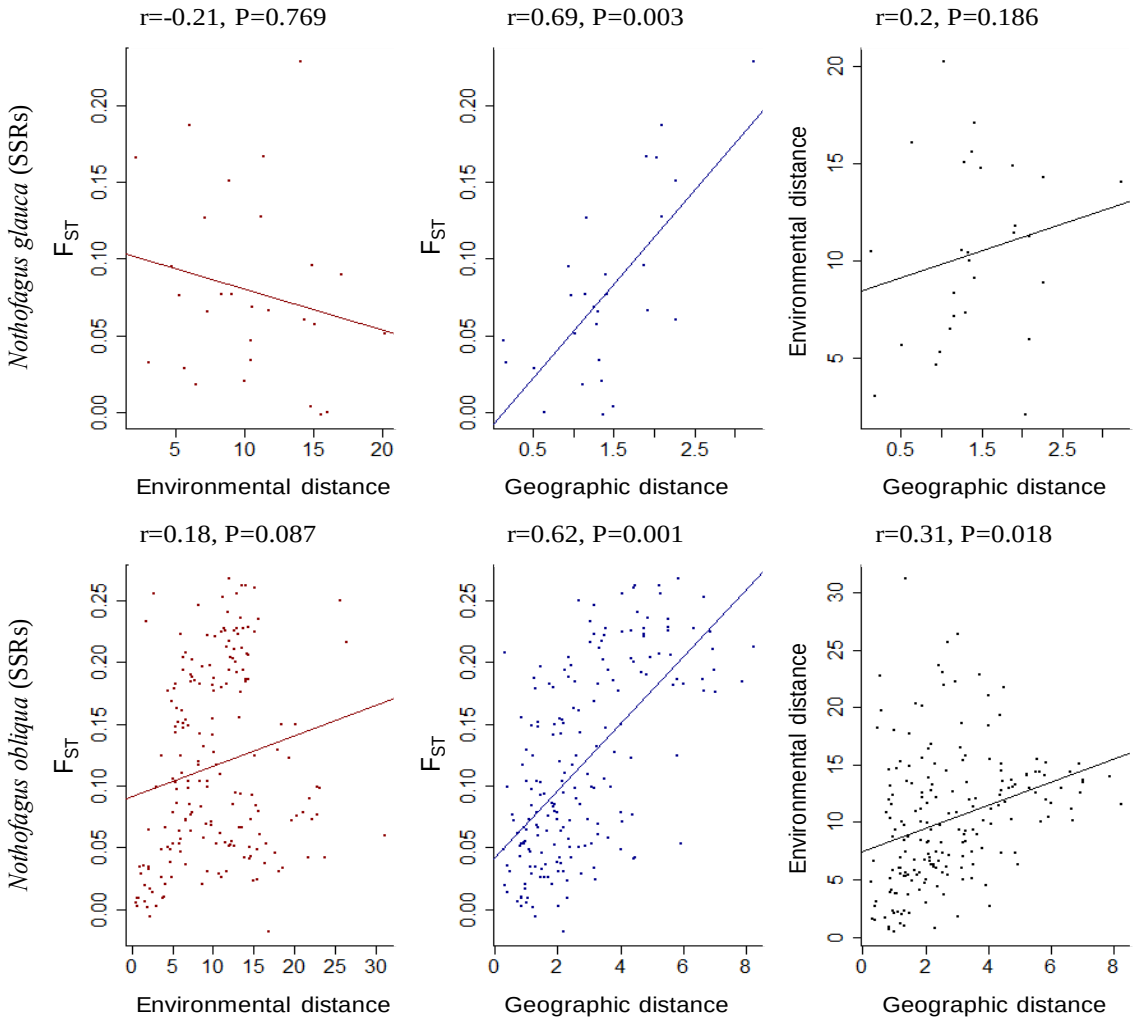


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

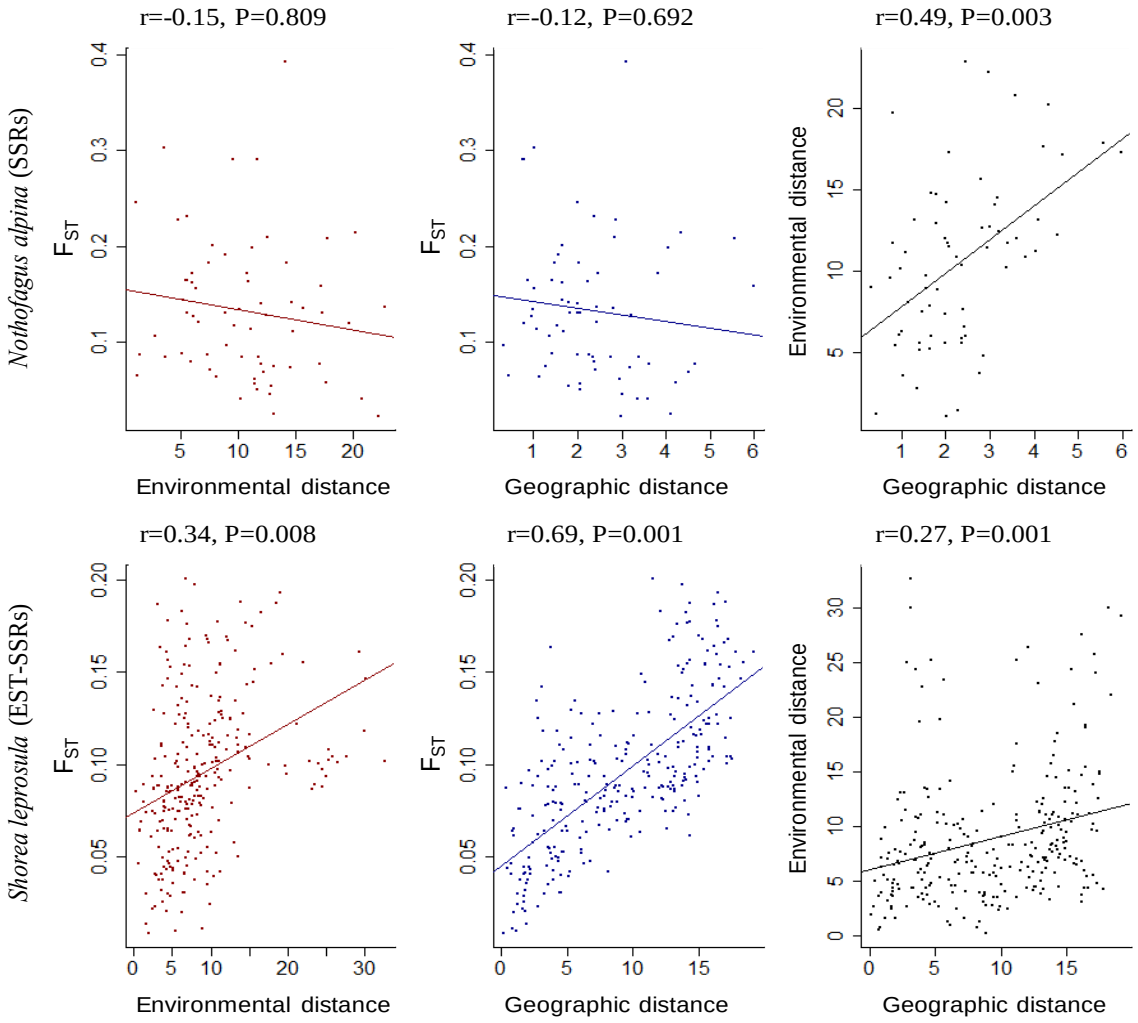


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

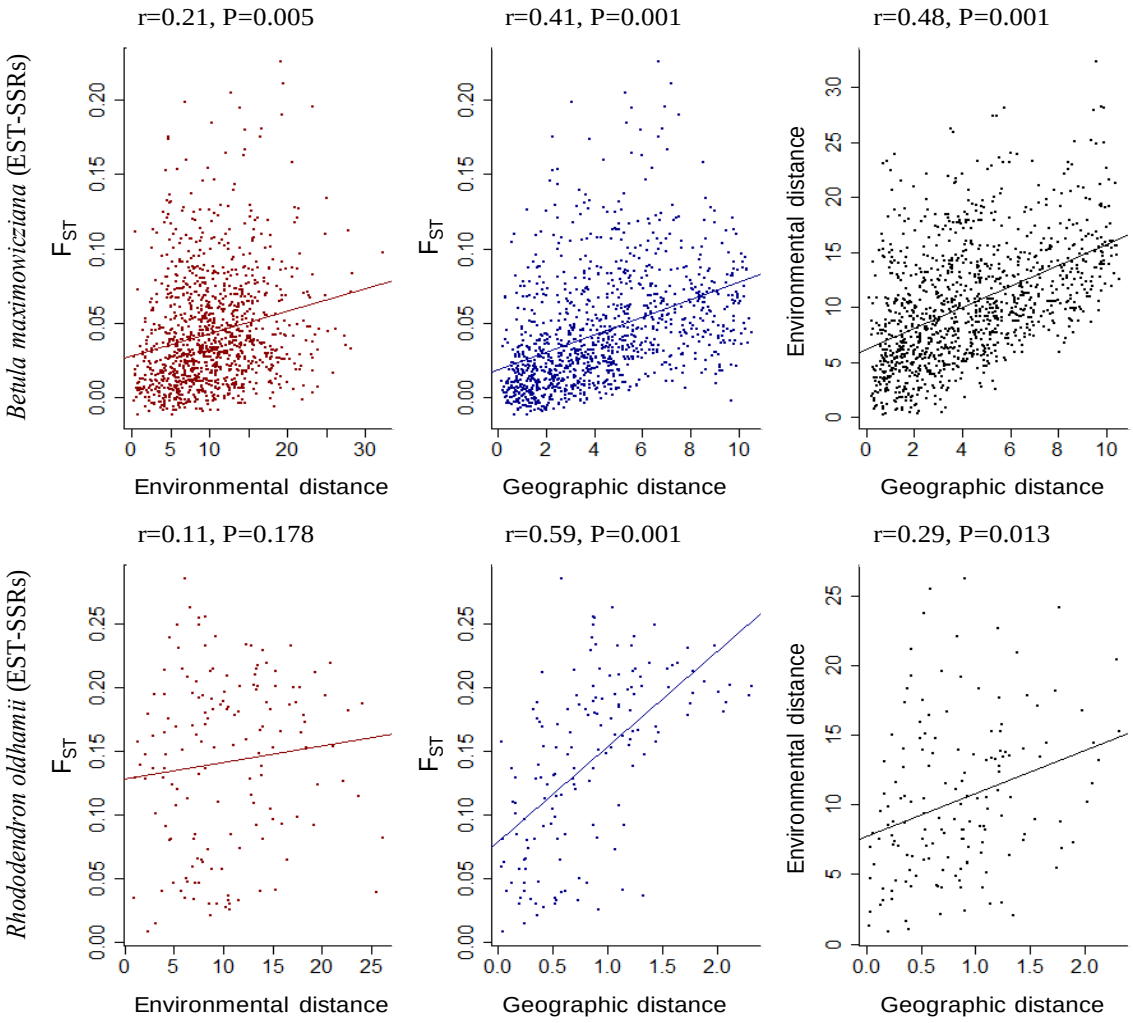


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

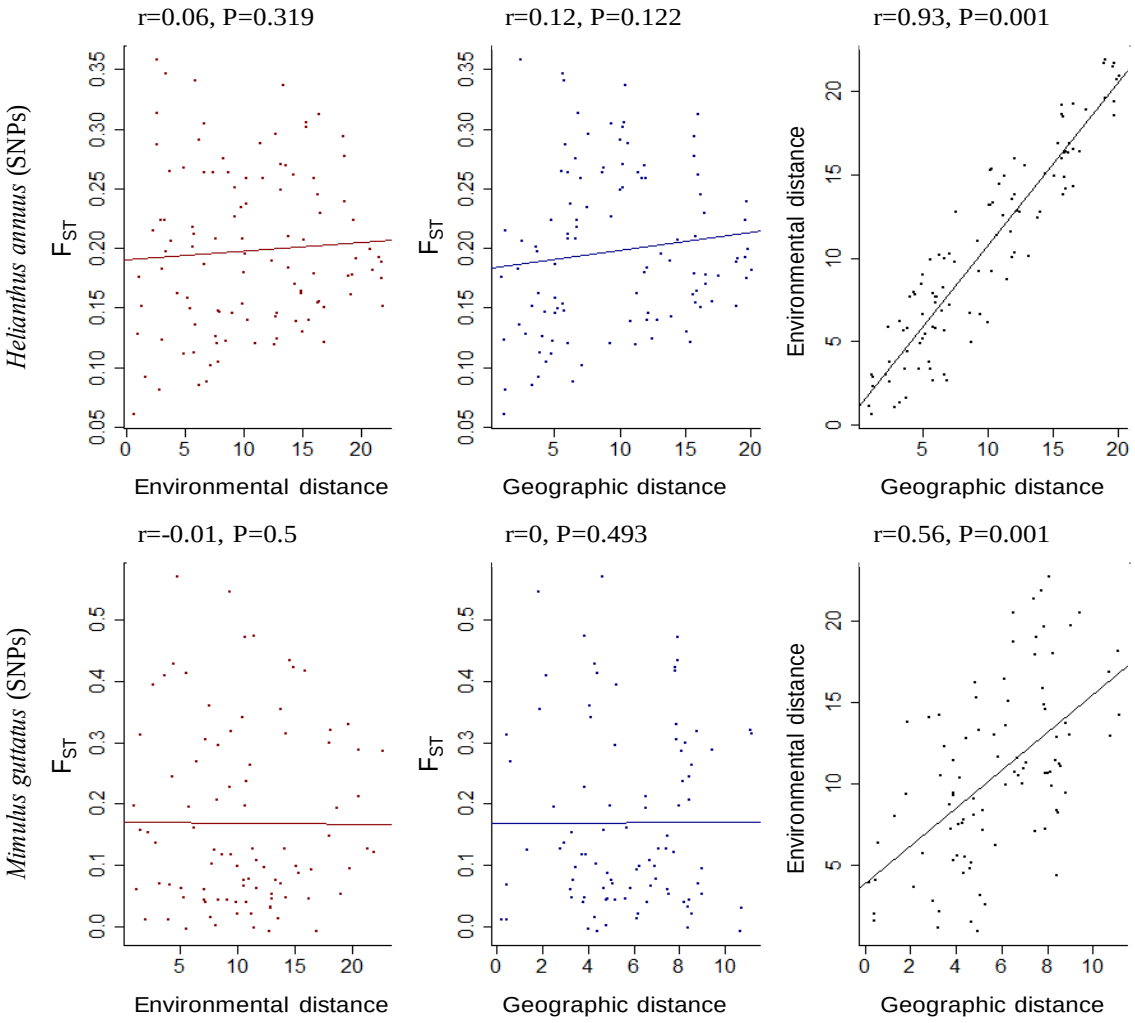


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

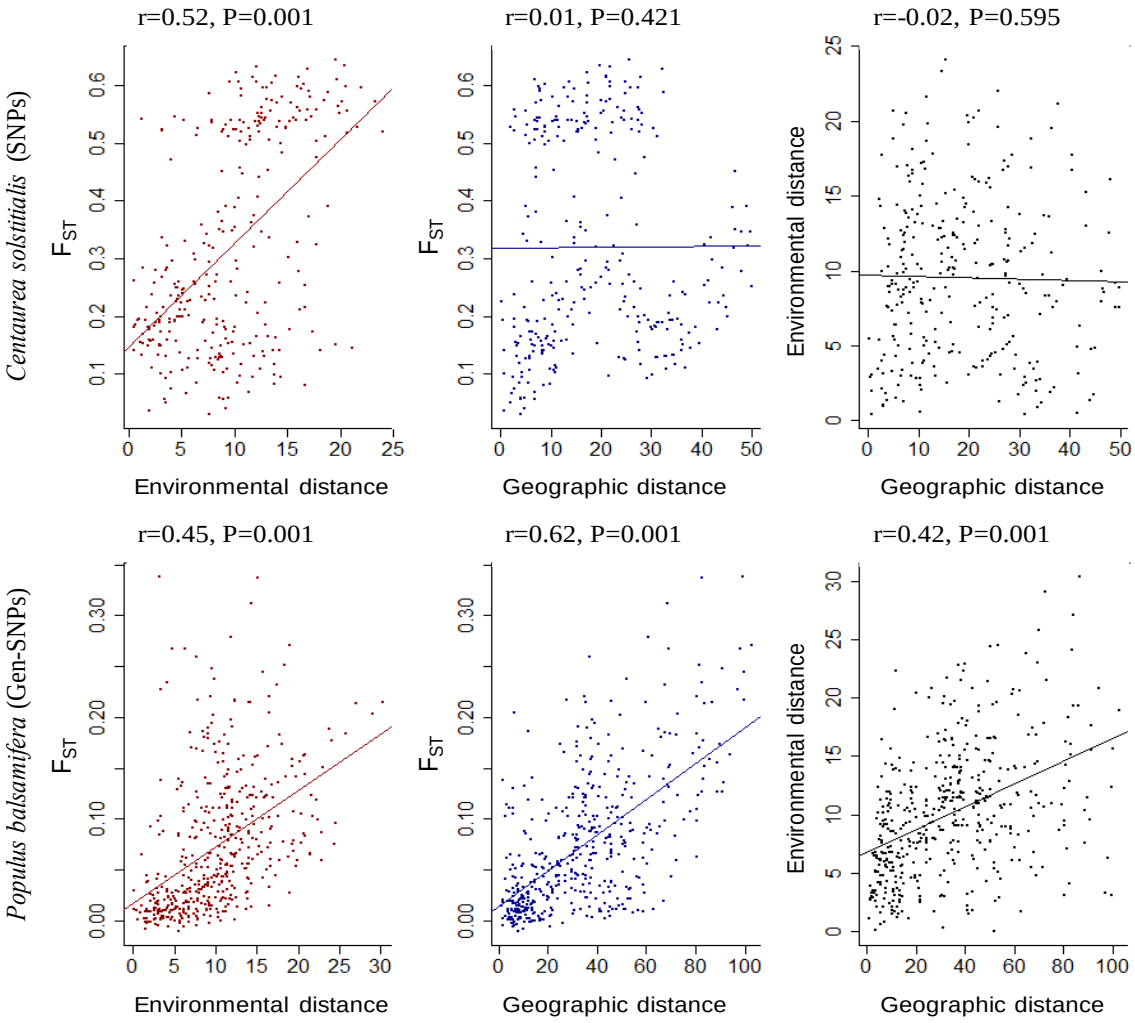


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

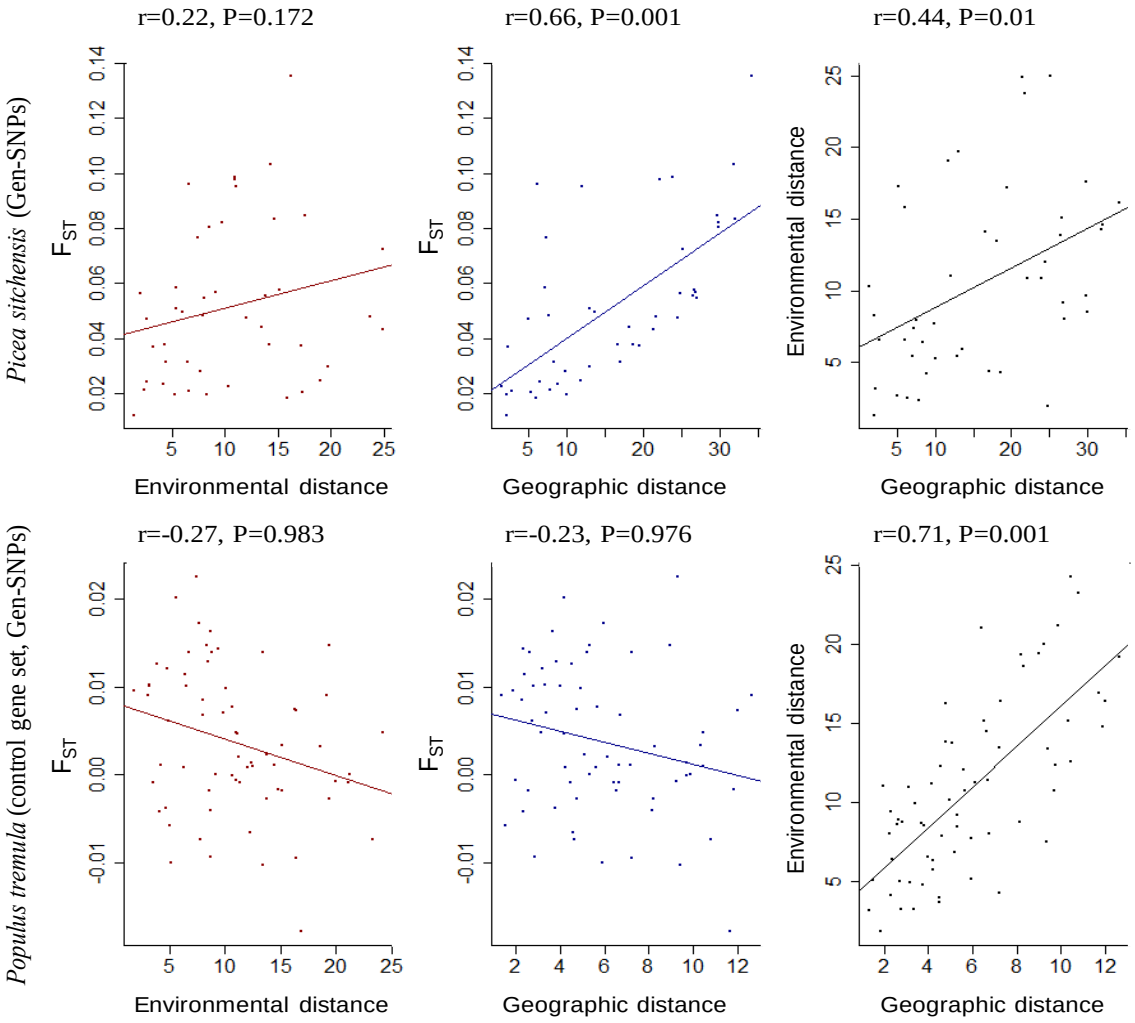


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

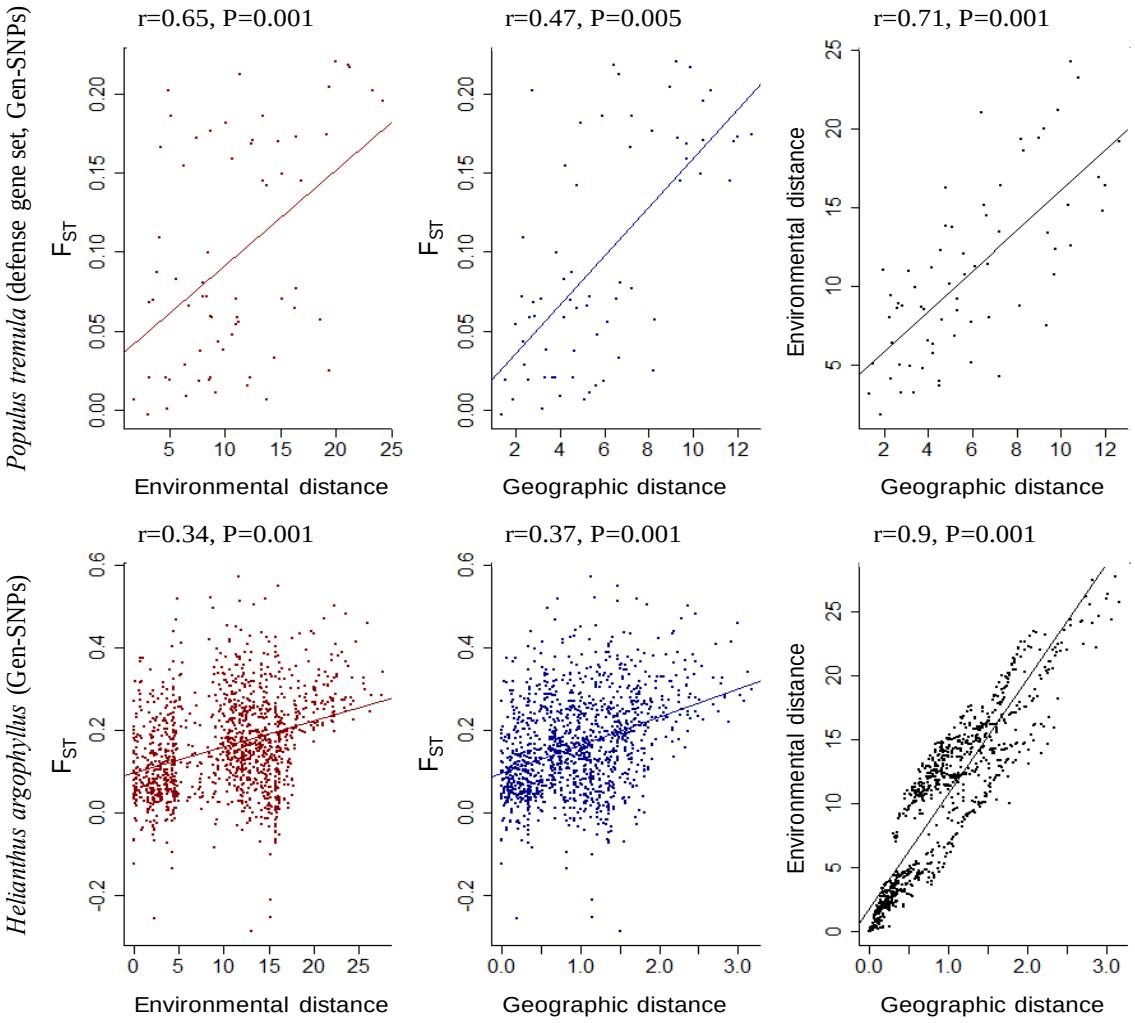


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

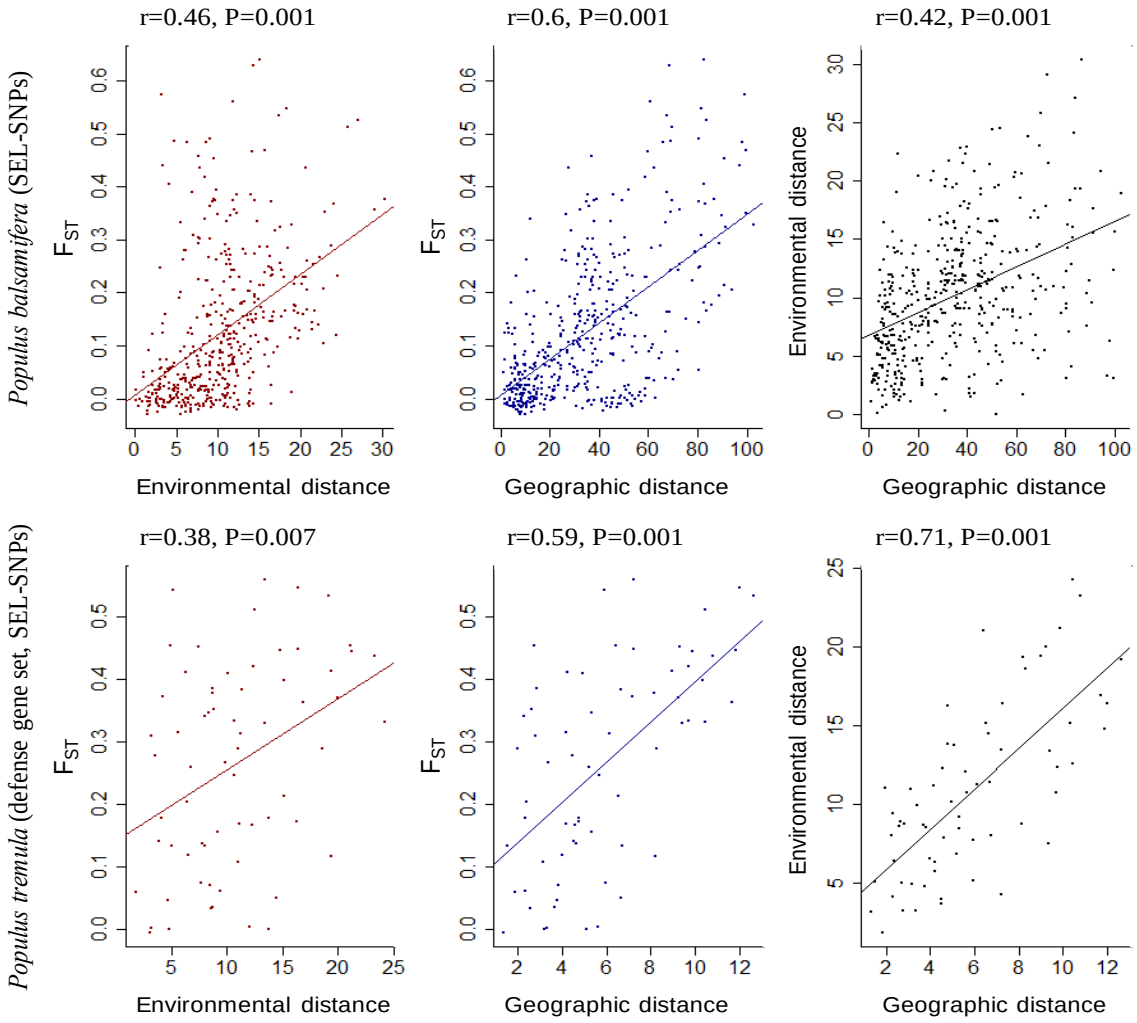


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

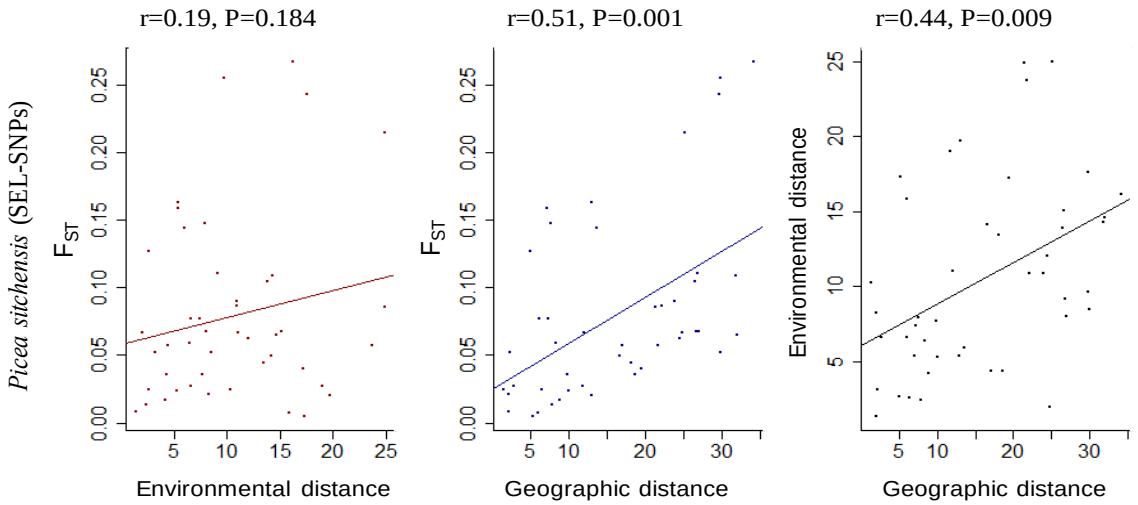


Figure A1. Scatterplots of the relationship between genetic differentiation, based on Nei's F_{ST}, and environmental distance (red), genetic differentiation and geographic distance (blue) as well as between environmental distance and geographic distance (black) for all 19 genetic and genomic datasets considered in this study (continued). Geographic distances were measured as the Euclidean distance between GPS coordinates of sampled populations (latitude, longitude), whereas environmental distances were measured as the Euclidean distance between the two major environmental principal components associated with populations. At the top of each plot figure the correlation coefficient (r) between distances assessed and its significance statistic (P), both estimated using a mantel test.

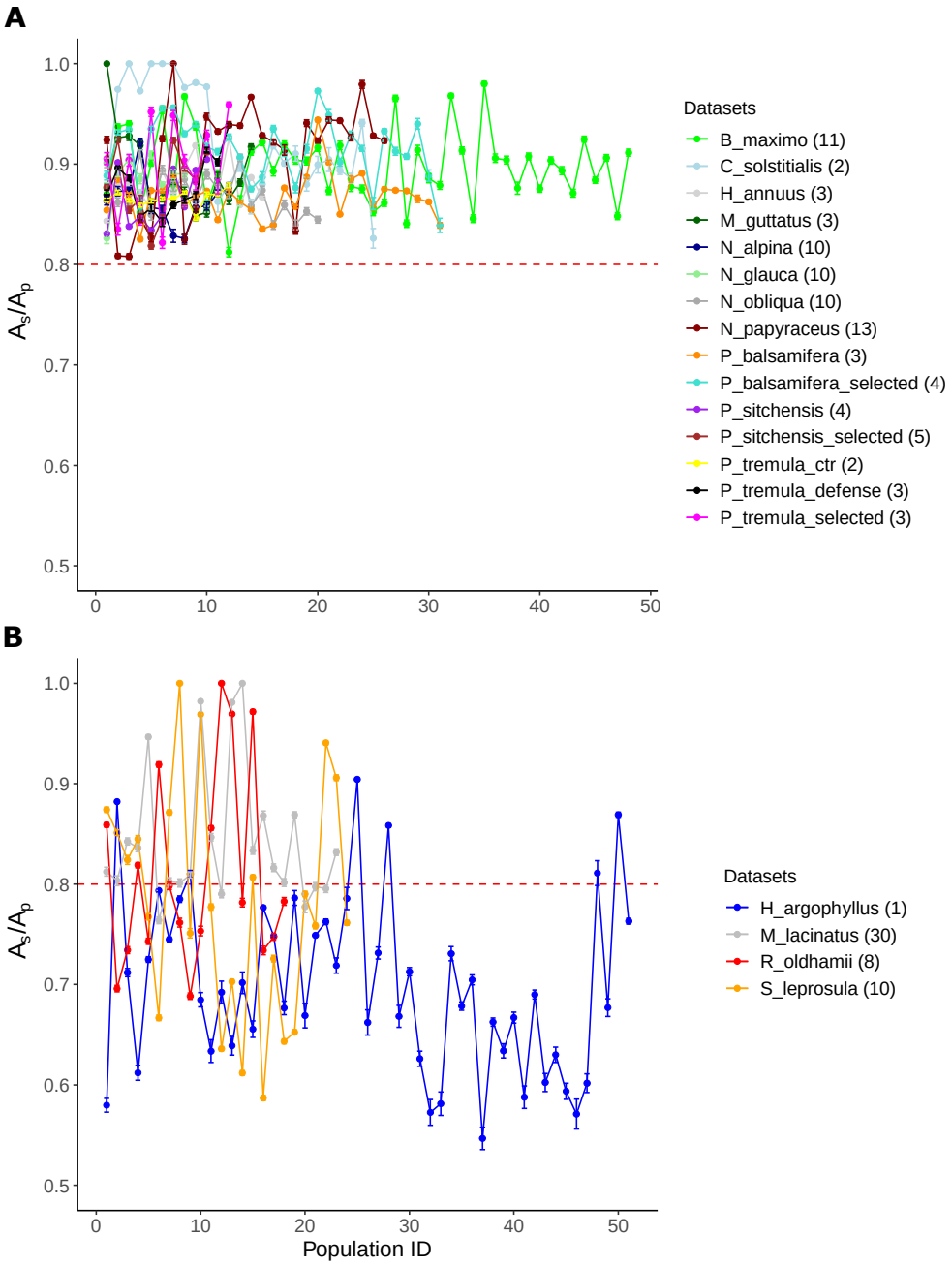


Figure A2. Proportion of allelic diversity captured within a population (one point illustrates one population) when *N* individuals were randomly sampled. Different colors represent different datasets. The legend indicates which color is associated with which dataset and provides the value of *N* used for simulations in brackets. Dashed red lines represent the threshold above which 80% or more of allelic diversity is captured and error bars represent Student 95% confidence intervals calculated from 500 iterations. (A) Datasets included in idealized and realistic simulations. For each dataset, $N = N_{80\%}$. (B) Datasets discarded from idealized and realistic simulations. *N* represents the size of the smallest population within each dataset.

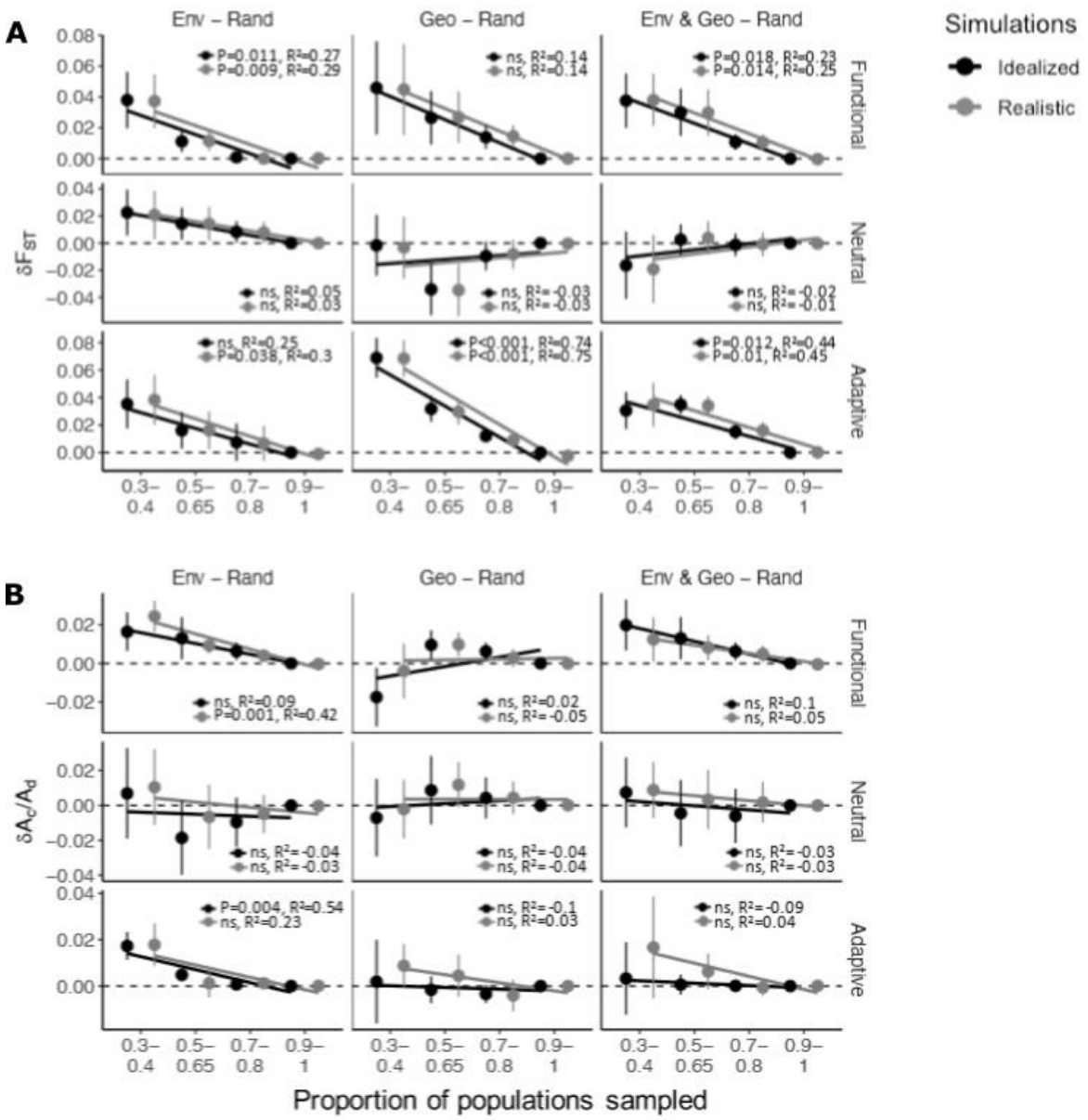


Figure A3. Average differences (±SE) across datasets in (A) population genetic differentiation (Nei's F_{ST}) and (B) allelic diversity (A_c/A_d) estimated from *ex situ* collections simulated using distance-informed (Env: environmental, Geo: geographic, and Env & Geo: environmental and geographic) and random (Rand) population sampling strategies separated by genetic marker classes (functional, neutral, and adaptive). Differences in genetic summary statistics were estimated for different proportions of populations sampled. Environmental distances between populations were estimated using the two major axes of differentiation obtained using a non-metric multidimensional scaling analysis (NMDS). Note that for both realistic and idealized simulations, *Np*=4 (proportion of population sampled = 0.5) was replaced by *Np*=5 (proportion of population sampled = 0.63) for *N. glauca* (SSRs) because at least one distance-informed population sampling strategy could not be simulated using the original *Np* value (Table A4). ns: non-significant.

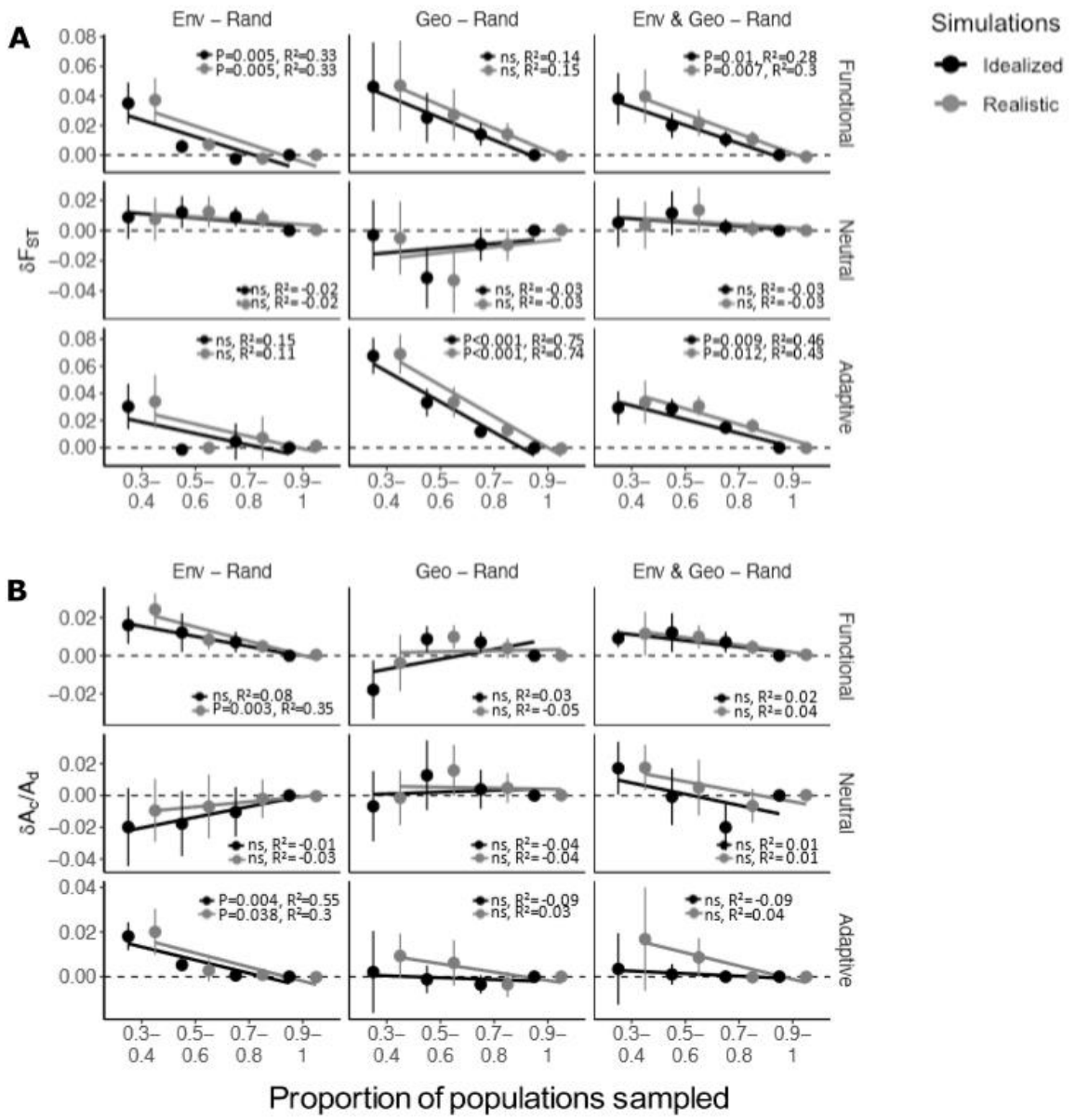


Figure A4. Average differences (±SE) across datasets in (A) population genetic differentiation (Nei's F_{ST}) and (B) allelic diversity (A_c/A_d) estimated from *ex situ* collections simulated using distance-informed (Env: environmental, Geo: geographic, and Env & Geo: environmental and geographic) and random (Rand) population sampling strategies separated by genetic marker classes (functional, neutral, and adaptive). Differences in genetic summary statistics were estimated for different proportions of populations sampled. Environmental distances between populations were estimated by the three major axes of differentiation obtained using a principal component analysis (PCA). Note that for both realistic and idealized simulations, *Np*=6 (proportion of population sampled = 0.3) was replaced by *Np*=7 (proportion of population sampled = 0.35) for *N. obliqua* (SSRs) because at least one distance-informed population sampling strategy could not be simulated using the original *Np* value (Table A4). ns: non-significant.

APPENDIX B. ADDITIONAL INFORMATION (CHAPTER 2)

Genetic and genomics datasets associated with 7 of the 15 studied species downloaded from Dryad (https://datadryad.org/) were modified to meet selection criteria for subsequent analysis. Modifications performed involved subsampling of original datasets through the removal of one or several individuals, populations, or genetic markers. In addition, geographic coordinates provided in Bernhardsson & Ingvarsson (2012) also needed to be adjusted to fully meet the second selection criterion. Below is a detailed list of modifications made to genetic, genomic, and geographic datasets.

Centaurea solstitialis – First, individuals belonging to another species that *C. solsitialis* (*Centaurea melitensis, Centaurea nicaeensis,* and *Centaurea pallescens*) were discarded because they did not span species geographic ranges or isolated part of their ranges and therefore violated the first selection criterion. Second, only the individuals of *C. solstitialis* distributed in Eurasia were kept because samples from the United States and South America were not representative of the species distribution or an isolated part of its distribution in these locations, also violating the first selection criteria. Third, individuals with the prefix "Ar-" were discarded because they could not be associated with a unique population. Two populations in Barker et al. (2017) were named "AR", on located in the United States (Lat 45.696˚, Long −118.871˚), the other located in South America (Lat 39.563˚, Long 67.007˚). Additionally, individuals labeled C1346, C1441, C1445, and C1448 as well as individuals with the prefix "Sie-" were removed from the dataset because they could not be assigned to any populations. Finally, individuals belonging to populations ETC and K113 as well as individuals labeled as C1309, C1310, and C1313, all three belonging to the same population, were discarded as they fell without the geographic range covered by ClimateEU and could not be associated with climatic variables. This would prevent variation

partitioning to be performed and environmental distance among populations to be estimated, impeding downstream statistical analyses. Overall, 225 individuals from 25 populations spanning the species native (Eurasian) distribution range met all three selection criteria and were used for analyses.

Helianthus argophyllus – First, all individuals from another species than *H. argophyllus* (*Helianthus annuus* and *Helianthus debilis*) were discarded as they were poorly sampled and covered only a small part of species distribution ranges, violating the first selection criterion. Second, individuals within populations labeled as ARG-1575, ARG-2623, HEL153, PI649866, PI490291_Moz, 448, 449 and 451 were removed from the dataset because no geographic coordinates were available for these populations, violating the second selection criterion. Overall, 554 individuals from 51 populations spanning the species native ancestral distribution range (Texas; Yatabe et al. 2007), which currently represents an isolated fraction of its distribution (Texas, Florida and North Carolina; https://plants.usda.gov) met all three selection criteria and were used for analyses.

Mimulus guttatus – Individuals belonging to populations labeled as ALA, WLB, CPB, HOC, HEC, ANR, LMC, WTB, DFAL and DAV were discarded as they were not representative of the species North American distribution nor representative of a fraction of the species range, violating the first selection criterion. Overall, 261 individuals from 14 populations sampled across the species British distribution met all three selection criteria and were used for subsequent analyses.

Mimulus lacinatus – In this dataset, 3 of 11 codominant genetic markers were removed (e617, e641, e423) as they are gene-intron-length markers and did not fall within the range of

genetic markers under study (SSR, EST-SSR, SNP, Gen-SNP, SEL-SNP), violating the third selection criterion. The remaining eight microsatellite markers were used for analyses.

Narcissus papyraceus – Individuals within 5 of the 31 populations studied in Simón-Porcar et al. (2015) (populations 1-5, region code CM) were discarded because they fell without the geographic range covered by ClimateEU. Consequently, no climatic variable could be retrieved for these sites, which would prevent variation partitioning to be performed and environmental distance among populations to be estimated, impeding downstream statistical analyses. Overall, 422 individuals from 26 populations sampled throughout most of the species European and North African distribution range met all three selection criteria and were used for analyses.

Picea sitchensis – First, individuals labeled with prefixes "10" and "15" were removed from the SNPs dataset as none of these individuals were assigned to one of the populations listed in Holliday et al. (2010). Second, of the 35 SNPs identified as putatively under selection in the study, only 34 were considered as SEL-SNPs because one SNP (273_98_NS) was absent from the original full genotypes' dataset. Finally, individuals belonging to Rocky Bay and Kodiak Island populations were discarded as no climatic data could be retrieved for these two populations using ClimateNA, which would prevent variation partitioning to be performed and environmental distance among populations to be estimated, impeding downstream statistical analyses. Overall, 286 individuals from 10 populations spanning most of the species native (North American) distribution range met all three selection criteria and were used for analyses.

Populus balsamifera – First, individuals from another species than P. balsamifera (Populus deltoids, Populus tremula, Populus tremuloides, Populus trichocarpa, and Populus angustifolia) were discarded as only a few individuals were sampled (1-8 individuals) per species

and thus neither represent whole nor part of species distribution ranges, violating the first selection criterion. Second, 7 genetic markers (ABi1D_183, CRY11_3201, GI5_5271, PHYB2_5048, GI2_10278, PIF31_1277, PIF31_2601) were removed from the genomic dataset as they represent indel and not SNP variation, violating the third selection criterion. Finally, to comply with the second selection criterion, only individuals belonging to the 31 populations studied in Keller et al. (2012) were kept as geographical coordinates could only be retrieved for these populations. Overall, the dataset we used for analyses matched the one used in Keller et al. (2012), including 443 individuals from 31 populations spanning most of P. balsamifera distribution range.

Populus tremula – In the study performed by Bernhardsson & Ingvarsson (2012), geographical coordinates are given per tree sampled. However, to fully comply with the second selection criterion and be able to both conduct variation partitioning and estimate environmental and geographical distances among populations, geographical coordinates per population are needed. To resolve this issue, we pooled all individuals belonging to the same population together and averaged values of latitude, longitude, and elevation to generate geographical coordinates and elevation data per population.

APPENDIX C. SUPPLEMENTARY MATERIAL (CHAPTER 3)

Table C1. Average estimates (±SE) of all 14 measured (A) and derived (B) seed traits summarized by maternal families for the island (n = 30) and the mainland (n = 80 and n = 30) population. Measurable traits: seed length (SL), seed width (SW), embryo length (EL), embryo width (EW), seed coat width (SCW), seed area (SA), endosperm area (ESA), and embryo area (EA). Derived traits: seed length/width ratio (SLW), embryo length/width ratio (ELW), relative embryo size (RES), relative endosperm size (REndS), seed coat area (SCA), and relative seed coat size (RSCS). Differences in seed morphology between mainland and island populations were significant (α=0.05) for all 14 seed traits, even after the mainland population was subset to 30 randomly selected maternal families.

A – Measured seed traits

Population	SL (cm)	SW (cm)	SCW (cm)	EL (cm)	EW (cm)	SA (cm^2)	ESA (cm^2)	EA (cm^2)
Mainland	1.415	0.768	0.089	1.082	0.155	0.887	0.548	0.169
80 families	±0.016	±0.009	±0.002	±0.012	±0.003	±0.023	±0.014	±0.004
Mainland	1.378	0.742	0.087	1.082	0.151	0.813	0.510	0.166
30 families	±0.022	±0.013	±0.002	±0.018	±0.003	±0.029	±0.017	±0.006
Island	1.674	0.966	0.117	1.198	0.186	1.410	0.807	0.232
	±0.020	±0.015	±0.002	±0.020	±0.004	±0.031	±0.021	±0.007

B – Derived seed traits

Population	SLW (SL/SW)	ELW (EL/EW)	SCA (cm^2) (SA-ESA)	RES (EA/ESA)	REndS (ESA/SA)	RSCS (SCA/SA)
Mainland	1.857	7.275	0.339	0.313	0.619	0.381
80 families	±0.013	±0.094	±0.010	±0.005	±0.003	±0.003
Mainland	1.872	7.342	0.302	0.328	0.629	0.371
30 families	±0.022	±0.132	±0.012	±0.005	±0.006	±0.006
Island	1.769	6.640	0.603	0.287	0.571	0.429
	±0.037	±0.136	±0.013	±0.004	±0.004	±0.004

Table C2. Proportion of variance in measured and derived seed morphological traits explained by populations (mainland, island; fixed effect), maternal families (random effect), and both populations and maternal families when 80 and 30 (referred below as 80 families) or 30 and 30 (referred below as 30 families) mainland and island maternal plants were used for analysis. SL, seed length [cm]; SW, seed width [cm]; EL, embryo length [cm]; EW, embryo width [cm]; SCW, seed coat width [cm]; SA, seed area [cm^2]; ESA, endosperm area [cm^2]; EA, embryo area [cm^2]; SLW, seed length/width ratio; ELW, embryo length/width ratio; RES, relative embryo size; REndS, relative endosperm size; SCA, seed coat area [cm^2]; RSCS, relative seed coat size; R^2_m, proportion of variance explained by fixed effects; R^2_c, proportion of variance explained by fixed and random effects.

Seed trait	Variance explained by populations (R^2_m) [a] (80 families/30 families)	Variance explained by maternal families within populations (R^2_c-R^2_m) [a] (80 families/30 families)	Total variance explained (R^2_c) (80 families/30 families)
SL	0.30/0.48	0.35/0.22	0.65/0.70
SW	0.36/0.48	0.24/0.16	0.60/0.64
SCW	0.32/0.48	0.28/0.20	0.60/0.68
EL	0.10/0.13	0.37/0.34	0.47/0.47
EW	0.11/0.23	0.16/0.15	0.27/0.38
SA	0.49/0.68	0.31/0.16	0.80/0.84
ESA	0.37/0.54	0.34/0.22	0.71/0.76
EA	0.22/0.32	0.29/0.27	0.51/0.59
SLW	0.02/0.02	0.07/0.06	0.09/0.08
ELW	0.03/0.07	0.12/0.12	0.15/0.19
SCA	0.57/0.73	0.24/0.12	0.81/0.85
RES	0.03/0.17	0.23/0.16	0.26/0.33
REndS	0.17/0.30	0.20/0.15	0.37/0.45
RSCS	0.17/0.30	0.20/0.15	0.37/0.45
Average	*0.23/0.35*	*0.24/0.18*	*0.48/0.53*

[a] The effect of population origin or maternal family within populations when 30 mainland maternal plants were used remained significant at α=0.05 for all seed traits listed, except for the ratio between seed length and seed width (SLW, random effect; P = 0.16).

Table C3. PC1 and PC2 loadings for all 14 measured and derived seed traits: seed length (SL, cm), seed width (SW, cm), embryo length (EL, cm), embryo width (EW, cm), seed coat width (SCW, cm), seed area (SA, cm^2), endosperm area (ESA, cm^2), embryo area (EA, cm^2), seed length/width ratio (SLW), embryo length/width ratio (ELW), relative embryo size (RES), relative endosperm size (REndS), seed coat area (SCA, cm^2), and relative seed coat size (RSCS).

Seed trait	PC1 –57.8% var. explained	PC2 –15.9% var. explained
SL	-0.33	-0.06
SW	-0.34	0.02
SCW	-0.29	0.16
EL	-0.27	0.17
EW	-0.27	0.31
SA	-0.34	-0.09
ESA	-0.33	-0.01
EA	-0.31	0.23
SLW	0.11	-0.15
ELW	0.13	-0.33
SCA	-0.34	-0.18
RES	0.09	0.46
REndS	0.18	0.45
RSCS	-0.18	-0.45

Table C4. Proportion of variance in measured and derived seed morphological traits explained by populations (mainland, island; fixed effect), maternal families (random effect), and both populations and maternal families. SL, seed length [cm]; SW, seed width [cm]; EL, embryo length [cm]; EW, embryo width [cm]; SCW, seed coat width [cm]; SA, seed area [cm^2]; ESA, endosperm area [cm^2]; EA, embryo area [cm^2]; SLW, seed length/width ratio; ELW, embryo length/width ratio; RES, relative embryo size; REndS, relative endosperm size; SCA, seed coat area [cm^2]; RSCS, relative seed coat size; R^2_m, proportion of variance explained by fixed effects; R^2_c, proportion of variance explained by fixed and random effects.

Seed trait	Variance explained by populations (R^2_m) [a]	Variance explained by maternal families within populations (R^2_c-R^2_m) [a]	Total variance explained (R^2_c)
SL	0.30	0.35	0.65
SW	0.36	0.24	0.60
SCW	0.32	0.28	0.60
EL	0.10	0.37	0.47
EW	0.11	0.16	0.27
SA	0.49	0.31	0.80
ESA	0.37	0.34	0.71
EA	0.22	0.29	0.51
SLW	0.02	0.07	0.09
ELW	0.03	0.12	0.15
SCA	0.57	0.24	0.81
RES	0.03	0.23	0.26
REndS	0.17	0.20	0.37
RSCS	0.17	0.20	0.37
Average	*0.23*	*0.24*	*0.48*

[a] The effect of population origin or maternal family within populations is significant at α=0.05 for all seed traits listed.

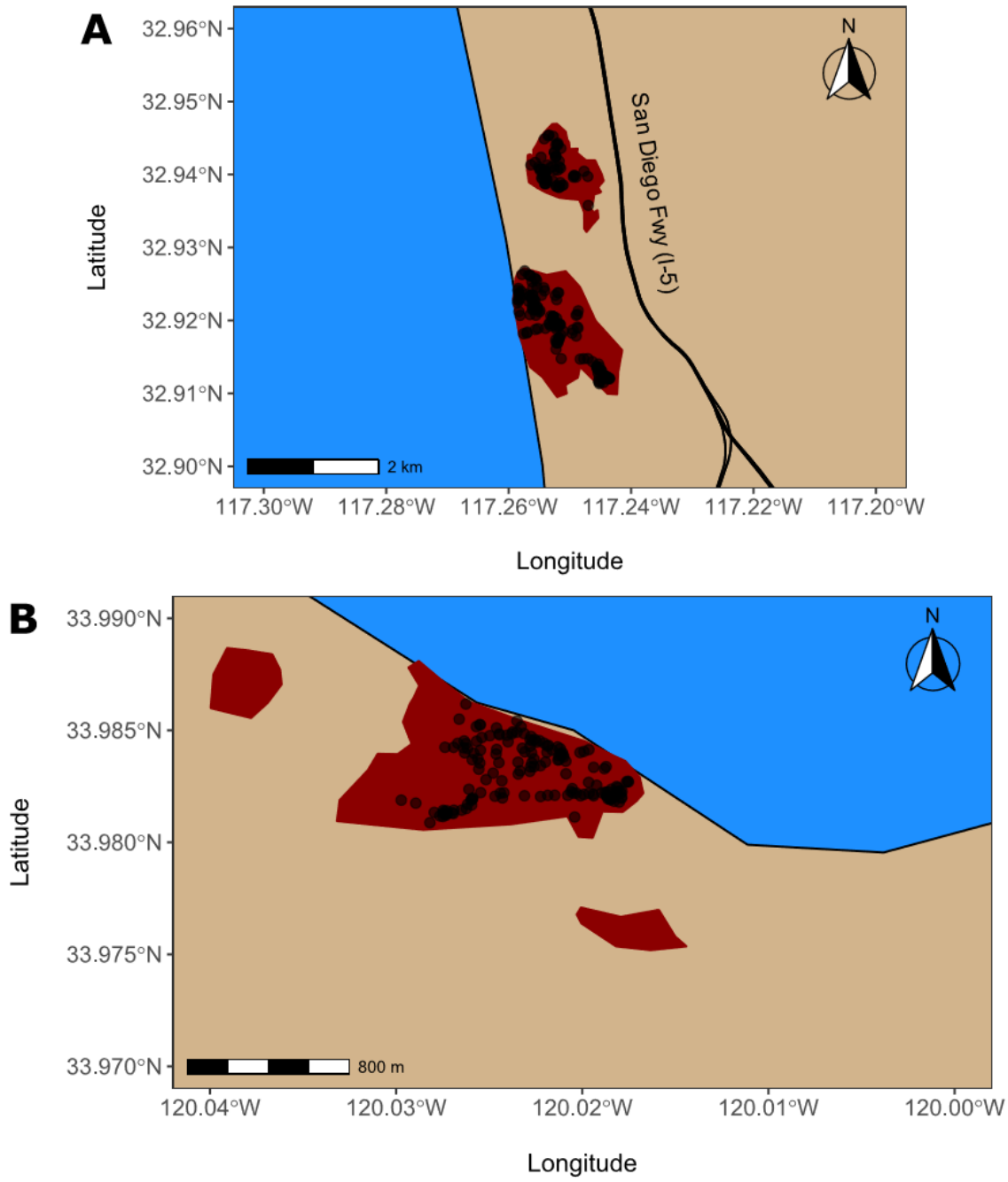
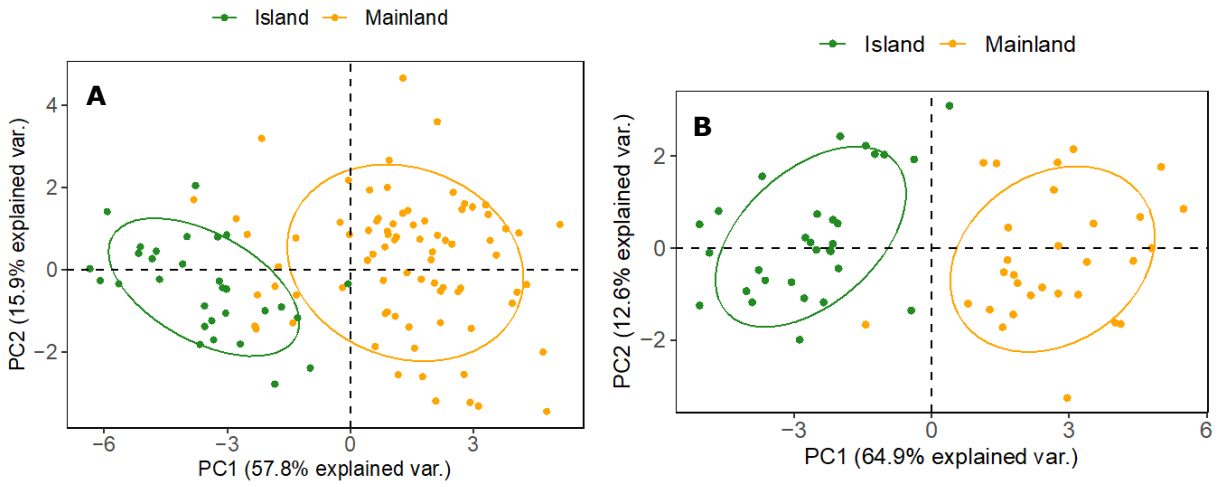


Figure C1. Torrey pine distribution (red shades) and position of maternal trees sampled (closed black circles). (A) Torrey Pine State Reserve, *Pinus torreyana* subsp *torreyana*, CA. (B) Santa Rosa Island, *Pinus torreyana* subsp *insularis*, CA.



Seed traits with highest loading on	Loading estimated using 80 mainland maternal families	Loading estimated using 30 mainland maternal families
PC1		
Seed length	0.33	0.31
Seed width	0.34	0.32
Seed area	0.34	0.33
Endosperm area	0.33	0.31
Seed coat area	0.34	0.33
PC2		
Relative seed coat size	0.45	0.44
Relative endosperm size	0.45	0.44
Relative embryo size	0.46	0.32
Embryo length	0.17	0.45

Figure C2. Principal component analysis using all 14 morphological and derived seed traits evaluated from (A) 80 mainland (orange) and 30 island (green) maternal families, and (B) 30 randomly selected mainland (orange) and 30 island (green) maternal families. Listed in the table are absolute values of PC1 and PC2 loadings for seed traits with highest loadings.

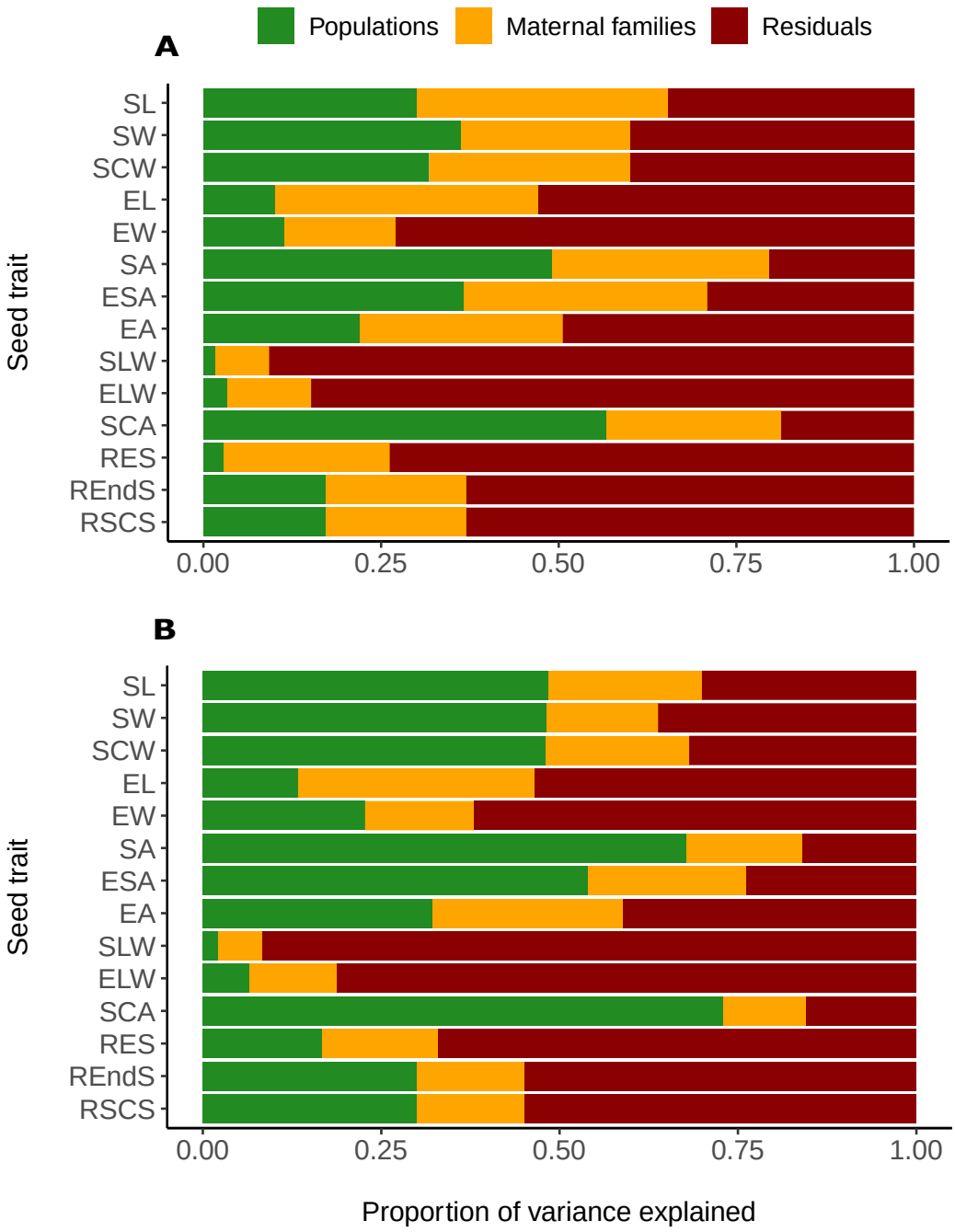


Figure C3. Proportion of variance in seed morphology explained by populations (green), maternal families within populations (orange), and other variables not accounted for in the model (residuals; dark red) for each of the 14 measured and derived seed traits when (A) 80 and 30, and (B) 30 random and 30 mainland and island individuals were considered for analysis, respectively. SL, seed length [cm]; SW, seed width [cm]; SCW, seed coat width [cm]; EL, embryo length [cm]; EW, embryo width [cm]; SA, seed area [cm^2]; ESA, endosperm area [cm^2]; EA, embryo area [cm^2]; SLW, seed length/width ratio; ELW, embryo length/width ratio; SCA, seed coat area [cm^2]; RES, relative embryo size; REndS, relative endosperm size; RSCS, relative seed coat size. See Table C2 for numerical estimates.

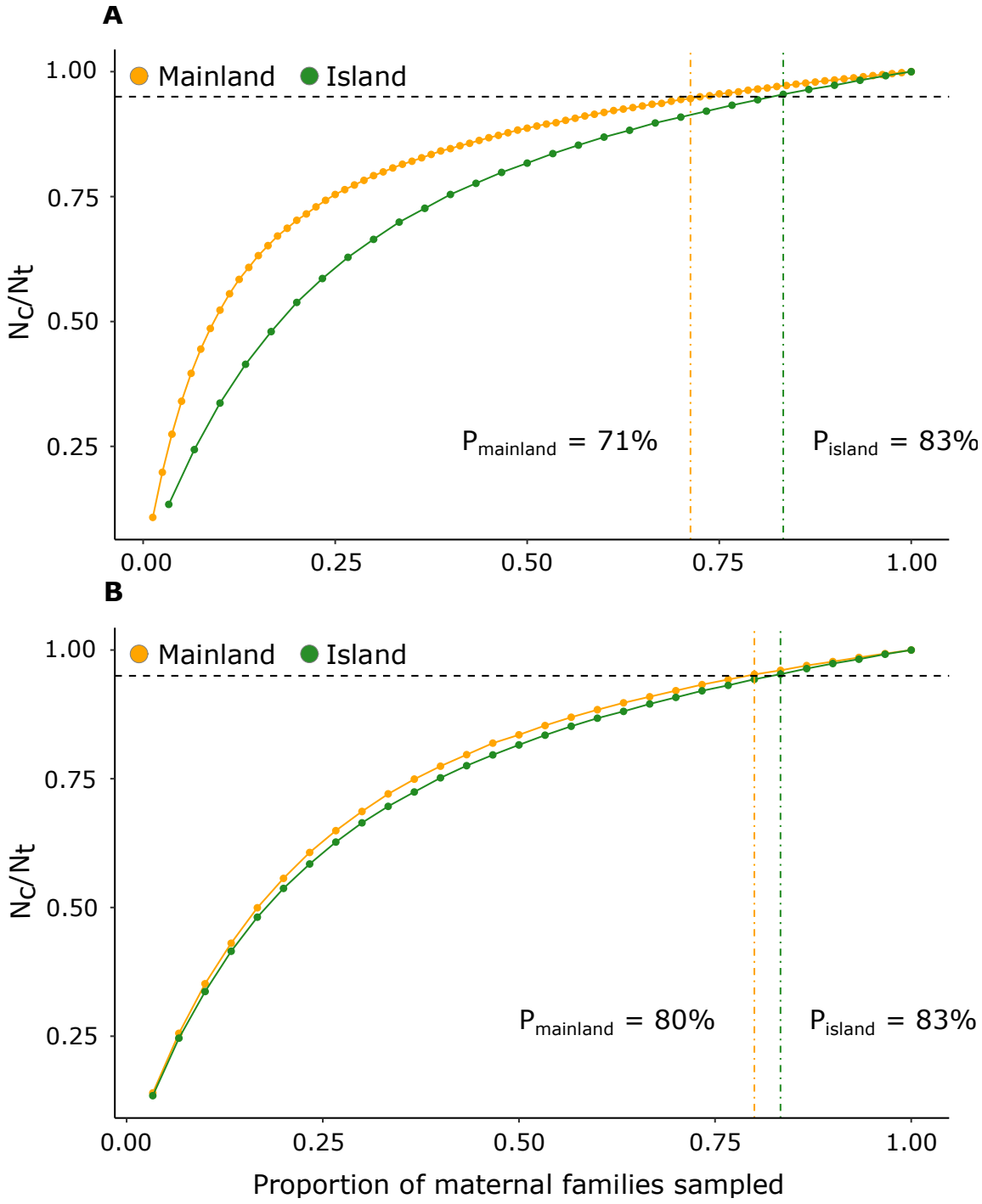


Figure C4. Phenotypic variation captured across seed traits in simulated collections (N_c) relative to total phenotypic variation present in seed populations (N_t) when (A) 80 mainland maternal families or (B) 30 randomly selected mainland maternal families were simulated. Average proportions of phenotypic variation captured (N_c/N_t) were estimated for various proportions of maternal families sampled. P_{island} and $P_{mainland}$ represent the proportion of maternal families required to capture 95% of morphological variation (horizontal dashed line) present in island (green) and mainland (orange) *ex situ* seed populations, respectively.

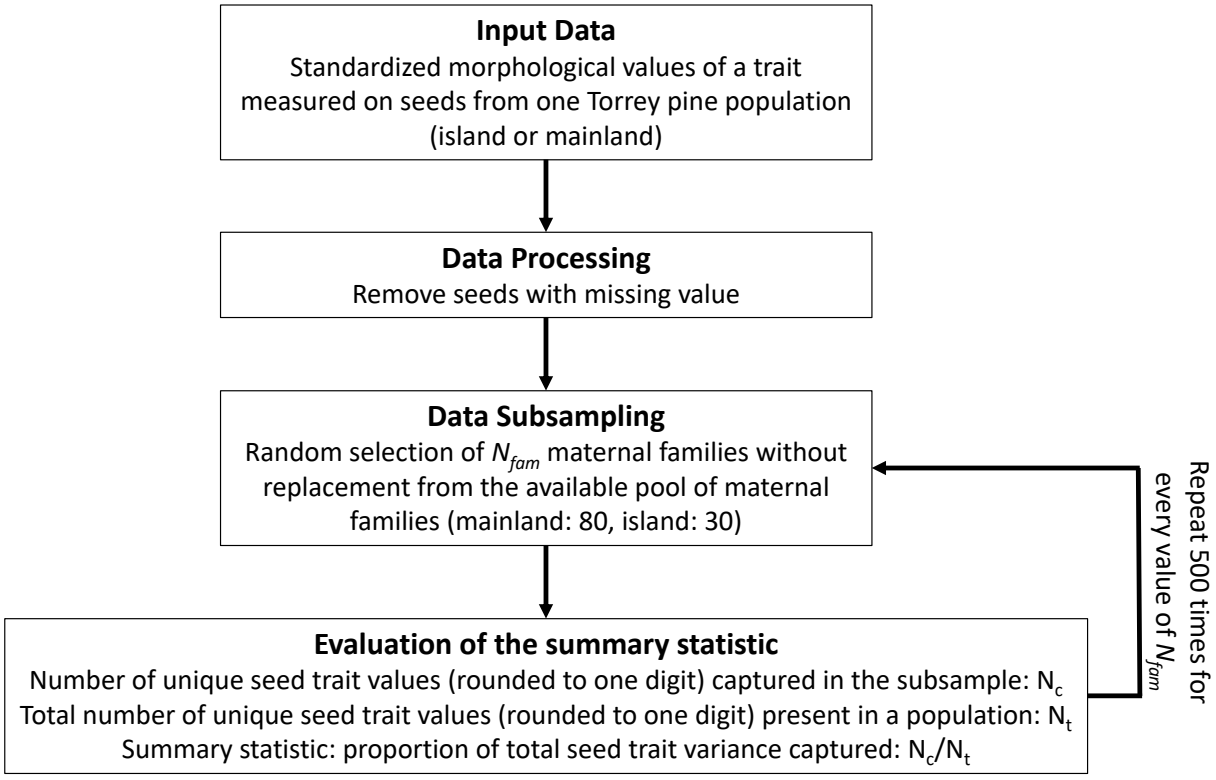


Figure C5. Simulation framework used to estimate phenotypic variation in seed morphology captured after resampling of contemporary Torrey pine *ex situ* seed collections. Simulations using this framework were conducted for each seed trait and Torrey pine population independently. Computation proceeds from top to bottom.

APPENDIX D. SUPPLEMENTARY MATERIAL (CHAPTER 4)

Table D1. *Pinaceae* species represented in the set of 80 putatively adaptive genes.

Species with BLAST's hits	Number of BLAST's hits
Larix gmelinii	3
Larix kaempferi	1
Picea abies	3
Picea glauca	26
Picea sitchensis	25
Pinus elliottii	1
Pinus lawsonii	1
Pinus massoniana	4
Pinus monticola	1
Pinus pinaster	2
Pinus pinea	1
Pinus radiata	3
Pinus sylvestris	2
Pinus tabuliformis	2
Pinus taeda	5
Total	80

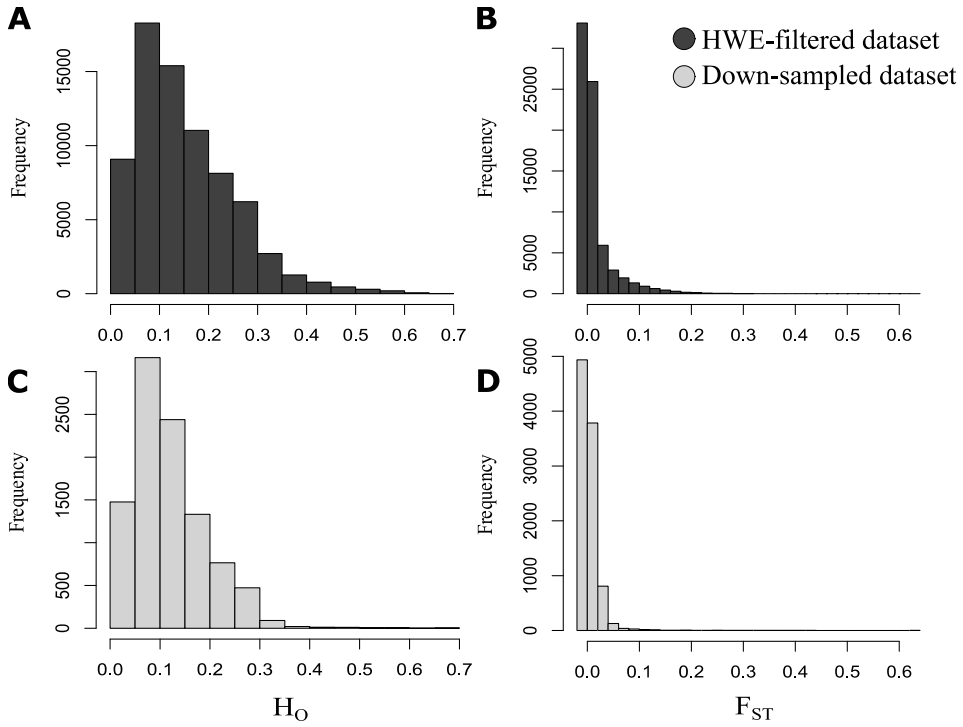


Figure D1. Distribution of genetic summary statistics in the HWE-filtered (73,928 variants, dark grey) and down-sampled (9,795 variants, light grey) datasets. (A, C) observed heterozygosity: H_O. (B, D) pairwise genetic distances: Nei's F_{ST}.

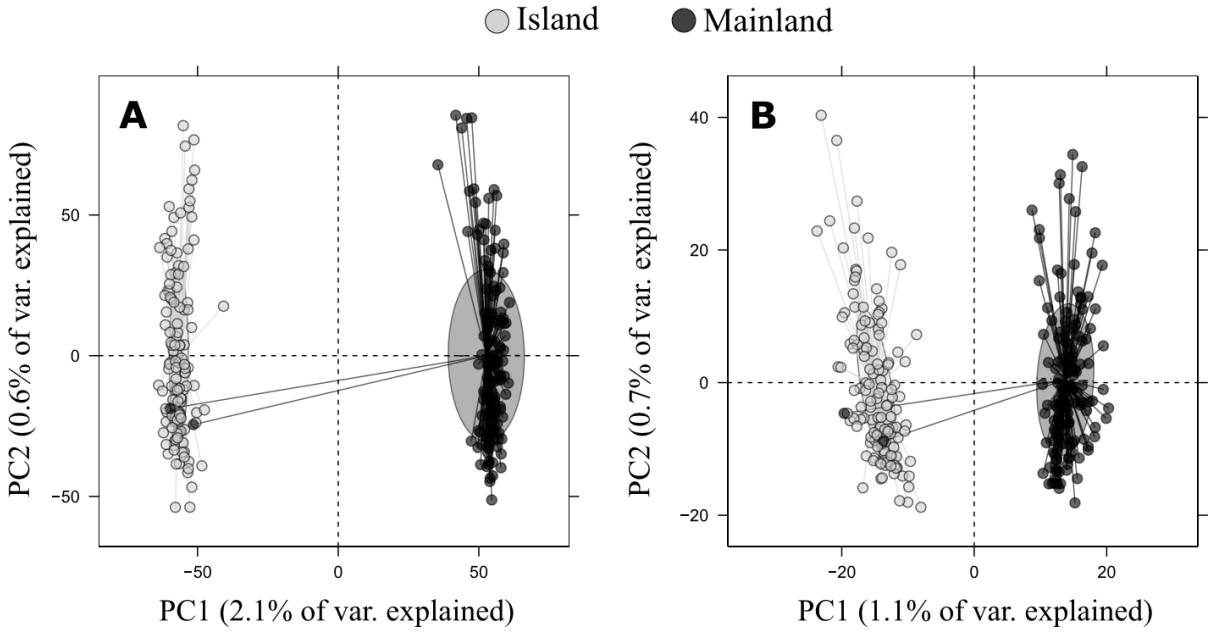


Figure D2. Principal component analyses conducted on two different genetic datasets. (A) HWE-filtered dataset, including 73,928 SNPs. (B) Down-sampled dataset, including 9,795 SNPs. Individuals were grouped by populations, including island (SRI, grey) and mainland (TPSR, black) populations. Variance explained by the first and second axis of differentiation are provided in parentheses.

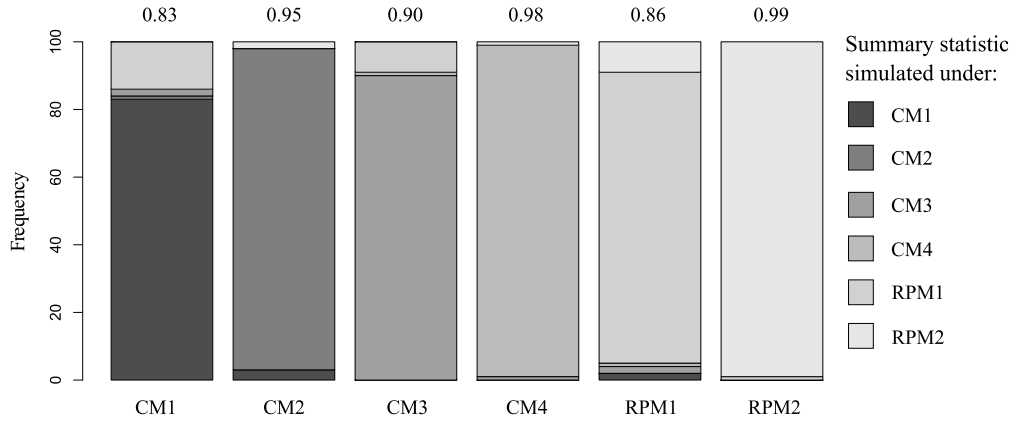


Figure D3. Misclassification proportions based on 100-fold cross-validation simulations with a tolerance rate of 0.01 for all six demographic scenarios. Different colors indicate summary statistics simulated under different demographic scenario. The proportion of simulations correctly assigned to their demographic model is provided above each bar. See text and Figure 12 for details on each scenario.

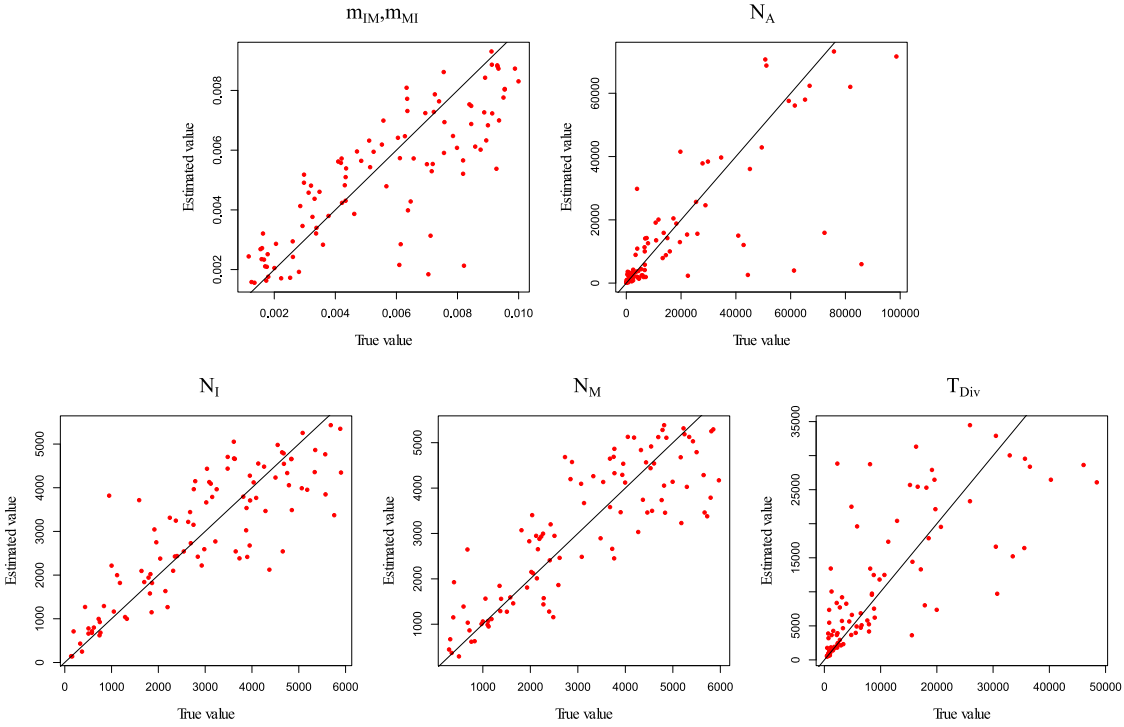


Figure D4. Cross-validation (100-fold) for demographic parameters inferred from the best model. True values (x-axis) are given against estimated values (y-axis) approximated using non-linear postsampling regression adjustment on log-transformed parameters and a tolerance rate of 0.01. The solid line represents the 1:1 (identity) relationship between true and estimated values. The closer to the line points are, the more accurate is the parameter inferred. N_A, ancestral effective population size; N_I, island effective population size; N_M, mainland effective population size; m_{IM}, migration probability from island to mainland; m_{MI}, migration probability from mainland to island; T_{Div}, time of population divergence.

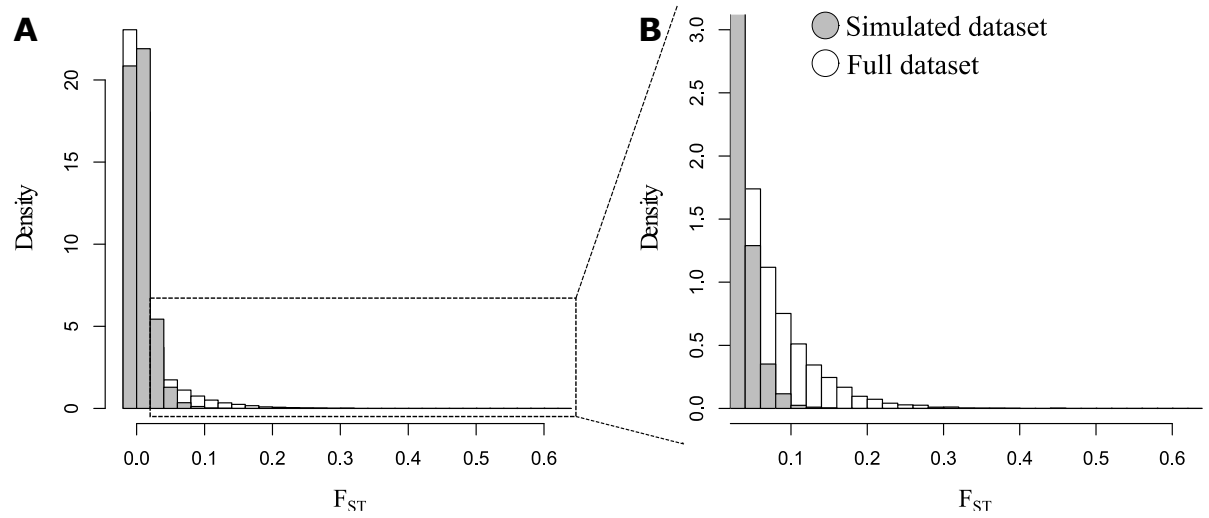


Figure D5. Distributions of pairwise Nei's F_{ST} values between the island and mainland populations estimated from the simulated (grey) and full (white) dataset. Both distributions were estimated using 93,085 SNPs. (A) Full range of observed and simulated F_{ST} values. (B) Right-hand tail of observed and simulated F_{ST} distributions. Neutral F_{ST} estimates were obtained by simulating the best demographic model (see Figure 15) in SIMCOAL2.

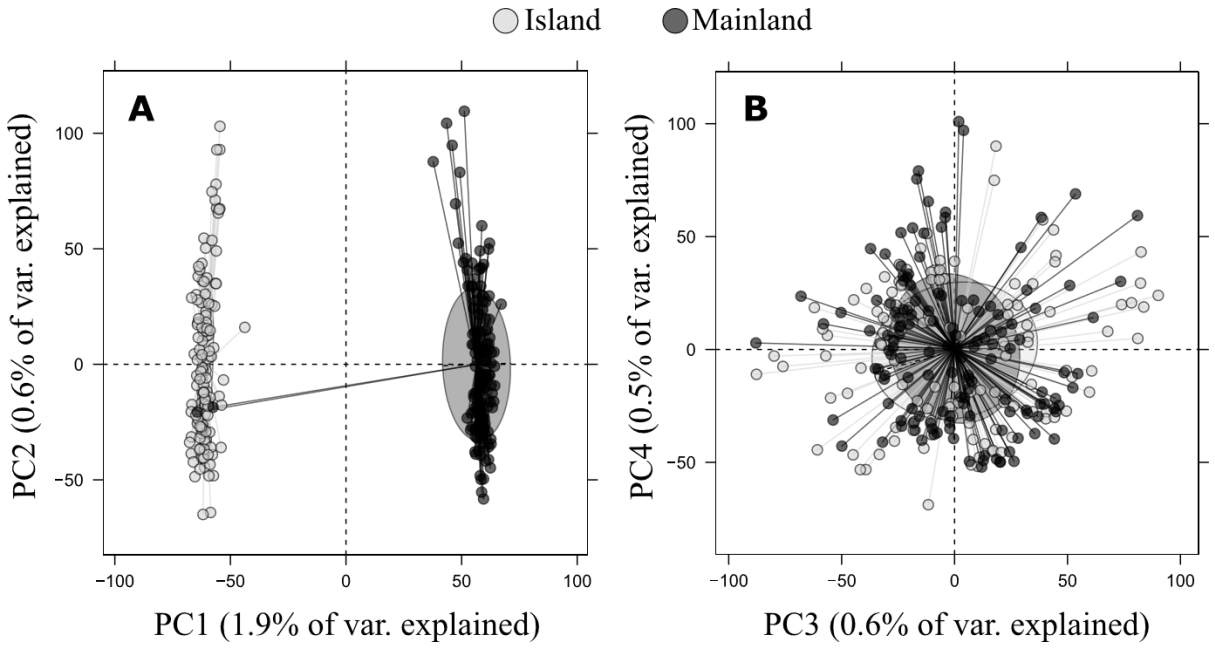


Figure D6. Principal component analysis conducted on the full dataset (93,085 variants) grouped by populations (island [SRI], grey; mainland [TPSR], black). (A) Distribution of individuals on the first and second axis of differentiation. (B) Distribution of individuals on the third and fourth axis of differentiation. Variation explained by each axis is given in parentheses.

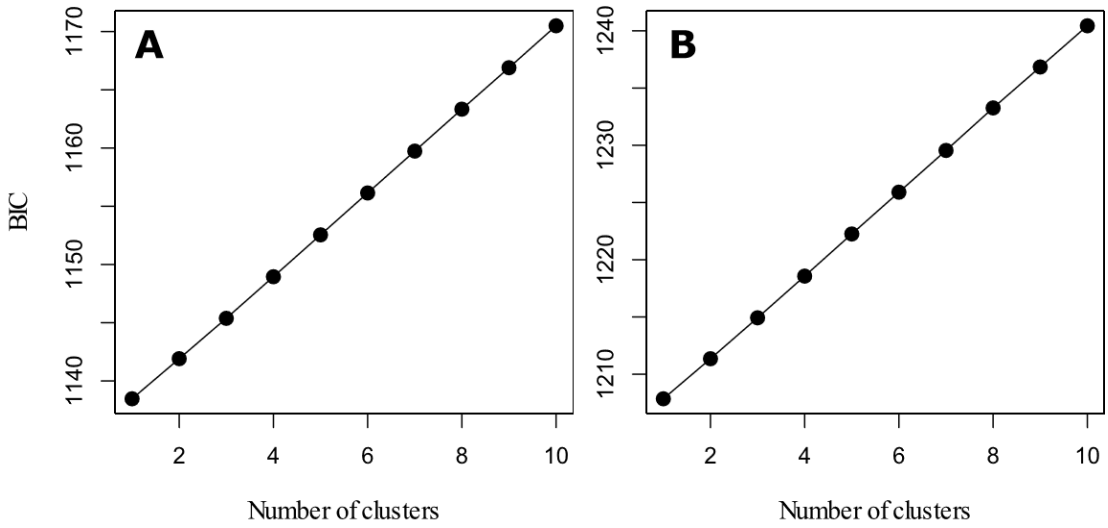


Figure D7. Bayesian Information Criterion (BIC) calculated for structure models assuming from 1 to 10 genetic clusters (*k*) within Torrey pine populations. (A) Island (SRI) population. (B) Mainland (TPSR) population. For both populations, $k = 1$ is associated with the lowest BIC value, identifying one as the most likely number of genetic clusters within populations.