Received: 2 July 2021 | Revised: 29 June 2022 | Accepted: 1 July 2022

DOI: 10.1111/mec.16615

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Reduced representation sequencing to understand the evolutionary history of Torrey pine (*Pinus torreyana* parry) with implications for rare species conservation

Lionel N. Di Santo[1] | Sean Hoban[2] | Thomas L. Parchman[3] | Jessica W. Wright[4] | Jill A. Hamilton[1,5]

[1]Department of Biological Sciences, North Dakota State University, Fargo, North Dakota, USA

[2]The Morton Arboretum, Lisle, Illinois, USA

[3]Department of Biology, University of Nevada Reno, Reno, Nevada, USA

[4]USDA- Forest Service, Pacific Southwest Research Station, Davis, California, USA

[5]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, Pennsylvania, USA

Correspondence
Lionel N. Di Santo, Department of Biological Sciences, North Dakota State University, Fargo, ND, USA.
Email: lionel.disanto@unige.ch

Funding information
Morton Arboretum Center for Tree Science Fellowship; North Dakota Experimental Program to Stimulate Competitive Research, Grant/Award Number: ND-EPSCoR NSF-IIA-1355466; North Dakota State University Environmental and Conservation Sciences Graduate Program; USDA Forest Service Health Protection Gene Conservation program; Western Wildlands Environmental Threat Assessment Center

Handling Editor: Mitchell Cruzan

Abstract

Understanding the contribution of neutral and adaptive evolutionary processes to population differentiation 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 interpopulation 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 identify loci possibly under selection, and evaluate genetic rescue as a potential conservation strategy. Overall, we observed exceedingly low standing variation within both Torrey pine populations, reflecting consistently low effective population sizes across time, and limited genetic differentiation, suggesting maintenance of gene flow between populations following divergence. However, genome scans identified more than 2000 candidate SNPs potentially under divergent selection. Combined with previous observations indicating population phenotypic differentiation, this indicates natural selection has probably contributed to the evolution of population genetic differences. Thus, while reduced genetic diversity, small effective population size, and genetic connectivity between populations suggest genetic rescue could mitigate the adverse effects of rarity, evidence for adaptive differentiation suggests 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.

KEYWORDS
adaptation, demographics, genetic rescue, genome scans, rare species conservation

1 | INTRODUCTION

Conservation biology aims to preserve rare species and determine the appropriate management strategies necessary for long-term persistence and maintenance of evolutionary potential (di Santo & Hamilton, 2020; Ralls et al., 2018; Swarts et al., 2009; Young et al., 1999). Rare species may have reduced effective population sizes (N_e), impeding populations' ability to adapt to change (e.g., $N_e < 1000$) or increasing probability of inbreeding (e.g., $N_e < 100$) (Frankham et al., 2014), ultimately increasing the risk of local extirpation. Combined, rarity and isolation are often associated with stochastic loss of genetic variation (Aguilar et al., 2008; Hague & Routman, 2016; Young et al., 1996). Genetic rescue is one conservation strategy that could be used in both animals and plants to mitigate this loss (Bossuyt, 2007; Hedrick et al., 2014; Johnson et al., 2010; Madsen et al., 1999; Westemeier et al., 1998; Willi et al., 2007). Genetic rescue introduces or restores gene flow between populations to alleviate 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 (Edmans, 2007; Goto et al., 2011; Hufford & Mazer, 2003). Thus, considering the contribution of natural selection to the evolution of population genetic differences is important for genetic rescue, as it may ultimately lead to maladaptation of migrants or translocated individuals (Lowry et al., 2008; Nosil et al., 2005). For rare species conservation, an understanding of contemporary effective population size is 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 establishing 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 (Cavender et al., 2015; Potter et al., 2017; Pritchard et al., 2012), providing a critical resource for the preservation of genetic variation, restoration, or reintroduction (Guerrant Jr et al., 2014; Potter et al., 2017). Ex situ sampling designs traditionally rely on neutral population genetic structure to guide sampling decisions (Caujapé-Castells & Pedrola-Monfort, 2004; Gapare et al., 2008; Hoban, 2019; Hoban & Schlarbaum, 2014). 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 (Bonin et al., 2007; Holderegger et al., 2006; McKay & Latta, 2002; Teixeira & Huber, 2021). To optimize neutral and adaptive genetic variation preserved when sampling ex situ, consideration of the impact different evolutionary processes have had on population genetic structure is needed (di Santo et al., 2021; di Santo & Hamilton, 2020). 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, the preservation of adaptive genetic differences should include of diverse population origins, discrete conservation of divergent populations ex situ, and population origin in the development of breeding programmes.

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, including both populations' demographic and adaptive history (Abebe et al., 2015; Liu et al., 2020; Wang et al., 2020; Xia et al., 2018). However, despite extensive use in characterizing genomes as well as neutral and adaptive variation in conifers (Eckert et al., 2010; Namroud et al., 2008; Nystedt et al., 2013; Stevens et al., 2016; Tyrmi et al., 2020; Wang et al., 2020), such data have only rarely been used to inform conservation management decisions.

Torrey pine (*Pinus torreyana* Parry) is a critically endangered pine (IUCN, 2021), endemic to California. One of the rarest pine species in the world (Critchfield & Little, 1966; Dusek, 1985), Torrey pine's distribution spans one island population (*Pinus torreyana* subsp. *insularis*) of approximately 3000 reproductive individuals (Santa Rosa Island, CA), and one mainland population (*Pinus torreyana* subsp. *torreyana*) of approximately 4000 reproductive individuals (Torrey Pine State Reserve in La Jolla, CA) (Franklin & Santos, 2011; Hall & 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 anthropogenic, abiotic, and biotic disturbances (Franklin & Santos, 2011; Hamilton et al., 2017; Ledig & Conkle, 1983; Waters & Schaal, 1991; Whittall et al., 2010). For these reasons, the species may be at imminent risk for population-scale extirpation 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 variation (Hamilton et al., 2017), and increase evolutionary potential necessary to respond to current and future ecological challenges (Carlson et al., 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 reduced genetic variation (Hamilton et al., 2017). However, if adaptive genetic differences have evolved between island and mainland populations, fitness consequences following the disruption of coadapted gene

complexes may not be observed in the first generation. Thus, although the combination of exceedingly low genetic diversity and conservation status suggest Torrey pine may be a candidate for genetic rescue, an evaluation of the species' demographic and adaptive evolutionary history will be necessary to inform management of this endangered conifer.

With this study, we use genomic data to quantify and model aspects of demographic and evolutionary history in *Pinus torreyana*, delineating the contribution of stochastic and deterministic processes to genomic differentiation. To that end, we ask 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 non-neutral processes within species collections, and the possible risks associated with genetic rescue. This study provides a framework for applying genomics to rare species management using an evaluation of historical and contemporary variation in effective population size, genetic connectivity, and adaptive population divergence. This framework provides information necessary to identify candidates for genetic rescue and assess the potential risks associated with artificially increasing or restoring gene flow between populations where conservation of species' evolutionary potential may be needed.

2 | MATERIALS AND METHODS

2.1 | 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 1). 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 and later processed to extract genomic DNA using between 25–30 mg of dry tissue and a modified CTAB protocol (Doyle & 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).

2.2 | 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 nonbarcoded (*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 noncoding 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 (>2 nM). 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 five lanes of an Illumina HiSeq 2500 using the 100 bp single-end sequencing protocol (1 × 100 bp).

2.3 | De novo assembly and SNP calling

Demultiplexing of sequence files was performed using IPYRAD version 0.9.12 (Eaton & Overcast, 2020) allowing one mismatch in the barcode sequence. Reads were filtered, assembled de novo, and used to call SNPs within the dDOCENT version 2.7.8 pipeline (Puritz, Hollenbeck, & Gold, 2014; Puritz, Matz, et al., 2014) (see Appendix S1 for details). This approach 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, 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 data set. 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 data set). 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.

2.4 | Population structure and genetic diversity analyses

To describe and estimate contemporary genetic differences between Torrey pine populations, we first assessed population

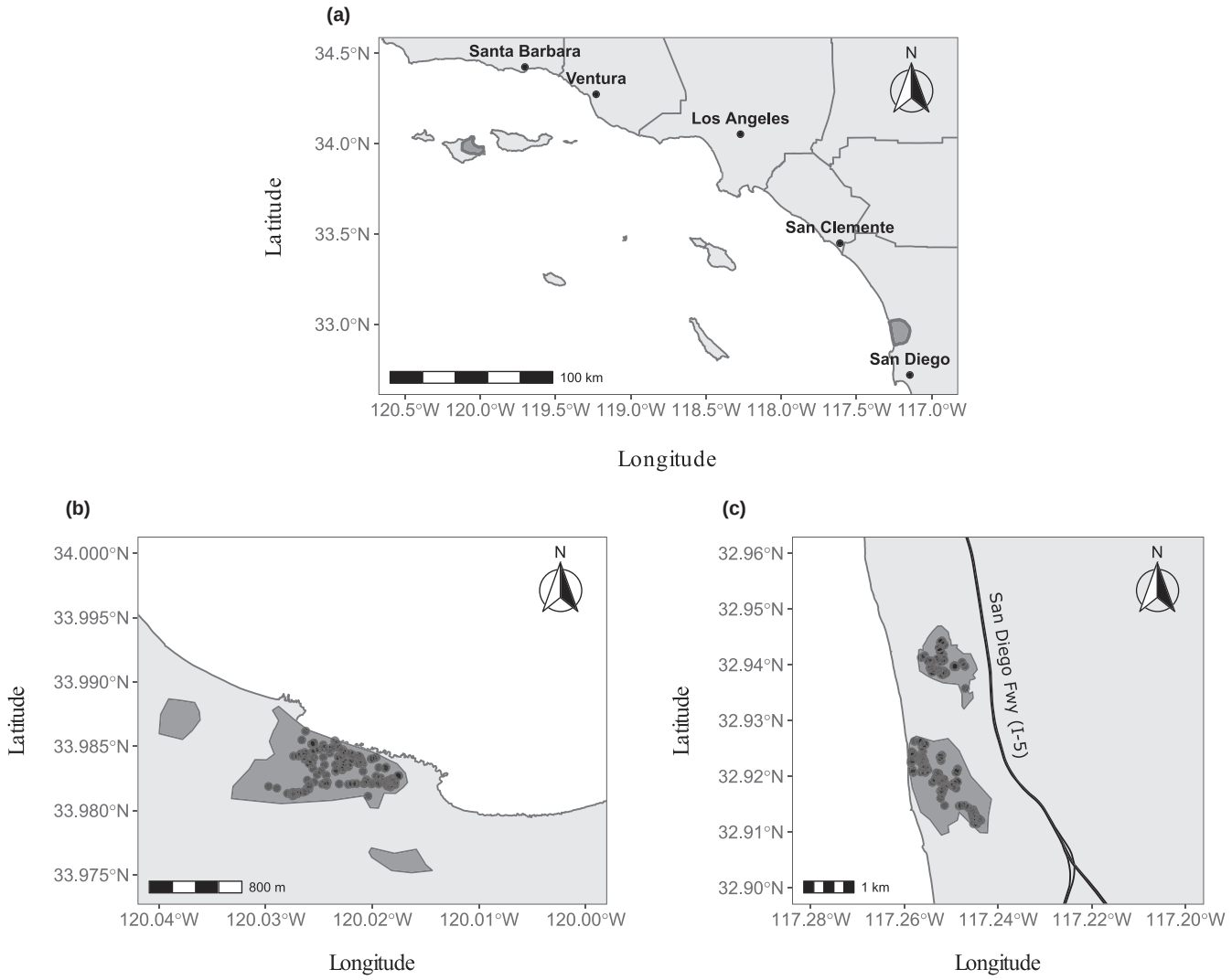


FIGURE 1 (a) Distribution of the last two remanent populations of Torrey pine (dark grey shades). Top left: Santa Rosa Island, Channel Islands, CA. Bottom right: Torrey pine state reserve, La Jolla, CA. (b and c) Population-specific distribution of Torrey pine (dark grey shades) and trees sampled for needle tissue (circles) on Santa Rosa Island (SRI, b) and at the Torrey pine state reserve (TPSR, c)

genetic structure using principal component analysis (PCA) implemented in the R package ADEGENET (Jombart, 2008; Jombart & Ahmed, 2011). Unless otherwise stated, analyses were performed in R version 3.6.2, 4.0.2, or 4.1.2 (R Core Team, 2019, 2020, 2021). To quantify genetic structure within (F_{IS}) and between (F_{ST}) island and mainland populations, we calculated Weir and Cockerham F-statistics (Weir & Cockerham, 1984) and constructed 95% confidence interval around estimates by bootstrapping sums of variance components over loci and deriving F-statistics 10,000 times using the R package HIERFSTAT (Goudet & Jombart, 2020).

To further examine 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 1–10 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 reran the analysis excluding the two individuals while maintaining the same range for *k* (1–10) and retaining 121 principal components (90.38% of variation explained after ordination).

Finally, to evaluate Torrey pine evolutionary potential, contemporary standing genetic variation within the species was estimated by calculating expected heterozygosity (H_E) and coancestry coefficients (θ) for both island and mainland populations independently. H_E was calculated for each SNP separately using the R package ADEGENET and averaged across loci to provide population-level

estimates. To evaluate θ, we used the R package RELATED (Pew et al., 2015) which 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 θ values across all individuals within TPSR and SRI provided population estimates of genetic relatedness. To build 95% confidence intervals around H_E and θ averages, the empirical distribution for each parameter within each population was bootstrapped 10,000 times in R using the BOOT (Canty & Ripley, 2021; Davison & Hinkley, 1997) and SIMPLEBOOT (Peng, 2019) packages.

2.5 | ABC demographic modelling

2.5.1 | Demographic models

To evaluate the impact genetic drift and gene flow have had on 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 2a), and (2) two-population demic expansion (Figures 2b,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 a single ancestral population of Torrey pine diverged to form one island and one mainland population following tectonic movement (Figure 2a). 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.

The remaining four models (CM1, CM2, CM3, CM4) tested two different hypotheses of land colonization where one population was founded by the other (Figure 2b,c). Models CM1 and CM2 specifically test the hypothesis that Santa Rosa Island was colonized by a subset of mainland individuals (Ledig & 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 (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 et al., 2009). Details on demographic parameters and their prior distributions are provided in Table 1.

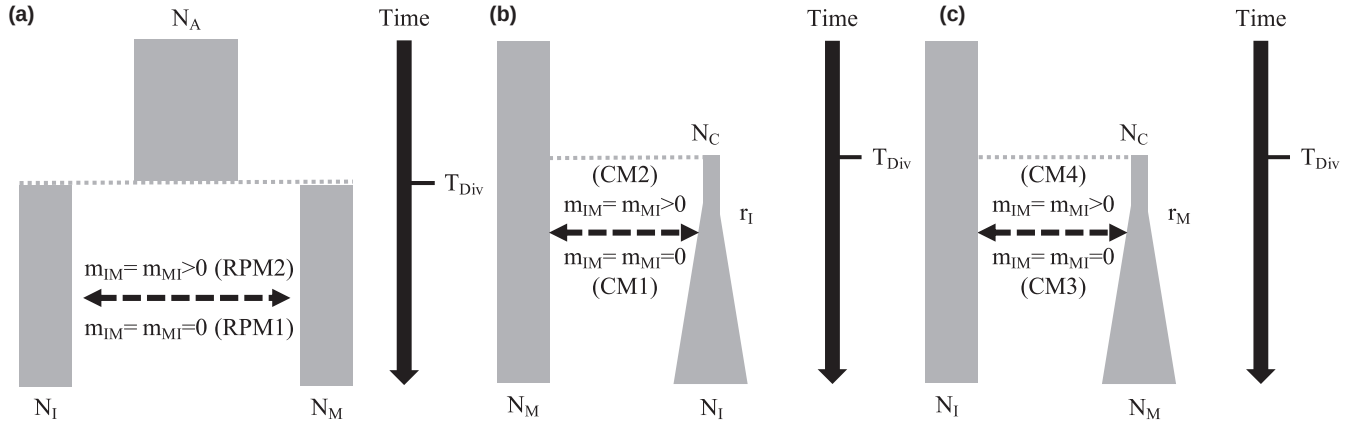


FIGURE 2 Schematics of demographic scenarios simulated. Rectangles represent current or ancestral populations, dashed arrows represent migration between populations, and solid arrows represent time. Note that two demographic scenarios (with and without migration between populations) are presented in each panel. (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

TABLE 1 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(100:6000)	ind.
All	Mainland effective population size	N_M[b]	U(100:6000)	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	Minimum derived allele frequency	maf[f]	0.01	–

[a]Census size of reproductively mature trees is of approximately 3000 to 4000 individuals in island and mainland population respectively (Franklin & Santos, 2011; Hall & Brinkman, 2015). We selected a wide prior range around those estimates to account for the fact that effective and not census population sizes were simulated.

[b]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).

[c]Time since populations' isolation was estimated between 4300 and 430,000 years ago (Ledig & Conkle, 1983). Generation time for Torrey pine being approximately 10 years, this translates into a time since divergence between 430 and 43,000 generations.

[d]Low genome-wide differentiation observed between populations in this study (F_{ST} = 0.0129) 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} = −0.1914).

[e]As little is known regarding the historical distribution of Torrey pine, a broad prior range in ancestral effective population size was used to simulate demographic scenarios, including scenarios starting with small, moderate, or large ancestral population sizes.

[f]Minimum derived allele frequency of 0.01 was chosen to reflect empirical genetic data, filtered for variants with minor allele frequencies of or above 0.01 (see Materials and Methods).

2.5.2 | Additional filtering and downsampling 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 data set 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 ($p < .05$) 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 data set).

For computational efficiency, we downsampled the HWE-filtered data set from 73,928 to 9795 variants (hereafter referred to as the downsampled data set) 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. This way, each bin is subsampled to reflect its contribution to the HWE-filtered data set (Appendix S2). Following subsampling, we conducted a principal component analysis using the downsampled data set to ensure patterns of genetic diversity and population structure were maintained between data sets (Appendix S3).

2.5.3 | 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 et al., 2013; Excoffier & Foll, 2011) to simulate 9795 unlinked SNPs with a minimum derived allele frequency of 0.01, mimicking the composition of the downsampled genetic data set. 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 & Lischer, 2010) was used to calculate 10 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 downsampled data

set. Characterizing and summarizing the amount and distribution of genetic variation within each data set, 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).

2.5.4 | ABC parameter estimation

Demographic parameters were estimated in R using the ABC package (Csilléry et al., 2012). First, we conducted cross-validation simulations to evaluate the ability of summary statistics to distinguish between demographic scenarios (Appendix S4a). Following this, we identified the best demographic model based on the highest posterior probability. In addition, we performed a goodness-of-fit test, as implemented in the function gfit(), to ensure the model provided a good fit to the data. Lastly, demographic parameters were estimated as the weighted medians of posterior distributions, generated using a nonlinear postsampling regression adjustment on log-transformed data (Blum & François, 2010), using the weighted. median() function implemented in the R package SPATSTAT (Baddeley et al., 2015). 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 was tested using additional cross-validation simulations (Appendix S4b). Cross-validation simulations, model selection, model validation, and parameter estimations 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 (Li & Jakobsson, 2012).

Finally, as census sizes for Torrey pine populations are available (Frankham et al., 2011; Hall & Brinkman, 2015), we calculated the proportion of the census size (N) to effective population size (N_e) 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).

2.6 | 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 Weir and Cockerham's F_{ST} estimated from the full SNP data set with a null distribution simulated using SIMCOAL2 version 12.09.07 (Laval & Excoffier, 2004). Neutral simulations were conducted using input parameters sampled from posterior distributions borrowed from the best demographic model (see Results) in proportion to their posterior probabilities. The null F_{ST} distribution was estimated from a total of 10,000 coalescent simulations, each generating a set of 2000 unlinked SNPs across 270 individuals. For each simulated SNP, the minimum frequency for the minor allele was set to 0.01 to reflect filters applied to the full data set. Locus-specific Weir and Cockerham' F_{ST} values were estimated for both the full and simulated data sets in R using the HIERFSTAT package (Appendix S5).

2.7 | Outlier detection analyses

To estimate the potential contribution of local adaptation to genomic differentiation between Torrey pine populations, we used the full data set of 93,085 SNPs to identify loci putatively under selection using three distinct methods: BAYESCAN version 2.1 (Foll & Gaggiotti, 2008), OUTFLANK version 0.2 (Whitlock & Lotterhos, 2014) and PCADAPT version 4.3.3 (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). A description of these methods and their implementation is provided in Appendix S6. To minimize the presence of false positives within our data set, only outlier SNPs common to all three approaches were considered as candidate loci and included in subsequent analyses. Finally, to assess genetic structure at putatively adaptive loci, we conducted a principal component analysis based only on these shared outlier candidate loci using the R package ADEGENET.

2.8 | 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 (The UniProt Consortium, 2021) 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). GO terms were mapped onto annotated sequences in R using the UNIPROTR package (Soudy et al., 2020).

3 | RESULTS

3.1 | 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 3). Nonetheless, PC1 unambiguously separated island from mainland individuals, suggesting some level of genetic differentiation exists

between populations. These results were further supported by the low coefficient of genetic differentiation observed between populations (Weir and Cockerham's F_{ST}), estimated at 0.0129 (95% CI: 0.0128–0.0130). 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 4). 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 $k = 1$) (Appendix S7). Interestingly, inbreeding ($F_{IS} = -0.1914$, 95% CI: −0.1919, −0.1910) and coancestry coefficients (Table 2) were both low within Torrey pine populations. While the lack of inbreeding supports the absence of observable within-population genetic structure, both low inbreeding and coancestry estimates indicate neither reproduction among relatives nor unequal reproductive success among individuals has probably contributed to reduced genetic diversity, measured as expected heterozygosity (H_E), estimated within populations (Table 2). Combined, our results indicate that Torrey pine exhibits exceedingly low genetic diversity, with most of the variation distributed within genetically unstructured populations.

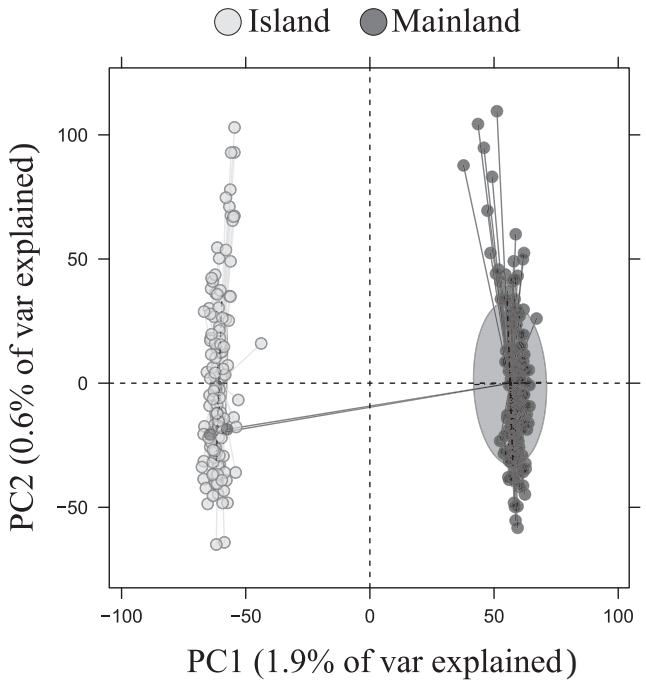


FIGURE 3 Principal component analysis using 93,085 SNPs for 270 Torrey pine individuals, including individuals from both mainland (black) and island (grey) populations. Variation explained by the first two principal components is provided in parentheses

3.2 | Demographic history of Torrey pine

Of the six demographic models evaluated (Figure 2), 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 (Appendix S4a). The goodness-of-fit test revealed that simulated summary statistics did not significantly differ from observed ones ($p = .76$), providing a good fit to the data.

Based on RPM2 model simulations, an ancestral Torrey pine population with an effective size of approximately 1124 individuals (95% CI: 939–1213) diverged during the early Pleistocene approximately 1.2 million YBP (95% CI: 1,195,367–1,296,186, assuming a generation time of 10 years) to form one island and one mainland population with effective sizes of approximately 2305 (95% CI: 2166–2338) and 1715 (95% CI: 1616–1759) individuals, respectively (Figure 5a–e). This resulted in a 0.75 ($N_I/N = 2305/3063$) proportion of the census to effective population size on the island, and a 0.45 ($N_M/N = 1715/3806$) 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: 8.15×10^{-3}–8.76×10^{-3}) per generation (Figure 5a,f). In general, cross-validation simulations indicated low prediction errors (Appendix S4b), suggesting high accuracy of inferred parameters. Nonetheless, note that for T_{Div}, the associated prediction error was higher than for other parameters.

3.3 | 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 data set of SNP variants (Appendix S5). For select SNPs, moderate to high empirical F_{ST} values (from approximately 0.2 to 0.62) could not or only very infrequently be generated through neutral simulations, suggesting they may be candidate for selection. In addition, PCADAPT, BAYESCAN, and OUTFLANK identified 3138 (3.37%), 2163 (2.32%), and 3942 (4.23%) outlier SNPs, respectively (Figure 6a). Of these outlier loci, 2053 (2.21%) were common to all three methods and contributed to genomic structure between Torrey pine populations (Figure 6b). 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.

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 GO terms. Overall, 110 (7.51%) contigs were annotated. After accounting

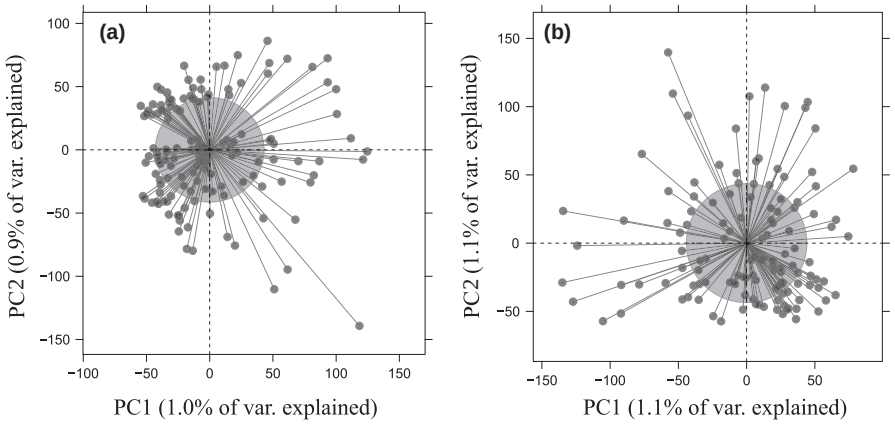


FIGURE 4 Population-specific principal component analysis based on 93,085 SNP variants. (a) Mainland population (TPSR, $n = 138$). (b) Island population (SRI, $n = 130$)

TABLE 2 Average genetic summary statistics across loci, including expected heterozygosity (H_E), and coancestry (relatedness) coefficient (θ) for island (SRI) and mainland (TPSR) Torrey pine populations

Population	H_E (95% CI)	θ (95% CI)
Island	0.1850 (0.1842, 0.1858)	0.0234 (0.0232, 0.0236)
Mainland	0.1837 (0.1829, 0.1845)	0.0224 (0.0222, 0.0225)

Note: Ninety-five percent confidence interval (95% CI) around mean estimates were obtained by bootstrapping.

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*'s species (Appendix S8) that may be targets of selection. Functionally, these genes appeared as primarily encoded in mitochondria and are involved in DNA integration or with processes associated with molecular functions such as RNA-directed DNA polymerase activity or nucleic acid binding (Table 3).

4 | 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 the potential consequences of genetic rescue (Frankham et al., 2011; Hufford & Mazer, 2003; Ralls et al., 2018). Here, we modelled demographic change and connectivity over time and explored the influence neutral and selective processes have had on contemporary population genomic structure in Torrey pine, a critically endangered endemic isolated to two populations in California. 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 modelling indicates Torrey pine has consistently suffered from low effective population size. Genome scans revealed over 2000 loci as candidates for selection. Consistent with previous observations of phenotypic differences between island and mainland populations, these data suggest adaptive genetic differences have evolved among populations (di Santo et al., 2021; Haller, 1986; Hamilton et al., 2017). From a conservation standpoint, these results lead to contrasting recommendations with respect to genetic rescue. Low genome-wide differentiation at neutral loci indicates the evolution of genetic incompatibilities following population-specific genetic drift is unlikely to impede a genetic rescue programme. However, previous observations of phenotypic differences paired with outlier loci potentially associated with divergent selection point towards the importance of adaptive evolution among Torrey pine populations. These results suggest increased genetic variation via interpopulation crosses may be needed in this species, but admixture should be evaluated first to quantify its fitness impact.

4.1 | Standing genetic variation

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 (N_e) were estimated at 0.1850 and 2305 individuals for the island population, and 0.1837 and 1715 individuals for the mainland population, respectively (Table 2; Figure 5a,c,d). Previous genetic analyses using allozymes and chloroplast DNA markers identified fixed genetic differences between populations, however, they failed to observe genetic variation within populations (Ledig & Conkle, 1983; Whittall et al., 2010). This study is thus one of the first to find genetic variability within island and mainland populations of Torrey pine, although this diversity remains low. Ledig and Conkle (1983) hypothesized reduced genetic diversity was attributable to drastically reduced mainland and island

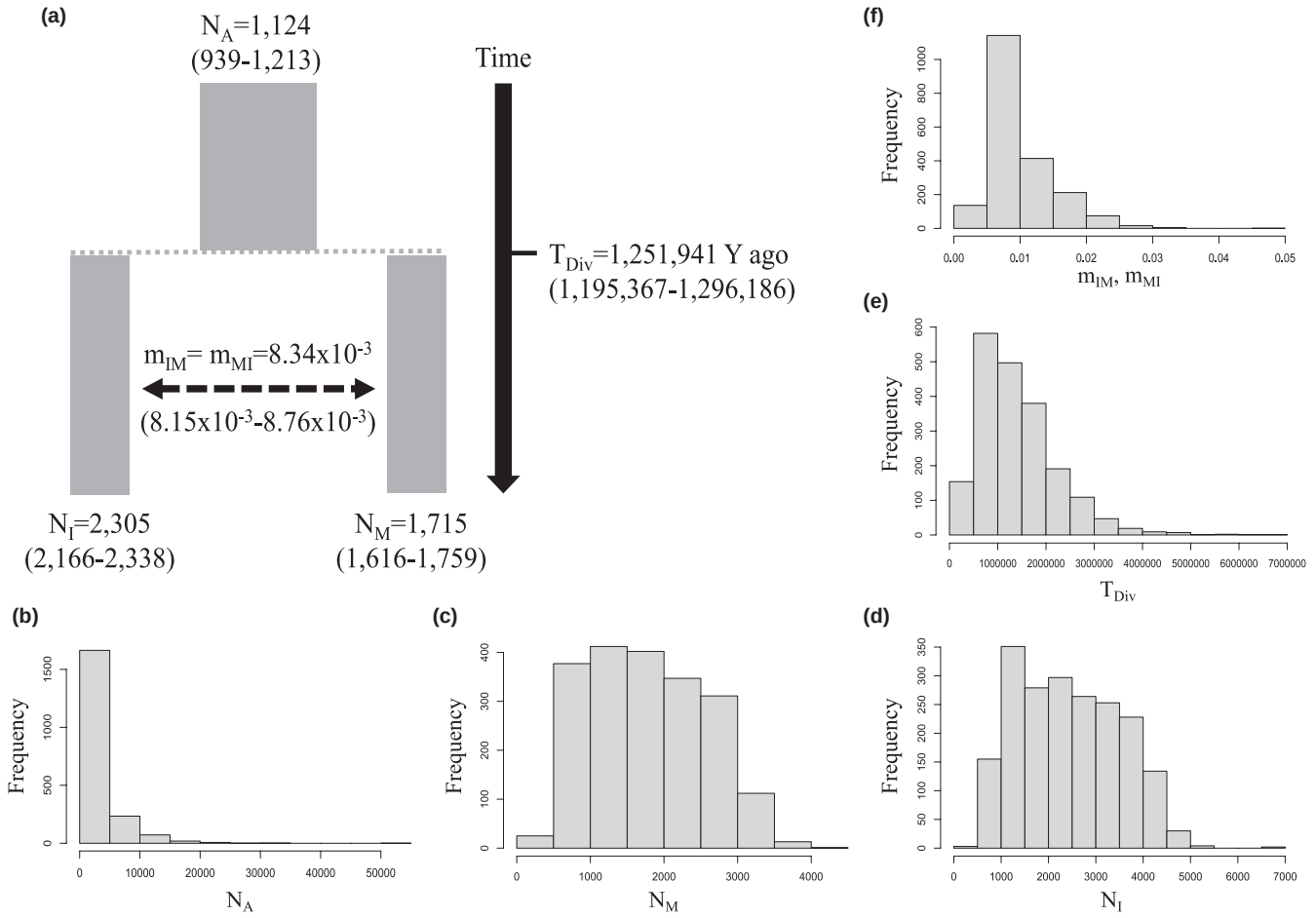
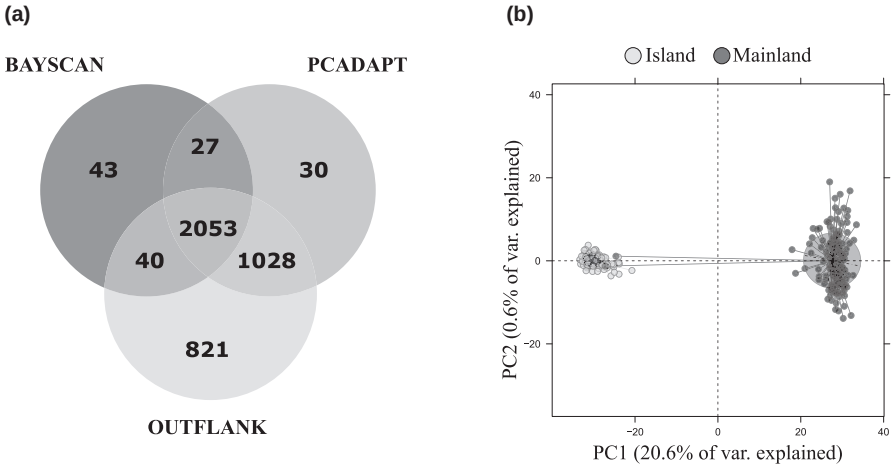


FIGURE 5 (a) Demographic parameters (weighted medians) and 95% confidence intervals (in parentheses) estimated for RPM2 model simulations. Each rectangle represents a population, either contemporary or ancestral. The solid arrow represents time, while the dashed arrow indicates gene flow between populations. (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

FIGURE 6 (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 2053 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



populations in the recent geological past. They suggested Torrey pine populations declined to fewer than 50 individuals to later recover and reach sizes of approximately 3000 to 4000 reproductively mature trees (Frankham et al., 2011; Hall & Brinkman, 2015). Demographic models herein, however, suggest that effective population size has always been low for Torrey pine (Figure 5a–d). Effective populations sizes of other pines have been estimated to be from a few dozens to a few thousand times greater than those inferred for island and mainland populations (Menon et al., 2018; Xia et al., 2018). The best fit demographic model predicted an ancestral effective population size (N_A) of 1124 and only a limited increase in population size following divergence. Given these observations,

TABLE 3 Functional categorization of 80 putatively adaptive genes between Torrey pine populations

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
	Defence response to bacterium	1.2
	Defence 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

Note: Listed are the 10 most frequent gene ontology (GO) terms within each of the three GO classes (biological process, molecular function, cellular component).

long-term reduced effective population size has probably exacerbated the consequences of genetic drift leading to a severe lack of genetic variation within Torrey pine populations.

Despite extremely low estimates of genetic variation and small effective population sizes, there was no evidence for inbreeding (F_{IS} = −0.1914) or excessive relatedness (θ) within populations (Table 2). These findings support high ratios between effective and census population size (N_e/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 et al., 2014). Indeed, it is not uncommon for plants to exhibit N_e/N ratios greater than 0.2 (Hoban et al., 2020; Waples et al., 2013). Combined with a lack of within population genetic structure (Figure 4, Appendix S7), these results also suggest reproduction among relatives or unequal reproductive success has not likely contributed to reduced standing genetic variation. Wind pollination and zoochorous seed dispersal have probably contributed to homogenizing the gene pool within populations (Johnson et al., 2003; Loveless & Hamrick, 1984). Pines also possess mechanisms that can reduce the probability of self-fertilization, including the embryo lethal system (Williams, 2009; Williams et al., 2001). This self-incompatibility system selectively induces death in embryos resulting from self-fertilization (Bramlett & Popham, 1971; Williams et al., 1999). Consequently, increased dispersal potential paired with post-zygotic barriers limiting the probability of mating among relatives have probably reduced within-population genetic structure in Torrey pine. Nonetheless, while dispersal and self-incompatibility may be important biological factors limiting genetic structure within populations, our bioinformatics approach may also have impacted the estimate of inbreeding. In pine genomes, which are large and highly repetitive (Stevens et al., 2016), paralogous reads can cluster into single loci, leading to excess heterozygosity that could misleadingly be interpreted as low inbreeding. Although de novo assembly and SNP filters were carefully selected to account for this issue (see Materials and Methods), we cannot rule out the potential impact of paralogous reads on our estimate of F_{IS}.

4.2 | Neutral genetic differences across populations over time

Demographic modelling using presumably neutral genomic variation indicated a simple model of isolation with migration best explained the contemporary distribution of genetic variation in Torrey pine. This supported the hypothesis formulated by Ledig and Conkle (1983) of island and mainland populations diverging from a single ancestral population. Geological deposits dating back to the Eocene and Miocene epochs suggest Santa Rosa Island uplifted from the ocean during the Tertiary age (55–5 Ma) (Muhs et al., 2014; Schumann et al., 2014), following which it would have been colonized by Torrey pine. Evidence suggests, however, that Santa Rosa Island may have been located closer to southern Californian coast during the mid-Miocene than it is today (Ledig & Conkle, 1983). This proximity may have promoted interpopulation gene exchange and the establishment of one, well-connected, ancestral metapopulation. In addition to Torrey pine, three other endemic plant species are found on Santa Rosa Island, including Munchkin dudleya (*Dudleya gnoma*), Brandegee's sage (*Salvia brandegeei*), and Soft leaved paintbrush

(*Castilleja mollis*) (Connaughton & Dussault-Martin, 2020). Although exclusively distributed on Santa Rosa Island today, these species may potentially share similarities with Torrey pine's evolutionary history, such as island colonization by ancestors and subsequent genetic differentiation.

Following population separation approximately 1.2 Ma, the best fit demographic model supports the maintenance of genetic connectivity between populations, estimating the probability of gene exchange at 8.34×10^{-3} per generation (Figures 5a,e,f). Despite geographic isolation among populations and reduced potential for inter-population gene flow, the contemporary estimate of $F_{ST} = 0.0129$ indicates only subtle genome-wide differentiation between island and mainland populations. Reduced genetic differentiation is typical of many conifers (Eckert et al., 2010; Namroud et al., 2008; Tyrmi et al., 2020), 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 (Johnson et al., 2003; Pesendorfer et al., 2016; Viana et al., 2016). Overall, our findings indicate that despite the increased probability of genetic drift due to 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 nonoverlapping 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. Additionally, migration was assumed in this study to be equal and of the same intensity between Torrey pine populations, which may have further biased our estimate of gene flow. Regionally asymmetric wind patterns are not uncommon along the southern Californian coast (Cole & Wahl, 2000), suggesting that unidirectional gene flow from Santa Rosa Island to the Torrey Pine State Reserve may dominate.

4.3 | Evidence for local adaptation

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 (di Santo et al., 2021; Haller, 1986; Hamilton et al., 2017). 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 hundred to a few thousand loci may be under divergent selection (Appendix S5). Genome scans further supported this observation, identifying 2053 (2.21%) candidate SNPs with accentuated divergence between island and mainland populations (Figure 6). Annotation of sequences containing these outlier SNPs suggested that adaptive evolution in Torrey pine might have caused genetic differentiation at both the nuclear and mitochondrial level (Table 3). Note, however, that the transfer of mitochondrial genes to the nucleus is common in plants (Adams et al., 2000; Adams & Palmer, 2003; Kan et al., 2021; Wu et al., 2017). To evaluate whether candidate SNPs were encoded in mitochondria and not the nuclear genome resulting from mitonuclear gene transfer, we assessed the mapping performance of candidate SNPs-bearing contigs to *Pinus taeda* genome and *Picea sitchensis* mitogenome (Appendix S9). While a significant portion of candidate contigs were annotated as mitochondrial loci, they most likely represent nuclear sequences as 98.36% (1140 out of 1464) of them mapped onto *P. taeda* genome compared to 0.2% (or 3 out of 1464) mapping onto *P. sitchensis* mitogenome. Although these sequences may be located in the nucleus following mito-nuclear gene transfers, recent work suggests these transfers are largely absent in *Pinaceae* (Kan et al., 2021). Consequently, while it remains possible that candidate SNPs represent genetic variation from recent mitonuclear gene transfer events, it is more likely the homology between candidate SNP-bearing contigs and the mitogenome detected in annotations reflects ancient paralogy between nuclear and mitochondrial protein sequences. As these loci represents only a tiny fraction of our genetic data set, we hesitate to make strong conclusions about them in either direction.

Overall, GO annotation indicated that genes important for mechanisms such as DNA integration, methylation, gene silencing, carbohydrate transport and metabolic processes, and defence against pathogens (bacteria and fungi) were candidates for selection (Table 3). 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 defence 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 outbreaks of the California five-spined ips (*Ips paraconfusus* Lanier) (Franklin & Santos, 2011; Shea & Neustein, 1995), whereas the island population has not been exposed to that selective pressure. Noteworthy with these results is the fact that pine genomes are exceedingly large (Grotkopp et al., 2004; Stevens et al., 2016), and our sequencing approach (reduced representation sequencing; see Materials and Methods) captures only a fraction of Torrey pine genome. Thus, despite differences observed, our estimate of variation important to adaptation is probably conservative and some variation has probably not yet been identified.

4.4 | Applying neutral and adaptive evolutionary processes to rare species conservation

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 (Frankham et al., 2014; Franklin & Frankham, 1998; Lynch & Lande, 1998). Torrey pine, critically endangered and endemic to just two native populations, suffers from extremely low effective population size ($N_I = 2305$, $N_M = 1715$) 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 Torrey pine may not retain the genetic variation within populations needed to adapt to change. Current monitoring within the mainland population suggests a lack of recruitment (personal observation by Lionel Di Santo, July 2017), infestation by *Ips* beetles (Steele et al., 2021), and climate warming (Diffenbaugh et al., 2015) may increase extinction risk. Thus, for Torrey pine, increased Ne and greater genetic diversity may be required for long-term persistence.

A genetic rescue programme would facilitate interpopulation gene flow to increase genetic variation and reduce potential load associated with isolation and rarity. Demographic modelling indicated some gene flow has been maintained between populations following isolation, contributing to reduced genome-wide estimates of differentiation ($F_{ST} = 0.0129$). However, the combination of observed phenotypic differences and number of genes that appear targets of selection suggest island and mainland populations of Torrey pine have undergone distinct adaptive evolutionary trajectories following isolation. Thus, a genetic rescue programme should be considered with caution as gene flow between populations may disrupt local adaptation and further reduce population performance (Goto et al., 2011; Hufford & Mazer, 2003; Montalvo & Ellstrand, 2001). 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 (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.

If ex situ preservation of existing genetic diversity distributed within natural populations of Torrey pine is desired, our results provide distinct recommendations based on neutral and adaptive processes. Low genome-wide differentiation among populations, a consistent history of low effective population size, and indications that some gene flow is maintained among populations suggest one population (island or mainland) could be targeted to effectively preserve neutral genetic variation. However, the combination of shared outlier candidate loci and previously observed phenotypic differences among populations suggests if the goal is to preserve adaptive genetic variation, a strategy favouring conservation efforts including both mainland and island populations will be necessary. Together, our results point towards the need to understand the distribution of neutral and adaptive genetic diversity to adequately meet conservation objectives for rare species (Bonin et al., 2007; Holderegger et al., 2006; Teixeira & Huber, 2021).

This study demonstrates that genomic data can provide substantial resolution in estimates of genetic diversity for rare species and is needed to understand the evolutionary processes underlying population persistence. In addition, relatedness among individuals approximated with such data can contribute knowledge necessary to designing, complementing, or completing pedigrees of wild or captive populations of endangered species (Galla et al., 2022). We also show that using genome-wide data to both model demographic changes over space and time and to tease apart evolutionary processes contributing to population structure can help prioritize ex situ collections and identify candidate populations for assisted gene flow. Rare species often occur in small, fragmented populations across different habitats, and management of such species should consider patterns of neutral and adaptive genetic diversity, historical divergence, and inbreeding. While assessment of parent and advanced-generation hybrid populations within a common environment is the "gold standard" to empirically test fitness consequences following population admixture (Hamilton et al., 2017; Irimia et al., 2021; Keller et al., 2000; Shi et al., 2018; VanWallendael et al., 2022), the analytical framework described here can provide critical insights into both the risks and benefits of establishing a genetic rescue management strategy.

AUTHOR CONTRIBUTIONS

Lionel N. D. Santo contributed to data collection, designing research, analysing the data, and led writing the manuscript. Sean Hoban contributed to data analysis and writing the manuscript. Thomas L. Parchman contributed to genomic data analysis and writing the manuscript. Jessica W. Wright contributed to data collection, study design, and writing the manuscript. Jill A. Hamilton contributed to data collection, designing the research, and writing the manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGEMENTS

This work was funded by USDA Forest Service Health Protection Gene Conservation Programme 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 Programme to Stimulate Competitive Research (ND-EPSCoR NSF-IIA-1355466) and the NDSU Environmental and Conservation Sciences Graduate Programme 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.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at 10.5061/dryad.83bk3j9v0.

DATA AVAILABILITY STATEMENT

Demultiplexed DNA sequences have been made available from the National Centre for Biotechnology Information database (BioProject: PRJNA840943). Final de novo sequence assembly, sample information (including sample name, population assignment, ecotype, and GPS coordinates), barcodes used during genomic libraries preparation, as well as the raw, full, HWE-filtered, and downsampled data sets are available from Dryad (doi: 10.5061/dryad.83bk3j9v0). Upon request to the corresponding author, R scripts may also be made available (di Santo et al., 2022a; di Santo et al., 2022b).

ORCID

Lionel N. Di Santo https://orcid.org/0000-0002-8288-4860
Sean Hoban https://orcid.org/0000-0002-0348-8449
Thomas L. Parchman https://orcid.org/0000-0003-1771-1514
Jill A. Hamilton https://orcid.org/0000-0002-3041-0508

REFERENCES

Abebe, T. D., Naz, A. A., & Léon, J. (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

Adams, K. L., Daley, D. O., Qiu, Y.-L., Whelan, J., & Palmer, J. D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature*, *408*(6810), 354–357. https://doi.org/10.1038/35042567

Adams, K. L., & Palmer, J. D. (2003). Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, *29*(3), 380–395. https://doi.org/10.1016/S1055-7903(03)00194-5

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., & Lobo, J. (2008). Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. *Molecular Ecology*, *17*(24), 5177–5188. https://doi.org/10.1111/j.1365-294X.2008.03971.x

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. CRC Press.

Blum, M. G. B., & François, O. (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

Bonin, A., Nicole, F., Pompanon, F., Miaud, C., & Taberlet, P. (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, B. (2007). Genetic rescue in an isolated metapopulation of a naturally fragmented plant species, Parnassia palustris. *Conservation Biology*, *21*(3), 832–841. https://doi.org/10.1111/j.1523-1739.2006.00645.x

Bramlett, D. L., & Popham, T. W. (1971). Model relating unsound seed and embryonic lethal alleles in self-pollinated pines. *Silvae Genet*, *20*, 192–193.

Campbell, I. D., McDonald, K., Flannigan, M. D., & Kringayark, J. (1999). Long-distance transport of pollen into the Arctic. *Nature*, *399*(6731), 29–30. https://doi.org/10.1038/19891

Canty, A., & Ripley, B. D. (2021). Boot: Bootstrap R (S-plus) functions. Retrieved from https://cran.r-project.org/package=boot

Carlson, S. M., Cunningham, C. J., & Westley, P. A. H. (2014). Evolutionary rescue in a changing world. *Trends in Ecology and Evolution*, *29*(9), 521–530. https://doi.org/10.1016/j.tree.2014.06.005

Caujapé-Castells, J., & Pedrola-Monfort, J. (2004). Designing ex-situ conservation strategies through the assessment of neutral genetic markers: Application to the endangered Androcymbium gramineum. *Conservation Genetics*, *5*(2), 131–144. https://doi.org/10.1023/B:COGE.0000029997.59502.88

Cavender, N., Westwood, M., Bechtoldt, C., Donnelly, G., Oldfield, S., Gardner, M., Rae, D., & McNamara, W. (2015). Strengthening the conservation value of ex situ tree collections. *Oryx*, *49*(3), 416–424. https://doi.org/10.1017/S0030605314000866

Cole, K. L., & Wahl, E. (2000). A late Holocene paleoecological record from Torrey pines state reserve, California. *Quaternary Research*, *53*(3), 341–351. https://doi.org/10.1006/qres.1999.2121

Connaughton, A., & Dussault-Martin, N. (2020). Santa Rosa Island species monitoring protocol and field guide visualization. Retrieved from https://scholarworks.calstate.edu/downloads/f1881p50z

Critchfield, W. B., & Little, E. L. (1966). *Geographic distribution of the pines of the world*. US Department of Agriculture, Forest Service.

Csilléry, K., François, O., & Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, *3*(3), 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Davison, A. C., & Hinkley, D. v. (1997). *Bootstrap methods and their application*. Cambridge University Press.

di Santo, L. N., & Hamilton, J. A. (2020). Using environmental and geographic data to optimize ex situ collections and preserve evolutionary potential. *Conservation Biology*, *35*(2), 733–744. https://doi.org/10.1111/cobi.13568

di Santo, L. N., Hoban, S., Parchman, T. L., Wright, J. W., & Hamilton, J. A. (2022a). Dataset: Reduced representation sequencing to understand the evolutionary history of Torrey pine (Pinus torreyana parry) with implications for rare species conservation. *Dryad*. https://doi.org/10.5061/dryad.83bk3j9v0

di Santo, L. N., Hoban, S., Parchman, T. L., Wright, J. W., & Hamilton, J. A. (2022b). *ddRAD HiSeq 2500 single-end Illumina sequences for Pinus torreyana*. NCBI; BioProject PRJNA840943.

di Santo, L. N., Polgar, M., Nies, S., Hodgkiss, P., Canning, C. A., Wright, J. W., & Hamilton, J. A. (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

Diffenbaugh, N. S., Swain, D. L., & Touma, D. (2015). Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13), 3931–3936. https://doi.org/10.1073/pnas.1422385112

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11–15.

Dusek, K. H. (1985). Update on our rarest pine. *American Forests*, *91*, 26–29, 61, 63.

Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L., Nelson, D. C., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics*, *185*(3), 969–982. https://doi.org/10.1534/genetics.110.115543

Edmans, S. (2007). Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, *16*(3), 463–475. https://doi.org/10.1111/j.1365-294X.2006.03148.x

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Foll, M. (2011). Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. https://doi.org/10.1093/bioinformatics/btr124

Excoffier, L., & Lischer, H. E. L. (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), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (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

Franklin, I. R., & Frankham, R. (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, J., & Santos, E. v. (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

Galla, S. J., Brown, L., Couch-Lewis Ngāi Tahu Te Hapū O Ngāti Wheke Ngāti Waewae Y, Cubrinovska, I., Eason, D., Gooley, R. M., Hamilton, J. A., Heath, J. A., Hauser, S. S., Latch, E. K., Matocq, M. D., Richardson, A., Wold, J. R., Hogg, C. J., Santure, A. W., & Steeves, T. E. (2022). The relevance of pedigrees in the conservation genomics era. *Molecular Ecology*, *31*(1), 41–54. https://doi.org/10.1111/mec.16192

Gapare, W. J., Yanchuk, A. D., & Aitken, S. N. (2008). Optimal sampling strategies for capture of genetic diversity differ between core and peripheral populations of Picea sitchensis (bong.) Carr. *Conservation Genetics*, *9*(2), 411–418. https://doi.org/10.1007/s10592-007-9353-8

Goto, S., Iijima, H., Ogawa, H., & Ohya, K. (2011). Outbreeding depression caused by intraspecific hybridization between local and nonlocal genotypes in Abies sachalinensis. *Restoration Ecology*, *19*(2), 243–250. https://doi.org/10.1111/j.1526-100X.2009.00568.x

Goudet, J., & Jombart, T. (2020). Hierfstat: Estimation and tests of hierarchical F-statistics. Retrieved from https://cran.r-project.org/package=hierfstat

Grotkopp, E., Rejmánek, M., Sanderson, M. J., & Rost, T. L. (2004). Evolution of genome size in pines (Pinus) and its life-history correlates: Supertree analyses. *Evolution*, *58*(8), 1705–1729. https://doi.org/10.1111/j.0014-3820.2004.tb00456.x

Guerrant, E. O., Jr., Havens, K., & Vitt, P. (2014). Sampling for effective ex situ plant conservation. *International Journal of Plant Sciences*, *175*(1), 11–20. https://doi.org/10.1086/674131

Hague, M. T. J., & Routman, E. J. (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

Hall, T., & Brinkman, A. (2015). *Population dynamics of the Island Torrey pine (Pinus torreyana ssp. insularis) on Santa Rosa Island, CA*. California State University Channel Islands.

Haller, J. 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, J. A., Royauté, R., Wright, J. W., Hodgskiss, P., & Ledig, T. F. (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), 7370–7381. https://doi.org/10.1002/ece3.3306

Hedrick, P. W., Peterson, R. O., Vucetich, L. M., Adams, J. R., & Vucetich, J. A. (2014). Genetic rescue in isle Royale wolves: Genetic analysis and the collapse of the population. *Conservation Genetics*, *15*(5), 1111–1121. https://doi.org/10.1007/s10592-014-0604-1

Hoban, S. (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, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjogren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S., Bertola, L. D., Bloomer, P., Breed, M., Rodriguez-Correa, H., Chris Funk, W., Grueber, C. E., Hunter, M. E., & Laikre, L. (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, S., & Schlarbaum, S. (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, R., Kamm, U., & Gugerli, F. (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

Hufford, K. M., & Mazer, S. J. (2003). Plant ecotypes: Genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, *18*(3), 147–155. https://doi.org/10.1016/S0169-5347(03)00002-8

Irimia, R. E., Hierro, J. L., Branco, S., Sotes, G., Cavieres, L. A., Eren, O., Lortie, C. J., French, K., Callaway, R. M., & Montesinos, D. (2021). Experimental admixture among geographically disjunct populations of an invasive plant yields a global mosaic of reproductive incompatibility and heterosis. *Journal of Ecology*, *109*, 2152–2162. https://doi.org/10.1111/1365-2745.13628

IUCN. (2021). The IUCN red list of threatened species. Version 2021-1. Retrieved April 19, 2021, from https://www.iucnredlist.org

Johnson, M., vander Wall, S. B., & Borchert, M. (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, W. E., Onorato, D. P., Roelke, M. E., Land, D. E., Cunningham, M., Belden, R. C., Belden, R. C., McBride, R., Jansen, D., Lotz, M.,

Howard, J., Wildt, D. E., Penfold, L. M., Hostetler, J. A., Oli, M. K., O'Brien, S. J., & Shindle, D. (2010). Genetic restoration of the Florida panther. *Science*, *329*(5999), 1641–1645. https://doi.org/10.1126/science.1192891

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kan, S.-L., Shen, T.-T., Ran, J.-H., & Wang, X.-Q. (2021). Both conifer II and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. *BMC Biology*, *19*(1), 146. https://doi.org/10.1186/s12915-021-01096-z

Keller, M., Kollmann, J., & Edwards, P. J. (2000). Genetic introgression from distant provenances reduces fitness in local weed populations. *Journal of Applied Ecology*, *37*(4), 647–659. https://doi.org/10.1046/j.1365-2664.2000.00517.x

Laval, G., & Excoffier, L. (2004). SIMCOAL 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, *20*(15), 2485–2487. https://doi.org/10.1093/bioinformatics/bth264

Ledig, T. F., & Conkle, T. M. (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

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, *30*(20), 2843–2851. https://doi.org/10.1093/bioinformatics/btu356

Li, S., & Jakobsson, M. (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

Liu, D., Zhang, L., Wang, J., & Ma, Y. (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

Loveless, M. D., & Hamrick, J. L. (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

Lowry, D. B., Rockwood, C. R., & Willis, J. H. (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, M., & Lande, R. (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, T., Shine, R., Olsson, M., & Wittzell, H. (1999). Restoration of an inbred adder population. *Nature*, *402*(6757), 34–35. https://doi.org/10.1038/46941

McKay, J. K., & Latta, R. G. (2002). Adaptive population divergence: Markers, QTL and traits. *Trends in Ecology & Evolution*, *17*(6), 285–291. https://doi.org/10.1016/S0169-5347(02)02478-3

Menon, M., Bagley, J. C., Friedline, C. J., Whipple, A. v, Schoettle, A. W., Leal-Sàenz, A., Wehenkel, C., Molina-Freaner, F., Flores-Rentería, L., Sniezko, R. A., Cushman, S. A., Waring, K. M., Eckert, A. J., Gonzalez-Elizondo, S. M. (2018). The role of hybridization during ecological divergence of southwestern white pine (Pinus strobiformis) and limber pine (P. flexilis). *Molecular Ecology*, *27*(5), 1245–1260. doi: https://doi.org/10.1111/mec.14505

Montalvo, A. M., & Ellstrand, N. C. (2001). Nonlocal transplantation and outbreeding depression in the subshrub Lotus scoparius (Fabaceae). *American Journal of Botany*, *88*(2), 258–269. https://doi.org/10.2307/2657017

Muhs, D. R., Simmons, K. R., Schumann, R. R., Groves, L. T., DeVogel, S. B., Minor, S. A., & Laurel, D. (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*, 209–238. https://doi.org/10.1016/j.quascirev.2014.09.017

Namroud, M.-C., Beaulieu, J., Juge, N., Laroche, J., & Bousquet, J. (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

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, *59*(4), 705–719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., … Alexeyenko, A. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, *497*(7451), 579–584. https://doi.org/10.1038/nature12211

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, A. C. (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, R. D. (2019). Simpleboot: Simple bootstrap routines. Retrieved from https://cran.r-project.org/package=simpleboot

Pesendorfer, M. B., Sillett, S. T., Koenig, W. D., & Morrison, S. A. (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), 215–237. https://doi.org/10.1650/CONDOR-15-125.1

Pew, J., Wang, J., Muir, P., & Frasier, T. (2015). Related: An R package for analyzing pairwise relatedness data based on codominant molecular markers. Retrieved from https://github.com/timothyfrasier/related

Potter, K. M., Jetton, R. M., Bower, A., Jacobs, D. F., Man, G., Hipkins, V. D., & Westwood, M. (2017). Banking on the future: Progress, challenges and opportunities for the genetic conservation of forest trees. *New Forests*, *48*, 153–180. https://doi.org/10.1007/s11056-017-9582-8

Pritchard, D. J., Fa, J. E., Oldfield, S., & Harrop, S. R. (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, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. B. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*, *37*(7), 2153–2154. https://doi.org/10.1093/molbev/msaa053

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (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, J. B., Matz, M. v, Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology*, *23*(24), 5937–5942. doi: https://doi.org/10.1111/mec.12965

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Austria. https://www.r-project.org/

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Austria. https://www.r-project.org/

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Austria. https://www.r-project.org/

Ralls, K., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., Sunnucks, P., & Frankham, R. (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

Schumann, R. R., Minor, S. A., Muhs, D. R., & Pigati, J. S. (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

Shea, P. J., & Neustein, M. (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* (pp. 39–43). USDA, Forest Service, General Technical Report INT: GTR-318.

Shi, J., Joshi, J., Tielbörger, K., Verhoeven, K. J. F., & Macel, M. (2018). Costs and benefits of admixture between foreign genotypes and local populations in the field. *Ecology and Evolution*, *8*(7), 3675–3684. https://doi.org/10.1002/ECE3.3946

Soudy, M., Anwar, A. M., Ahmed, E. A., Osama, A., Ezzeldin, S., Mahgoub, S., & Magdeldin, S. (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

Steele, S. E., Ryder, O. A., & Maschinski, J. (2021). RNA-seq reveals adaptive genetic potential of the rare Torrey pine (Pinus torreyana) in the face of Ips bark beetle outbreaks. *Conservation Genetics*, *22*(6), 1035–1050. https://doi.org/10.1007/s10592-021-01394-7

Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., Paul, R., Gonzalez-Ibeas, D., Koriabine, M., Martínez-García, P. J., Sezen, U. U., Marçais, G., Jermstad, K., Mcguire, P. E., Loopstra, C. A., Davis, J. M., Eckert, A., de Jong, P., Yorke, J. A., … Holtz-Morris, A. (2016). Sequence of the sugar pine megagenome. *Genetics*, *204*(4), 1613–1626. https://doi.org/10.1093/genetics/204.4.NP

Swarts, N. D., Sinclair, E. A., Krauss, S. L., & Dixon, K. W. (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

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), e2015096118. https://doi.org/10.1073/pnas.2015096118

The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*, D480–D489.

Tsumura, Y., Uchiyama, K., Moriguchi, Y., Ueno, S., & Ihara-Ujino, T. (2012). Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, Cryptomeria japonica. *Heredity*, *109*(6), 349–360. https://doi.org/10.1038/hdy.2012.50

Tyrmi, J. S., Vuosku, J., Acosta, J. J., Li, Z., Sterck, L., Cervera, M. T., Savolainen, O., & Pyhäjärvi, T. (2020). Genomics of clinal local adaptation in Pinus sylvestris under continuous environmental and spatial genetic setting. *G3: Genes, genomes, Genetics*, *10*(8), 2683–2696. https://doi.org/10.1534/g3.120.401285

VanWallendael, A., Lowry, D. B., & Hamilton, J. A. (2022). One hundred years into the study of ecotypes, new advances are being made through large-scale field experiments in perennial plant systems. *Current Opinion in Plant Biology*, *66*, 102152. https://doi.org/10.1016/J.PBI.2021.102152

Varis, S., Pakkanen, A., Galofré, A., & Pulkkinen, P. (2009). The extent of south-north pollen transfer in Finnish scots pine. *Silva Fennica*, *43*(5), 717–726. https://doi.org/10.14214/sf.168

Viana, D. S., Gangoso, L., Bouten, W., & Figuerola, J. (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

Wang, X., Bernhardsson, C., & Ingvarsson, P. K. (2020). Demography and natural selection have shaped genetic variation in the widely distributed conifer Norway spruce (Picea abies). *Genome Biology and Evolution*, *12*(2), 3803–3817. https://doi.org/10.1093/gbe/evaa005

Waples, R. S., Luikart, G., Faulkner, J. R., & Tallmon, D. A. (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

Waters, E. R., & Schaal, B. A. (1991). No variation is detected in the chloroplast genome of Pinus torreyana. *Canadian Journal of Forest Research*, *21*(12), 1832–1835. https://doi.org/10.1139/x91-253

Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Using ABCtoolbox. Retrieved from http://cmpg.unibe.ch/software/ABCtoolbox/

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 1–7. https://doi.org/10.1186/1471-2105-11-116

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Westemeier, R. L., Brawn, J. D., Simpson, S. A., Esker, T. L., Jansen, R. W., Walk, J. W., Kershner, E. L., Bouzat, J. L., & Paige, K. N. (1998). Tracking the long-term decline and recovery of an isolated population. *Science*, *282*(5394), 1695–1698. https://doi.org/10.1126/science.282.5394.1695

Whitlock, M. C., & Lotterhos, K. (2014). OutFLANK: Fst outliers with trimming. Retrieved from https://github.com/whitlock/OutFLANK

Whittall, J. B., Syring, J., Parks, M., Buenrostro, J., Dick, C., Liston, A., & Cronn, R. (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, Y., van Kleunen, M., Dietrich, S., & Fischer, M. (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), 2357–2364. https://doi.org/10.1098/rspb.2007.0768

Williams, C. G. (2009). *Conifer reproductive biology*. Springer.

Williams, C. G. (2010). Long-distance pine pollen still germinates after meso-scale dispersal. *American Journal of Botany*, *97*(5), 846–855. https://doi.org/10.3732/ajb.0900255

Williams, C. G., Barnes, R. D., & Nyoka, I. (1999). Embryonic genetic load for a neotropical conifer, Pinus patula Schiede et Deppe. *Journal of Heredity*, *90*(3), 394–398. https://doi.org/10.1093/jhered/90.3.394

Williams, C. G., Zhou, Y., & Hall, S. E. (2001). A chromosomal region promoting outcrossing in a conifer. *Genetics*, *159*(3), 1283–1289. https://doi.org/10.1093/genetics/159.3.1283

Wu, Z., Sloan, D. B., Brown, C. W., Onica Rosenblueth, M., Palmer, J. D., & Chuan Ong, H. (2017). Mitochondrial Retroprocessing promoted functional transfers of rpl5 to the nucleus in grasses. *Molecular Biology and Evolution*, *34*(9), 2340–2354. https://doi.org/10.1093/molbev/msx170

Xia, H., Wang, B., Zhao, W., Pan, J., Mao, J. F., & Wang, X. R. (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), 1931–1945. https://doi.org/10.1111/eva.12697

Young, A. G., Boyle, T., & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants.

Trends in Ecology & Evolution, *11*(10), 413–418. https://doi.org/10.1016/0169-5347(96)10045-8

Young, A. G., Brown, A. H. D., & Zich, F. A. (1999). Genetic structure of fragmented populations of the endangered daisy Rutidosis leptorrhynchoides. *Conservation Biology*, *13*(2), 256–265. https://doi.org/10.1046/j.1523-1739.1999.013002256.x

SUPPORTING INFORMATION

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

How to cite this article: Di Santo, L. N., Hoban, S., Parchman, T. L., Wright, J. W., & Hamilton, J. A. (2022). Reduced representation sequencing to understand the evolutionary history of Torrey pine (*Pinus torreyana* parry) with implications for rare species conservation. *Molecular Ecology*, *31*, 4622–4639. https://doi.org/10.1111/mec.16615