1   **Seed morphological traits as a tool to quantify variation maintained in *ex situ* collections: a**

2   **case study in *Pinus torreyana* (Parry)**

3

4   Lionel N Di Santo[a,d], Monica Polgar[a], Storm Nies[a], Paul Hodgkiss[b], Courtney A Canning[c],

5   Jessica W Wright[b], and Jill A Hamilton[a]

6

7   [a] North Dakota State University, Department of Biological Sciences, Fargo, ND, USA

8   [b] USDA Forest Service, Pacific Southwest Research Station, Davis, CA, USA

9   [c] USDA Forest Service, Pacific Southwest Research Station, Placerville CA, USA

10   [d] Correspondence: lionel.disanto@ndsu.edu

11

12 **Abstract**

13 Background: Understanding the within- and among-population distribution of trait variation

14 within seed collections may provide a means to approximate standing genetic variation and

15 inform plant conservation.

16 Aims: This study aimed to estimate population- and family-level seed trait variability for existing

17 seed collections of Torrey pine (*Pinus torreyana*), and to use these data to guide sampling of

18 future collections.

19 Methods: We quantified variation in 14 seed morphological traits and seedling emergence within

20 and among Torrey pine populations. Using a simulation-based approach, we used estimates of

21 within-population variance to assess the number of maternal families required to capture 95% of

22 trait variation within each existing seed collection.

23 Results: Substantial structure was observed both within and among Torrey pine populations, with

24 island and mainland seeds varying in seed size and seed coat thickness. Despite morphological

25 differences, seedling emergence was similar across populations. Simulations revealed that 83%

26 and 71% of all maternal families within island and mainland seed collections respectively needed

27 to be resampled to capture 95% of seed trait variation within existing collections.

28 Conclusions: From a conservation perspective, our results indicate that to optimize genetic

29 diversity captured in Torrey pine seed collections, maximizing the number of maternal families

30 sampled within each population will be necessary.

31

32 **Keywords**

Introduction

35

36 Ex situ seed collections preserve species genetic diversity outside of their native range, providing

37 the raw material for species reintroductions and germplasm to augment restoration (Guerrant Jr

38 *et al.* 2014; Potter *et al.* 2017). Ensuring ex situ collections represent genetic variation found in

39 natural populations is critical to both contemporary conservation and potential future restoration

40 efforts (Schaal and Leverich 2004; Basey *et al.* 2015). An invaluable conservation resource,

41 particularly for rare species, ex situ collections protect against biodiversity loss in the wild, while

42 preserving species' evolutionary potential. However, the cost and logistical constraints

43 associated with seed collection pose a significant challenge. Given this challenge, means are

44 needed to optimize ex situ sampling efforts (Hoban and Schlarbaum 2014; Di Santo and

45 Hamilton 2020).

46

47 One approach may be to use the distribution of trait variation existing within contemporary *ex*

48 *situ* seed collections as a proxy for quantifying standing genetic variation within and among

49 populations. Although multiple factors influence plant phenotypes (Monty *et al.* 2013; Villellas

50 *et al.* 2014), seed morphological variation is often considered highly heritable. For example, seed

51 length, seed width, and seed mass have a heritability (or repeatability) estimated between 0.33

52 and 0.98 in conifers, including maritime pine (*Pinus pinaster*), chir pine (*Pinus roxburghii*), and

53 white spruce (*Picea glauca*) (Roy *et al.* 2004; Carles *et al.* 2009; Zas and Sampedro 2015). In

54 addition, traits such as seed shape, seed coat thickness, or embryo length also exhibit high

55 heritability, with values estimated between 0.59 and 0.96 for agronomic species, including

56 soybean (*Glycine max*), narrow-leafed lupin (*Lupinus angustifolius*) and rice (*Oryza sativa*)

57 (Pandey *et al.* 1994; Cober *et al.* 1997; Mera *et al.* 2004; Hakim and Suyamto 2017). Given these

58   observations, variation in seed morphological traits likely has a genetic basis and may reflect

59   standing genetic variation within and among populations. In addition, morphological variation of

60   seeds stored ex situ may reflect variation attributable to the maternal environment (Platenkamp

61   and Shaw 1993; Singh *et al.* 2017). However, for rare species where existing genetic data are

62   limited, quantifying within and between population variation for traits largely considered

63   heritable within existing seed collections may be invaluable to optimizing future collections,

64   even if estimates of genetic variation do not control for maternal environment.

65

66   The distribution of heritable genetic variation estimated via common garden experiments –

67   experimental approaches used to understand the genetic contribution to phenotypic variation

68   under common environmental conditions – (Weber and Kolb 2014; Hamilton *et al.* 2017; Yoko

69   *et al.* 2020) or molecular genetic data (Zhang and Zhou 2013; Hausman *et al.* 2014; Tamaki *et*

70   *al.* 2018) can be used to quantify the distribution of standing genetic variation. However, when

71   common garden experiments or molecular genetic data are unavailable, quantifying trait

72   variation within and among ex situ seed population collections may provide a reasonable proxy

73   for the distribution of genetic variation. Millions of seed accessions have been stored in gene

74   banks internationally (FAO 2010), representing a large conservation and research resource.

75   Although common garden experiments are preferred when available, heritability of seed

76   morphological traits and ease of access to seeds through existing ex situ collections suggests that

77   quantifying seed morphological variation may provide a timely approach to estimating variation

78   preserved in collections. In addition, where the goal is to limit the loss of biodiversity and

79   preserve evolutionary potential for rare species, existing seed morphological data may be

80   leveraged to optimize supplemental conservation collections.

81

82   *Pinus torreyana* Parry (Torrey pine), is one of the rarest pines in the world (Critchfield and Little

83   1966; Dusek 1985), endemic to two discrete natural populations in California. Torrey pine

84   occupies one mainland population (*Pinus torreyana* subsp. *torreyana*) of approximately 6,000

85   trees at the Torrey Pine State Reserve in La Jolla, CA, and an island population (*Pinus torreyana*

86   subsp. *insularis*) of approximately 3,000 reproductively mature trees on Santa Rosa Island, CA,

87   one of the Channel Islands (Ledig and Conkle 1983; Haller 1986; Hamilton *et al.* 2017) (Fig. 1).

88   Listed as critically endangered by the IUCN (2020), Torrey pine is of critical conservation

89   concern due to multiple factors, including low population size (Franklin and Santos 2011; Hall

90   and Brinkman 2015), low genetic diversity (Ledig and Conkle 1983; Whittall *et al.* 2010),

91   climate change, and environmental and human-mediated disturbances (Franklin and Santos

92   2011; Hamilton *et al.* 2017). While in situ conservation has preserved the whole of the species'

93   range, with fewer than 10,000 reproductively mature individuals in native populations, there are

94   substantial risks for population-level extirpation events. To mitigate potential losses in the wild,

95   conservation efforts have focused on preservation of seed ex situ. While ex situ seed collections

96   provide an invaluable conservation resource, they may also be used to quantify species' trait

97   variation needed to inform future conservation efforts.

98

99   In this study, we evaluate morphological trait variation in a large ex situ conservation collection

100 of Torrey pine seed sourced from the two native extant populations. Specifically, we quantify the

101 distribution of variation for 14 seed morphology traits and assess differences in emergence

102 between island and mainland seedlings. In addition, we use existing ex situ collection data to

103 provide supplemental population sampling guidance for future Torrey pine collections. For this

104 latter objective, we use simulations to estimate the number of maternal families required to

105 capture 95% of seed morphological variation existing in contemporary ex situ collections, for

106 both island and mainland population independently. This study evaluates the distribution of seed

107 morphological variation in ex situ collections as a proxy for standing genetic diversity,

108 quantifying variation attributable to within and between population differences. These data are

109 then used to inform population sampling necessary to meet conservation objectives in future seed

110 collections. Although presented here using Torrey pine, our approach is broadly applicable for ex

111 situ collections within species with largely heritable seed trait variation.

112

113 **Materials and methods**

114 ***Cone collection and seed processing***

115 Mature, open-pollinated Torrey pine (*Pinus torreyana* Parry) cones were collected from native

116 extant populations as part of a large ex situ conservation collection between June and July of

117 2017. Cones were collected from 157 trees on Santa Rosa Island (Channel Islands National

118 Park), CA (island population) and 201 trees at the Torrey Pine State Reserve in La Jolla, CA

119 (mainland population), representing the species' entire natural distribution (Fig. 1; See

120 Supporting Information Figure S1). Where possible, we collected between five to ten cones per

121 maternal tree at each location. Sampling of reproductive maternal trees was evenly spaced;

122 however, adjacent maternal trees were occasionally sampled to ensure enough cones were

123 collected. On average, mainland and island trees sampled were separated by approximately 714

124 (range = 0 – 2,092) and 397 (range = 1 – 1,131) meters, respectively. Seeds, organized by

125 individual maternal tree, were then extracted from cones using a combination of mallet and pliers

126 and processed for inclusion in a long-term ex situ conservation collection (see below).

127

128   *Seed viability tests*

129   Estimating viability of seeds preserved *ex situ* is necessary given their potential role in

130   restoration, reforestation, or reintroduction. Given this, the potential viability of Torrey pine

131   seeds was tested using two complementary approaches prior to inclusion in the final ex situ

132   collection. A float test was first used as a rapid, low-cost approach to assess seed viability.

133   Floating seeds were presumed to lack an endosperm or embryo, while seeds that sunk were

134   presumed filled. Seeds were dropped into water for approximately 15 seconds to differentiate

135   presumed non-viable, floating seeds from presumed viable, sinking seeds (Gribko and Jones

136   1995; Morina *et al.* 2017). Those seeds classed as likely viable were organized by maternal tree

137   using paper bags, and then placed in a Blue M drying oven (Thermal Product Solutions, White

138   Deer, Pennsylvania, USA) maintained at 37°C for 24 hours to remove potential surface moisture.

139   Following this, seeds from a haphazard sample of maternal families were x-rayed at the

140   Placerville Nursery, CA. In addition to visualizing seed morphological variation, x-ray

141   photographs were used to verify viability based on float tests. Acrylic seed trays [20.3 cm x 25.4

142   cm x 0.48 cm], with a 9 x 11 array of wells, were used to separate and position each Torrey pine

143   seed over the x-ray film. Kodak x-OMAT HBT film (20.3 cm x 25.4 cm) was placed in a

144   lightproof x-ray film cassette which was positioned in the x-ray machine with the seed tray

145   centered on top of the film, with a shelf height of 55.9 cm. The x-ray was taken using a 17 kVP

146   exposure for a total of two minutes, based on standardized conditions established previously for

147   *Pinus coulteri* (Sara Wilson, USDA Forest Service, pers. comm.). X-ray images were digitized

148   using a Nikon D40 digital camera mounted on a tripod over a light box.

149

150   *Morphological measurement of seed traits*

151   Using ImageJ (Abràmoff *et al.* 2004), eight seed morphological traits were measured across 80

152   mainland maternal families and 30 island maternal families, representing a haphazard subset of

153   the complete collection (Fig. 2; Table 1). Each x-ray picture was scaled using the diameter of a

154   seed tray well (1.87 cm) to express pixels as trait values in centimeters. Directly measured seed

155   traits included seed length (SL, cm), seed width (SW, cm), embryo length (EL, cm), embryo

156   width (EW, cm), seed coat width (SCW, cm), seed area (SA, cm^2), endosperm area (ESA, cm^2),

157   and embryo area (EA, cm^2). We selected these traits as they can readily be measured from x-ray

158   pictures of seeds and provide a ubiquitous means to evaluate morphological variation for plants

159   preserved ex situ. Using measured seed traits, six additional traits were derived (Table 1),

160   including seed length/width ratio (SLW), embryo length/width ratio (ELW), relative embryo size

161   (RES), relative endosperm size (REndS), seed coat area (SCA, cm^2), and relative seed coat size

162   (RSCS). These traits were derived as they provide a means to relate different morphological

163   traits to each other and can provide a fine-scale estimate of the relative contribution of growth

164   and size traits within individual seeds. We measured five randomly selected seeds per maternal

165   tree, including three technical replicates per seed for each trait (the same seed was measured

166   three times for any given morphological trait). Measurements were averaged across technical

167   replicates to summarize the mean trait value per seed. In total, 550 seeds were measured from

168   across 110 maternal trees spanning the two Torrey pine populations.

169

170   *Seedling emergence test*

171   Within a restoration or reintroduction context, concurrent seedling emergence is often preferred

172   for nursery plantings. To evaluate the timing and probability of seedling emergence within

173 Torrey pine populations, a trial was conducted in January 2018 using a random subset of seeds

174 from the ex situ collection, including seeds sourced from Torrey Pine State Reserve and Santa

175 Rosa Island, CA. Following x-ray, seeds were stored at 4°C in sealed mylar bags (USA

176 emergency supply, Beaverton, Oregon, USA) placed in plastic boxes; each box contained

177 desiccant crystals to decrease ambient moisture and reduce likelihood of mold. Seeds from eight

178 maternal families per population were selected for the emergence trial. Between eight to ten

179 seeds per maternal tree were weighed and then stratified under cold, moist conditions for 30 days

180 (placed in plastic boxes on a moist paper towel at 4°C). Seeds were sown directly into a 164 mL

181 Ray Leach "Cone-tainer" ™ (Stuewe & Sons, Tangent, Oregon, USA) filled with Sunshine®

182 Mix #4 (Sungro horticulture, Agawam, Massachusetts, USA), pressed halfway into the soil, and

183 then covered with a thin layer of gravel. For approximately one month following planting, seeds

184 were misted for one minute at hourly intervals over a daily eight-hour period (9am – 4pm).

185 Following emergence, seedlings were hand watered to saturation weekly to biweekly as needed.

186 Emergence was quantified across three separate timepoints (Feb 06/2018, Feb 16/2018, and Mar

187 07/2018) per maternal family as the proportion of seeds that successfully developed into living

188 seedlings from the total initially planted.

189

190 *Evaluating the distribution of seed trait variation*

191 We conducted a principal component analysis (PCA) using all 14 measured and derived seed

192 traits averaged by maternal family to evaluate population-specific differentiation in seed

193 morphology. Prior to performing the PCA, to account for differences in measurement units, all

194 seed traits were standardized using the *scale()* function in R implementing the z-score

195 standardization: $z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$ (1), where x_{ij} is the non-transformed trait value, μ_j is the mean of

196   a given seed trait across populations, and σ_j the standard deviation of the same seed trait across

197   populations. Subsequently, to test for seed trait differences between population means, we used

198   either Student's two-sample test or its non-parametric equivalent when normality was not met,

199   Wilcoxon's two-sample test, within the R package "exactRankTests" (Hothorn and Hornik

200   2019). Normality was assessed using Shapiro-Wilk's test of normality within each population. In

201   total, four of the fourteen traits were distributed normally in both Torrey pine populations,

202   including seed width (mainland: W = 0.97, P = 0.06; island: W = 0.97, P = 0.52), embryo length

203   (mainland: W = 0.98, P = 0.29; island: W = 0.95, P = 0.21), embryo width (mainland: W = 0.97,

204   P = 0.09; island: W = 0.93, P = 0.05), and embryo area (mainland: W = 0.98, P = 0.45; island: W

205   = 0.96, P = 0.26).

206

207   To evaluate the distribution of morphological trait variation within and between Torrey pine

208   populations, we quantified the proportion of variation attributed to population and maternal tree

209   families using measured and derived morphological traits summarized by seeds. For each trait,

210   we fit a linear mixed model using the R package "lme4" (Bates *et al.* 2015) with population

211   considered a fixed effect and maternal families within populations considered a nested random

212   effect: $Y_{ij} = \mu + \pi_i + r_{i/j} + e_{ij}$, where Y_{ij} is the observed seed trait value, μ is the seed trait

213   overall mean, π_i is the effect of population origin on the seed trait mean, $r_{i/j}$ is the effect of

214   maternal family within populations on the observed seed trait value, and e_{ij} are the effects on the

215   seed trait value of any other variables unaccounted for in the model (residual error). For each

216   model, normality of residual errors was visually assessed and significance of fixed- and random-

217   effect terms was tested using the functions *anova()* and *ranova()* respectively, implemented in

218   the R package "stats" (R Core Team 2020) and "lmerTest" (Kuznetsova *et al.* 2017). Proportions

219   of seed morphological variance explained by populations (marginal R^2, R^2_m), both populations

220   and maternal families (conditional R^2, R^2_c), and maternal families alone ($R^2_c - R^2_m$) were

221   quantified for each model independently using the function *r.squaredGLMM()* implemented in

222   the R package "MuMIn" (Bartoń 2020).

223

224   ***Assessing differences in seedling emergence across populations***

225   To test for differences in the probability and the timing of seedling emergence in Torrey pine, we

226   evaluated the proportion of seeds that produced seedlings both within and between populations

227   across timepoints. First, we used Friedman's rank sum test (non-parametric repeated measures

228   ANOVA) followed by Wilcoxon' paired two-sample test, both implemented in the R package

229   "rstatix" (Kassambara 2020), to assess differences in the proportion of emerged seedlings

230   between timepoints within populations. We used a non-parametric approach for both Torrey pine

231   populations because normality could not be assumed at select timepoints due to high frequency

232   of zero values. We accounted for multiple testing using Benjamini and Hochberg (1995)'s False

233   Discovery Rate (FDR) correction implemented in the *wilcox_test()* function. Following this, we

234   evaluated timepoint-specific population differences in seedling emergence. We used Shapiro-

235   Wilk's test to assess populations' deviation from normality at each timepoint and either

236   Student's (for timepoints passing the normality test) or Wilcoxon's two-sample test (for

237   timepoints failing the normality test) to evaluate differences in population emergence.

238   Timepoints Feb 16/2018 (mainland: $W = 0.94$, $P = 0.57$; island: $W = 0.9$, $P = 0.28$) and Mar

239   07/2018 (mainland: $W = 0.96$, $P = 0.85$; island: $W = 0.88$, $P = 0.21$) passed the normality test,

240   while timepoint Feb 06/2018 (mainland: $W = 0.52$, $P < 0.001$; island: $W = 0.73$, $P = 0.004$) failed

241   the normality test. All statistical analyses were performed using R version 4.0.2 and 4.0.5 (R

242   Core Team 2020, 2021).

243

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

245   For each of the 14 measured and derived seed traits, we conducted a separate simulation

246   quantifying morphological variation captured when increasing the number of maternal families

247   sampled from contemporary Torrey pine seed collections. Simulations were conducted in R

248   version 3.6.3 (R Core Team 2020) using a customized script [See Supporting Information Figure

249   S2]. Resampling of ex situ collections were performed for island and mainland Torrey pine

250   populations independently, using between one and the total number of maternal families

251   available within each ex situ population collection (mainland: 80 maternal families, island: 30

252   maternal families) (N_{fam}). Maternal trees were sampled randomly without replacement from the

253   pool of available families. All seeds within each selected maternal family were sampled as part

254   of this simulation, except those with missing values for the trait simulated. Overall, between two

255   to five seeds per maternal family were sampled within each population.

256

257   To evaluate the number of maternal families needed to capture 95% of seed trait variation in both

258   island and mainland populations, we estimated the number of unique seed trait values captured in

259   a sample of N_{fam} maternal families (N_c) relative to the total number of unique seed trait values

260   present in a seed population (N_t). Here, we define "unique seed trait values" as the number of

261   non-redundant standardized measurements for the seed trait simulated rounded to the first digit.

262   Seed morphological measurements were rounded to the first digits as we believe that seed trait

263   variation estimated using additional digits is more likely to fail to capture meaningful biological

264   variation. Standardization of the data was performed so that all seed traits share the same unit

265   (the number of standard deviations a value is from the overall trait mean across populations, see

266   equation (1) above) and become comparable. Sampling of maternal families and estimation of

267   the summary statistic, defined as the proportion of total seed trait variance captured (N_c/N_t), were

268   repeated 500 times for each seed morphological trait and Torrey pine population. In this way,

269   N_c/N_t accounts for potential variation in number of seeds sampled per maternal family or

270   variation in sampled maternal families included.

271

272   Finally, for each number of maternal trees sampled (N_{fam}), we averaged the summary statistic

273   across all 500 replicates. This process was repeated for each of the 14 seed morphological traits

274   and performed for each Torrey pine population separately. Following this, the summary statistic

275   was averaged across all seed traits and separated by populations (see Results below). Proportions

276   of total seed trait variance captured (N_c/N_t) are provided based on proportions of maternal

277   families sampled (instead of the number of maternal families sampled) as sample sizes varied

278   across Torrey pine populations.

279

280   **Results**

281   *Island-mainland differentiation in seed morphology*

282   A principal component analysis (PCA) using all 14 measured and derived seed traits averaged by

283   maternal family revealed substantial differences in seed morphology between island and

284   mainland populations of Torrey pine (Fig. 3). The first PC axis explained 57.8% of variation in

285   seed morphological traits, primarily separating the island from the mainland population. Seed

286   length, seed width, seed area, endosperm area, and seed coat area exhibited the five highest

287   loadings (absolute values) on PC1 [See Supporting Information Table S1], indicating that seed

288   size and seed coat thickness can largely discriminate island from mainland individuals. On

289   average, seeds collected on island trees were longer, wider, larger, and thicker than seeds

290   collected on mainland trees (Table 1). The second PC axis explained 15.9% of seed trait

291   variation and summarizes within population variability in seed morphology (Fig. 3). Relative

292   seed coat size, relative endosperm size, and relative embryo size had the three highest loadings

293   (absolute values) on PC2 [See Supporting Information Table S1]. This suggests that once

294   corrected for seed size, seed coat thickness, endosperm size, and embryo size are traits

295   contributing to within-population variation.

296

297   ***Contribution of population origin and maternal family to seed trait variation***

298   Consistent with our principal component analysis, linear mixed models constructed for each of

299   the 14 measured and derived seed traits demonstrated that considerable variation in seed

300   morphology in Torrey pine is explained by population origin (Fig. 4). On average, population

301   origin explained 23% (range = 0.02–0.57) of variation across the species' distribution [See

302   Supporting Information Table S2]. Traits associated with seed size and seed coat thickness

303   exhibited the highest proportion of variance explained by population origin. These include seed

304   coat area (0.57; $F_{1,107.60}$ = 221.91, P < 0.001), seed area (0.49; $F_{1,107.56}$ = 156.45, P < 0.001),

305   endosperm area (0.37; $F_{1,108.07}$ = 100.58, P < 0.001), seed width (0.36; $F_{1,108}$ = 126.04, P <

306   0.001), seed coat width (0.32; $F_{1,108}$ = 96.04, P < 0.001), and seed length (0.30; $F_{1,108.50}$ = 78.92,

307   P < 0.001). Overall, this suggests seed size and seed coat thickness are major discriminants of

308   island and mainland Torrey pine seeds.

309

310   While population origin explained substantial variation across populations, assessment of

311   maternal seed families within populations indicated substantial family structure to seed trait

312   variation (Fig. 4). On average, maternal seed family explained 24% (range = 0.07–0.37) of

313   variation within populations [See Supporting Information Table S2]. Embryo length (0.37; $\chi^2 =$

314   124.82, df = 1, P < 0.001), seed length (0.35; $\chi^2 = 180.72$, df = 1, P < 0.001), endosperm area

315   (0.34; $\chi^2 = 211.87$, df = 1, P < 0.001), seed area (0.31; $\chi^2 = 256.71$, df = 1, P < 0.001), embryo

316   area (0.29; $\chi^2 = 100.16$, df = 1, P < 0.001), and seed coat width (0.28; $\chi^2 = 126.93$, df = 1, P <

317   0.001) exhibited the highest proportion of seed trait variation explained by within-population

318   maternal families. This suggests that there is substantial family-level structure to seed size,

319   endosperm size, embryo size, and seed coat thickness within Torrey pine populations.

320

321   ***Impact of population seed trait differentiation on seedling emergence***

322   The proportion of emerged seedlings increased over time for both island (Q = 15.5, df = 2, P <

323   0.001) and mainland (Q = 15.2, df = 2, P < 0.001) populations (Fig. 5). However, we found no

324   significant differences in the proportion of individuals emerging between populations across

325   observed time points. On average, 7% and 9% of mainland and island seedlings emerged a

326   month after sowing (Feb 06/2018; W = 28, P = 0.64), 63% and 53% of mainland and island

327   seedlings emerged a month and a half after sowing (Feb 16/2018; t = 0.81, df = 14, P = 0.43),

328   and 78% of mainland and island seedlings emerged two months after sowing (Mar 07/2018; t =

329   −0.06, df = 14, P = 0.95). Overall, this indicates that under controlled conditions, timing and

330   probability of emergence may not be impacted by population differences in seed morphology for

331   Torrey pine seedlings.

332

333   *Morphological variation captured in simulated seed collections*

334   Simulations revealed that to capture 95% of seed trait variation present in our existing ex situ

335   collections, on average 83% (25 out of 30) and 71% (57 out of 80) of all island and mainland

336   families would need to be resampled, respectively (Fig. 6). This indicates that both island and

337   mainland populations harbor considerable within-population structure for seed morphological

338   traits. Interestingly, capturing equal morphological variation across seed collections always

339   required a higher proportion of island maternal families to be collected relative to the mainland

340   population.

341

342   **Discussion**

343   Evaluating between- and within-population morphological variation in contemporary ex situ seed

344   collections may aid in understanding the distribution of variation needed to guide future

345   conservation efforts. Here, we quantified the distribution of trait variation within an ex situ

346   collection of Torrey pine seeds, with an aim to optimize future supplemental collections.

347   Morphologically, island and mainland seeds were significantly different from each other. Island

348   seeds were larger on average with thicker seed coats relative to their mainland counterparts.

349   These morphological differences may be explained by a combination of stochastic and

350   deterministic factors associated with population origin, including genetic bottlenecks following

351   island colonization, genetic drift, and selection associated with unique biotic pressures.

352   Interestingly, despite substantial morphological differentiation, seedling emergence did not vary

353   among populations, suggesting that either the probability and timing of emergence under

354   controlled conditions is not impacted by differences in seed morphology or that island and

355   mainland seeds respond similarly to an artificial germination protocol. In addition to population

356 origin, a considerable proportion of seed trait variation within each population was explained by

357 maternal family. This suggests that there is substantial within-population variation that will be

358 important to conserve and maintain populations' evolutionary potential. Finally, simulations

359 demonstrated that 83% and 71% of all maternal families within island and mainland seed

360 collection respectively were necessary to capture 95% of morphological variation within the

361 existing collections. To ensure we maintain representative levels of trait variation in future seed

362 collections, our results suggest that the number of maternal families sampled within natural

363 populations should be maximized, with the island population potentially requiring additional

364 sampling efforts.

365

366 Species distributed between island and mainland origins often exhibit marked among-population

367 phenotypic differentiation, including differentiation in seed morphology (Burns *et al.* 2012; Lens

368 *et al.* 2013; Kavanagh and Burns 2014; Burns 2016). Our results revealed considerable seed

369 morphological differences between Torrey pine populations (Fig. 3), primarily in seed size and

370 seed coat thickness (Fig. 4; See Supporting Information Table S1). On average, islands seeds

371 were larger and had thicker seed coats than seeds collected on the mainland (Table 1). These

372 results are consistent with previous studies of island-mainland systems that noted island

373 populations exhibited larger seeds relative to mainland populations (Kavanagh and Burns 2014;

374 Burns 2016; Biddick *et al.* 2019). A combination of different factors could contribute to

375 morphological variation among seed populations, including both stochastic and deterministic

376 forces.

377

378   On islands, seeds traits associated with long-distance dispersal may be selected against as they

379   can increase the probability an individual would disperse beyond an island's limits (Cody and

380   Overton 1996; Kavanagh and Burns 2014; Ottaviani *et al.* 2020, but see Burns 2018). For Torrey

381   pine, increased seed size on the island may have evolved to limit potential seed losses via wind-

382   dispersal, as seed mass negatively correlates with dispersal distance in pines (Greene and

383   Johnson 1993; Debain *et al.* 2003, but see Wyse and Hulme 2020). Nonetheless, Torrey pine

384   seeds possess degenerated wings (Ledig and Conkle 1983), suggesting that other mechanisms

385   likely contribute to seed dispersal in this species. Rodents and birds both feed on Torrey pine,

386   suggesting that seeds may undergo animal-mediated dispersal (Johnson *et al.* 2003). Thus, seed

387   predation may contribute to differences in seed size observed between populations. On the

388   island, *Peromyscus maniculatus* (Deer mouse) is the only rodent present to predate on Torrey

389   pine seeds (Johnson *et al.* 2003). This contrasts with the mainland, where multiple seed predators

390   have been documented; including *Peromyscus boylei* (Brush mice), *Peromyscus maniculatus*

391   (Deer mice), *Peromyscus eremicus* (Cactus mice)*, Chaetodipus californicus* (California pocket

392   mice), *Spermophilus beecheyi* (California ground squirrels), or *Aphelocoma californica* (Scrub

393   jays) (Johnson *et al.* 2003). If large seeds are preferentially targeted by seed predators (Reader

394   1993; Gómez 2004), reduced seed size on the mainland may have evolved as a consequence of

395   the trade-off between attracting predators to promote seed dispersal and mitigating fitness loss

396   due to seed consumption.

397

398   While selection may contribute to population differences, differentiation in seed morphology

399   may result from stochastic evolutionary forces. Founder effects associated with the colonization

400   of Santa Rosa island by mainland individuals, and genetic drift in the face of limited gene flow,

401   may have led to morphological differentiation between Torrey pine populations (Ledig and

402   Conkle 1983). Alternatively, a more complex demographic history of the two populations,

403   including colonization, extinction, and recolonization events may have led to the differences

404   observed between populations (Haller 1986). While both stochastic and deterministic factors

405   may contribute to population differences in seed morphology, additional experiments are

406   required to test mechanistic hypotheses. Seeds evaluated in this manuscript were collected from

407   natural populations. To tease apart the contribution of environment and genetics to seed trait

408   differences observed among populations, a common garden experiment is required. Indeed, a

409   reciprocal transplant experiment would be the most effective test of the action of natural

410   selection in shaping morphological differences between island and mainland seeds.

411

412   Despite significant differences in seed morphology between populations, timing and probability

413   of emergence was similar across populations (Fig. 5). Emergence rates were high throughout the

414   trial, with 78% of island and mainland seedlings emerging two months after sowing. The absence

415   of differences in seedling emergence between populations was surprising, as seed size often

416   negatively correlates with time to germination (Daws *et al.* 2005; Tanveer *et al.* 2013). However,

417   seed coat thickness can also influence rates of emergence. Seeds with thick seed coats relative to

418   their mass often germinate later than seeds with thinner seed coats (Daws *et al.* 2005; Hamilton

419   *et al.* 2013). For Torrey pine, Hamilton *et al.* (2017) found that island seeds germinate on

420   average two days after mainland seeds. Interestingly, island seeds were not only larger, but also

421   had thicker seed coats relative to mainland seeds (Table 1). Even after correcting for differences

422   in seed size, seed coat thickness (relative seed coat size; RSCS) remained moderately higher in

423   island seeds. Together, these results predict that island seedlings should emerge at similar or later

424   timepoints relative to mainland seedlings, which is consistent with current and previous

425   observations.

426

427   Similar emergence rates may also result from our experimental design. Abe and Matsunaga

428   (2011), in a mainland-island comparison study, observed that cold stratification attenuates

429   differences in germination rates between populations of *Rhaphiolepis umbellata*. Additionally,

430   complete and rapid germination of pine seeds is generally observed when pretreated under cold

431   and moist conditions (Krugman and Jenkinson 2008). Overall, this suggests that cold

432   stratification may mask population-specific differences in seedling emergence. Concurrent

433   seedling emergence from both Torrey pine populations coupled with high emergence success

434   suggests a cold stratification protocol is valuable for Torrey pine, particularly where

435   simultaneous emergence for nursery-grown seedlings is desired. Note, however, that variation in

436   the proportion of emerged seedlings within populations across timepoints may have concealed

437   population-specific differences in emergence rates. Consequently, weak differences in the timing

438   and probability of seedling emergence observed between island and mainland populations may

439   be an artifact of small numbers of seeds and maternal families used during emergence trials.

440

441   Although population origin explained a substantial proportion of seed trait variation, linear

442   mixed models demonstrated that maternal seed families within populations explained as much

443   variation (Fig. 4; See Supporting Information Table S2). Given generally high heritability for

444   seed morphological traits and the half-sib design of our collection (Pandey *et al.* 1994; Cober *et*

445   *al.* 1997; Mera *et al.* 2004; Roy *et al.* 2004; Carles *et al.* 2009; Zas and Sampedro 2015; Hakim

446   and Suyamto 2017), family-level seed trait variation likely provides a useful proxy for assessing

447   within-population genetic diversity. With nearly 25% of variation explained on average by

448   maternal families [See Supporting Information Table S2], this suggest there is substantial genetic

449   structure within Torrey pine populations. These results were notable as previous studies using

450   allozymes and chloroplast DNA suggested that the species exhibits little to no within-population

451   genetic variability (Ledig and Conkle 1983; Waters and Schaal 1991; Whittall *et al.* 2010).

452   However, the common garden experiment conducted by Hamilton and colleagues (2017)

453   indicated substantial family-level variation in tree height within both island and mainland

454   populations. Overall, these results indicate that Torrey pine populations may possess within-

455   population genetic variation necessary for natural selection to act upon. From a conservation

456   perspective, these findings suggest that a strategy maximizing the number of maternal families

457   sampled would optimize genetic diversity preserved in future ex situ seed collections and

458   increased distance among individuals may limit relatedness among maternal trees.

459

460   Generally, ex situ seed collections aim to capture 95% of genetic diversity present throughout a

461   species' distribution (Marshall and Brown 1975; Brown and Marshall 1995; Li *et al.* 2002;

462   Gapare *et al.* 2008). Simulations revealed that, in order to capture 95% of morphological

463   variation currently maintained ex situ, 25 (83% of island collection) and 57 (71% of mainland

464   collection) maternal families within each seed collection would need to be sampled (Fig. 6).

465   These data indicate that sampling more maternal families from the island population may be

466   necessary to achieve the same level of representation of morphological variation. Assuming

467   increased phenotypic variation observed on the island results from higher allelic diversity,

468   capturing 95% of genetic variation within the island population will always require more

469   maternal families relative to the mainland population. For these simulations, we assumed that

470   contemporary ex situ collections captured all morphological variation both within and between

471   populations, including seed phenotype frequencies. However, if this is not the case, these

472   recommendations may result in suboptimal sampling of standing variation within targeted

473   populations. This caveat is important because the number of x-rayed maternal families differed

474   between island (30 maternal families) and mainland (80 maternal families) seed collections. To

475   address this caveat, it will be important to have a general understanding of the fraction of natural

476   morphological variation captured across ex situ seed populations and adapt sampling efforts

477   accordingly.

478

479   Practical and cost-effective, long-term storage of seeds ex situ is widely used to capture and

480   maintain rare species genetic diversity. These seed collections represent an invaluable resource

481   to quantify within and between population trait variation that may be used to guide future ex situ

482   sampling efforts. Using Torrey pine as a model, we demonstrate that incorporating existing

483   information from ex situ collections offers a unique opportunity to monitor and optimize

484   conservation objectives, particularly important for rare species. While our results and

485   recommendations may be specific to Torrey pine, the empirical, statistical, and simulation-based

486   approaches presented here are broadly applicable to heritable traits across ex situ seed

487   collections.

488

489   **Data**

490   The data for this article, including seed morphological measurements and R scripts used are

491   available from GitHub: https://github.com/lnds-anonymous/AoBP2021.

492

Acknowledgements

References

Abe T, Matsunaga M. **2011**. Geographic variation in germination traits in Melia azedarach and Rhaphiolepis umbellata. *American Journal of Plant Sciences* **2**: 52–55.

Abràmoff MD, Magalhães PJ, Ram SJ. **2004**. Image processing with ImageJ. *Biophotonics international* **11**: 36–42.

Bartoń K. **2020**. MuMIn: Multi-Model Inference. R package version 1.43.17.

Basey AC, Fant JB, Kramer AT. **2015**. Producing native plant materials for restoration: 10 rules to collect and maintain genetic diversity. *Native Plants Journal* **16**: 37–53.

516   **Bates D, Mächler M, Bolker B, Walker S**. **2015**. Fitting linear mixed-effects models using

517   lme4. *Journal of Statistical Software* **67**: 1–48.

518   **Benjamini Y, Hochberg Y**. **1995**. Controlling the false discovery rate: a practical and powerful

519   approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*

520   **57**: 289–300.

521   **Biddick M, Hendriks A, Burns KC**. **2019**. Plants obey (and disobey) the island rule.

522   *Proceedings of the National Academy of Sciences* **116**: 17632–17634.

523   **Brown AHD, Marshall DR**. **1995**. A basic sampling strategy: theory and practice In: Guarino L,

524   Ramanatha Rao V, Reid R, eds. *Collecting plant genetic diversity: technical guidelines*.

525   Wallingford: CAB International, 75–91.

526   **Burns KC**. **2016**. Size changes in island plants: independent trait evolution in Alyxia ruscifolia

527   (Apocynaceae) on Lord Howe Island. *Biological Journal of the Linnean Society* **119**: 847–855.

528   **Burns KC**. **2018**. Time to abandon the loss of dispersal ability hypothesis in island plants: A

529   comment on García☐Verdugo, Mairal, Monroy, Sajeva and Caujapé☐Castells (2017). *Journal of*

530   *Biogeography* **45**: 1219–1222.

531   **Burns KC, Herold N, Wallace B**. **2012**. Evolutionary size changes in plants of the south☐west

532   Pacific. *Global Ecology and Biogeography* **21**: 819–828.

533   **Carles S, Lamhamedi MS, Beaulieu J, Stowe DC, Colas F, Margolis HA**. **2009**. Genetic

534   variation in seed size and germination patterns and their effect on white spruce seedling

535   characteristics. *Silvae genetica* **58**: 152–161.

536   **Cober ER, Voldeng HD, Fregeau-Reid JA**. **1997**. Heritability of seed shape and seed size in

537   soybean. *Crop Science* **37**: 1767–1769.

538   **Cody ML, Overton JM**. **1996**. Short-term evolution of reduced dispersal in island plant

539   populations. *Journal of Ecology* **84**: 53–61.

540   **Critchfield WB, Little EL**. **1966**. *Geographic distribution of the pines of the world*.

541   Washington, DC: US Department of Agriculture, Forest Service.

542   **Daws MI, Garwood NC, Pritchard HW**. **2005**. Traits of recalcitrant seeds in a

543   semi☐deciduous tropical forest in Panamá: some ecological implications. *Functional Ecology*

544   **19**: 874–885.

545   **Debain S, Curt T, Lepart J**. **2003**. Seed mass, seed dispersal capacity, and seedling

546   performance in a Pinus sylvestris population. *Ecoscience* **10**: 168–175.

547   **Dusek KH**. **1985**. Update on our rarest pine. *American Forests* **91**: 26–29, 61, 63.

548   **FAO**. **2010**. *The second report on the state of the world's plant genetic resources for food and*

549   *agriculture*. Rome: Food & Agriculture Org.

550   **Franklin J, Santos E V**. **2011**. A spatially explicit census reveals population structure and

551   recruitment patterns for narrowly endemic pine, Pinus torreyana. *Plant Ecology* **212**: 293–306.

552   **Gapare WJ, Yanchuk AD, Aitken SN**. **2008**. Optimal sampling strategies for capture of

553   genetic diversity differ between core and peripheral populations of Picea sitchensis (Bong.) Carr.

554   *Conservation Genetics* **9**: 411–418.

555   **Gómez JM**. **2004**. Bigger is not always better: conflicting selective pressures on seed size in

556   Quercus ilex. *Evolution* **58**: 71–80.

557   **Greene DF, Johnson EA**. **1993**. Seed mass and dispersal capacity in wind-dispersed diaspores.

558   *Oikos* **67**: 69–74.

559   **Gribko LS, Jones WE**. **1995**. Test of the float method of assessing northern red oak acorn

560   condition. *Tree Planters' Notes* **46**: 143–147.

561   **Guerrant Jr EO, Havens K, Vitt P**. **2014**. Sampling for effective ex situ plant conservation.

562   *International Journal of Plant Sciences* **175**: 11–20.

563   **Hakim L, Suyamto S**. **2017**. Gene action and heritability estimates of quantitative characters

564   among lines derived from varietal crosses of soybean. *Indonesian Journal of Agricultural*

565   *Science* **18**: 25–32.

566   **Hall T, Brinkman A**. **2015**. *Population Dynamics of the Island Torrey Pine (Pinus torreyana*

567   *ssp. insularis) on Santa Rosa Island, CA*. Undergraduate thesis, California State University

568   Channel Islands, United States.

569   **Haller JR**. **1986**. Taxonomy and Relationships of the Mainland and Island Populations of Pinus

570   torreyana (Pinaceae). *Systematic Botany* **11**: 39–50.

571   **Hamilton KN, Offord CA, Cuneo P, Deseo MA**. **2013**. A comparative study of seed

572   morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern

573   Australian rainforest species. *Plant Species Biology* **28**: 51–62.

574   **Hamilton JA, Royauté R, Wright JW, Hodgskiss P, Ledig TF**. **2017**. Genetic conservation

575   and management of the California endemic, Torrey pine (Pinus torreyana Parry): Implications of

576   genetic rescue in a genetically depauperate species. *Ecology and Evolution* **7**: 7370–7381.

577   **Hausman CE, Bertke MM, Jaeger JF, Rocha OJ**. **2014**. Genetic structure of green ash

578   (Fraxinus pennsylvanica): implications for the establishment of ex situ conservation protocols in

579   light of the invasion of the emerald ash borer. *Plant Genetic Resources* **12**: 286–297.

580   **Hoban S, Schlarbaum S**. **2014**. Optimal sampling of seeds from plant populations for ex-situ

581   conservation of genetic biodiversity, considering realistic population structure. *Biological*

582   *Conservation* **177**: 90–99.

583   **Hothorn T, Hornik K**. **2019**. exactRankTests: Exact Distributions for Rank and Permutation

584   Tests. R package version 0.8-31.

585 **IUCN**. **2020**. *The IUCN Red List of Threatened Species. Version 2020-2.*

586 https://www.iucnredlist.org. 9 Jul. 2020.

587 **Johnson M, Vander Wall SB, Borchert M**. **2003**. A comparative analysis of seed and cone

588 characteristics and seed-dispersal strategies of three pines in the subsection Sabinianae. *Plant*

589 *Ecology* **168**: 69–84.

590 **Kassambara A**. **2020**. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package

591 version 0.4.0.

592 **Kavanagh PH, Burns KC**. **2014**. The repeated evolution of large seeds on islands. *Proceedings*

593 *of the Royal Society B: Biological Sciences* **281**: 20140675.

594 **Krugman SL, Jenkinson JL**. **2008**. Pinus L. pine In: Bonner FT, Karrfalt RP, eds. *The woody*

595 *plant seed manual*. Washington, DC: US Department of Agriculture, Forest Service, 809–847.

596 **Kuznetsova A, Brockhoff PB, Christensen RHB**. **2017**. Package: Tests in Linear Mixed

597 Effects Models. *Journal of Statistical Software* **82**: 1–26.

598 **Ledig TF, Conkle TM**. **1983**. Gene diversity and genetic structure in a narrow endemic, Torrey

599 Pine (Pinus torreyana Parry ex Carr.). *Evolution* **37**: 79–85.

600 **Lens F, Davin N, Smets E, del Arco M**. **2013**. Insular woodiness on the Canary Islands: a

601 remarkable case of convergent evolution. *International Journal of Plant Sciences* **174**: 992–

602 1013.

603 **Li Q, Xu Z, He T**. **2002**. Ex situ genetic conservation of endangered Vatica guangxiensis

604 (Dipterocarpaceae) in China. *Biological Conservation* **106**: 151–156.

605 **Marshall DR, Brown AHD**. **1975**. Optimum sampling strategies in genetic conservation In:

606 Frankel OH, Hawks JG, eds. *Crop Genetic Resources for Today and Tomorrow*. London:

607 Cambridge University Press, Cambridge, 53–80.

608 **Mera M, Jerez R, Miranda H, Rouanet JL**. **2004**. Seed coat specific weight in Lupinus

609 angustifolius: influence of genotype and environment and relationship with seed coat proportion.

610 *Australian journal of agricultural research* **55**: 1189–1195.

611 **Monty A, Bizoux J-P, Escarré J, Mahy G**. **2013**. Rapid plant invasion in distinct climates

612 involves different sources of phenotypic variation. *PLoS one* **8**: e55627.

613 **Morina DL, Lashley MA, Chitwood MC, Moorman CE, DePerno CS**. **2017**. Should we use

614 the float test to quantify acorn viability? *Wildlife Society Bulletin* **41**: 776–779.

615 **Ottaviani G, Keppel G, Götzenberger L, *et al.* 2020**. Linking plant functional ecology to

616 island biogeography. *Trends in Plant Science* **25**: 329–339.

617 **Pandey MP, Seshu D V, Akbar M**. **1994**. Genetics of embryo size and its relationship with seed

618 and seedling vigour in rice (Oryza sativa L.). *Indian Journal of Genetics and Plant Breeding* **54**:

619 258–268.

620 **Platenkamp GAJ, Shaw RG**. **1993**. Environmental and Genetic Maternal Effects on Seed

621 Characters in Nemophila menziesii. *Evolution* **47**: 540–555.

622 **Potter KM, Jetton RM, Bower A, *et al.* 2017**. Banking on the future: progress, challenges and

623 opportunities for the genetic conservation of forest trees. *New Forests* **48**: 153–180.

624 **R Core Team**. **2020**. R: A language and environment for statistical computing.

625 **R Core Team**. **2021**. R: A language and environment for statistical computing.

626 **Reader RJ**. **1993**. Control of seedling emergence by ground cover and seed predation in relation

627 to seed size for some old-field species. *Journal of Ecology* **81**: 169–175.

628 **Roy SM, Thapliyal RC, Phartyal SS**. **2004**. Seed source variation in cone, seed and seedling

629 characteristic across the natural distribution of Himalayan low level pine Pinus roxburghii Sarg.

630 *Silvae Genetica* **53**: 116–123.

631   **Di Santo LN, Hamilton JA**. **2020**. Using environmental and geographic data to optimize ex situ

632   collections and preserve evolutionary potential. *Conservation Biology* **35**: 733–744.

633   **Schaal B, Leverich W**. **2004**. Population genetic issues in ex situ plant conservation In:

634   Guerrant EO, Havens-Young K, Maunder M, eds. *Ex Situ Plant Conservation: Supporting*

635   *Species Survival In The Wild*. Washington, DC: Island Press, 267–284.

636   **Singh J, Clavijo Michelangeli JA, Gezan SA, Lee H, Vallejos CE**. **2017**. Maternal Effects on

637   Seed and Seedling Phenotypes in Reciprocal F1 Hybrids of the Common Bean (Phaseolus

638   vulgaris L.). *Frontiers in Plant Science* **8**: 42.

639   **Tamaki S, Isoda K, Takahashi M, Yamada H, Yamashita Y**. **2018**. Genetic structure and

640   diversity in relation to the recently reduced population size of the rare conifer, Pseudotsuga

641   japonica, endemic to Japan. *Conservation Genetics* **19**: 1243–1255.

642   **Tanveer A, Tasneem M, Khaliq A, Javaid MM, Chaudhry MN**. **2013**. Influence of seed size

643   and ecological factors on the germination and emergence of field bindweed (Convolvulus

644   arvensis). *Planta daninha* **31**: 39–51.

645   **Villellas J, Berjano R, Terrab A, García MB**. **2014**. Divergence between phenotypic and

646   genetic variation within populations of a common herb across Europe. *Ecosphere* **5**: 1–14.

647   **Waters ER, Schaal BA**. **1991**. No variation is detected in the chloroplast genome of Pinus

648   torreyana. *Canadian Journal of Forest Research* **21**: 1832–1835.

649   **Weber A, Kolb A**. **2014**. Differences in heritable trait variation among populations of varying

650   size in the perennial herb Phyteuma spicatum. *Conservation genetics* **15**: 1329–1337.

651   **Whittall JB, Syring J, Parks M, *et al.*** **2010**. Finding a (pine) needle in a haystack: chloroplast

652   genome sequence divergence in rare widespread pines. *Molecular Ecology* **19**: 100–114.

653   **Wyse S V, Hulme PE**. **2020**. Limited evidence for a consistent seed mass☐dispersal trade☐off

654   in wind☐dispersed pines. *Journal of Ecology*.

655   **Yoko ZG, Volk KL, Dochtermann NA, Hamilton JA**. **2020**. The importance of quantitative

656   trait differentiation in restoration: landscape heterogeneity and functional traits inform seed

657   transfer guidelines. *AoB Plants* **12**: plaa009.

658   **Zas R, Sampedro L**. **2015**. Heritability of seed weight in Maritime pine, a relevant trait in the

659   transmission of environmental maternal effects. *Heredity* **114**: 116–124.

660   **Zhang D-Q, Zhou N**. **2013**. Genetic diversity and population structure of the endangered conifer

661   Taxus wallichiana var. mairei (Taxaceae) revealed by Simple Sequence Repeat (SSR) markers.

662   *Biochemical systematics and ecology* **49**: 107–114.

663

664 **Tables**

665 **Table 1.** Morphological measurements of Torrey pine seeds sourced from Santa Rosa Island (Island, n=30) and Torrey Pine State

666 Reserve (Mainland, n=80), CA. Listed are population mean estimates ($\pm$SE) of measured (A) and derived (B) seed traits summarized

667 by maternal families. Measurable traits: seed length (SL), seed width (SW), embryo length (EL), embryo width (EW), seed coat width

668 (SCW), seed area (SA), endosperm area (ESA), and embryo area (EA). Derived traits: seed length/width ratio (SLW), embryo

669 length/width ratio (ELW), relative embryo size (RES), relative endosperm size (REndS), seed coat area (SCA), and relative seed coat

670 size (RSCS). Differences in seed morphology between mainland and island populations were significant (α=0.05) for all 14 seed traits.

671 A – Measured seed traits

Population	SL (cm)	SW (cm)	SCW (cm)	EL (cm)	EW (cm)	SA (cm^2)	ESA (cm^2)	EA (cm^2)
Mainland	1.41$\pm$0.016	0.77$\pm$0.009	0.09$\pm$0.002	1.08$\pm$0.012	0.15$\pm$0.003	0.89$\pm$0.023	0.55$\pm$0.014	0.17$\pm$0.004
Island	1.67$\pm$0.02	0.97$\pm$0.015	0.12$\pm$0.002	1.2$\pm$0.02	0.19$\pm$0.004	1.41$\pm$0.031	0.81$\pm$0.021	0.23$\pm$0.007

672 B – Derived seed traits

| Population	SLW (SL/SW)	ELW (EL/EW)	SCA (cm^2) (SA-ESA)	RES (EA/ESA)	REndS (ESA/SA)	RSCS (SCA/SA)
Mainland	1.86$\pm$0.013	7.28$\pm$0.094	0.34$\pm$0.01	0.31$\pm$0.005	0.62$\pm$0.003	0.38$\pm$0.003
Island	1.77$\pm$0.037	6.64$\pm$0.136	0.6$\pm$0.013	0.29$\pm$0.004	0.57$\pm$0.004	0.43$\pm$0.004

673   **Figures**
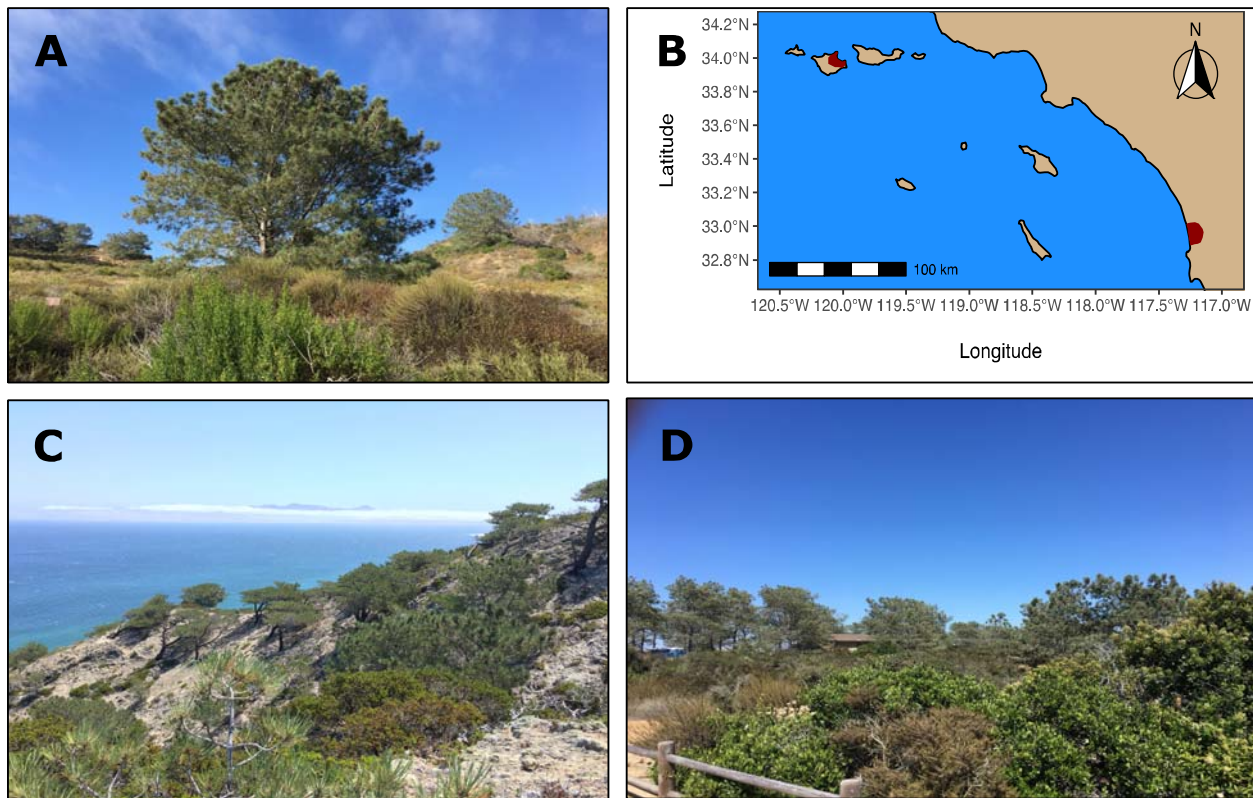
674

675   **Fig. 1.** (A) *Pinus torreyana* individual. (B) *Pinus torreyana* distribution map, including Torrey

676   pine distribution on Santa Rosa Island, CA (*Pinus torreyana* subsp. *Insularis*, top left red-shaded

677   area) and at the Torrey Pine State Reserve, CA (*Pinus torreyana* subsp. *torreyana*, bottom right

678   red-shaded area). (C) Torrey pine stand on Santa Rosa Island, CA. (D) Torrey pine stand at the

679   Torrey Pine State Reserve, CA.

680
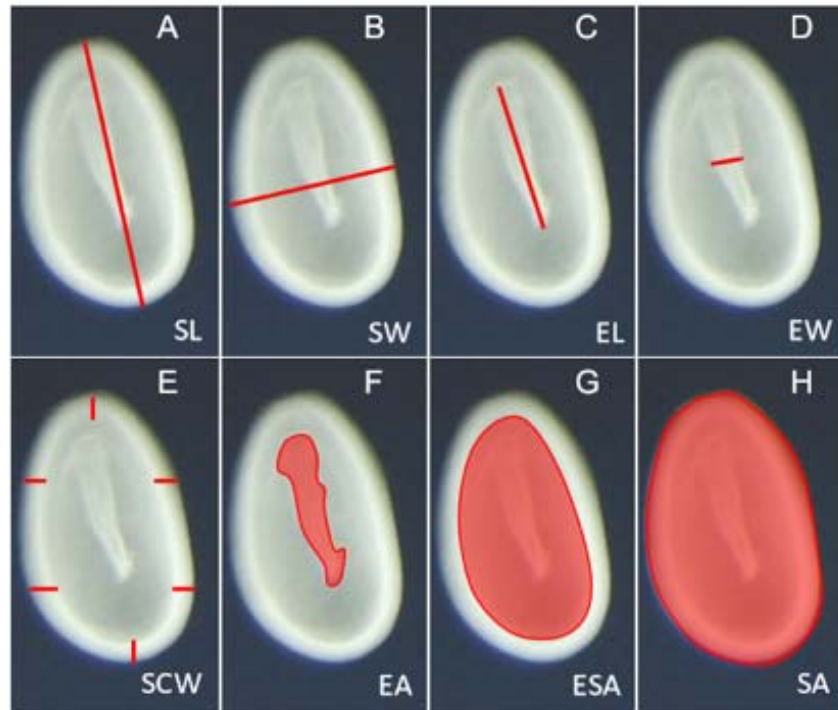
Fig. 2. Visual of morphological measurements taken using ImageJ for seeds collected on Santa

Rosa Island and at the Torrey pine State Reserve. (A) Seed length [cm]. (B) Seed width [cm]. (C)

Embryo length [cm]. (D) Embryo width [cm]. (E) Seed coat width [cm]. (F) Embryo area [cm^2].

(G) Endosperm area [cm^2]. (H) Seed area [cm^2].

685
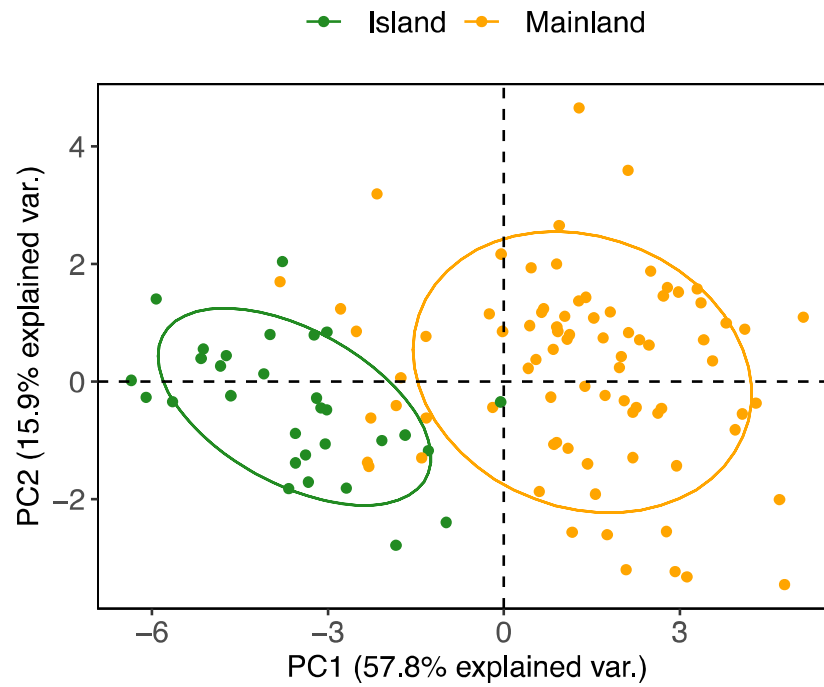
Fig. 3. Principal components analysis (PCA) using all 14 seed morphological traits measured and derived from maternal plants collected on Santa Rosa Island (green) and at the Torrey Pine State Reserve (orange).

689

690
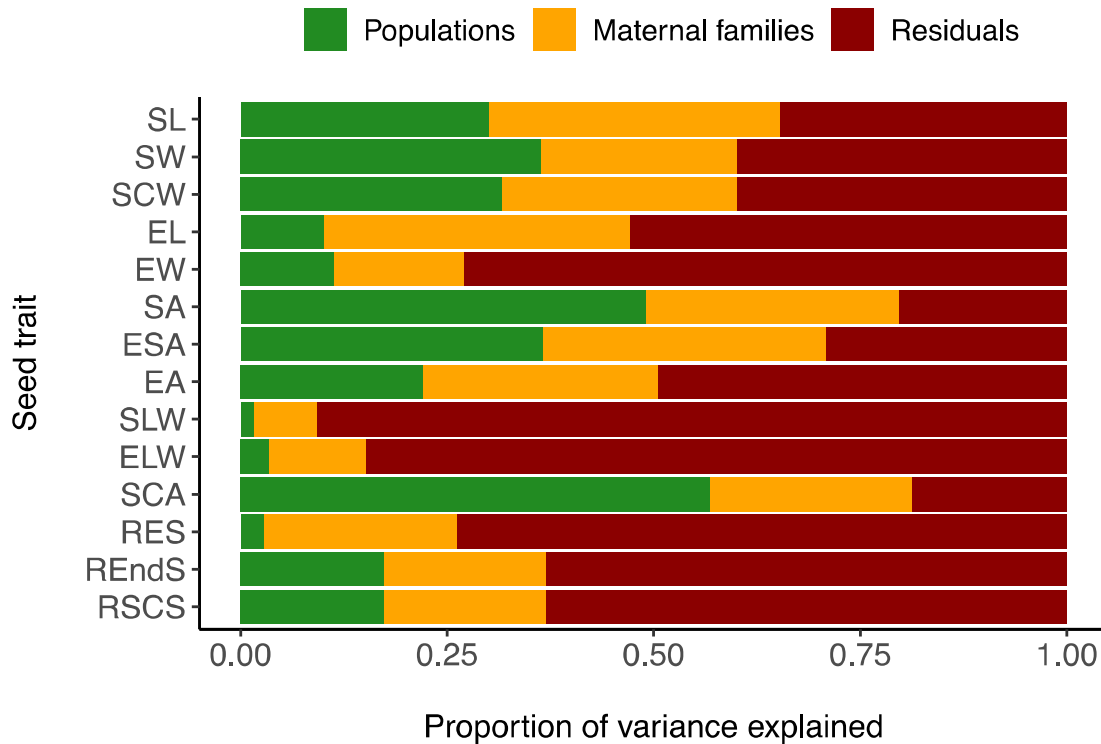
Fig. 4. Proportion of variance in seed morphology explained by populations (green), maternal

families within populations (orange), and other variables not accounted for in the model

(residuals; dark red) for each of the 14 measured and derived seed traits. SW, seed width (cm);

SLW, seed length/width ratio; SL, seed length (cm); SCW, seed coat width (cm); SCA, seed coat

area (cm^2); SA, seed area (cm^2); RSCS, relative seed coat size; RES, relative embryo size;

REndS, relative endosperm size; EW, embryo width (cm); ESA, endosperm area (cm^2); ELW,

embryo length/width ratio; EL, embryo length (cm); EA, embryo area (cm^2). See Supporting

Information Table S2 for numerical estimates.

699
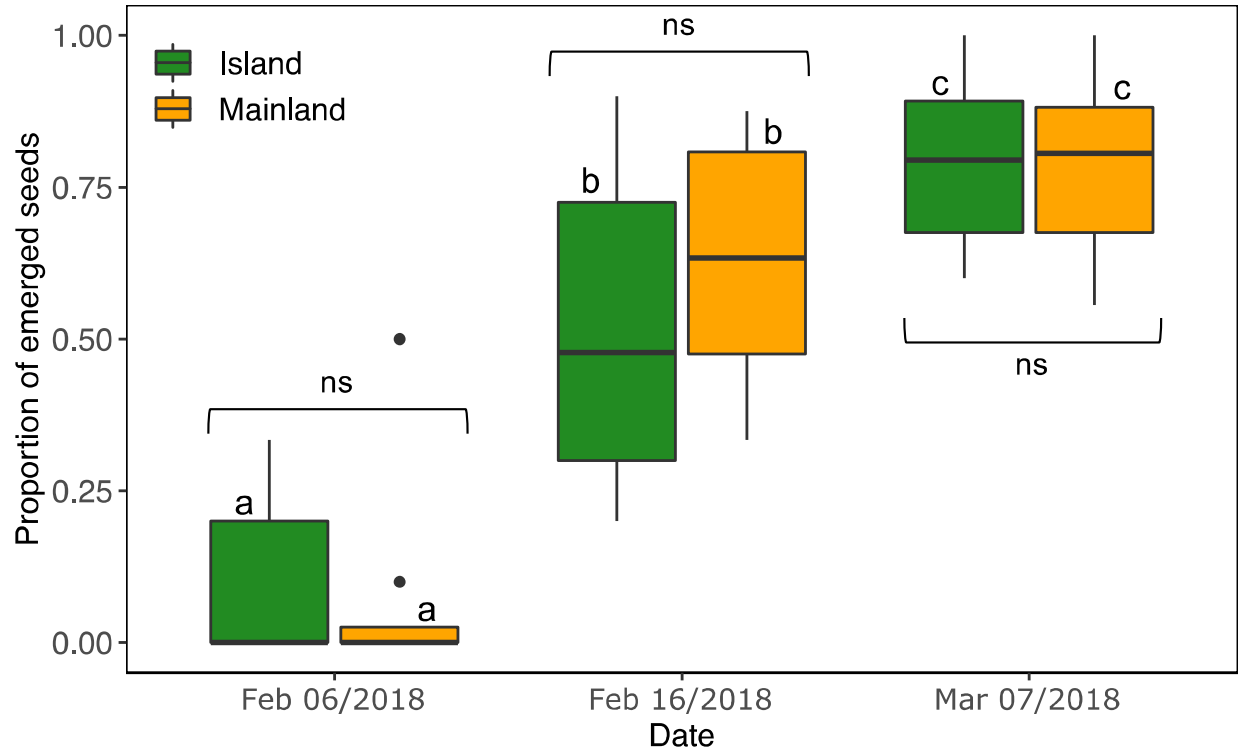
Fig. 5. Proportion of emerged seedlings (y axis) recorded at three different timepoints (x axis) for seeds sampled on Santa Rosa Island (green) and at the Torrey Pine State Reserve (orange). Significant differences in emergence time across timepoints within populations are indicated with different letters. Comparisons between populations at each timepoint is indicated with square brackets. ns, non-significant differences ($\alpha=0.05$).

705
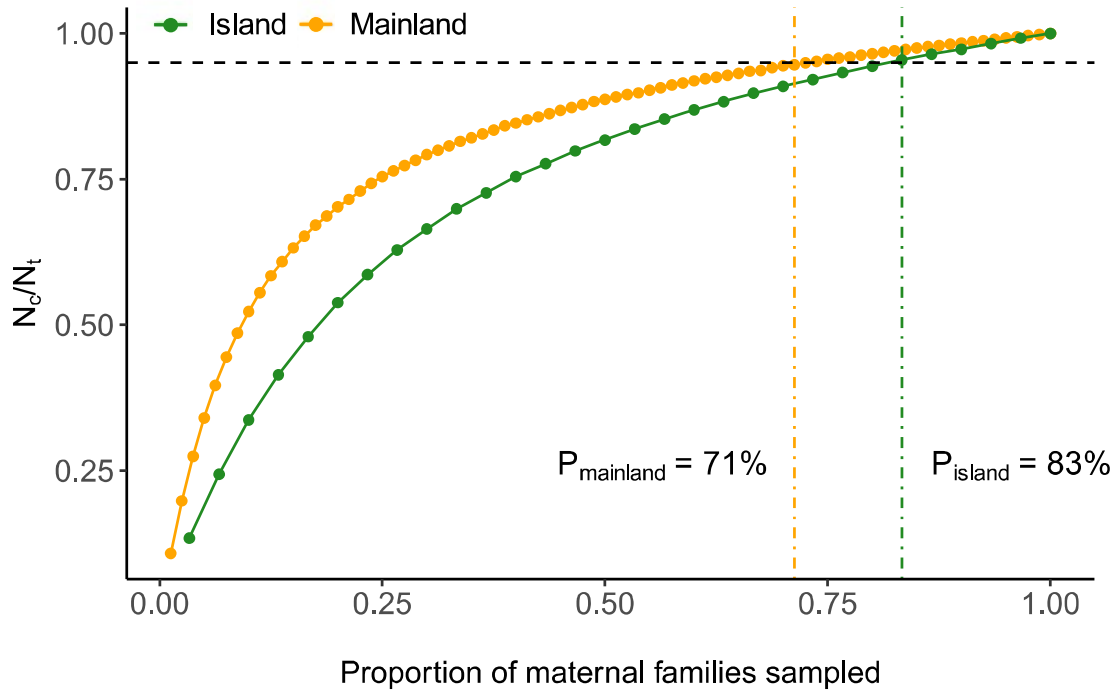
Fig. 6. Phenotypic variation captured across seed traits in simulated collections (N_c) relative to total phenotypic variation present in seed populations (N_t). Average proportion of phenotypic variation captured (N_c/N_t) was estimated for various proportions of maternal families sampled. P_{island} and $P_{mainland}$ represent the proportion of maternal families required to capture 95% of morphological variation (horizontal dashed line) present in island (green) and mainland (orange) ex situ seed populations, respectively.