Evolutionary rate and genetic load in an emblematic Mediterranean tree following an ancient and prolonged population collapse

Juan Jaramillo-correa, Francesca Bagnoli, Delphine Grivet, Bruno Fady, Filippos Aravanopoulos, Giovanni Vendramin, Santiago C. Gonzalez-Martinez

Evolutionary rate and genetic load in an emblematic Mediterranean tree following an ancient and prolonged population collapse

Juan P. Jaramillo-Correa[1], Francesca Bagnoli[2], Delphine Grivet[3], Bruno Fady[4], Filippos A. Aravanopoulos[5], Giovanni G. Vendramin[2], Santiago C. González-Martínez[6,*]

1. Department of Evolutionary Ecology, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico
2. Institute of Biosciences and Bioresources, Division of Florence, National Research Council, Sesto Fiorentino (FI), Italy
3. Department of Forest Ecology and Genetics, Forest Research Centre, INIA-CIFOR, Madrid, Spain
4. INRAE, Unité de Recherche Écologie des Forêts Méditerranéennes (URFM), Avignon, France
5. Laboratory of Forest Genetics and Tree Breeding, Department of Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece
6. INRAE, Université de Bordeaux, UMR BIOGECO, Cestas, France

*Corresponding author: santiago.gonzalez-martinez@inrae.fr

Abstract

Severe bottlenecks significantly diminish the amount of genetic diversity and the speed at which it accumulates (i.e. evolutionary rate). They further compromise the efficiency of natural selection to eliminate deleterious variants, which may reach fixation in the surviving populations. Consequently, expanding and adapting to new environments may pose a significant challenge when strong bottlenecks result in genetic pauperization. Herein, we surveyed the patterns of nucleotide diversity, molecular adaptation and genetic load across 177 gene-loci in a circum-Mediterranean conifer (*Pinus pinea* L.) that represents one of the most extreme cases of genetic pauperization in widespread outbreeding taxa. We found very little genetic variation in both hypervariable nuclear microsatellites (SSRs) and gene-loci, which translated into genetic diversity estimates one order of magnitude lower than those previously reported for pines. Such values were consistent with a strong population decline that began some ~1Ma. Comparisons with the related and parapatric maritime pine (*Pinus pinaster* Ait.) revealed reduced rates of adaptive evolution (α and ω_a) and a significant accumulation of genetic load. It is unlikely that these are the result from differences in mutation rate or linkage disequilibrium between the two species; instead they are the presumable outcome of contrasting demographic histories affecting both the speed at which these taxa accumulate genetic diversity, and the global efficacy of selection. Future studies, and programs for conservation and management, should thus start testing for the effects of genetic load on fitness, and integrating such effects into predictive models.

Introduction

Evolutionary rate, the speed at which genetic diversity accumulates within a species' genome, depends on both the rate at which mutations appear, and the likelihood that they reach fixation (Charlesworth et al. 1995; Eyre-Walker and Keightley 2007; Charlesworth 2009). This likelihood is conditioned by each mutation's effect on fitness, and stochastic processes that may eliminate them. Consequently, evolution (and adaptation) indirectly relies on species' biological features like population size (or its change over time), migration rate and mating system, which condition both the efficacy of selection and the extent of genetic drift (e.g. Charlesworth 2009; Chen et al. 2017). For instance, deleterious mutations accumulate at very low frequencies in large populations (often in heterozygous states), while they can more easily reach fixation in smaller populations, because of a diminished efficiency of purifying selection (Charlesworth et al. 1993; Eyre-Walker and Keightley 2007). Similarly, advantageous variants may increase in frequency at faster rates in large populations or under elevated gene flow regimes, than in smaller or isolated stands, where they can disappear through the effect of drift (Charlesworth and Eyre-Walker 2007; Eyre-Walker and Keightley 2009).

The consequences of drastic population size reductions on the efficacy of selection and the accumulation of genetic load (the reduction of mean fitness in a population caused by deleterious variation relative to a mutation-free population) are thus fundamental aspects of population genetics (e.g. Kimura et al. 1968; Ohta 1992; Gillespie 2001). They are also of high interest in current conservation, domestication, improvement and clinical genetics projects (e.g. Marsden et al. 2016; Robinson et al. 2016; Pedersen et al. 2017). Theory predicts that neutral variation accumulates proportionally to mutation rate, and independently of population size (Lanfear et al. 2014). As a result, the (neutral) evolutionary rate of population/species should not be

affected after a bottleneck if mutation rate remains unchanged. However, a negative relationship is expected between the accumulation of deleterious mutations and population size, with larger populations/species purging more efficiently disadvantageous variants than smaller ones (e.g. Chen et al. 2017). Consequently, comparing how different sorts of mutations (neutral vs. putatively deleterious) accumulate in populations/species with contrasting demography may be used as a proxy for inferring adaptive evolutionary rates (e.g. Böndel et al. 2015; González-Martínez et al. 2017; Grivet et al. 2017).

Some studies (e.g. Balloux and Lehmann 2012; Wei et al. 2015) have nevertheless suggested that mutation rate can also be affected by demographic fluctuations, particularly when generations overlap, as in most long-lived taxa, including mammals and forest trees. While such a possibility may flaw the comparisons above, other authors have claimed that it is the generation time that is affected by demographic changes, rather than mutation rate itself (Lanfear et al. 2014). A recent empirical survey on two pines with contrasting demography, and similar generation times and life-history traits, (Grivet et al. 2017) pointed in that direction, as it reports no differences in the rate of evolution between these taxa.

Forest trees are reputed for their resilience to stochastic processes, particularly those derived from population size changes (Petit and Hampe 2006). They have long generation times, extensive gene flow and large reproductive output over the years, which, combined with strong selective pressures at early life stages, reduce the impact of inbreeding depression and maintains (ancestral) genetic diversity (e.g. Ledig et al. 2000). On the other hand, such features also result in a lowered accumulation of mutations per unit of time, which suggests a slow recovering of molecular genetic variation after a strong and prolonged bottleneck (Lanfear et al. 2014; Jaramillo-Correa

et al. 2015a). Interestingly, these features can also make them more impermeable to the accumulation of genetic load and favour the spread of new advantageous mutations (Petit and Hampe 2006; Lanfear et al. 2014).

Stone pine (*Pinus pinea* L.) is an emblematic Mediterranean tree that has been associated to humans for thousands of years. For instance, fatty acids related to stone pine nuts have been identified in calculus from hominid teeth remnants more than 300 ka old (Hardy et al. 2015), and charcoal fragments assigned to stone pine have been reported in several human settlements dating back to the last ~50 ka (Prada et al. 1997; Badal 1998; Carrión et al. 2008). Archaeological evidence indicates that this species has been cultivated for over 6,000 years (Prada et al. 1997; Pérez-Jordà et al. 2017), and its nuts have been traded since the Phoenician civilization (i.e. 3,500 BP; Wilcox 1977; Popova and Hristova 2017). Stone pine is thus considered an archaeophyte (i.e. a species distinguished by unrecorded early man introductions) currently distributed around the whole Mediterranean basin, where it forms small populations growing on different types of soil, from the sea level up to 1,000 meters (Richardson 1998). Despite this distribution, stone pine is one of the most genetically depauperate outbreeding plants in the world, with virtually no genetic variation at allozyme or chloroplast microsatellite loci (Fallour et al. 1997; Vendramin et al. 2008). These features make stone pine an ideal model for testing hypotheses related to evolutionary rate changes and the accumulation of genetic load after strong and prolonged bottlenecks.

Herein, we collected nuclear microsatellite and DNA sequence data to estimate more accurately the genomic diversity of stone pine and infer its demographic history. For comparison, we used data available for the parapatric sister taxon maritime pine (*Pinus pinaster* Ait.), which inhabits similar environments but has much higher amounts of genetic diversity (Bucci et al. 2007; Jaramillo-Correa et al. 2015b; Grivet et al. 2017). We

120 then determined the accumulation of putatively neutral and deleterious genetic
121 variation since its divergence from maritime pine, by using the distantly related (and
122 native to North America) loblolly pine (*Pinus taeda* L.) as outgroup. We inferred various
123 evolutionary parameters by taking into account both the shared and exclusive portion of
124 genetic variation to each taxon. We hypothesized that the low effective population size
125 of stone pine resulted in a diminished evolutionary rate and in higher genetic load when
126 compared to maritime pine. However, such values should be different from those
127 expected under a simple bottleneck model with no selection. That is, we expected that
128 purifying selection have managed to purge most of the deleterious part of the genetic
129 load and allowed for the retention of some advantageous variants in stone pine (i.e. the
130 number of fixed adaptive mutations should be different from zero).
131

132 **Materials and Methods**

133 *Sampling, genotyping and sequencing*

134 Two datasets were used to infer demographic and population genetic patterns in stone
135 pine. First, nuclear microsatellite (nuSSR) data were obtained for 735 individuals from
136 33 populations distributed along the species range (Fig. 1A; Table S1). Sample size per
137 population varied between 16 and 30 individuals; all of which were at least 50 m apart.
138 DNA was extracted from foliage using DNeasy Plant Mini Kits (QIAGEN, Valencia, CA,
139 USA) and amplified using stone pine specific primers for 13 nuSSRs. Primer sequences
140 and PCR conditions are described in Pinzauti et al. (2012). PCR amplification was carried
141 out in a GeneAmp PCR 9700 thermal-cycler (Applied Biosystems, Foster City, California,
142 USA), and PCR products were separated in an ABI 3130xl automatic sequencer (Applied
143 Biosystems). Electrophoregrams were scored using GeneMapper version 4.0 (Applied
144 Biosystems).

Second, DNA sequences for 264 gene loci originally identified on loblolly pine were obtained for 12 individuals from as many populations dispersed across the full species' range (Fig. 1A; Table S1). Sanger resequencing was performed on haploid seed megagametophytes, which allows for direct phase inference and identification of co-amplified gene paralogs. DNA sequences were visually checked and manually edited with SEQUENCHER 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA). Low-quality sequences and those exhibiting double peaks (indicative of paralog co-amplification) were removed, which resulted in a final dataset of 177 gene loci scattered over most pine chromosomes (see Westbrock et al. 2015; Plomion et al. 2016), and covering up to 56 Kbp. The retained sequences were aligned with their maritime pine and loblolly pine orthologs, and annotated from homology with EST contigs from the last species and the NCBI reference protein database using GENEIOUS version 6.1 (Biomatters, Auckland, New Zealand).

Genetic diversity and population structure

For nuSSR data, standard genetic diversity statistics were estimated for each population using GENALEX version 6.503+ (Peakall and Smouse 2006). These included the mean (A) and effective (Ae) number of alleles, the number of private alleles, and the observed (H_O) and expected (H_E) heterozygosity, from which F-statistics were computed with FSTAT (Goudet 2002). Deviations from Hardy-Weinberg (H-W) expectations and linkage disequilibrium between loci were also evaluated with GENALEX. Only one locus (*pEST2669*) showed deviations from H-W in over 70% of the populations and was removed from all analyses. No linkage between pairs of loci was systematically observed. Population clustering and differentiation were evaluated using a principal coordinates analysis (PCoA, performed with GENALEX; Peakall and Smouse 2006), and

with Bayesian software STRUCTURE (Pritchard et al. 2000). For the latter analysis, ten runs of 1,000,000 iterations of an admixture model with correlated allele frequencies and no priors on population locations were performed for *K*-values ranging from 1 to 13, after a burn-in period of 500,000 steps. Convergence among runs with the same *K*-value was verified using CLUMPP (Jakobsson and Rosenberg 2007), and the most plausible number of clusters was determined in STRUCTURE HARVESTER (Earl and von Holdt 2011) following Evanno et al. (2005) and Janes et al. (2017).

For the DNA sequence dataset, standard diversity estimates, such as the number of segregating sites (S) and haplotypes (h), π (Nei and Li 1979) and Watterson's θ_W (Watterson 1975) were calculated using DnaSP v.6 (Rozas et al. 2017) for all sites, and separately for synonymous and non-synonymous sites. In addition, the pairwise divergence (D_{xy}) from loblolly pine was computed together with the number of synonymous substitutions per synonymous site (K_s) and of non-synonymous substitutions per non-synonymous site (K_a), both using the Jukes-Cantor correction. To ease some of the demographic and genetic-load inferences below, the two last parameters (K_s and K_a) were also determined between maritime and loblolly pines. Sequence alignments were also used to separate divergence counts into those that were exclusive to each maritime and stone pines (i.e. that arose after their divergence), and those that were shared between these two taxa (i.e. that appeared before their split but after the separation from their common ancestor with loblolly pine).

Phylogeny and demographic inferences

The phylogenetic relationships and divergence times between stone, maritime and loblolly pines were co-inferred from a random subset of 50 gene-sequence alignments. This subset was selected to reduce computing time and facilitate convergence among

partitions and runs. First, the best-fit gene partition, with the corresponding molecular evolution models, was selected based on a Bayesian Information Criterion with PARTITIONFINDER 1.1 (Lanfear et al. 2012) using linked branch lengths and a greedy algorithm. Then, phylogenies were inferred using the Bayesian framework implemented in BEAST v1.7 (Drummond et al. 2013) by assuming a *log*-normal relaxed molecular clock and specifying *log*-normal-distributed priors for the parameters, and a Calibrated Yule model for the tree. Following previous phylogenetic work on pines (e.g. Gernandt et al. 2005; Saladin et al. 2017), the divergence between North American loblolly pine (subsect. *Autrales*) and the two Mediterranean pines (stone and maritime pines; subsect. *Pinaster*) was calibrated using a *log*-normal distribution centred at 35 million years ago (Ma) and a 95% probability interval of ± 10 Ma; such a calibration covers the subgenus *Pinus* split, which was dated back to 45Ma based on *P. bailey* fossils (Saladin et al. 2017). Analyses consisted of ten runs of 100 million iterations each; samples were collected every 1,000 steps after removing the first 25% of trees as burn-in. Chain mixing, likelihood stability, and convergence across runs were surveyed using TRACER v1.6 (Drummond et al. 2012). Subsampled trees were finally summarized by Maximum Clade Credibility with common ancestor heights as node heights on TREEANNOTATOR 1.8.0 (Drummond et al. 2012).

The demographic history of stone pine was initially surveyed using STAIRWAYPLOT (Liu and Fu 2015), an approach that uses the site frequency spectrum (SFS) to fit a demographic model consisting of population size changes over various periods of time. This analysis was performed from the unfolded SFS derived from all DNA sequences (55,833 sites) and using the median mutation rate inferred on the phylogenetic analyses above (i.e. 4.18×10^{-8} per site per generation); as for other pines (Brown et al. 2004; Willyard et al. 2006), a generation time of 25-40 years was assumed. Median values and

220 confidence intervals for population size estimates were derived from 200 replicates of
221 the input file using the built-in bootstrap function. Stairway-plots were generated for the
222 whole sample and for each genetic cluster separately (see below).
223 In addition, and to gain a more detailed insight of the most recent demographic
224 history of stone pine, we tested several potential evolutionary scenarios based on
225 present-day population structure (involving two differentiated genetic clusters for
226 eastern and western Mediterranean populations, see Results), using the approximate
227 Bayesian procedure implemented in DIYABC (Cornuet et al. 2010).
228 The tested scenarios were as follows (see also Fig. S1, provided as Supplementary
229 Information):
230 *Scenario1 (null hypothesis)*: no population genetic structure and no population size
231 changes;
232 *Scenario 2 (ancient bottleneck)*: no population genetic structure and ancient and
233 severe bottleneck;
234 *Scenario 3 (ancient split + bottleneck 1)*: ancient divergence followed by collapse of
235 both eastern and western Mediterranean genetic clusters;
236 *Scenario 4 (ancient split + bottleneck 2)*: ancient divergence followed by collapse of
237 the eastern genetic cluster;
238 *Scenario 5 (ancient bottleneck + split)*: ancient and severe bottleneck predating the
239 divergence of the two genetic clusters;
240 *Scenario 6 (ancient bottleneck + split + recent bottleneck)*: identical to *Scenario 5* but
241 including a collapse for the eastern genetic cluster after divergence.
242 After fine-tuning prior parameters (all of which had uniform distributions, except for
243 mutation rate, which had a *log*-uniform distribution), a reference table with one million
244 simulated datasets per scenario was built. Then, the 1% of these datasets that were the

closest to the observed data were used to infer the posterior probability of each scenario. Goodness-of-fit was assessed for each scenario by model checking using the Principal Component Analysis (PCA) implemented in DIYABC, which measures the discrepancy between simulated and real data. Confidence in scenario choice (i.e. Type I and II errors) was estimated by simulating 500 datasets under each scenario to determine the probability of not choosing each scenario when it was the true one, or to choosing it when it was not the true scenario. The most likely scenario was finally employed to estimate demographic parameters for the two genetic clusters: time in number of generations, and effective sizes for the current and ancestral populations (see Figs. 2 and S1).

Adaptive evolution and accumulation of genetic load

The proportion of nonsynonymous substitutions fixed by adaptive evolution (α) and the rate of adaptive substitutions scaled by the rate of neutral substitutions (ω_a) were inferred from DNA sequence data using methods available in DOFE 4.0 (http://www.sussex.ac.uk/lifesci/eyre-walkerlab/documents/dofe-31-for-windows.zip). To ease the results interpretation, statistics were estimated and compared between stone and maritime pines by using loblolly pine as an outgroup; only the non-shared part of the divergence (i.e. the fixed exclusive differences for each taxon determined above) from loblolly pine was included in these calculations. Briefly, α was estimated based on Smith and Eyre-Walker (2002), and ω_a was determined using the non-parametric method of Gossman et al. (2010) and method II of Eyre-Walker and Keightley (2009). This last statistic incorporates putative demographic changes that may affect the patterns of nucleotide polymorphism by comparing the observed SFS at neutral sites, with the one expected in a stationary population at equilibrium. This

method was also used to infer the distribution of fitness effects (DFE) for 0-fold degenerate mutations exclusive to each species. For stone pine, DFE was estimated for the whole species and for each genetic cluster, and summarized in three bins of increasing purifying selection ($0<N_e s<1$; $1<N_e s<10$; $N_e s>10$).

In addition, the relative proportion of new deleterious mutations that had accumulated since the divergence of each taxa was approximated in two ways. First, the ratio π_a/π_s (a *proxy* of the efficacy of selection) was estimated, using loblolly pine as outgroup, from the diversity of 0-fold and 4-fold positions of all polymorphic coding sequences with only two variants per polymorphic site. Second, the functional effects of all variants (fixed and polymorphic) exclusive to each stone or maritime pines were inferred using PROVEAN (Choi and Chan 2015). This tool uses alignments from protein blasts against public databases to determine a score for any given site change, by assuming that the rarer a mutation is, the most likely it is to be deleterious. It gives particularly low scores to nonsense replacements and to those resulting in substantial biochemical changes, especially in conserved protein regions (Choi and Chan 2015). Following a similar survey in spruce (Conte et al. 2017), variants with scores below -2.282 were retained as deleterious candidates (see also Zhang et al. 2016 for poplar). Candidate counts per species, per gene and per individual were finally compared with non-parametric tests.

Results

Microsatellite variation and genetic structure

Genetic diversity estimated with nuSSRs was extremely low in stone pine. Four out of the 13 loci surveyed were monomorphic; the nine polymorphic nuSSRs carried between two and five alleles, totalling only 29 size variants (mean = 2.23 alleles per locus). The

mean number of alleles per population ranged between 1.00 and 1.69 (mean = 1.38), and the average number of effective alleles per population between 1.00 and 1.41 (mean = 1.19; Table S1). In addition, allele frequencies were highly homogeneous across populations. Most loci only displayed one frequent allele common to all populations (frequency > 0.8) and a few rare variants scattered across isolated populations with no apparent geographic trend. Ten out of the 29 alleles observed (34.5%) were private; four of these were observed in populations from the Iberian Peninsula, although some were also found in France, Tunisia, Lebanon and Turkey (Fig. 1A); two of these private alleles were in homozygous state. Heterozygosity was also very low, ranging from 0.014 and 0.176 for H_0 (mean = 0.083) and between 0 and 0.221 for H_E (mean = 0.114; Fig. 1A; Table S1). Mean population divergence was high (F_{ST} = 0.242), and the most plausible number of genetic clusters was two (Fig. 2). These clusters were mostly located in the eastern or western parts of the Mediterranean Basin. Increasing *K*-value to three revealed an additional subdivision for the eastern genetic cluster, separating populations from Israel/Lebanon from those of Greece/Turkey; stands from Cyprus were genetically intermixed. *K*-values 4 and 5 revealed no further subdivisions (Fig. S2). For all examined *K*-values, assignment of the Portuguese population and of a Spanish marginal population (Garrovillas, see Fig. 2B) was dubious; they both had a large genetic component from the eastern genetic cluster, suggesting a recent anthropogenic introduction. These populations were removed from further analyses. Similarly weak population genetic structure was observed with the first two coordinates of the PCoA, which explained 44% and 23.7% of the data variance, respectively (Fig. S2).

Species divergence and demography

Only one tree topology was supported by the phylogenetic analyses (posterior probability = 1.0 for all nodes), indicating (as expected) that stone pine is more closely related to maritime pine than to loblolly pine. Divergence between stone pine and its common ancestor with maritime pine dated back to ~29 Ma (95%CI: 35-12Ma), while its separation from the common ancestor with loblolly pine was inferred in ~40Ma (95%CI: 25-52Ma; Fig. 3A).

The demographic history of stone pine, as inferred with STAIRWAYPLOT (on DNA sequence variation), was marked by a deep trend of effective population size decrease starting ~1Ma. This trend was observed for the whole species (Fig. 3C), and for each independent genetic cluster (Fig. S3). For the whole species, a first initial collapse reduced the stone pine population size by about 80%, while two more recent episodes narrowed it down to 1-5% of its ancestral size during the last ~200Ka (Fig. 3C). Population collapse seemed stronger for the eastern genetic cluster than for the western one (Fig. S3). Interestingly, maritime pine suffered a population size decline of a similar magnitude during this period, from which it was able to recover to approximately ancestral population sizes before the decline (Fig. 3B). No population size recovery was inferred for stone pine.

A more detailed exploration of the recent demographic dynamics of stone pine was performed by testing several hypotheses by approximate Bayesian inference on the nuSSR data. The most likely scenario (*scenario 5*, see Materials and Methods; Posterior Probability = 0.8) involved a severe bottleneck in the ancestral population, and a more recent population subdivision between the western and eastern genetic clusters, which currently have highly dissimilar population sizes (much larger in the west; Fig. 3D). These episodes were respectively inferred 5,510 and 238 generations into the past. If we consider a generation time of 25-40 years, as for other pines, the ancestral population

collapse would date back to between 220Ka and 135 Ka, which roughly coincides with the previous interglacial and the second to last glacial period (i.e. the Mindel-Riss interglacial and Riss Glaciation in the Alps), respectively, and with the second population decline inferred from the DNA-sequence data (Fig. 3C). Moreover, population divergence would have occurred during the Holocene (9.5 – 6Ka), roughly coinciding with the earliest fossil evidence of stone pine cultivation (see Discussion).

Nucleotide diversity, adaptive evolution and genetic load

DNA sequence variation was also extremely low for stone pine (even lower than for the nuSSRs, in comparative terms); only 15 of the 177 gene loci were polymorphic (8.5%), and contained 21 variable sites, which represents one SNP every ~2.5 Kbp. For a comparison, 171 of the same gene loci were polymorphic in maritime pine, and harbored 548 SNPs (one SNP every ~102 bp). This translated into a mean nucleotide diversity that was one order of magnitude lower in stone pine (π = 0.000212 [95%CI: 0.000184–0.000246]; θ_W = 0.000149 [95%CI: 0.000126–0.000178]) than in maritime pine (π = 0.002423 [95%CI: 0.002065–0.002778]; θ_W = 0.002305 [95%CI: 0.001993–0.002621]; Table 1). Both π and θ_W estimates were equally low and not significantly different between the two stone pine genetic clusters (Table S2).

Despite this very low nucleotide diversity, the mean rate of total (D_{xy}), synonymous (K_s) and non-synonymous substitutions (K_a) accumulated in stone pine since its divergence from loblolly pine (0.0168 [95%CI: 0.0147–0.0185], 0.0382 [95%CI: 0.0338–0.0426], 0.0099 [95%CI: 0.0073–0.0116]; respectively) were virtually identical to those observed for maritime pine (D_{xy} = 0.0164 [95%CI: 0.0146–0.0182]; K_s = 0.0384 [95%CI: 0.0336–0.0432]; K_a = 0.0101 [95%CI: 0.00836–0.0118]). Decomposing these figures revealed that they were mostly due to the number of shared substitutions

between species (Fig. 4). The non-shared portion was significantly smaller in stone pine than in maritime pine, and it was composed of a relatively similar number of synonymous and non-synonymous substitutions (152 *vs.* 113). In maritime pine, the non-shared variants contained almost twice as many silent than replacement substitutions (325 *vs.* 188; Fig. 4). Such values translated in a larger K_a/K_s ratio (estimated per gene) for stone pine than for maritime pine (0.46 [95%CI: 0.41–0.52] *vs.* 0.33 [95%CI: 0.24–0.40]), suggesting relaxed purifying selection in stone pine (Table 1). The efficacy of selection, as determined with the π_a/π_s ratio, pointed in the same direction, as estimates were again higher in stone pine (0.448 [95%CI: 0.371–0.526]) than in maritime pine (0.256 [95%CI: 0.186–0.326]; Table 1). As for nucleotide diversity, no differences were observed for K_a/K_s and π_a/π_s ratios between stone pine genetic clusters (Table S2).

In terms of adaptive evolution, stone pine also seemed hampered when compared to maritime pine, not only because the proportion of (non-shared) non-synonymous substitutions fixed by adaptive evolution (α) and the relative rate of (non-shared) adaptive substitutions (ω_a) were both significantly lower, but because they were also not significantly different from zero; estimates of α and ω_a were respectively 0.171 (95%CI: -0.027–0.309) and 0.062 (95%CI: -0.001–0.119) for stone pine, while they were 0.494 (95%CI: 0.346–0.609) and 0.304 (95%CI: 0.219–0.391) for maritime pine (Fig. 4). Estimates of α and ω_a were virtually identical for both stone pine genetic clusters (Table S2). On the other hand, the general shape of the distribution of fitness effects (DFE) was similar for both stone and maritime pines. However, the first species had a significantly higher proportion of least deleterious mutations (Fig. 4), while maritime pine had a higher proportion of highly deleterious mutations. Again, no differences were observed between stone pine eastern and western genetic clusters. It must be reminded that all

394 calculations above for stone and maritime pines (i.e. DFE, α, ω_a, K_a/K_s and π_a/π_s) were
395 performed using only exclusive polymorphisms and replacements accumulated after the
396 divergence from loblolly pine.
397 Finally, genetic load, evaluated as the predicted number and frequency of non-shared
398 replacements with putatively deleterious effects according to PROVEAN (Choi and Chan
399 2015), was higher in stone pine than in maritime pine. The first species accumulated
400 relatively less replacements (116) and in less genes (20) than maritime pine (166
401 replacements in 38 genes; Mann-Whitney U-test; $P = 0.0394$ and $P = 0.0142$ for the
402 number of replacements and genes, respectively). However, 34 of the replacements
403 observed in stone pine (29.3% of the total) were putatively deleterious, and most of
404 them (27 out of the 34; or 23.3% of the total substitutions) were fixed in the individuals
405 analysed (Fig. 4). On the other hand, 48 (28.9%) putatively harmful replacements were
406 inferred for maritime pine, while only 15 of them (9% of the total substitutions) were
407 fixed (Fig. 4). Both the total number of putatively deleterious mutations and the number
408 of fixed deleterious mutations were significantly different between the two pines
409 (Mann-Whitney U-test; $P = 0.0271$ and $P = 0.0128$, respectively).
410 Differences in genetic load were more evident when evaluated at the individual level.
411 In stone pine, individuals had on average 0.22 putatively deleterious mutations per
412 gene, while they were significantly less (mean = 0.15) in maritime pine (Mann-Whitney
413 U-test; $P = 0.0184$). After scaling by the number of synonymous mutations per species,
414 the mean number of putatively deleterious mutations per gene per individual was still
415 significantly higher for stone pine (0.042) than for maritime pine (0.019; Mann-Whitney
416 U-test; $P = 0.0036$), indicating relaxed purifying selection in the first species.
417 Interestingly, deleterious variants were not equally distributed across polymorphic
418 and monomorphic genes in stone pine (Table S3), particularly those that were fixed for

significantly more replacement substitutions than the average gene (i.e. five or more replacements; *n*=39). More than half of the polymorphic genes (53%) contained putatively deleterious variants, while only a fifth of the monomorphic genes bearing five or more replacements exhibited such harmful variants (Mann-Whitney U-test; P = 0.0083). These monomorphic genes with reduced fixation of deleterious mutations were further biased towards regulation and stress response processes and have often been reported as candidate genes for adaptation in other pines (see Discussion), while the polymorphic genes showed no particular function enrichment (Table S3).

Discussion

In this study, we demonstrate that the widespread and outbreeding stone pine has unprecedentedly low genome-wide levels of genetic diversity. This near absence of genomic variation is the product of a deep and prolonged demographic collapse, rather than changes in the speed at which it accumulates genetic diversity. This demographic framework further resulted in a lowered efficacy of purifying selection and the fixation of deleterious variants.

Low levels of genome-wide genetic diversity and phylogeographic inferences

In line with previous studies on allozymes (Fallour et al. 1997) and chloroplast microsatellites (cpSSRs) (Vendramin et al. 2008), we found very little genetic variation in both hypervariable nuclear microsatellites (29 alleles in 13 nuSSRs) and gene loci (21 SNPs in 177 gene loci spanning more than 55 Kbp). While estimation accuracy of individual parameters might be at stake because of our relatively small dataset (both in terms of markers and sample size), it is unlikely that such low diversity values are the result of limited population or genome sampling. Indeed, both population genetics and

444 coalescence theory indicates that much of the genetic variability within a species can be
445 captured with a limited number of individuals, while strong demographic changes
446 affecting the whole genome can be detected even when surveying a modest part of the
447 genome (Tajima 1983; Nielsen and Slatkin 2013).

448 In plants, low diversity values are usually observed in self-fertilizing taxa or in species
449 with very restricted distributions, but are unusual in widespread outbreeding taxa
450 (Table 2). Within the genus *Pinus*, and forest trees in general, estimates of genomic
451 diversity are usually much higher than those obtained herein for stone pine (π =
452 0.0002), even for endemic species, such as *P. balfouriana* (π = 0.0028) or *P. longaeva* (π =
453 0.0021; Eckert et al. 2013). Indeed, to our knowledge, the only conifer for which such an
454 equally low nucleotide diversity has been reported is the tertiary relict *Picea breweriana*
455 (π = 0.0001; Chen et al. 2010), which is distributed in small scattered stands at the
456 border between California and Oregon.

457 Previous works have suggested an ancient and prolonged bottleneck to account for
458 the low genetic variation in stone pine, but without providing further insights on times
459 or population sizes (Fallour et al. 1997; Vendramin et al. 2008). Our demographic
460 framework indicated a general and strong trend of population collapse that could have
461 started as early as ~1Ma, reduced the ancestral stone pine population to 1-10% of its
462 original effective size, and continued well within the Holocene (Fig. 3). The number and
463 timing of decline pulses vary according to the type of marker (SSRs vs. gene-loci) and
464 method used (ABC vs. SFS-based), but they all coincide with glacial periods of the
465 Pleistocene (e.g. Günz, Riss, and Würm; Cohen and Gibbard 2011; Fig. 3). While such
466 differences might be the result of our limited sample size, variation in demographic
467 parameters is rather common when using different markers and methods (e.g. Patton et

al. 2019). This deserves a more detailed study, likely including simulations, which is out of the scope of the present work.

In addition to demographic history, some biological features of stone pine, such as its inability to disperse its seeds by wind (contrarily to other Mediterranean pines) or the 3-year cone maturation (all other Mediterranean pines take only two years), may have also contributed to its unusually low genetic diversity. For example, because of restricted gene flow, all mutations that appeared after or during the population decline would have been only locally dispersed, increasing their chances to be lost by genetic drift. Ancient alleles would have suffered a similar fate, being differentially fixed or lost, but never dispersed among distant populations. Indeed, such a feature can also explain the genetic structure of stone pine, which is mostly driven by private nuSSR alleles, whose spatial distribution (together with that of polymorphic SNPs and rare cpDNA haplotypes; see Vendramin et al. 2008) could be mostly the result of restricted gene flow among populations.

According to our Bayesian simulations, the main genetic clusters of stone pine did not diverge until the Holocene (9 – 6.5 Ka; Fig. 3), while none of the main genetic clusters showed signs of expansion. This implies that stone pine should have colonized its current distribution very quickly. However, its low seed dispersal ability hints that a circum-Mediterranean colonization could not have occurred without the support of idiosyncratic dispersers (Vendramin et al. 2008). Evidence indicates that stone pine seeds have been consumed by humans since at least ~50 Ka BP, that it has been cultivated for over 6,000 years, and its nuts traded during no less than three millennia (Prada et al. 1997; Badal 1998; Carrión et al. 2008; Pérez-Jordà et al. 2017). Estimated divergence times between genetic clusters roughly coincide with the beginning of stone pine cultivation, which points at anthropogenic factors as the main drivers of recent

493 expansion, divergence and ecological success (Barbero et al. 1998; Richardson, 1998;
494 Vendramin et al. 2008).
495 Archeological remnants and analyses of ancient trade routes further suggest that the
496 ancestral stone pine population (from where colonization took place) was likely located
497 within the Iberian Peninsula (Carrión et al. 2008; Mutke et al. 2019; Fig. 3.4 from
498 Rubiales et al. 2011); a view supported by the distribution of private nuSSR alleles (Fig.
499 1) and rare cpDNA types (Vendramin et al. 2008), and by the higher effective population
500 sizes estimated herein for the western genetic cluster (Fig. 3D). Unfortunately, the low
501 genetic diversity of this species hampers for the time being more detailed
502 phylogeographic and genomic analyses (including the putative effects of incipient
503 domestication for nut taste and size).
504
505 *Evolutionary rates and adaptive variation*
506 Our estimates of adaptive evolution (α, ω_a, and DFE) pointed to a diminished (adaptive)
507 evolutionary rate in stone pine when compared to maritime pine, as well as a relaxation
508 of purifying selection (Table 1; Fig. 4). These results agree with theoretical expectations,
509 implying less potential for adaptive evolution in species with small effective population
510 size (i.e. stone pine). Although limited by our rather modest genome sampling, π_a/π_s, α
511 and ω_a values, and the DFE, obtained herein for stone pine are not unusual for pines or
512 plants in general (Gossman et al. 2010; Chen et al. 2017; Grivet et al. 2017). The nearly
513 neutral theory of molecular evolution predicts a simple relationship between adaptive
514 evolution and effective population size, which can be derived into a *log–log* linear
515 correlation between π_s and π_a/π_s (Kimura 1968; Welch et al. 2008). Such a relationship
516 was indeed reported for a wide variety of organisms (Chen et al. 2017), and the values
517 obtained herein for stone pine fit well within this regression line (Fig. S4). Thus, current

518 effective population size resulting from the peculiar demography of this species seems
519 to account for both the rate at which it is accumulating genetic diversity, and the global
520 efficacy of selection. Because, the sequenced gene-loci were widespread within the stone
521 pine genome, it is unlikely that sequencing additional genes would reverse the patterns
522 observed herein.

523 It could be argued that other elements, such as variation in mutation rate (μ) and
524 linkage disequilibrium (LD), or spatiotemporal changes of fitness effects, may underlie
525 observed differences between stone and maritime pines. For instance, it has been
526 suggested that demographic fluctuations can affect μ, particularly in species with
527 overlapping generations (Balloux and Lehmann 2012; Wei et al. 2015). Although we did
528 not formally estimate this parameter, no significant differences were observed between
529 the two pines for D_{xy}, K_a or K_s, which suggest that their mutation rates should be at least
530 in the same order of magnitude. Indeed, to our knowledge, no differences in mutation
531 rates have been reported so far among pine taxa, including those from different
532 subsections (Willyard et al. 2009).

533 Linkage disequilibrium can cause selection at one locus to affect the fixation of
534 mutations in neighbouring genomic regions (i.e. interference; Lanfear et al. 2014).
535 Conifers are among the plants with the lowest genome-wide recombination rates
536 (Jaramillo-Correa et al. 2010b), which indicates that interference should be indeed taken
537 into account. Unfortunately, because of its low diversity, no reliable recombination
538 inference could be performed herein for stone pine. However, given that fixations
539 between loci (including both nuSSRs and gene-sequences) appeared independent from
540 each other, and that even modest amounts of recombination can be sufficient to alleviate
541 interference (Presgraves 2005; Weissman and Barton 2012), the contribution of LD
542 variation to explain observed differences across stone and maritime pines should be

543 negligible. Future studies with more complete genome data are however necessary to
544 test this hypothesis more formally.
545 Selective pressures may also vary across a species range and over time, having
546 heterogeneous effects on substitutions rates, and thus on evolutionary rate (Bell 2010).
547 However, if our demographic and phylogeographic inferences are correct, such changes
548 should be affecting more maritime pine than stone pine; given that the first species is
549 composed by six ancient genetic clusters distributed across contrasting environments
550 (Bucci et al. 2007; Jaramillo-Correa et al. 2015b), while population structure in stone
551 pine is less marked and more recent (Fig. 3). While there is still little information on how
552 spatiotemporal variation of fitness affects evolutionary rate (Lanfear et al. 2014), no
553 differences in the proportion of adaptive substitutions were observed across the more
554 ancient maritime pine gene pools (Grivet et al. 2017). Thus, suggesting again that
555 differences in demography appear as the most likely explanation for the contrasting
556 rates at which these two parapatric pines are accumulating genetic diversity.
557 Genomes of heavily bottlenecked species usually exhibit overall genetic depletion,
558 with isolated and randomly distributed peaks of diversity (Robinson et al. 2016). Stone
559 pine seems to fit such expectations. The few SNPs detected were distributed in only 15
560 genes that were not particularly biased towards specific biological processes. They
561 participated in functions as varied as transport, ATP binding and secondary metabolism,
562 and are scattered in various pine linkage groups (Westbrock et al. 2015; Plomion et al.
563 2016). More than half of these genes (53%) contained putatively deleterious variants
564 (either fixed or polymorphic; Table S3).
565 Interestingly, among the monomorphic genes, those that were fixed for significantly
566 more replacement substitutions than the average gene (i.e. five or more replacements;
567 *n*=39) tended to have fewer deleterious variants. These genes were further enriched

568 towards regulation and stress response processes, and most of them (60%) have been
569 reported as candidate genes for local adaptation in other pines and/or are upregulated
570 under drought stress in stone pine (Perdiguero et al. 2013; Table S3). Although these
571 observations are preliminary, given the modest gene sampling of our study and the lack
572 of current knowledge on gene function in conifers, they do provide an interesting
573 starting point to understand how adaptive variation arises and reaches fixation in these
574 taxa.
575
576 *Long-term persistence in the absence of genetic variation*
577 An interesting question that arises from our study is how a species genetically depleted
578 has been able to survive, thrive, and successfully compete with taxa that occupy a
579 similar ecological niche. Extrapolating from the limited number of genes surveyed,
580 maritime pine bore significantly more silent and replacement substitutions and more
581 putatively harmful mutations than stone pine. The large majority of these changes
582 (91%) were in a heterozygous state in maritime pine, whereas most of the non-
583 synonymous substitutions detected for stone pine, including many evaluated as
584 deleterious, were homozygous (Fig. 4). These results fit the expectations of a differential
585 accumulation of neutral and putatively deleterious mutations within a species after a
586 bottleneck (Gillespie 2001; Lanfear et al. 2014; Chen et al. 2017), and it is unlikely that
587 this trend could be reversed when additional gene-regions are sequenced.
588 The accumulation of partially deleterious mutations in stone pine also points to
589 reduced inbreeding depression. In forest trees, this phenomenon generally results in the
590 elimination of virtually all inbred progeny at early life-stages (Petit and Hampe 2006),
591 which is believed to be homozygous for recessive deleterious mutations. When
592 populations are large (as for maritime pine), deleterious mutations tend to accumulate

593 in heterozygous states. However, after a bottleneck (as for stone pine), these mutations
594 are exposed and can be purged more efficiently, which leads to inbreeding tolerance (i.e.
595 inbreds are no longer selected against), and to a more accelerated diversity loss by
596 genetic drift (Lanfear et al. 2014). Early work in stone pine and other genetically
597 depleted conifers seem to point in that direction (Ammannati 1989; Ledig et al. 2000).
598 It must be noted, however, that genetic load in stone pine is likely underestimated
599 because of our limited genome sampling, and because virtually all mutations evaluated
600 as deleterious came from bioinformatics predictions. That is, they were either very rare
601 within conifer databases or produced biochemical changes in the proteins. No strongly
602 harmful mutations, like premature stop codons or insertions changing the protein
603 reading frames, were observed in stone pine. While such types of mutations might be
604 detected with a more in-depth genome survey, they are also expected to be more
605 common in species with lower inbreeding depression than forest trees, such as annual
606 plants or vertebrates, including humans (Robinson et al. 2016; González-Martínez et al.
607 2017; Pedersen et al. 2017; Laenen et al. 2018). Should this trend be confirmed after a
608 more extensive exome re-sequencing, it will indicate that purifying selection is still
609 operating for highly-deleterious variants in stone pine, as could be inferred from the
610 DFE (Fig. 4).
611 Another mechanism fostering stone pine long-term persistence despite depleted
612 genetic variation might be the presence of islands of genetic diversity within its genome,
613 and which may be associated to relevant adaptive quantitative genetic variation and/or
614 phenotypic plasticity (Carrasquinho and Gonçalves 2013; Mutke et al. 2013; Sánchez-
615 Gómez et al. 2011). Typically, this type of variation is produced by small allele frequency
616 changes in multiple loci across the genome (Pritchard et al. 2010; Boyle et al. 2017;
617 Csilléry et al. 2018); variation that is better preserved during bottlenecks by balancing

or frequency-dependent selection than newly arisen mutations (Lynch 1996). If this is the case for stone pine, a more detailed genomic survey (including full genome or transcriptome sequencing) should reveal further diversity-rich regions upon which selection may operate. Another possibility is methylation diversity, which has already been shown to be substantial in stone pine (Sáez-Laguna et al. 2014). Such variability is often inheritable, and can rapidly accumulate after bottlenecks and contribute to phenotypic plasticity in adaptive traits (Chinnusamy and Zhu 2009; Johaness et al. 2009; Balao et al. 2018).

Finally, future studies should formally survey the effects on fitness (e.g. via association with adaptive phenotypes, as in Zhang et al. 2016) of the putatively deleterious variants inferred in this study; given that these are merely bioinformatics predictions (Choi and Chan 2015; Conte et al. 2017). Integrating such effects, and genetic load, into predictive models should become a priority for the improvement of commercially important traits in stone pine, such as those related to nut production (see Wallace et al. 2018). Considering deleterious variants is also important for conservation and management programs in forest trees (Holliday et al. 2017). Indeed, avoiding deleterious alleles, especially in homozygosity, might account for a larger part of the phenotypic variance of species than the avidly-searched adaptive genes usually reported in most population genomic studies.

Acknowledgements

We are grateful to Diana Barba and Denis Vauthier for sampling, and Carmen García-Barriga and Mario Zabal-Aguirre for laboratory assistance and DNA sequence edition. We further thank Yoshiaki Tsuda for help with ABC analyses and Michele Bozzano for stimulating and passionate discussions on stone pine demographic history. Samples

were collected from common gardens in France and Spain, installed within the FAO-Silva Mediterranea collaborative framework. French common gardens belong to the GEN4X network (http://www.efpa.inra.fr/Outils-et-Ressources/Systemes-d-experimentation-et-d-observation/Reseau-GEN4X/). This research was supported by funding from EVOLTREE (www.evoltree.eu), a former EU Network of Excellence, and currently a European Research Group facilitated by the European Forest Institute (EFI), and B4EST, a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 773383. JPJ-C was supported by the "Dirección General de Asuntos del Personal Académico" (DGAPA-UNAM) through a PASPA fellowship, for a sabbatical leave during which this study was elaborated.

Data accessibility

Nuclear microsatellite (nuSSR) data, DNA sequence alignments, gene annotations, and site type counts are available in Dryad, with doi: 10.5061/dryad.59zw3r23r.

Author contributions

JPJC, GGV and SCGM conceived the study. GGV, BF, SCGM, DG and FAA produced the data. DG and FB contributed to data analyses. JPJC and SCGM analysed the data and drafted the manuscript. All authors revised and edited the text, and produced the final manuscript.

References

Ammannati, R. (1988). Effeti dell'autoimpollinazione sulla crescita in alteza in *Pinus pinea* L. *Monti e Boschi*, *3*, 50–52.

667 Badal, E. (1998). *Interés económico del pino piñonero para los habitantes de la Cueva de*
668 *Nerja*. In J.L. Sanchidrián, & M.D. Simón (Eds.), *Las culturas del Pleistoceno Superior en*
669 *Andalucía* (pp. 287-300). Málaga: Patronato de la Cueva de Nerja.
670 Balao, F., Paun, O., & Alonso, C. (2018). Uncovering the contribution of epigenetics to
671 plant phenotypic variation in Mediterranean ecosystems. *Plant Biology*, *20*, 38-49.
672 Balloux, F., & Lehmann, L. (2012). Substitution rates at neutral genes depend on
673 population size under fluctuating demography and overlapping generations.
674 *Evolution*, *66*, 605–611.
675 Barbéro, M., Loisel, R., Quézel, P., Richardson, D.M., & Romane, F. (1998). *Pines of the*
676 *Mediterranean basin*. In D.M. Richardson (Ed.), *Ecology and biogeography of Pinus* (pp.
677 153–170). Cambridge: Cambridge University Press.
678 Bazile-Robert, E. (1981). Le pin pignon (*Pinus pinea* L.) dans le Würm récent de
679 Provence. *Géobios*, *14*, 395–397.
680 Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable
681 environments. *Philosophical Transactions of the Royal Society of London, Series B:*
682 *Biological Sciences*, *365*, 87–97.
683 Böndel, K.B., Lainer, H., Nosenko, T., Mboup, M., Tellier A., & Stephan, W. (2015). North–
684 South colonization associated with local adaptation of the wild tomato species
685 *Solanum chilense*. *Molecular Biology and Evolution*, *32*, 2932-2943.
686 Boraks, A., & Broders, K.D. (2016). Population genetic diversity of the rare hardwood
687 butternut (*Juglans cinerea*) in the northeastern USA. *Tree Genetics & Genomes*, *12*, 43.
688 Boyle, E.A., Li, Y.I., & Pritchard, J.K. (2017) An expanded view of complex traits: from
689 polygenic to omnigenic. *Cell*, *169*, 1177-1186.
690 Boys, J., Cherry, M., & Dayanandan, S. (2005). Microsatellite analyses reveals genetically
691 distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of*
692 *Botany*, *92*, 833-841.
693 Bromham, L., Hua, X., Lanfear, R., & Cowman, P.F. (2015). Exploring the relationships
694 between mutation rates, life history, genome size, environment, and species richness
695 in flowering plants. *The American Naturalist*, *185*, 507-524.
696 Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H., & Neale, D.B. (2004). Nucleotide
697 diversity and linkage disequilibrium in loblolly pine. *Proceeding of the National*
698 *Academy of Sciences of USA*, *101*, 15255–15260.

699 Bucci, G., González-Martínez, S.C., Le Provost, G., Plomion, C., Ribeiro, M.M., Sebastiani, F.,
700 Alia, R.,& Vendramin, G.G. (2007). Range-wide phylogeography and gene zones in
701 *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology*,
702 *16*, 2137–2153.
703 Carrasquinho, I., & Gonçalves, E. (2013). Genetic variability among *Pinus pinea* L.
704 provenances for survival and growth traits in Portugal. *Tree Genetics & Genomes*, *9*,
705 855–866.
706 Carrión, J.S., Finlayson, C., Fernández, S., Finlayson, G., Allué, E., López Sáez, J.A., López
707 García, P., Gil-Romera, G., Bailey G., & González-Sampériz, P. (2008). A coastal
708 reservoir of biodiversity for Upper Pleistocene human populations: palaeoecological
709 investigations in Gorham's Cave (Gibraltar) in the context of the Iberian Peninsula.
710 *Quaternary Sciences Reviews*, *27*, 2118-2135.
711 Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size
712 and patterns of molecular evolution and variation. *Nature Reviews, Genetics*, *10*, 195–
713 205.
714 Charlesworth, B., Morgan, M.T., & Charlesworth, D. (1993). The effect of deleterious
715 mutations on neutral molecular variation. *Genetics*, *134*, 1289-1303.
716 Charlesworth, D., Charlesworth, B., & Morgan, M.T. (1995). The pattern of neutral
717 molecular variation under the background selection model. *Genetics*, *141*, 1619–1632.
718 Charlesworth, J., & Eyre-Walker, A. (2007). The other side of the nearly neutral theory,
719 evidence of slightly advantageous backmutations. *Proceedings of the National*
720 *Academy of Sciences of USA*, *104*, 16992–16997.
721 Chen, J., Källman, T., Gyllenstrand, N., & Lascoux, M. (2010). New insights on the
722 speciation history and nucleotide diversity of three boreal spruce species and a
723 Tertiary relict. *Heredity*, *104*, 3–14.
724 Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic diversity and the efficacy of purifying
725 selection across plant and animal species. *Molecular Biology and Evolution*, *34*, 1417–
726 1428.
727 Chinnusamy, V., & Zhu, J.K. (2009). Epigenetic regulation of stress responses in plants.
728 *Current Opinion in Plant Biology*, *12*, 133–139.
729 Choi, Y., & Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect
730 of amino acid substitutions and indels. *Bioinformatics*, *31*, 2745-2747.

731 Cohen, K.M., & Gibbard, P. (2011). *Global chronostratigraphical correlation table for the*
732 *last 2.7 million years*. Cambridge: Subcommission on Quaternary Stratigraphy
733 (International Commission on Stratigraphy). Retrieved from
734 http://quaternary.stratigraphy.org/charts/
735 Conte, G.L., Hodgins, K.A., Yeaman, S., Degner, J.C., Aitken, S.N., Rieseberg, L.H., &
736 Whitlock, M.C. (2017). Bioinformatically predicted deleterious mutations reveal
737 complementation in the interior spruce hybrid complex. *BMC Genomics*, *18*, 970.
738 Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., … Estoup, A.
739 (2014). DIYABC v2.0: A software to make approximate Bayesian computation
740 inferences about population history using single nucleotide polymorphism, DNA
741 sequence and microsatellite data. *Bioinformatics*, *30*, 1187– 1189.
742 Csilléry, K., Rodríguez-Verdugo, A., Rellstab, C., & Guillaume F. (2018). Detecting the
743 genomic signal of polygenic adaptation and the role of epistasis in evolution.
744 *Molecular Ecology*, *27*, 606-612.
745 Cuenca, A. (2003) *Evidencia de dos linajes genéticos en Pinus cembroides revelada por*
746 *microsatélites de cloroplasto* (Unpublished MSc thesis). Universidad Nacional
747 Autónoma de México.
748 Drummond, A.J., Suchard, M.A., Xie, D., & Rambaut, A. (2012). Bayesian Phylogenetics
749 with BEAUti and the BEAST 1.7, *Molecular Biology and Evolution*, *29*, 1969–1973.
750 Earl, D.A., & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program
751 for visualizing STRUCTURE output and implementing the Evanno method.
752 *Conservation Genetics Resources*, *4*, 359-361.
753 Echt, C.S, De Verno, L.L., Anzidei, M., & Vendramin, G. G. (1998). Chloroplast
754 microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait.
755 *Molecular Ecology*, *7*, 307-316.
756 Eckert, A. J., Bower, A. D., Jermstad, K. D., Wegrzyn, J. L., Knaus, B. J., Syring, J. V., & Neale,
757 D. B. (2013). Multilocus analyses reveal little evidence for lineage-wide adaptive
758 evolution within major clades of soft pines (*Pinus subgenus* Strobus). *Molecular*
759 *Ecology*, *22*, 5635–5650.
760 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of
761 individuals using the software structure: a simulation study. *Molecular Ecology*, *14*,
762 2611–2620

763 Eyre-Walker, A., & Keightley, P.D. (2007). The distribution of fitness effects of new
764 mutations. *Nature Reviews, Genetics*, *8*, 610-618.
765 Eyre-Walker, A., & Keightley, P.D. (2009). Estimating the rate of adaptive molecular
766 evolution in the presence of slightly deleterious mutations and population size
767 change. *Molecular Biology and Evolution*, *26*, 2097–2108.
768 Fallour, D., Fady, B., & Lefèvre, F. (1997). Study on isozyme variation in *Pinus pinea* L.:
769 evidence for low polymorphism. *Silvae Genetica*, *46*, 201–207.
770 Gaut, B.S., Díez, C.M, & Morrell, P.L. (2015). Genomics and the contrasting dynamics of
771 annual and perennial domestication. *Trends in Genetics*, *31*, 709-719.
772 Gernandt, D.S., López, G.G., García, S.O. & Liston, A. (2005). Phylogeny and classification
773 of *Pinus. Taxon*, *54*, 29-42.
774 Gillespie, J. H. (2001). Is the population size of a species relevant to its evolution?
775 *Evolution*, *55*, 2161–2169.
776 González-Martínez, S.C., Ridout, K., & Pannell, J.R. (2017). Range expansion compromises
777 adaptive evolution in an outcrossing plant. *Current Biology*, *27*, 2544-2551.
778 Gossmann, T.I., Song, B.-H., Windsor, A.J., Mitchell-Olds, T., Dixon, C.J., Kapralov, M.V., …
779 Eyre-Walker, A. (2010). Genome wide analyses reveal little evidence for adaptive
780 evolution in many plant species. *Molecular Biology and Evolution*, *27*, 1822–1832.
781 Goudet J. (2002). FSTAT, A program to estimate and test gene diversities and fixation
782 indices. Retrieved from http://www2.unil.ch/popgen/softwares/fstat.htm
783 Grivet, D., Avia, K., Vaattovaara, A., Eckert, A.J., Neale, D. B., Savolainen, O., & González-
784 Martínez, S.C. (2017). High rate of adaptive evolution in two widespread European
785 pines. *Molecular Ecology*, *26*, 6857–6870.
786 Hardy, K., Radini, A., Buckley, S., Sarig, R., Copeland, L., Gopher, A., & Barkai, R. (2016).
787 Dental calculus reveals potential respiratory irritants and ingestion of essential plant-
788 based nutrients at Lower Palaeolithic Qesem Cave Israel. *Quaternary International*,
789 *398*, 129-135.
790 Holliday, J.A., Aitken, S.N., Cooke, J.E.K., Fady, B., González-Martínez, S.C., Heuertz, M.,
791 Jaramillo-Correa, J.P., Lexer, C., Staton, M., Whetten, R.W., & Plomion, C. (2017).
792 Advances in ecological genomics in forest trees and applications to genetic resources
793 conservation and breeding. *Molecular Ecology*, *26*, 706-717.

794 Jakobsson, M., & Rosenberg, N.A. (2007). CLUMPP: a software matching and permutation
795 program for dealing with label switching and multimodality in analysis of population
796 structure. *Bioinformatics*, *23*, 1801–1806.
797 Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorell, J.C., Cullingham, C.I., &
798 Andrew, R.L. (2017). The *K*=2 conundrum. *Molecular Ecology*, *26*, 3594-3602.
799 Jaramillo-Correa, J.P., Gérardi, S., Beaulieu, J., Ledig, F.T., & Bousquet, J. (2015a). Inferring
800 and outlining past population declines with linked microsatellites: a case study in two
801 spruce species. *Tree Genetics & Genomes*, *11*, 9.
802 Jaramillo-Correa, J.P., Rodríguez-Quilón, I., Grivet, D., Lepoittevin, C., Sebastiani, F.,
803 Heuertz, M., Garnier-Géré, P., Alía, R., Plomion, C., Vendramin, G.G., & González-
804 Martínez, S.C. (2015b). Molecular proxies of climate maladaptation in a long-lived
805 tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*, *199*, 793-807.
806 Jaramillo-Correa, J.P., Grivet, D., Terrab, A., Kurt, Y., de Lucas, A.I., Wahid, N., Vendramin,
807 G.G., & González-Martínez, S.C. (2010a). The Strait of Gibraltar as a major
808 biogeographic barrier in Mediterranean conifers: a comparative phylogeographic
809 survey. *Molecular Ecology*, *19*, 5452-5468.
810 Jaramillo-Correa, J.P., Verdú, M., & González-Martínez, S.C. (2010b). The contribution of
811 recombination to heterozygosity differs among plant evolutionary lineages and life-
812 forms. *BMC Evolutionary Biology*, *10*, 22.
813 Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon M, Agier, N., … Colot,
814 V. (2009). Assessing the impact of transgenerational epigenetic variation on complex
815 traits. *PLoS Genetics*, *5*, e1000530.
816 Kern, A.D., & Hahn M.W. (2018). The Neutral Theory in Light of Natural Selection.
817 *Molecular Biology and Evolution*, *35*, 1366-1371.
818 Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, *217*, 624-626.
819 Laenen, B., Tedder, A., Nowak, M.D., Toräng, P., Wunder, J., Wötsel, S., … Slotte, T. (2018).
820 Demography and mating system shape the genome-wide impact of purifying selection
821 in *Arabis alpina*. *Proceedings of the National Academy of Sciences of USA*, *115*, 816-821.
822 Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of
823 evolution. *Trends in Ecology and Evolution*, *29*, 33–41.
824 Ledig, F.T, Bermejo-Velázquez, B., Hodgskiss, P.D., Johnson, D.R., Flores-López, C., &
825 Jacob-Cervantes, V. (2000). The mating system and genic diversity in Martínez spruce,
826 an extremely rare endemic of México's Sierra Madre Oriental: an example of

827 facultative selfing and survival in interglacial refugia. *Canadian Journal of Forest*
828 *Research*, *30*, 1156-1164.
829 Lemieux, M.J., Beaulieu, J., & Bousquet, J. (2011). Chloroplast DNA polymorphisms in
830 eastern hemlock: range-wide genogeographic analyses and implications for gene
831 conservation. *Canadian Journal of Forest Research*, *41*, 1047-1059.
832 Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency
833 spectra. *Nature Genetics*, *47*, 555–559.
834 Lynch, M. (1996). *A quantitative-genetic perspective on conservation issues*. In J. C. Avise
835 &J. L. Hamrick (Eds). *Conservation genetics: case histories from nature* (pp. 471–501).
836 New York: Chapman and Hall.
837 Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., …
838 Lohmuller, K.E. (2016). Bottlenecks and selective sweeps during domestication have
839 increased deleterious genetic variation in dogs. *Proceedings of the National Academy*
840 *of Sciences of USA*, *133*, 152–157.
841 Moyers, B.T., Morrell, P.L., & McKay, J.K. (2018). Genetic costs of domestication and
842 improvement. *Journal of Heredity*, *109*, 103–116.
843 Mutke, S., Gordo, J., Khouja, M., & Fady, B. (2013). Low genetic and high environmental
844 diversity at adaptive traits in *Pinus pinea* from provenance tests in France and Spain.
845 *Options Méditerranées*, *105*, 73–79.
846 Mutke, S., Vendramin, G.G., Fady, B., Bagnoli, F., González-Martínez, S.C. 2019. *Molecular*
847 *and quantitative genetics of stone pine (*Pinus pinea *L.)*. In D. Nandwani (Ed.). *Genetic*
848 *diversity in horticultural plants* (pp. 61–84). Switzerland: Springer Nature.
849 Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of
850 restriction endonucleases. *Proceedings of the National Academy of Sciences of USA*, *76*,
851 5269–5273.
852 Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of*
853 *Ecology and Systematics*, *23*, 263–286.
854 Peakall, R., & Smouse, P.E. (2006). Genalex 6: genetic analysis in Excel. Population
855 genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288-295.
856 Pedersen, C.-E. T., Lohmueller, K.E., Grarup, N., Bjerregaard, P., Hansen, T., Siegismund,
857 H.R., Moltke, I., & Albrechtsen, A. (2017). The effect of an extreme and prolonged
858 population bottleneck on patterns of deleterious variation: insights from the
859 Greenlandic Inuit. *Genetics*, *205*, 787–801.

860 Perdiguero, P., Barbero M.C., Cervera, M.T., Collada, C., & Soto, A. (2013). Molecular
861 response to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and
862 *Pinus pinea*). *Plant Physiology and Biochemistry*, *67*, 199-208.
863 Pérez-Jordà, G., Peña-Chocarro, L., García Fernández, M., & Vera-Rodríguez, J.C. (2017).
864 The beginnings of fruit tree cultivation in the Iberian Peninsula: plant remains from
865 the city of Huelva (southern Spain). *Vegetation History and Archaeobotany*, *26*, 527-
866 538.
867 Petit, R.J., & Hampe A. (2006). Some evolutionary consequences of being a tree. *Annual*
868 *Review of Ecology, Evolution, and Systematic*s, *37*, 187-214.
869 Pinzauti, F., Sebastiani, F., Budde, K.B., Fady, B., González-Martínez, S.C. & Vendramin,
870 G.G. (2012). Nuclear microsatellites for *Pinus pinea* (Pinaceae), a genetically
871 depauperate tree, and their transferability to *P. halepensis*. *American Journal of*
872 *Botany*, *99*, e362-e365.
873 Plomion, C., Bartholomé, J., Lesur, I., Boury, C., Rodríguez-Quilón, I., Lagraulet, H., …
874 González-Martínez, S.C. (2016). High-density SNP assay development for genetic
875 analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources*, *16*, 574-587.
876 Popova, T., & Hristova, H. (2018). Trees of eternity-*Pinus pinea* L. in daily life, rituals,
877 religion and symbolism. Archaeobotanical evidence from the territory of Bulgaria.
878 *Journal of Archaeological Science: Reports*, *19*, 987-991.
879 Potter, K.M., Jetton, R.M., Dvorak, W.S., Hipkins, V.D., Rhea, R., & Whittier, W.A. (2012).
880 Widespread inbreeding and unexpected geographic patterns of genetic variation in
881 eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer.
882 *Conservation Genetics*, *13*, 475-498.
883 Prada, M.A., Gordo, J., de Miguel, J., Mutke, S., Catalán-Bachiller, G., Iglesias, S., & Gil, L.
884 (1997). *Las regiones de procedencia de Pinus pinea L. en España*. Madrid: Ministerio de
885 Medio Ambiente, Organismo Autónomo Parques Nacionales.
886 Presgraves, D.C. (2005). Recombination enhances protein adaptation in *Drosophila*
887 *melanogaster*. *Current Biology*, *15*, 1651–1656.
888 Pritchard, J.K., Stephens, M., & Donnelly, P. (2000) Inference of population structure
889 using multilocus genotype data. *Genetics*, *155*, 945–959.
890 Pritchard, J.K., Pickrell, J.K., & Coop, G. (2010). The genetics of human adaptation: hard
891 sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208-R215.

892 Richardson, D.M. (ed.). (1998). *Ecology and Biogeography of Pinus*. Cambridge:
893 Cambridge University Press.
894 Robinson, J.A, Ortega-Del-Vecchyo D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D.,
895 Lohmuller, K.E., & Wayne, R.K. (2016). Genomic flatlining in the endangered island
896 fox. *Current Biology*, *26*, 1183-1189.
897 Ross-Davis, A., Ostry, M., & Woeste, K.E. (2008). Genetic diversity of butternut (*Juglans*
898 *cinerea*) and implications for conservation. *Canadian Journal of Forest Research*, *38*,
899 899-907.
900 Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-
901 Onsins, S.E, & Sánchez-Gracia, A. (2017). DnaSP v6: DNA sequence polymorphism
902 analysis of large datasets. *Molecular Biology and Evolution*, *34*, 3299-3302.
903 Rubiales, J.M., García-Álvarez, S., García-Amorena, I., Hernández, L., Morales-Molino, C.,
904 Moreno, E., & Gómez-Manzaneque, F. (2011) Palaeobiogeographical perspectives on
905 *Pinus pi*nea L.: a controversial and enigmatic Mediterranean pine. 5th Biennial
906 Conference of the International Biogeography Society, Irakleion, Greece.
907 Sáez-Laguna, E., Guevara, M.-Á., Díaz, L.-M., Sánchez-Gómez, D., Collada, C., Aranda, I., &
908 Cervera, M.T. (2014). Epigenetic variability in the genetically uniform forest tree
909 species *Pinus pinea* L. *PLoS ONE*, *9*, e103145.
910 Saladin, B., Leslie, A.B., Wüest R.O., Litsios, G., Conti, E., Salamin, N. & Zimmermann, N.E.
911 (2017). Fossils matter: improved estimates of divergence times in *Pinus* reveal older
912 diversification. *BMC Evolutionary Biology*, *17*, 95.
913 Sánchez-Gómez, D., Velasco-Conde, T., Cano-Martín, F.J., Ángeles-Guevara, M.A., Cervera,
914 M.T., & Aranda, I. (2011). Inter-clonal variation in functional traits in response to
915 drought for a genetically homogeneous Mediterranean conifer. *Environmental and*
916 *Experimental Botany*, *70*, 104–109.
917 Smith, N. G., & Eyre-Walker, A. (2002). Adaptive protein evolution in Drosophila. *Nature*,
918 *415*, 1022–1024.
919 Tajima, F. (1983). 3. Evolutionary relationship of DNA sequences in finite populations.
920 *Genetics*, *105*, 437-460.
921 Vendramin, G.G., Fady, B., González-Martínez, S.C, Hu, F.-S., Scotti, I., Sebastiani, F., Soto,
922 A., & Petit, R.J. (2008). Genetically depauperate but widespread: the case of an
923 emblematic Mediterranean pine. *Evolution*, *62*, 680–688.

Wallace, J.G., Rodgers-Melnick, E., & Buckler, E.S. (2018). On the road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of crop quantitative genomics. *Annual Review of Genetics, 52,* 421–444.

Walter, R., & Epperson, B.K. (2005). Geographic pattern of genetic diversity in Pinus resinosa: contact zone between descendants of glacial refugia. *American Journal of Botany*, *92*, 92-100.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276.

Wei, X., Zhao, L., Lascoux, M., & Waxman, D. (2015). Population structure and the rate of evolution. *Journal of Theoretical Biology*, *365*, 486–495.

Weissman, D.B. & Barton, N.H. (2012). Limits to the rate of adaptive substitution in sexual populations. *PLoS Genetics*, *8*, e1002740.

Welch, J.J., Eyre-Walker, A., & Waxman, D. (2008). Divergence and polymorphism under the nearly neutral theory of molecular evolution. *Journal of Molecular Evolution*, *67*, 418–426.

Westbrook, J.W., Chhatre, V.E., Wu, L.-S., Chamala, S., Gomide-Neves, L. Muñoz, P., ... Echt, C.S. (2015). A consensus genetic map for *Pinus taeda* and *Pinus elliottii* and extent of linkage disequilibrium in two genotype-phenotype discovery populations of *Pinus taeda*. *G3: Genes, Genomes, Genetics*, *5*, 1685-1694.

Wilcox, G.H. (1977). Exotic plants from Roman waterlogged sites in London. *Journal of Archaeological Science*, *4*, 269-282.

Willyard, A., Cronn, R., & Liston, A. (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution*, *52*, 498–511.

Zhang, M., Zhou, L., Bawa, R., Suren, H., & Holliday, J.A. (2016). Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar. *Molecular Biology and Evolution*, *33*, 2899–2910.

Legends to Figures

FIGURE 1 Stone pine (*P. pinea*) distribution (green areas) and location of populations sampled for nuSSR (circles) and DNA sequence analyses (triangles). Circle color indicates the number of private alleles (A_{PT}) found for nuSSRs.

FIGURE 2 (**A**) Geographic distribution of mean genetic-cluster membership for stone pine (*P. pinea*) populations, as obtained from STRUCTURE analysis of nuSSR variation when assuming two genetic clusters ($K = 2$). (**B**) Individual bar-plot of genetic-cluster membership coefficients for the same analysis. (**C**) Distribution of ΔK values for the various numbers of genetic clusters (K) assumed in STRUCTURE analyses. See Table S1 for population codes and additional details.

FIGURE 3 Genetic divergence and demographic history of stone pine (*P. pinea*). (**A**) Phylogenetic relationship and divergence times from maritime (*P. pinaster*) and loblolly (*P. taeda*) pines as determined from the Bayesian analyses of 50 nuclear gene-loci. Numbers and shaded rectangles over branches represent mean and 95% credible intervals (CI) for divergence time estimates (in Ma), respectively. All nodes had posterior probabilities of 1.0. (**B, C**) Changes of population size over time for maritime (**B**) and stone (**C**) pines inferred with a stairway-plot derived from the site frequency spectrum (SFS) of 55,833 DNA-sequence sites. Dark and light lines represent the median and 95% CI, respectively. Estimates based on the assumption of a mutation rate of 1.64×10^{-9} per site per year, and a generation time of 25 years. Blue areas denote glacial periods; those mentioned in the text are indicated on top of **B**. (**D**) Schematic representation of the best demographic scenario for stone pine according to DIYABC

976 (based on the variation of 13 nuSSRs), and population parameters estimated from this
977 scenario (see Fig. S1 for all scenarios).
978
979 **FIGURE 4** (**A**) Shared and non-shared (exclusive) mean number of SNPs found in 177
980 gene-loci in stone (*P. pinea*) and maritime (*P. pinaster*) pines after their divergence from
981 loblolly pine (*P. taeda*). Exclusive SNPs are further subdivided in synonymous and non-
982 synonymous (replacements); non-synonymous variants are additionally separated in
983 putatively neutral and deleterious, according to functional effects inferred with PROVEAN
984 (Choi and Chan 2015). The mean number of fixed deleterious SNPs per species is also
985 indicated. (**B**) Proportion of (non-shared) non-synonymous substitutions fixed by
986 adaptive evolution (α) and relative rate of (non-shared) adaptive substitutions (ω_a) per
987 species. (**C**) Distribution of fitness effects (DFE) of exclusive mutations in each taxa.
988

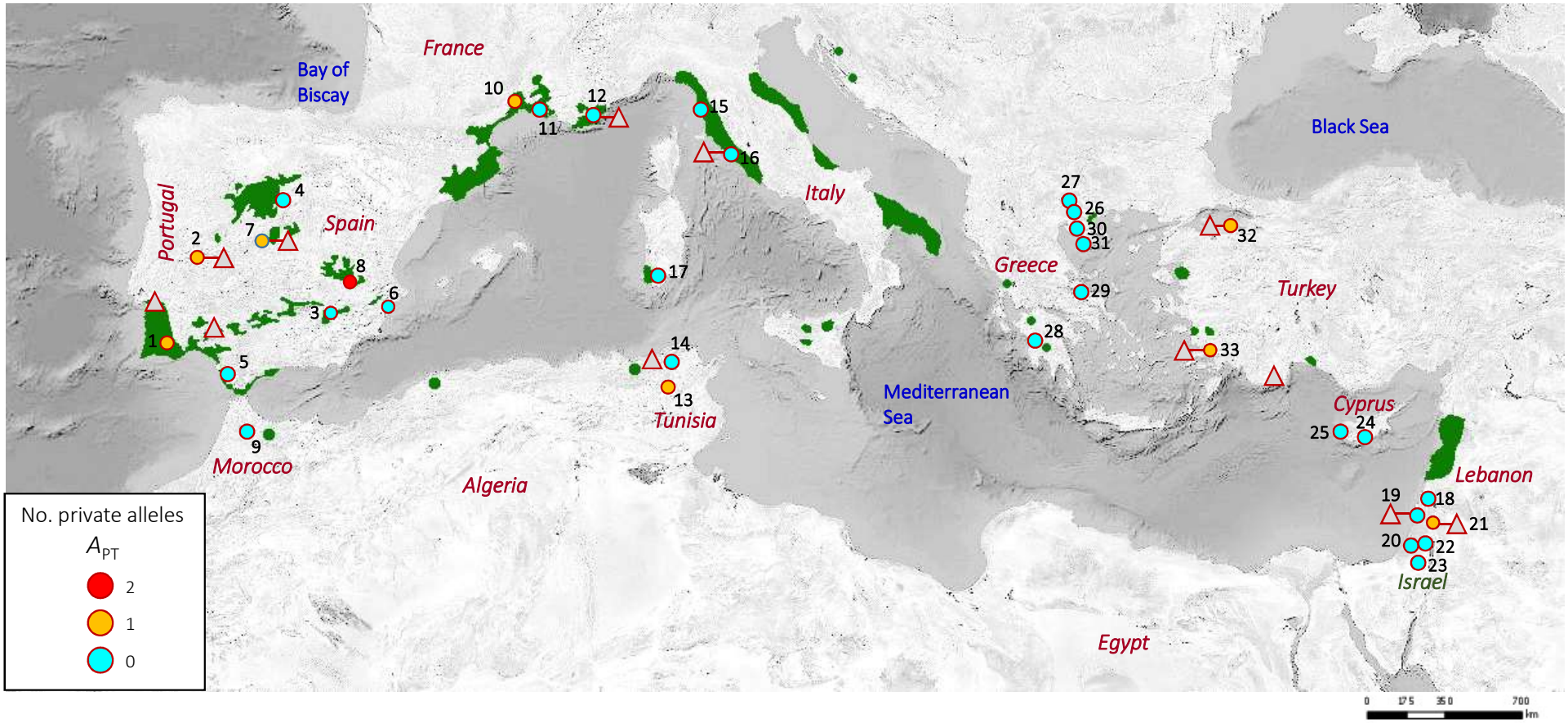
TABLE 1 Mean DNA sequence diversity and divergence from loblolly pine (*P. taeda*) estimated from 177 common gene-loci (representing 55,833 bp) in stone (*P. pinea*) and maritime (*P. pinaster*) pines

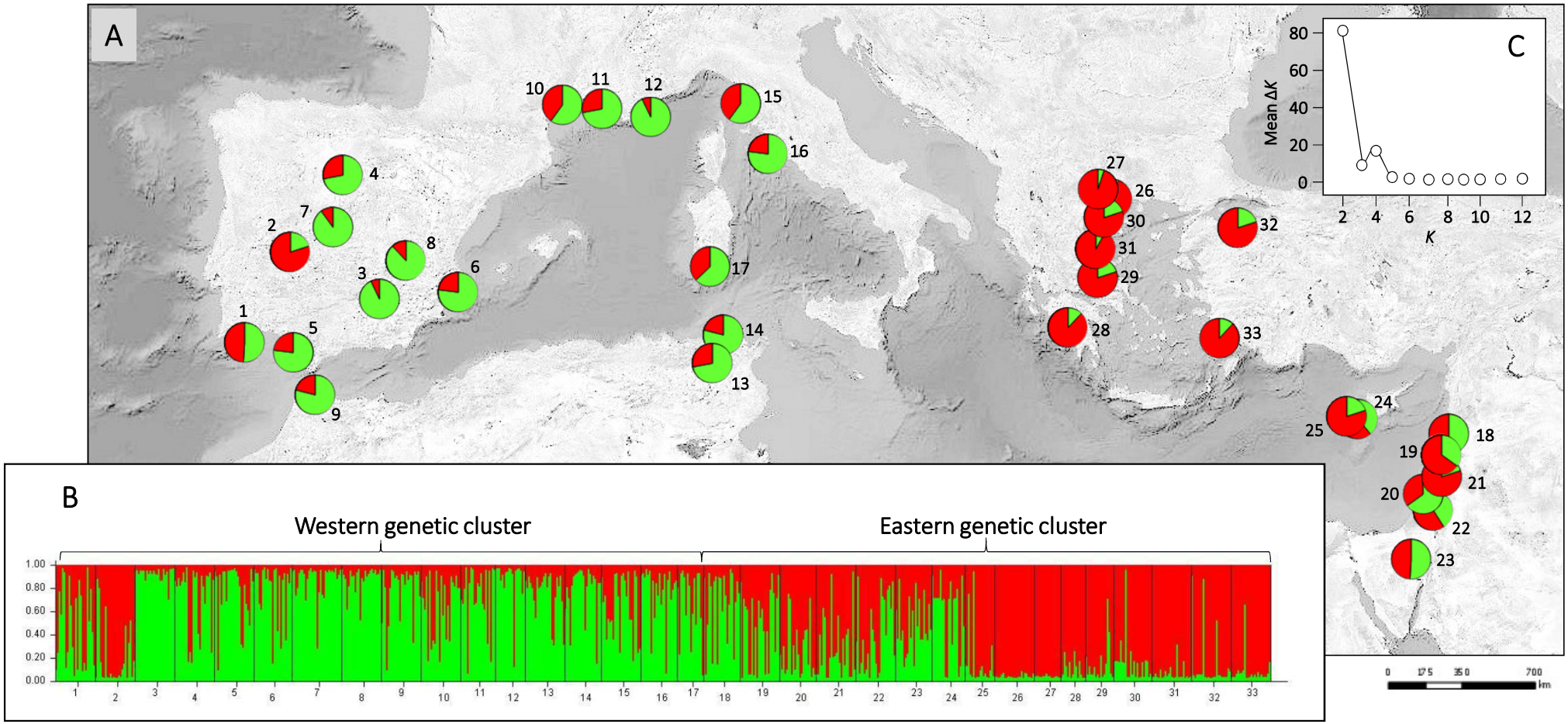
Parameter*	Stone pine	Maritime pine
θ_W (*SD*)	0.00015 (0.00055)	0.00231 (0.00212)
π (*SD*)	0.00021 (0.00125)	0.00242 (0.00243)
π_a (*SD*)	0.00019 (0.00126)	0.00029 (0.00407)
π_s (*SD*)	0.00043 (0.00353)	0.00454 (0.00474)
π_a/π_s (*SD*)	0.4487 (0.5238)	0.2561 (0.4781)
D_{xy} (*SD*)	0.0168 (0.0115)	0.0164 (0.0117)
K_a (*SD*)	0.0099 (0.0114)	0.0101 (0.0115)
K_s (*SD*)	0.0382 (0.0289)	0.0384 (0.0317)
K_a/K_s (*SD*)	0.4598 (1.0666)	0.3315 (0.5126)

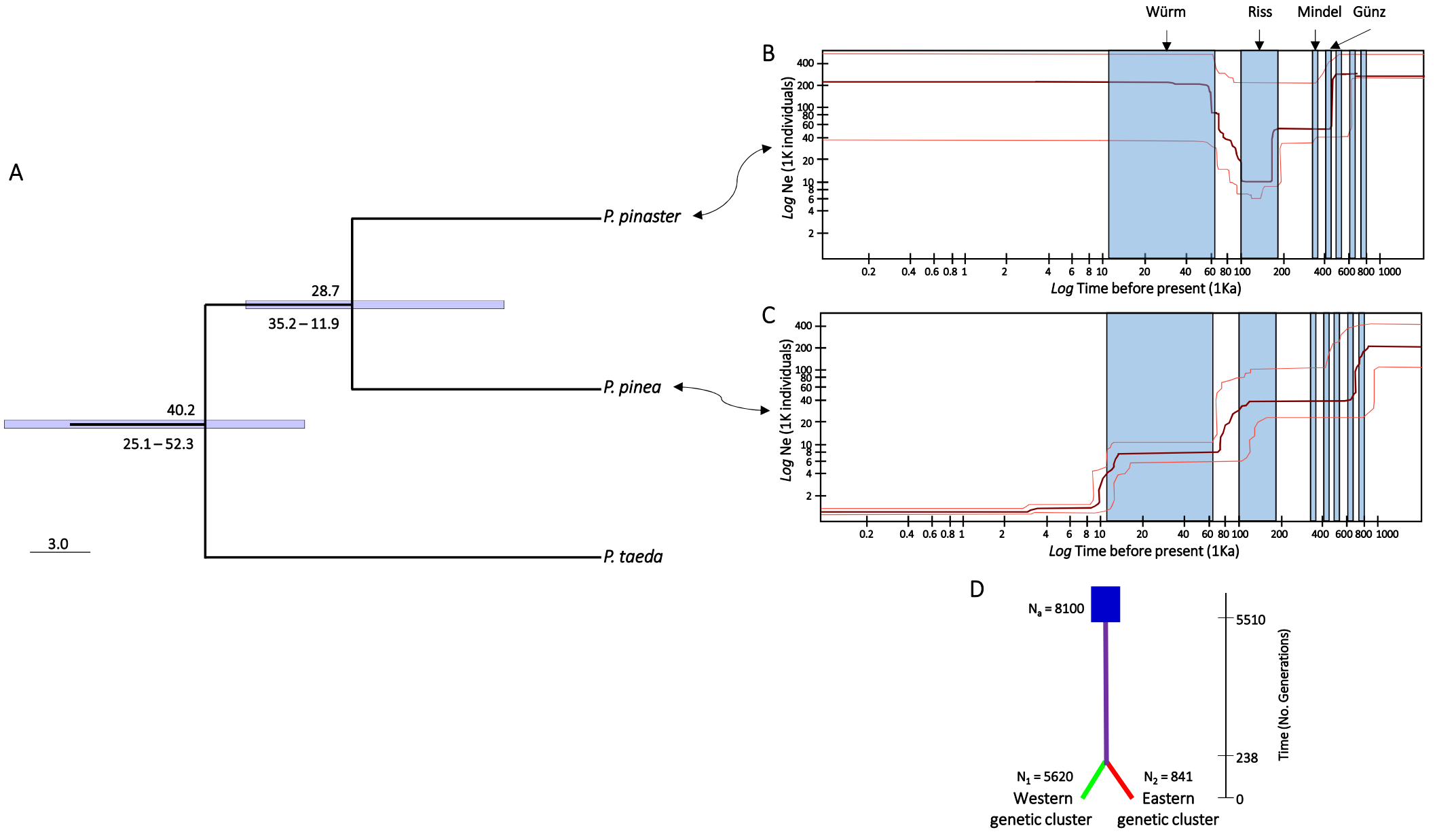
* Abbreviations: *SD*, standard deviation; θ_W, average Watterson's nucleotide diversity per site; π, average Tajima's nucleotide diversity per site; π_a, average π for non-synonymous sites; π_s, average π for synonymous sites; π_a/π_s, mean per gene ratio of non-synonymous to synonymous nucleotide diversity; D_{xy}, average pairwise divergence; K_a, average number of non-synonymous substitutions per non-synonymous site; K_s, average number of synonymous substitutions per synonymous site; K_a/K_s, mean per gene ratio of non-synonymous to synonymous substitutions. The last three estimates were determined using the Jukes-Cantor correction.

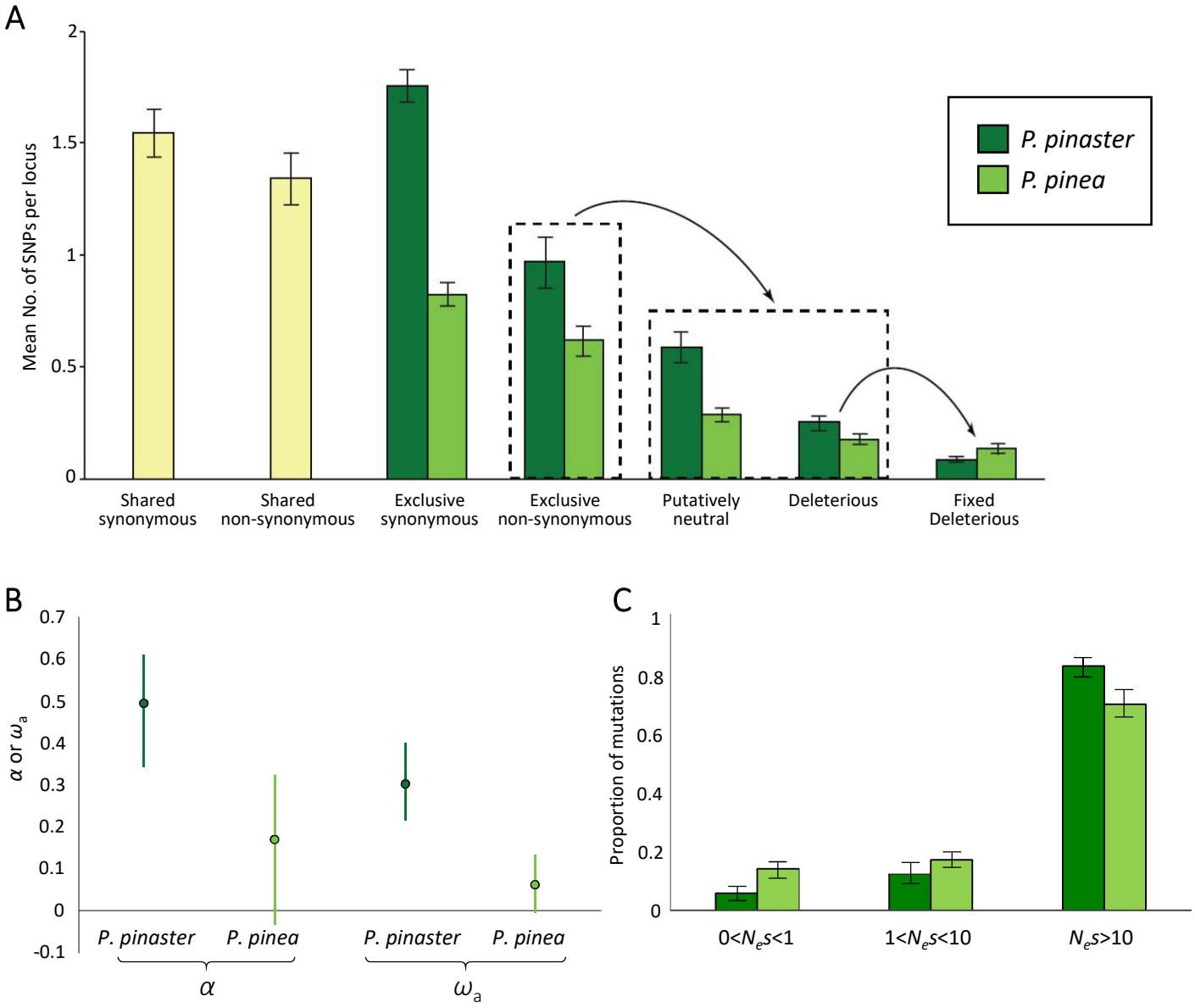
TABLE 2 Genetic diversity (cpDNA, cpSSR, and nuSSR) in genetically depauperate but widespread plants, including stone pine. Clonal and self-fertilizing plants were excluded.

Species	Markers	H_0	H_E	H_S	H_T	References
Pinus resinosa	9 cpSSR			0.543	0.618	Echt et al. 1998
	10 cpSSR		0.152			Walter and Epperson 2005
	5 nSSR	0.185	0.508			Boys et al. 2005
Pinus pinea	3 cpSSR				0.019	Vendramin et al. 2008
	9 nSSR	0.083	0.114			*This study*
Juglans cinerea	7 nSSR	0.760	0.750		0.762	Ross-Davis et al. 2008
	7 nSSR	0.659	0.691			Boraks and Broders 2016
Tsuga canadensis	4 cpDNA loci		0.727			Lemieux et al. 2011
	3 cpSSR		0.672			Lemieux et al. 2011
	13 nSSR	0.566	0.526			Potter et al. 2012


Bay of Biscay
France
Portugal
Spain
Italy
Black Sea
Greece
Turkey
Mediterranean Sea
Tunisia
Algeria
Morocco
Cyprus
Lebanon
Israel
Egypt
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
No. private alleles
A_{PT}
2
1
0
0
175
350
700
km


A
10
11
12
15
16
4
2
7
8
3
6
17
1
5
14
13
9
27
26
30
31
29
32
28
33
24
25
18
19
21
20
22
23
C
Mean ΔK
80
60
40
20
0
2
4
6
8
10
12
K
B
Western genetic cluster
Eastern genetic cluster
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
0
175
350
700
km


A
28.7
35.2 – 11.9
40.2
25.1 – 52.3
3.0
P. pinaster
P. pinea
P. taeda
B
Würm
Riss
Mindel
Günz
Log Ne (1K individuals)
Log Time before present (1Ka)
C
Log Ne (1K individuals)
Log Time before present (1Ka)
D
Na = 8100
N1 = 5620
Western
genetic cluster
N2 = 841
Eastern
genetic cluster
5510
238
0
Time (No. Generations)


A
Mean No. of SNPs per locus
2
1.5
1
0.5
0
P. pinaster
P. pinea
Shared synonymous
Shared non-synonymous
Exclusive synonymous
Exclusive non-synonymous
Putatively neutral
Deleterious
Fixed Deleterious
B
α or ω_a
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
P. pinaster
P. pinea
P. pinaster
P. pinea
α
ω_a
C
Proportion of mutations
1
0.8
0.6
0.4
0.2
0
0<N_es<1
1<N_es<10
N_es>10