

RNA-Seq Reveals Adaptive Genetic Potential of the Rare Torrey Pine (*Pinus torreyana*) in the Face of Ips Bark Beetle Outbreaks

Stephanie E. Steele

San Diego Zoo Wildlife Alliance

Oliver Ryder

San Diego Zoo Wildlife Alliance

Joyce Maschinski ([✉ jmaschinski@sdzwa.org](mailto:jmaschinski@sdzwa.org))

San Diego Zoo Wildlife Alliance

Research Article

Keywords: adaptive variation, evolutionary potential, plant resistance, conservation genomics, rare conifer, Ips paraconfusus

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-391478/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Conservation Genetics on August 18th, 2021. See the published version at <https://doi.org/10.1007/s10592-021-01394-7>.

Abstract

The ability of tree species to adapt to water stress and increased frequency of bark beetle outbreaks with climate change may increase with population size and standing genetic variation, calling into question the resilience of small, rare plant populations. The Torrey pine (*Pinus torreyana*) is a rare, genetically depauperate conifer that occurs naturally in a mainland and island population in southern California. Due to recent declines in the mainland population coinciding with drought and *Ips paraconfusus* bark beetle outbreaks, the species would benefit from an assessment of adaptive genetic diversity. Here, we use RNA-Seq to survey gene-coding diversity across 40 individuals to 1) characterize patterns of genetic diversity in the species and 2) test for genetic differentiation between trees that succumbed to beetle attack or survived following an outbreak. Consistent with previous studies, we found few genetic variants, with most SNPs occurring as fixed differences between populations. However, we found structure within the mainland and polymorphisms segregating in both populations. Interestingly, we found differentiation in genotypes between attacked and surviving trees, 11 SNPs associated with survival status, and an enrichment of defense-related functions among the top 10% of SNPs. While low diversity suggests limited adaptive capacity, genetic associations with survival in functionally relevant genes suggest adaptive potential for bark beetle defense. This initial study prompts future research to explore the genetic basis of putative resistance and suggests conservation efforts should protect surviving genotypes and the full spectrum of genetic diversity across populations to preserve the evolutionary potential of the species.

Introduction

A major question in evolutionary and conservation biology is whether species will be able to adapt to changing environmental conditions. This question is particularly relevant for long-lived, immobile trees that may not migrate rapidly enough to track suitable climates (Aitken et al. 2008). In many conifer forests in western North America, prolonged drought and warmer temperatures due to climate change have weakened trees and created favorable conditions for bark beetle outbreaks, leading to large-scale tree mortality (Raffa et al. 2008; Meddens et al. 2012; Anderegg et al. 2015). The ability of tree species to adapt to ongoing beetle outbreaks and concomitant evapotranspirative stress may increase with the strength of selection, population size, and the degree of standing genetic variation (Alberto et al. 2013). Large, highly outcrossing populations typical of trees tend to harbor intermediate to high standing genetic variation (Petit and Hampe 2006; Savolainen and Pyhäjärvi 2007), and may contain at least some resistant individuals, allowing potential adaptation to increasing pressure from bark beetles (Budde et al. 2016). However, for rare plants with small populations, evolutionary potential may be limited by the eventual loss of genetic variants to drift (Leimu et al. 2006). Thus, the resilience of small, threatened tree populations in the face of increasing pressure from climate-driven bark beetle outbreaks is of particular conservation concern, warranting an assessment of their adaptive genetic potential.

For tree populations to adapt to increasing beetle pressure, there must be variation in the population for resistance and such variation must be heritable (Telford et al. 2015). In conifers, resistance includes constitutive or induced mechanical or chemical cues or defenses that deter, kill, or limit the spread of beetles (Franceschi et al. 2005; Telford et al. 2015) or their fungal (Paine et al. 1997) and bacterial (Raffa 2014) symbionts. Conifer secondary chemical profiles signal species, age, stress, nutritional quality (Raffa et al. 2015), and defensive capacity (Emerick et al. 2008; Schiebe et al. 2012) and are used by beetles to locate and select suitable hosts (Raffa et al. 2016). For trees identified as prospective hosts, viscous, terpene-rich oleoresin is a key defense that can physically and chemically defend against beetle attack (Franceschi et al. 2005; Keeling and Bohlmann 2006). Studies have shown that multiple chemical and resin-related traits are genetically variable and heritable (Rosner and Hannrup 2004; Yanchuk et al. 2008; Westbrook et al. 2013), indicating that genes underlying chemical biosynthesis pathways, defense structures, or pathogen detection may be subject to selection. Additionally, prolonged drought can reduce tree defenses against bark beetles (Gaylord et al. 2013; Netherer et al. 2015), suggesting that genes involved in water regulation can further contribute to host resistance if defenses can be maintained under water-limiting conditions.

Field studies have indeed demonstrated that selective tree mortality by bark beetles can shift phenotypic (Schiebe et al. 2012; Ferrenberg et al. 2014; Zhao and Erbilgin 2019) and genetic variation (Six et al. 2018) of natural conifer populations. In a recent study of two pine species that experienced a mountain pine beetle outbreak, Six et al. (2018) identified a genetic basis for survival, illustrating an evolutionary response to beetle pressure. While the genetic markers in the study were unable to pinpoint the survival mechanism, genome-wide sequence data may be able to identify putative loci and associated functions differentiating trees that withstand or succumb to beetle attack.

For conifers lacking extensive genomic resources, detecting genetic associations with tree survival following a beetle outbreak presents a challenge. While whole genome sequencing is becoming increasingly tractable for many species (Fuentes-Pardo & Ruzzante, 2017), conifer genomes are exceptionally large (18 – 35 Gb) (Mackay et al. 2012) and consist predominately of repetitive elements (82% of the *Pinus taeda* genome) (Wegrzyn et al. 2014), complicating assembly for even well-studied species. RNA-Seq, however, offers a feasible approach for obtaining single nucleotide polymorphisms (SNPs) (*sensu* Geraldes et al. 2011; Gugger et al. 2016), by acting as sequence reduction method and capturing variants in expressed genes that may underlie adaptive phenotypes. RNA-Seq is a suitable means for studying adaptation in species of conservation interest that often lack genome assemblies.

Torrey pine (*Pinus torreyana* Parry) is an iconic species of pressing conservation concern that would benefit from RNA-Seq studies to assess its adaptive potential to survive bark beetle outbreaks. The rarest pine in North America, it occurs naturally in only two localities in Southern California; in coastal San Diego County (mainland population) and on Santa Rosa Island off the coast of Santa Barbara (island population). The native California five-spined ips beetle (*Ips paraconfusus* Lanier, hereafter *Ips*) poses a substantial threat to the mainland population, although beetles are absent on the island. The Torrey Pines State Natural Reserve, which protects a substantial portion of mainland trees, experienced a 12% population decline from 1988 to 1991 (Shea and Neustein 1995) and a 12% decline from 2006 to 2018 with approximately 3700 individuals remaining (D. Smith, *personal communication*). Each decline was a likely consequence of major bark beetle outbreaks, the latter of which coincided with a period of prolonged drought. The capacity for Torrey pines to adapt to re-occurring beetle outbreaks with changing climate remains in question. Allozymes (Ledig and Conkle 1983) and the chloroplast genome (Whittall et al. 2010) and markers (Waters and Schaal 1991; Provan et al. 1999; Whittall et al. 2010) indicated complete within-population monomorphism and few between-population differences. However, surveying putative adaptive variation across the genome will be essential to understanding the evolutionary potential of the species.

Our goal is to conduct a preliminary study to evaluate the potential of Torrey pines to adapt to increased bark beetle pressure amid climate change. While traditionally measured as additive genetic variance, genome-wide measures of genetic diversity may be reasonable proxies for evolutionary potential, especially for polygenic traits (e.g. resistance) (Harrisson et al. 2014; Ørsted et al. 2019). However, assessing regions of the genome underlying adaptive phenotypes will be most informative (Hoffmann et al. 2017; Mable 2019). To this end, our goals were to use RNA-Seq to: 1) assess patterns of genome-wide functional genetic variation within and among Torrey pine populations as a general indicator of evolutionary potential, and 2) test for genetic differentiation between ‘case’ trees that succumbed to *Ips* attack and living ‘control’ trees to understand the capacity to adapt to ongoing beetle pressure. For the latter goal, we test whether cases and controls are genetically diverged at multivariate genotypes, identify specific variants associated with case/control status, and test for enrichment of functionally relevant gene ontology terms.

Materials And Methods

Field Site and Sampling Scheme

We sampled trees at the Torrey Pines State Natural Reserve (TPSNR, the mainland population) near La Jolla/Del Mar in San Diego County, California. The reserve occurs along coastal bluffs and comprises multiple habitat types including coastal pine forest, chaparral, and coastal sage scrub. A lagoon separates the reserve into two sites including the main reserve and the Torrey Pines Extension (hereafter TPSR and TPEx populations, respectively). Our sampling design included pairs of proximate trees with and without signs of *Ips* in order to reduce environmental variation that may influence beetle attack. Characteristic signs of *Ips* infestation included small entrance holes exuding red pitch and frass on the trunk and branches, and needle yellowing or browning throughout the tree crown as infestation progressed. We sampled tree pairs throughout the reserve to capture a representative pool of variation for diversity and structure estimates.

We sampled a total of 40 trees from October to November 2017 and from November 2018 to April 2019 (Fig. 1, S1, Table S1). Fall of 2017 had high *Ips* activity with widespread tree mortality in clusters throughout the reserve, while the season beginning in Fall of 2018 had lower *Ips* activity in scattered localities. In 2017, we sampled 22 mainland trees including seven pairs of trees ($n = 14$) with and without signs of *Ips*, four trees throughout the reserve to capture additional genotypes, and four trees with a Santa Rosa Island origin planted prior to 1960 at the nearby Sanford Burnham Prebys Discovery Institute (SBPDI). Beginning in Fall of 2018, we sampled nine mainland tree pairs ($n = 18$) with and without signs of *Ips*. Because mortality can proceed rapidly once needle yellowing begins, we sampled beetle-colonized trees with sufficient living tissue opportunistically as we identified infested trees. Where possible, we matched beetle-colonized trees to nearby asymptomatic trees of similar diameter at breast height (DBH), and in the second season, we matched environmental variables between tree pairs using 2018 census data provided by TPSNR and geographic information system

data. Tree pairs occurred within approximately 30 – 50 m of one another (or up to 150 m for more isolated trees). A multiple factor analysis indicated comparable size, spatial, and environmental conditions between groups (Fig. S2) despite sampling across a heterogeneous landscape.

In November 2019, we re-surveyed all 40 trees, assigning dead trees with signs of *Ips* infestation as ‘cases’ and living trees as ‘controls.’ Some controls had potential signs of past beetle attack limited to isolated branches, though no evidence of ongoing infestation. One *Ips*-infested individual identified in 2017 was still living in 2019 and one asymptomatic individual had succumbed to beetle attack by 2018, and they were reassigned as a ‘control’ and ‘case’, respectively. A second bark beetle, *Dendroctonus valens* LeConte (red turpentine beetle) also occurred extensively throughout the reserve, although it is largely restricted to the lower tree bole. Prior to sampling, 21 of 40 trees displayed signs of *Dendroctonus* including 13 case trees and 7 control trees (infestation could not be determined for three trees whose lower bole was covered by shrubbery).

For each of the 40 trees, we sampled both twig and needle tissue to maximize transcript diversity. We selected small, terminal branches healthy in appearance when possible, although case trees were in various stages of *Ips* colonization. From each branch, we sampled needles and twig segments 15 – 20 cm behind the oldest needle growth. We preserved tissues on liquid nitrogen or dry ice prior to transporting and storing at –80°C at the San Diego Zoo Institute for Conservation until further processing.

RNA Extraction, Library Preparation, and Sequencing

We extracted total RNA from twig and needle tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO) according to the manufacturer’s protocol. For twigs, we used a razor blade to remove the outer bark layer to reduce lichen and other potential contaminants. We verified RNA quality and quantity using a Nanodrop One spectrophotometer (Thermo Scientific, Waltham, MA), Qubit BR Assay kit (Life Technologies, Waltham, MA) and an Agilent 2100 Bioanalyzer, and cleaned extracts with AMPure XP beads as needed. Of two cases with dry twigs, needles were exclusively (sample tpsr0195) or predominately (sample tpsr0912) used for sequencing. However, RNA integrity number (RIN) scores did not differ significantly between groups (cases: 7.1 – 9.9, mean 8.43; controls: 7.2 – 9.2, mean 8.35; $F_{1,30}=3.16$, $P=0.578$).

We prepared mRNA libraries with a TruSeq Stranded mRNA Kit (Illumina, San Diego, CA) and unique dual index adapters following the manufacturer’s protocol (October 2013). Libraries were PCR amplified for 15 cycles, quantified by Qubit, and validated on an Agilent 2100 Bioanalyzer. We performed double size selection with AMPure XP beads as needed to reduce minor secondary peaks in the Bioanalyzer trace. We sequenced final libraries as paired end 75 bp reads across four runs on an Illumina NextSeq500 platform at SBPDI (San Diego, CA). We included six libraries as technical replicates across runs to control for potential batch effects and re-sequenced seven low coverage samples (Table S1).

Read Filtering, Mapping, and Variant Calling

We filtered raw, demultiplexed fastq files using the HTStream pre-processing pipeline (<https://bioinformatics.ucdavis.edu/software>). We removed 28s and 18s *Pinus taeda* rRNA, PCR duplicates, adapters, and poor quality sequence from 3’ read ends (with quality score < 20 in a sliding window of 10 bp). Finally, we removed remaining N characters and any processed reads under 50 bp.

Following the Genome Analysis Tool Kit (GATK) best practice guidelines for RNA-Seq short variant discovery (<https://gatk.broadinstitute.org/>), we aligned reads to the *Pinus taeda* genome v2.01 (Zimin et al. 2017) downloaded from the TreeGenes database (Falk et al. 2018) with modified steps as follows. Due to the large and fragmented nature of the genome assembly (22 Gb, 1.76 million contigs), we used a custom perl script (S. Fitz-Gibbon, *personal communication*) to sort contigs by size and concatenate them into 222 ~100 Mb ‘pseudocontigs’ with contigs separated by 10 N’s. Using two-pass mode with the splice aware aligner STAR v2.6.0 (Dobin et al. 2013), we aligned processed reads to the *P. taeda* pseudocontigs with settings – genomeSuffixLengthMax 1000, –outSJfilterCountTotalMin 9 3 3 3, and –alignIntronMax 160000, where maximum intron length was based on estimates for *P. taeda* (Wegrzyn et al. 2014). We used Samtools v1.9 (Li et al. 2009) to retain only uniquely mapping reads and to remove PCR/optical duplicates, secondary or supplementary alignments, reads failing QC, and unpaired reads.

To call variants, we ran two iterations of GATK v4.1.4.1 HaplotypeCaller and GenotypeGVCFs with option –stand-call-conf 30 and base quality score recalibration (BQSR). During the first variant calling iteration, we applied hard filtering thresholds in GATK to remove SNPs

with the following INFO fields: QD < 5.0, FS > 30.0, ReadPosRankSum < -4.0 or > 6.0, BaseQRankSum < -4.0. We then used VCFtools (Danecek et al. 2011) to remove SNPs with minimum mean depth < 5 and call rate < 50%, resulting in over 419 K SNPs for BQSR. During the second iteration, we performed hard filtering as above with an additional filter to remove SNPs with ExcessHet > 16.43 (Z-score 2). Using VCFtools, we set genotypes with depth (DP) < 5 and genotype quality (GQ) < 20 to missing, and removed sites with mean DP < 15, minor allele frequency < 0.05, > 2 alleles, mean GQ < 44, and call rate < 95%. We modified further filtering steps from the dDocent pipeline (<http://www.ddocent.com/filtering/>) to remove sites with high read depth that may contain potential paralogs (all loci with DP > 2 standard deviations (SD) from the mean; loci with DP > 1 SD from the mean if QUAL < 2 * DP). Finally, we used BCFtools v1.9 to prune SNPs in linkage disequilibrium ($r^2 > 0.5$ in a window of 5 SNPs) using the original genome coordinates. We subset SNPs with and without LD-pruning by population for use in downstream analyses (Table 1).

Kinship, population structure, and genetic diversity

Our first goal was to assess functional genetic diversity within and among populations of Torrey pines. To test for family groups that could result from sampling proximate case/control pairs, we estimated relatedness among samples with the KING inference method in VCFtools (Manichaikul et al. 2010). We then assessed genetic differentiation controlling for kinship through Principal Components Analysis in Related samples (PC-AiR) (Conomos et al. 2015) in the R package GENESIS v.2.12.4 (Gogarten et al. 2019) for all individuals, mainland individuals, and case/control pairs using the LD-pruned SNP set. PC-AiR first identifies an unrelated subset of individuals in the data representing all ancestries. It then uses those individuals to infer principal components (PCs) and projects the remaining samples onto those PCs based on genetic similarity. We ran one to two iterations of PC and kinship estimation with a scaling constant of 1 when estimating kinship, a kinship threshold of 0.0884 (second degree relatives) for identifying unrelated individuals, and two ancestry informative PCs to estimate kinship accounting for ancestry.

For each sample group, we calculated genetic diversity and structure estimates for all individuals and for unrelated individuals identified by PC-AiR. We estimated the mean expected (H_E) and observed multi-locus heterozygosity (H_0) per population with the R package Adegenet v2.1.1 (Jombart et al. 2010) and the average inbreeding coefficient (F) per individual within population with VCFTools. Using the R package HIERFSTAT (Goudet 2005), we calculated pairwise F_{ST} based on Weir and Cockerham's q (1984). We performed all analyses in R v.3.5.2 (R Core Team, 2018) using the LD-pruned SNP set.

Multivariate analysis of genetic divergence between cases and controls

To test for divergence between multivariate genotypes of cases and controls, we employed Discriminant Analysis of Principal Components (DAPC) in Adegenet. To control for population structure, we first performed DAPC on the LD-pruned SNP set with population (TPEX or TPSR) as group for the 32 paired samples. We then regressed the full SNP set on the first discriminant function (DF) summarizing 2 PCs from the population model. We used residuals from the regression in a second DAPC model with case/control as group and missing data set to the mean. To select the optimal number of PCs for the DF, we performed cross validation, assigning 80% of the data to a training set and the remaining 20% to a validation set in a stratified random manner, with 50 replicates. Highest mean success and lowest root mean squared error estimates indicated an optimum of 12 PCs. To determine whether the difference between case and control groups was significant, we ran a MANOVA Pillai test with 12 PCs as the dependent variable. We repeated DAPC for subsets of the data dropping a PCA outlier or samples with high missing data. All analyses were limited to SNPs with a 95% call rate across the case/control dataset.

Association of individual SNPs with case/control status

To identify individual SNPs associated with case/control status, we ran the linear mixed model GEMMA which calculates an exact association test statistic (Zhou and Stephens 2012). We selected a 'centered' covariance matrix to control for hidden kinship and population stratification and included latitude, longitude, and population (TPEX, TPSR) as additional covariates. Only SNPs with a 95% call rate among the 32 paired samples were included, amounting to 854 LD-pruned SNPs to estimate the relatedness matrix and 1526 SNPs from the full SNP set for the association test. We repeated the association test separately for TPSR, TPEX, and subsets of the data excluding a PCA outlier or samples with high missing data. Using the R package qvalue v.2.14.1, we adjusted P -values from the likelihood ratio test by employing the false discovery rate method to correct for multiple testing (Storey and Tibshirani 2003).

Additionally, PCs for differing tiers of ‘top SNPs’ in the tail of the distribution were visualized based on estimates from the R package SNPRelate v1.16.0 (Zheng et al. 2012). We used a dDocent script to test deviations from Hardy-Weinberg equilibrium in top SNPs within TPEX, TPSR, and the combined mainland population.

Functional annotation and gene ontology enrichment analysis of top SNPs

We matched SNPs to functional annotations for 51,751 gene models in the *Pinus taeda* reference genome from the TreeGenes Database (Pita.2_01.final_annotations_lvl0.tsv downloaded April 2020:

<https://treegenesdb.org/FTP/Genomes/Pita/v2.01/annotation/>). SNPs lacking a functional annotation in the reference were supplemented with custom annotations. Briefly, we used Stringtie v2.1.1 (Pertea et al. 2015) to assemble transcripts for each sample using the *P. taeda* gene models as a guide and 15' minimum read coverage (-c 15 -s 15). We then merged individual assemblies to create a single transcriptome assembly. To functionally annotate transcripts containing top SNPs, we ran EnTAP v0.9.1 (Hart et al. 2020) with blastx similarity search across the Uniprot-SwissProt and Plant RefSeq databases (downloaded May 2020) with the following parameters: E-value threshold 10^{-6} , 30% query and target coverage, and *Pinaceae* taxon prioritization. Finally, we identified gene ontology (GO) terms for all SNPs tested in GEMMA by running EggNOG-mapper (Huerta-Cepas et al. 2016) through EnTAP.

To identify GO terms overrepresented among SNPs associated with case/control status, we performed singular enrichment analysis in AgriGO v2.0 (Tian et al. 2017). We ran two sets of enrichment tests, using either GO terms from the *P. taeda* reference annotations or from custom annotations with EnTAP (633 vs 1170 of 1526 SNPs with at least one GO term, respectively). If a SNP matched multiple alternative transcripts, we selected GO terms from the longest transcript. We separated gene ids and corresponding GO terms into a query set (SNPs identified in the top 5% or 10% of the distribution of the GEMMA association test) and a reference set (the non-overlapping set of the remaining SNPs tested in GEMMA). Only one SNP per gene was considered. We performed Complete GO enrichment analysis separately for biological process, cellular component, and molecular function levels. To assess significance, we used a Chi-squared test for GO terms matching at least 5 transcripts and adjusted for multiple testing with the Benjamini and Hochberg (1995) false discovery rate (FDR) method.

Results

Raw reads, mapping, and variant calling

We obtained approximately 14 – 54 million paired-end reads per sample after quality filtering. On average, $70.2 \pm 3.8\%$ of reads mapped uniquely to the *Pinus taeda* reference genome per library (excluding one degraded library), covering 23.84 million base pairs at greater than 15' average coverage per sample. GATK Haplotype Caller produced over 4.75 million raw SNPs, of which 9599 total biallelic SNPs and 4192 LD-pruned SNPs remained after filtering (Table 1). Minimum depth (DP), genotype quality (GQ), and minor allele frequency filters had the largest effect on number of SNPs retained. The final SNP set exhibited a transition/transversion ratio of 1.51, high mean coverage ($71.7 \pm 21.9'$) and generally high mean GQ (86.43 ± 8.14) and low mean number of missing SNPs per sample ($1.04 \pm 1.49\%$, Fig. S3). The number of missing SNPs increased with decreasing DP and GQ, but was not associated with heterozygosity (results not shown). There was no significant difference in DP, GQ, and number of missing SNPs between cases and controls (results not shown), although two cases had an elevated number of missing SNPs (less than 6.5%; Fig. S3).

Kinship, genetic diversity and population structure

Consistent with previous studies, Torrey pines displayed low genetic variation within and between populations. Of the 9599 SNPs called across samples, most were fixed differences between the island and mainland populations (Table 1). Only 1888 SNPs were present among 36 mainland individuals and 750 SNPs were present among four island individuals with a minor allele count > 1 (Table 1). Interestingly, TPSR had fewer SNPs (1332 vs. 1424), lower expected heterozygosity (H_E , 0.044 vs. 0.053) and observed heterozygosity (H_O , 0.039 vs. 0.053), and higher mean inbreeding (F , 0.137 vs. 0.053) than TPEX (Table 1), despite greater sampling across a wider spatial area (Fig. 1). Surprisingly, although SRI had the lowest H_E among groups (0.036), it had higher H_O (0.064) and lower inbreeding (-0.565) than either mainland population (Table 1). F_{ST} between the island and mainland was extremely high (0.915) and was high between TPEX and TPSR (0.523), indicating restricted gene flow even between adjacent populations (Table 2).

To account for potential bias due to hidden kinship, we tested for relatedness among samples and re-estimated genetic diversity and structure among unrelated individuals. The KING inference method indicated relatedness between sample pairs within the island and mainland and within TPEX and TPSR (Fig. S4). However, limiting estimates to unrelated individuals (beyond second degree relatives) identified by PC-AiR had little effect on genetic diversity statistics (Table S2) and F_{ST} (Table S3). Furthermore, PC-AiR plots accounting for hidden kinship revealed distinct clustering within the island and mainland populations along PC1 (84.66% of variation explained; Fig. 2a). When only the mainland was considered, TPSR and TPEX separated into distinct populations along PC1 (44.37% of variation explained; Fig. 2b), with TPEX more diffusely distributed along PC2.

We next examined genetic structure and diversity in cases and controls to ensure that case/control status was not confounded with population structure. PC-AiR indicated no structure between cases and controls (Fig. 2c), which was similarly reflected by low F_{ST} (-0.013, Table 2). H_E , H_0 , and inbreeding were similar for cases and controls, while greater H_E than H_0 in both groups reflected population structure between TPEX and TPSR (Table 1). When we considered only unrelated individuals, similar patterns emerged (Table S2, S3), although cases had slightly lower H_0 than controls (0.032 vs 0.041).

Multivariate analysis of genetic divergence between cases and controls

The discriminant analysis showed differentiation between multivariate genotypes of cases and controls after controlling for population (Fig. 3). Discriminant function (DF) 1 was largely non-overlapping between groups (Fig. 3a) and the 12 PCs summarized by DF1 were significantly different between cases and controls ($F(12, 19) = 2.567, P = 0.0322$, Pillai's trace = 0.619) though the assumption of no outliers was not met. Generally, membership probability to group reflected the case/control status of the sample (Fig. 3b). Individuals were correctly assigned to group with an 84.38% probability, with higher assignment probability to controls (93.75%) than cases (75%). Results were largely similar after dropping a PCA outlier alone or in combination with two samples with higher missing data (results not shown).

Association of individual SNPs with case/control status

The GEMMA mixed model identified 11 SNPs significantly associated with case/control status after correction for multiple testing (Table 3, Fig. S5a). It should be noted, however, that the P -value distribution was not uniformly distributed (Fig. S5b-c), which could indicate genome-wide differences unaccounted for between cases or controls or alternatively, could reflect small sample size or number of SNPs. Nevertheless, 15 SNPs were shared in the top 5% of the P -value distribution for each of five GEMMA models (the full model or models based on subsets of the data: TPEX, TPSR, or excluding a PCA outlier alone or in combination with two samples with high missing SNPs). Many of these 15 SNPs overlapped the 11 significant SNPs, amounting to 19 SNPs (Table 3). All SNPs were within HWE in TPEX ($P > 0.05$), TPSR ($P > 0.05$), and the mainland population ($P > 0.005$). PCA indicated clear separation between cases and controls when plotting one SNP per gene among the top 19 SNPs (Fig. 4a) and among SNPs in the top 5% of the distribution (Fig. 4b), indicating their utility for differentiating cases and controls. Among the latter SNPs, population separated individuals along PC2, although it explained less variation in the data than case/control status along PC1. Among top SNPs, one group (cases or controls) tended to be homozygous at a given locus while the other group was predominately heterozygous (Fig. S6).

Functional annotation and gene ontology enrichment analysis of top SNPs

The top 19 SNPs occurred within 11 putative genes, six of which matched annotated gene models in the *P. taeda* reference and four of which were supplemented with functional annotations in the EnTAP pipeline (Table 3). The most significant SNP occurred in a probable aquaporin gene with GO terms including 'water transport' and 'response to osmotic stress' among others. Notably, genes with potential defense roles included cytochrome P450, DMR6-like oxygenase, peroxidase, and a major allergen (a pathogenesis-related protein). Remaining genes were either uncharacterized or displayed GO terms related to metabolic processes.

Gene ontology enrichment analysis further suggested that defense response genes may be related to case/control status. Among the top 10% of custom EnTAP-annotated genes, six biological process GO terms related to 'death' (a now obsolete term) and response/defense response to bacteria, fungi, and other organisms were significantly enriched after correction for multiple testing (Fig. 5). Similar functions were also among the most enriched terms in the top 5% of SNPs, although they were not significantly overrepresented. Only a molecular function term related to oxidoreductase activity was significantly enriched among the top 5% of

SNPs (FDR = 0.018). No cellular component terms were enriched in the top 5% or 10% of SNPs. Despite fewer GO annotations among transcripts matching *Pinus taeda* gene models, we found similar patterns as above, with five biological process terms related to biotic response significantly overrepresented in the top 10% of SNPs (results not shown).

Discussion

With changing climate, conifer populations are predicted to face increased evapotranspirative stress and frequency of bark beetle outbreaks, warranting an evaluation of their ability to adapt to such challenges. Using RNA-Seq to survey functional regions across the genome of the rare Torrey pine, we found low genetic diversity within the species, predicting low adaptive capacity. However, our finding of genetic associations with *Ips*-colonization status and an enrichment for defense-related functions among top SNPs suggests potential selection by bark beetles on genetically-based traits and at least some adaptive potential for resistance. These initial findings prompt further controlled studies to dissect genetic variation underlying potential host resistance phenotypes and have implications for management of the species to support its persistence into the future.

Low genetic diversity for a pine with high population structure

Consistent with previous studies on Torrey pines, we found few genetic variants (9599 SNPs) expressed in transcripts across populations, with most SNPs found as fixed differences between the island and mainland. Expected and observed heterozygosity were low within the island and two mainland populations (H_E of 0.036 – 0.053, H_O of 0.039 – 0.064) while F_{ST} was extremely high between the island-mainland and TPEX-TPSR groups (0.915 and 0.523, respectively). This structuring of genetic variation is unusual for outcrossing trees, typically characterized by high standing genetic variation partitioned largely within rather than among populations (Hamrick et al. 1992). For example, various locally and widely distributed conifer species (Dillon et al. 2013; Mosca et al. 2014; Nadeau et al. 2015) displayed H_E and H_O of 0.20 – 0.40 per population, with among-population structure below 10% based on EST-derived SNPs. On the other hand, *Pinus resinosa*, another genetically depauperate conifer, displayed high population divergence ($F_{ST} = 0.28$) in microsatellites but still less than Torrey pine (Boys et al. 2005). The lack of diversity and high structure in Torrey pines may be due to one or more extreme bottlenecks, a founder event leading to colonization of the island, and genetic drift following population divergence (Ledig and Conkle 1983).

Novel to our study were findings that 1) polymorphisms are segregating within populations albeit at a low level, and 2) there is population structure between TPEX and TPSR within the mainland. The discovery of within-population polymorphisms is not wholly unexpected, given that we examined gene-coding regions on a genome-wide scale. Previous studies examining 59 allozyme loci (Ledig and Conkle 1983) and the chloroplast genome or markers (Waters and Schaal 1991; Provan et al. 1999; Whittall et al. 2010) found complete monomorphism within populations. Fewer variants are expected in allozymes and in the haploid, non-recombining chloroplast that evolves at slower rates than the nuclear genome (Wolfe et al. 1987). Our finding of genetic structure between TPEX and TPSR suggests that the ~1.3 km wide lagoon separating these populations may present a substantial barrier to gene flow. Alternatively, genetic structure between these sites may reflect restoration efforts in the mainland in the early 1900's and more recently in 1994 to replace trees lost during a preceding beetle outbreak (Ledig et al. 1998). If limited maternal sources were reintroduced, increased kinship in TPSR could elevate structure between populations. Indeed, TPEX had higher H_E and H_O and lower inbreeding than TPSR despite fewer individuals sampled and may thus represent an older, undisturbed pool of individuals. Structure on the mainland may reflect a combination of restricted gene flow across the lagoon and elevated kinship among planted trees.

Genetic differentiation between groups in relevant genes suggests selection by beetles

The discriminant analysis and association test indicated multivariate and single SNP differences between cases and controls, suggesting that beetles are responding in part to genetically controlled differences in the tree population. Beetles may be selectively avoiding well-defended trees or may fail to colonize resistant trees if they attempt attack. Supporting this hypothesis, gene ontology (GO) analysis found an overrepresentation of genes related to defense responses among the top 10% of genes associated with case/control status.

Although controlled experiments would be required to elucidate the mechanisms underlying potential host resistance, functional annotations of top genes suggest that defenses and water regulation may play a role. In the GO analysis, responses against bacteria, fungi, and other organisms were significantly enriched among top genes (Fig. 5). Moreover, four of the top 11 genes from the association test have roles in plant immunity. Cytochrome P450, for example, belongs to a well-known gene family involved in synthesis of insect-defensive chemicals through the terpenoid and phenylpropanoid biosynthesis pathways (Keeling and Bohlmann 2006). DMR6-like oxygenase, peroxidase, and major allergen, alternatively, are involved in active microbial defenses (Kolosova and Bohlmann 2012; Zeilmaker et al. 2015). Because bark beetles have known associations with fungi (Paine et al. 1997) and bacteria (Boone et al. 2013) that facilitate host colonization, microbial defenses in conifers could contribute to collective insect resistance. In a previous study in Torrey pines, Zavarin et al. (1967) found little indication that constitutive oleoresin chemistry contributes to resistance variation within populations. However, resistance can be achieved through multiple mechanisms, and our findings suggest other pathways could account for genetic differences between groups.

Notably, the SNP most significantly associated with case/control status was found in a putative gene encoding an aquaporin, a membrane channel protein involved in water, solute, and gas transport. In sweet orange, multiple aquaporins were down-regulated in susceptible genotypes infected with a common citrus pathogen, suggesting that aquaporins moderate disease progression by altering water and nutrient transport involved in normal growth (de Paula Santos Martins et al. 2015). Aquaporins may also facilitate the transport of H₂O₂ into the cytoplasm, triggering immunity pathways (Tian et al. 2016). In general, drought is known to predispose trees to beetle attack (Netherer et al. 2015), suggesting a benefit to maintaining water potential and defenses under water limitation.

While our findings indicate a genetic basis contributing to bark beetle attack, a few caveats are warranted. Genetic differentiation between cases and controls could occur in merged paralogs, given that aquaporins and many defense-related genes belong to large gene families. Mapping transcripts to a reference genome of another species, particularly a fragmented genome like that of *Pinus taeda*, can present challenges for multi-gene families. To account for this, we retained only uniquely mapped reads and applied stringent SNP filters to minimize the incidence of merged paralogs. As an additional test for multi-locus contigs (*sensu* McKinney et al. 2017; Rellstab et al. 2019), we ran a second SNP filtering pipeline with several pre-filtering steps followed by exclusion of 1) entire exons containing SNPs with excess heterozygosity (EH) and 2) SNPs with allele balance (AB) deviations from the expected 0.50 ratio (Supplemental methods). We then ran a GO analysis of SNPs that failed the EH and AB filters relative to the background set of pre-filtered SNPs. Five biological process terms were significantly enriched among failed SNPs, including terms related to defense response and cell death (Table S5). To make meaningful interpretations of our results possible, we retained our original set of 9599 SNPs to avoid discarding biologically relevant defense-related SNPs that are central to our question. While we can better assess a large suite of genetically-controlled traits to which beetles may respond, our results should be interpreted cautiously. Genes involved in putative adaptive responses to insects are also likely to be differentially expressed between cases and controls. If these differentially expressed genes belong to multi-gene families that map to the same locus, it could generate false genotypic differences between cases and controls. In order to disentangle true SNPs from potential artifacts of differential gene expression and paralogs, a high-quality reference genome and *de novo* transcriptome assembly for Torrey pine supplemented with long reads would be beneficial.

It is also worth noting that changes in bark beetle dynamics over time due to environmental and biotic factors can alter selection on hosts (Raffa et al. 2008) and thus the genetic composition of cases and controls. Given the larger *Ips* outbreak among Torrey pines in 2017 than 2018 and the positive relationship between beetle abundance and the ability to overwhelm vigorous hosts (Raffa 2014), selection may have favored trees with greater resistance in 2017. Additionally, biotic interactions can influence beetle dynamics and host selection. *Dendroctonus valens* was another bark beetle prevalent throughout the reserve that may have affected *Ips* colonization. In a long-term study of *Pinus resinosa* (Aukema et al. 2010), probability of attack by two *Ips* species increased with the number of *Dendroctonus* colonizing a given tree. While all trees with successful *Ips* colonization eventually died, mortality was unlikely if *Dendroctonus* was present without *Ips*. In Torrey pines, *Dendroctonus* may similarly affect the tree ability to withstand *Ips* attack and given more time, Torrey pines colonized by *Dendroctonus* may eventually succumb to *Ips*. However, we found no further *Ips*-related mortality among control trees through late 2019, although we re-assigned one control that died by 2018 as a case. Because mortality proceeds rapidly following successful *Ips* colonization, it is unlikely that there were other asymptomatic, successfully *Ips*-infested trees among our control samples. Nevertheless, our study assessed selection over a finite outbreak period, and changes in the environment and biotic interactions over time can influence beetle behavior, host susceptibility, and the consequent genetic pool of individuals that survive or succumb to beetle attack. Future studies should link genetic variation to specific resistant phenotypes under different treatments to better understand tree capacity to respond to various degrees of selection pressure.

A final caveat is our study's small sample size. While human genome-wide association studies assessing the genetic basis of disease typically include thousands of individuals (Zondervan and Cardon 2007), studies of adaptive traits in plants may require many fewer samples since positively selected alleles underlying these traits are likely to occur at higher frequency (Ingvarsson and Street 2011; Kloth et al. 2012). This is especially true for pathogen resistance which is often controlled by a few genes of large effect, although the genetic architecture of insect resistance is less well characterized (Broekgaarden et al. 2011). Using a moderate sample size of 96 to 192 inbred *Arabidopsis* lines, for example, Atwell et al. (2010) identified common alleles of large effect associated with 107 adaptive phenotypes, including 23 pathogen and insect defense-related traits. Similarly, in a study of mountain pine beetle (MPB) susceptibility in *Pinus contorta*, Cullingham et al. (2020) identified one candidate gene (of 17 initial candidates) significantly associated with attack status in a sample of 356 attacked and unattacked trees. A study by Six et al. (2018) also identified genetic differentiation between survivors and the general population of younger trees among 66 *Pinus albicaulis* and 40 *P. contorta* trees sampled after an MPB outbreak. While Atwell et al. (2010) argues for greater sample sizes to increase statistical power, these studies demonstrate that even moderate sample sizes have utility to detect meaningful genetic associations for further study. Here, using a small sample of 32 Torrey pines, we identified genetic variation associated with survival following an *Ips* outbreak. Our preliminary work prompts large-scale follow-up studies to uncover and validate additional putative variants that may contribute to resistance.

Adaptive potential of Torrey pines

A principal goal of our study was to assess the adaptive genetic potential of Torrey pines, both to general selection pressures and ongoing bark beetle outbreaks. In the absence of information on specific traits involved in adaptation, genome-wide diversity can serve as a proxy for adaptive potential (Ørsted et al. 2019), and is especially relevant for polygenic traits (Harrisson et al. 2014). Low genetic variation across gene-coding loci in Torrey pines relative to other conifer species suggests limited capacity to respond to selection in a changing environment (Ørsted et al. 2019). TPEX, with the highest expected heterozygosity, may have the highest adaptive capacity of the populations. However, given strong genetic differentiation among TPEX, TPSR, and the island, the full evolutionary potential of the species is captured among individuals across all populations.

Our findings suggest that despite low genetic diversity, Torrey pines may have some capacity to respond adaptively to ongoing bark beetle pressure. Genetic differences between case and control trees in functionally relevant genes suggest that there is heritable variation in the population for traits that influence beetle attack. Other systems have revealed that variation in adaptively important genes can be maintained even when neutral genetic diversity is low (Robinson et al. 2016; Mable 2019). In the case of defense adaptations, costly resistance variants may be maintained at low frequency due to past selection by bark beetles or exposure to similar herbivorous insects and pathogens (Budde et al. 2016). In theory, as beetle outbreaks increase in frequency and severity, selection against susceptible individuals can increase the frequency of resistance alleles and the resulting average fitness of the population (McKinney et al. 2014). However, if only a small percentage of the population has resistance, high mortality during outbreaks may lead to population bottlenecks, drift, and inbreeding (McKinney et al. 2014; Budde et al. 2016). Torrey pines have an uncertain coevolutionary history with *Ips paraconfusus*, and the native bark beetle is only suspected of being a significant mortality agent in pines since the 1940's (Murray et al. 2013). It is possible that unknown historic selection events by *Ips* or other insects and pathogens could have maintained resistance variants in Torrey pines that can increase in frequency with new outbreaks. Potential resistance variation may be lower in the naïve island population where bark beetles have been historically absent, unless other important selective agents have co-evolved with this population.

Future directions and conservation implications

We present an initial study using RNA-Seq as a conservation tool to assess gene-coding diversity in the rare Torrey pine as a proxy for adaptive potential to respond to environmental stressors, with a focus on pressure from *Ips paraconfusus* bark beetles. Low genetic diversity across the genome indicates limited evolutionary potential for the species. However, genetic differences between case and control trees associated with beetle attack suggests that beetles may be responding to genetically controlled traits, indicating potential for the population to shift towards more adaptive phenotypes. Putative candidate genes for resistance are related to defense against microorganisms and water regulation.

We caution that our findings are preliminary but provide reason to further explore the putative genetic basis of resistance in a common garden setting. A long-term common garden would allow screening for multiple resistance phenotypes, assessing their genetic basis,

and evaluating the prevalence and degree of resistance in the island and mainland populations with contrasting *Ips* outbreak histories. Additionally, water treatments could disentangle the contribution of genotype and the environment to resistance phenotypes. A robust genome assembly for Torrey pine, when more technically and economically feasible, would benefit these analyses. It would further facilitate the discrimination of large gene families and allow one to test whether differential gene expression and paralogs are confounding genetic variation in the present study.

From a conservation standpoint, our finding of population structure suggests that genotypes across all populations are needed to preserve the evolutionary potential of the species. Given low genetic diversity and small population size, mixing individuals from TPEX and TPSR may minimize inbreeding and loss of alleles to drift. Fitness consequences of experimental crosses can be evaluated in a common garden. Interestingly, Hamilton et al. (2017) have demonstrated increased fitness among island-mainland F₁ hybrids relative to either parent population in growth and fecundity. Future studies can explore to what extent growth-defense trade-offs affect resistance. Furthermore, given that the mainland population has experienced multiple severe beetle outbreaks, the surviving pool of individuals may theoretically contain an increased proportion of resistance variants. Following further genomic assessment and validation, resistance variants from a large number of maternal sources should be represented in seed bank collections and reintroduction efforts. Finally, methods of long-term storage of viable germplasm for regenerating populations can provide increased assurance that the full extent of genome-wide diversity of the species is available in the future. By maximizing genome-wide and adaptive diversity, this and future studies can help preserve the evolutionary potential of the rare and iconic Torrey pine to respond to continued bark beetle outbreaks and other stresses with climate change.

Declarations

Acknowledgments:

We thank Darren Smith and Cara Stafford from California State Parks for permission to carry out this work at the Torrey Pines State Natural Reserve. Our gratitude to Clarice Mendoza, Emily Burson, Christa Horn, Charlie de la Rosa, Ruby Iacuaniello, Tony Rowe, Kathleen Thommes, Ray Lish, Javier Quiroz, the San Diego Zoo Safari Park horticulture team, and numerous San Diego Zoo Wildlife Alliance staff and volunteers for field assistance. We thank John Clark for advising this project, Jill Hamilton and Jessica Wright for valuable discussions, two anonymous reviewers for improving the manuscript, and Brian James at Sanford Burnham Prebys Discovery Institute for sequencing. This work was supported by a postdoctoral fellowship to S.E.S. and grants through SDZWA.

Funding:

This research was funded by a postdoctoral fellowship to S.E.S. and grants through SDZWA.

Conflicts of interest/Competing interests:

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval:

Not applicable

Consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and material:

The datasets generated during the current study will be available in public repositories upon manuscript acceptance.

Code availability:

Not applicable

References

1. Aitken SN, Yeaman S, Holliday JA, et al (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1:95–111. <https://doi.org/10.1111/j.1752-4571.2007.00013.x>
2. Alberto FJ, Aitken SN, Alía R, et al (2013) Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology* 19:1645–1661. <https://doi.org/10.1111/gcb.12181>
3. Anderegg WRL, Hicke JA, Fisher RA, et al (2015) Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytol* 208:674–683. <https://doi.org/10.1111/nph.13477>
4. Aukema BH, Zhu J, Møller J, et al (2010) Predisposition to bark beetle attack by root herbivores and associated pathogens: Roles in forest decline, gap formation, and persistence of endemic bark beetle populations. *Forest Ecology and Management* 259:374–382. <https://doi.org/10.1016/j.foreco.2009.10.032>
5. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
6. Boone CK, Keefover-Ring K, Mapes AC, et al (2013) Bacteria Associated with a Tree-Killing Insect Reduce Concentrations of Plant Defense Compounds. *J Chem Ecol* 39:1003–1006. <https://doi.org/10.1007/s10886-013-0313-0>
7. Boys J, Cherry M, Dayanandan S (2005) Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany* 92:833–841. <https://doi.org/10.3732/ajb.92.5.833>
8. Budde KB, Nielsen LR, Ravn HP, Kjær ED (2016) The Natural Evolutionary Potential of Tree Populations to Cope with Newly Introduced Pests and Pathogens—Lessons Learned From Forest Health Catastrophes in Recent Decades. *Curr Forestry Rep* 2:18–29. <https://doi.org/10.1007/s40725-016-0029-9>
9. Conomos MP, Miller MB, Thornton TA (2015) Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. *Genet Epidemiol* 39:276–293. <https://doi.org/10.1002/gepi.21896>
10. Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
11. de Paula Santos Martins C, Pedrosa AM, Du D, et al (2015) Genome-Wide Characterization and Expression Analysis of Major Intrinsic Proteins during Abiotic and Biotic Stresses in Sweet Orange (*Citrus sinensis* L. Osb.). *PLoS ONE* 10:e0138786. <https://doi.org/10.1371/journal.pone.0138786>
12. Dillon SK, Nolan MF, Matter P, et al (2013) Signatures of adaptation and genetic structure among the mainland populations of *Pinus radiata* (D. Don) inferred from SNP loci. *Tree Genetics & Genomes* 9:1447–1463. <https://doi.org/10.1007/s11295-013-0650-8>
13. Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
14. Emerick JJ, Snyder AI, Bower NW, Snyder MA (2008) Mountain Pine Beetle Attack Associated with Low Levels of 4-Allylanisole in Ponderosa Pine. *ENVIRONMENTAL ENTOMOLOGY* 37:5
15. Falk T, Herndon N, Grau E, et al (2018) Growing and cultivating the forest genomics database, TreeGenes. *Database (Oxford)* 2018:. <https://doi.org/10.1093/database/bay084>
16. Ferrenberg S, Kane JM, Mitton JB (2014) Resin duct characteristics associated with tree resistance to bark beetles across lodgepole and limber pines. *Oecologia* 174:1283–1292. <https://doi.org/10.1007/s00442-013-2841-2>

17. Franceschi VR, Krokene P, Christiansen E, Krekling T (2005) Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist* 167:353–376. <https://doi.org/10.1111/j.1469-8137.2005.01436.x>
18. Fuentes-Pardo AP, Ruzzante DE (2017) Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology* 26:5369–5406. <https://doi.org/10.1111/mec.14264>
19. Gaylord ML, Kolb TE, Pockman WT, et al (2013) Drought predisposes piñon–juniper woodlands to insect attacks and mortality. *New Phytol* 198:567–578. <https://doi.org/10.1111/nph.12174>
20. Geraldes A, Pang J, Thiessen N, et al (2011) SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Molecular Ecology Resources* 11:81–92. <https://doi.org/10.1111/j.1755-0998.2010.02960.x>
21. Gogarten SM, Sofer T, Chen H, et al (2019) Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* 35:5346–5348. <https://doi.org/10.1093/bioinformatics/btz567>
22. Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
23. Gugger PF, Cokus SJ, Sork VL (2016) Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes* 12:15. <https://doi.org/10.1007/s11295-016-0975-1>
24. Hamilton JA, Royauté R, Wright JW, et al (2017) Genetic conservation and management of the California endemic, Torrey pine (*Pinus torreyana* Parry): Implications of genetic rescue in a genetically depauperate species. *Ecol Evol* 1–12. <https://doi.org/10.1002/ece3.3306>
25. Hamrick JL, Godt MJW, Sherman-Broyles SLS (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests (Netherlands)* 95
26. Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol Appl* 7:1008–1025. <https://doi.org/10.1111/eva.12149>
27. Hart AJ, Ginzburg S, Xu M (Sam), et al (2020) EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources* 20:591–604. <https://doi.org/10.1111/1755-0998.13106>
28. Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting Adaptive Potential, Population Size, and Conservation. *Trends in Ecology & Evolution* 32:506–517. <https://doi.org/10.1016/j.tree.2017.03.012>
29. Huerta-Cepas J, Szklarczyk D, Forslund K, et al (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286–D293. <https://doi.org/10.1093/nar/gkv1248>
30. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94. <https://doi.org/10.1186/1471-2156-11-94>
31. Keeling Cl, Bohlmann J (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens*. *New Phytologist* 170:657–675. <https://doi.org/10.1111/j.1469-8137.2006.01716.x>
32. Kolosova N, Bohlmann J (2012) Conifer defense against insects and fungal pathogens. In: Matyssek R, Schnyder H, Oßwald W, et al. (eds) *Growth and Defence in Plants*. Springer Berlin Heidelberg, pp 85–109
33. Ledig FT, Conkle MT (1983) Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex Carr.). *Evolution* 37:79–85. <https://doi.org/10.2307/2408176>
34. Ledig FT, Vargas-Hernández JJ, Johnsen KH (1998) The Conservation of Forest Genetic Resources: Case Histories from Canada, Mexico, and the United States. *Journal of Forestry* 96:32–41
35. Leimu R, Mutikainen P, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology* 94
36. Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
37. Mable BK (2019) Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conserv Genet* 20:89–100. <https://doi.org/10.1007/s10592-018-1129-9>
38. Mackay J, Dean JFD, Plomion C, et al (2012) Towards decoding the conifer giga-genome. *Plant Mol Biol* 80:555–569. <https://doi.org/10.1007/s11103-012-9961-7>
39. Manichaikul A, Mychaleckyj JC, Rich SS, et al (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26:2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>

40. Mata R de la, Hood S, Sala A (2017) Insect outbreak shifts the direction of selection from fast to slow growth rates in the long-lived conifer *Pinus ponderosa*. PNAS 114:7391–7396. <https://doi.org/10.1073/pnas.1700032114>
41. McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Molecular Ecology Resources 17:656–669. <https://doi.org/10.1111/1755-0998.12613>
42. McKinney LV, Nielsen LR, Collinge DB, et al (2014) The ash dieback crisis: genetic variation in resistance can prove a long-term solution. Plant Pathology 63:485–499. <https://doi.org/10.1111/ppa.12196>
43. Meddens AJH, Hicke JA, Ferguson CA (2012) Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. Ecological Applications 22:1876–1891. <https://doi.org/10.1890/11-1785.1>
44. Mosca E, González-Martínez SC, Neale DB (2014) Environmental versus geographical determinants of genetic structure in two subalpine conifers. New Phytologist 201:180–192. <https://doi.org/10.1111/nph.12476>
45. Murray TA, Kohler GR, Willhite EA (2013) *Ips paraconfusus* Lanier (Coleoptera: Curculionidae): New Records of the California Fivespined Ips from Washington State and the Columbia River Gorge in Oregon. The Coleopterists Bulletin 67:28–31. <https://doi.org/10.1649/072.067.0106>
46. Nadeau S, Godbout J, Lamothe M, et al (2015) Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: A comparison between eastern and western North American postglacial colonization histories. American Journal of Botany 102:1342–1355. <https://doi.org/10.3732/ajb.1500160>
47. Netherer S, Matthews B, Katzensteiner K, et al (2015) Do water-limiting conditions predispose Norway spruce to bark beetle attack? New Phytol 205:1128–1141. <https://doi.org/10.1111/nph.13166>
48. Ørsted M, Hoffmann AA, Sverrisdóttir E, et al (2019) Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. PLoS Genet 15:e1008205. <https://doi.org/10.1371/journal.pgen.1008205>
49. Ott DS, Yanchuk AD, Huber DPW, Wallin KF (2011) Genetic Variation of Lodgepole Pine, *Pinus contorta* var. *latifolia*, Chemical and Physical Defenses that Affect Mountain Pine Beetle, *Dendroctonus ponderosae*, Attack and Tree Mortality. J Chem Ecol 37:1002. <https://doi.org/10.1007/s10886-011-0003-8>
50. Paine TD, Raffa KF, Harrington TC (1997) Interactions among scolytid bark beetles, their associated fungi, and live host conifers. Annual Review of Entomology 42:179–206. <https://doi.org/10.1146/annurev.ento.42.1.179>
51. Pertea M, Pertea GM, Antonescu CM, et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295. <https://doi.org/10.1038/nbt.3122>
52. Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. Annual Review of Ecology, Evolution, and Systematics 37:187–214
53. Provan J, Soranzo N, Wilson NJ, et al (1999) A Low Mutation Rate For Chloroplast Microsatellites. Genetics 153:943–947
54. Raffa (2014) Terpenes Tell Different Tales at Different Scales: Glimpses into the Chemical Ecology of Conifer - Bark Beetle - Microbial Interactions. J Chem Ecol 40:1–20. <https://doi.org/10.1007/s10886-013-0368-y>
55. Raffa, Andersson MN, Schlyter F (2016) Host Selection by Bark Beetles. In: Advances in Insect Physiology. Elsevier, pp 1–74
56. Raffa, Aukema BH, Bentz BJ, et al (2008) Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. BioScience 58:501–517. <https://doi.org/10.1641/B580607>
57. Raffa KF, Grégoire J-C, Lindgren BS (2015) Natural history and ecology of bark beetles. In: Bark Beetles: Biology and Ecology of Native and Invasive Species. Elsevier Inc, pp 1–40
58. Rellstab C, Dauphin B, Zoller S, et al (2019) Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of *Pinus cembra*. Mol Ecol Resour 19:536–551. <https://doi.org/10.1111/1755-0998.12986>
59. Robinson JA, Ortega-Del Vecchyo D, Fan Z, et al (2016) Genomic Flatlining in the Endangered Island Fox. Current Biology 26:1183–1189. <https://doi.org/10.1016/j.cub.2016.02.062>
60. Rosner S, Hannrup B (2004) Resin canal traits relevant for constitutive resistance of Norway spruce against bark beetles: environmental and genetic variability. Forest Ecology and Management 200:77–87. <https://doi.org/10.1016/j.foreco.2004.06.025>
61. Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. Current Opinion in Plant Biology 10:162–167. <https://doi.org/10.1016/j.pbi.2007.01.011>
62. Schieber C, Hammerbacher A, Birgersson G, et al (2012) Inducibility of chemical defenses in Norway spruce bark is correlated with unsuccessful mass attacks by the spruce bark beetle. Oecologia 170:183–198. <https://doi.org/10.1007/s00442-012-2298-8>

63. Schuler MA (1996) The role of cytochrome P450 monooxygenases in plant-insect interactions. *Plant Physiol* 112:1411–1419
64. Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* 4:326–337. <https://doi.org/10.1111/j.1752-4571.2010.00157.x>
65. Shea PJ, Neustein M (1995) Protection of a rare stand of Torrey pine from *Ips paraconfusus*. In: Application of semiochemicals for management of bark beetle infestations: Proceedings of an informal conference: Annual meeting of the Entomological Society of America, December 12-16, 1993. [Ogden, UT]: U.S. Dept. of Agriculture, Forest Service, Intermountain Research Station
66. Six DL, Vergobbi C, Cutter M (2018) Are Survivors Different? Genetic-Based Selection of Trees by Mountain Pine Beetle During a Climate Change-Driven Outbreak in a High-Elevation Pine Forest. *Front Plant Sci* 9:. <https://doi.org/10.3389/fpls.2018.00993>
67. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100:9440–9445. <https://doi.org/10.1073/pnas.1530509100>
68. Telford A, Cavers S, Ennos RA, Cottrell JE (2015) Can we protect forests by harnessing variation in resistance to pests and pathogens? *Forestry (Lond)* 88:3–12. <https://doi.org/10.1093/forestry/cpu012>
69. Tian S, Wang X, Li P, et al (2016) Plant Aquaporin AtPIP1;4 Links Apoplastic H₂O₂ Induction to Disease Immunity Pathways. *Plant Physiology* 171:1635–1650. <https://doi.org/10.1104/pp.15.01237>
70. Tian T, Liu Y, Yan H, et al (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* 45:W122–W129. <https://doi.org/10.1093/nar/gkx382>
71. Waters ER, Schaal BA (1991) No variation is detected in the chloroplast genome of *Pinus torreyana*. *Can J For Res* 21:1832–1835. <https://doi.org/10.1139/x91-253>
72. Wegrzyn JL, Liechty JD, Stevens KA, et al (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196:891–909. <https://doi.org/10.1534/genetics.113.159996>
73. Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38:1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
74. Westbrook JW, Resende MFR, Munoz P, et al (2013) Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential. *New Phytol* 199:89–100. <https://doi.org/10.1111/nph.12240>
75. Whittall JB, Syring J, Parks M, et al (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19:100–114. <https://doi.org/10.1111/j.1365-294X.2009.04474.x>
76. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *PNAS* 84:9054–9058
77. Yanchuk AD, Murphy JC, Wallin KF (2008) Evaluation of genetic variation of attack and resistance in lodgepole pine in the early stages of a mountain pine beetle outbreak. *Tree Genetics & Genomes* 4:171–180. <https://doi.org/10.1007/s11295-007-0098-9>
78. Zavarin E, Hathaway W, Reichert T, Linhart YB (1967) Chemotaxonomic study of *Pinus torreyana* Parry turpentine. *Phytochemistry* 6:1019–1023. [https://doi.org/10.1016/S0031-9422\(00\)86055-0](https://doi.org/10.1016/S0031-9422(00)86055-0)
79. Zeilmaker T, Ludwig NR, Elberse J, et al (2015) DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in *Arabidopsis*. *The Plant Journal* 81:210–222. <https://doi.org/10.1111/tpj.12719>
80. Zhao S, Erbilgin N (2019) Larger Resin Ducts Are Linked to the Survival of Lodgepole Pine Trees During Mountain Pine Beetle Outbreak. *Front Plant Sci* 10:1459. <https://doi.org/10.3389/fpls.2019.01459>
81. Zheng X, Levine D, Shen J, et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>
82. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824. <https://doi.org/10.1038/ng.2310>
83. Zimin A, Stevens KA, Crepeau MW, et al (2014) Sequencing and assembly of the 22-Gb Loblolly pine genome. *Genetics* 196:875–890. <https://doi.org/10.1534/genetics.113.159715>
84. Zimin AV, Stevens KA, Crepeau MW, et al (2017) An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience* 6:. <https://doi.org/10.1093/gigascience/giw016>

Tables

Table 1

Genetic diversity summary statistics for Torrey pine populations and case/control groups. Counts for total and LD-pruned, biallelic SNPs with a 95% call rate in the full dataset ($N = 40$) are reported. SNPs in sub-populations had a minor allele count of at least 1. H_E , H_0 , and F were calculated from 4192 LD-pruned SNPs for each population.

Population	<i>N</i>	Total SNPs	LD-pruned SNPs	H_E	H_0	F
All	40	9599	4192	0.208	0.046	0.783
Mainland	36	1888	1027	0.073	0.044	0.409
TPEX	13	1424	790	0.053	0.053	0.053
TPSR	23	1332	776	0.044	0.039	0.137
Island (SRI)	4	750	346	0.036	0.064	-0.565
Cases/Controls	32	1762	994	0.074	0.043	0.423
Controls	16	1654	936	0.070	0.045	0.403
Cases	16	1708	962	0.075	0.042	0.443

Table 2

F_{ST} statistics (Weir & Cockerham 1984) for population comparisons of Torrey pines based on 4192 LD-pruned SNPs.

Population 1	Population 2	F_{ST}
Island (SRI)	Mainland	0.915
Island (SRI)	TPEX	0.938
Island (SRI)	TPSR	0.947
TPEX	TPSR	0.523
Case	Control	-0.013

Table 3

Functional annotations for top 19 SNPs from GEMMA association tests

PITA Contig	Transcript/ Gene ID ¹	GenBank/UniProt ID	Description ¹	SNP	Alleles	GEMMA Q- values ²	Shared GEMMA Tests ³
C4983731	PITA_01744	XP_027164094.1	protein DMR6-like oxygenase 2-like isoform X3	45607	T/C	2.35E-02	*
				45608	C/G	2.35E-02	*
scaffold101749	PITA_27080	XP_006829063.1	probable polygalacturonase	78303	T/A	1.47E-02	*
				78312	T/A	4.33E-03	
				78342	A/T	1.47E-02	
				78561	T/C	2.35E-02	*
scaffold145484	PITA_02067	Q8GVE8 CAPP4_ARATH	phosphoenolpyruvate carboxylase 4	206818	G/A	1.75E-03	*
scaffold156333	PITA_46475	P49243 FABH_ARATH	3-oxoacyl-[acyl-carrier-protein] synthase III, chloroplastic	128636	T/C	3.34E-01	*
scaffold20	MSTRG.12247	XP_002884019.1	probable aquaporin PIP2-8	280734	T/C	4.91E-07	*
scaffold225151	PITA_19888	O81973 C93A3_SOYBN	cytochrome P450 93A3	70681	G/A	1.43E-01	*
				70699	T/A	3.21E-01	*
				70706	G/A	3.21E-01	*
scaffold49411	MSTRG.13305		unknown	31872	C/T	7.38E-01	*
scaffold69794	MSTRG.6166	XP_028762983.1	uncharacterized protein	46824	T/A	3.34E-01	*
scaffold7116	MSTRG.28184	XP_006841522.1	major allergen Pru ar 1	128586	A/T	4.57E-01	*
				128618	G/C	4.22E-01	*
				128710	A/T	6.56E-04	*
scaffold95502	MSTRG.1076	XP_021284657.1	peroxidase 12	49473	G/A	2.35E-02	
super4429	PITA_38813	O81884 GALDH_ARATH	L-galactose dehydrogenase	198209	G/A	1.47E-02	

¹PITA Gene IDs and descriptions are listed for SNPs matching annotated regions of the *Pinus taeda* reference genome. For SNPs without PITA Gene IDs, transcript IDs from the Stringtie assembly (beginning MSTRG) are listed, with descriptions from EnTAP functional annotations (see Methods).

²Bold highlighting indicates 11 SNPs with significant associations with case-control status from the GEMMA mixed model after correcting for multiple testing.

³Asterisks indicate 15 SNPs in the top 5% of the distribution for each association test in GEMMA with subsets of the data (see Methods).

Figures

Figure 1.

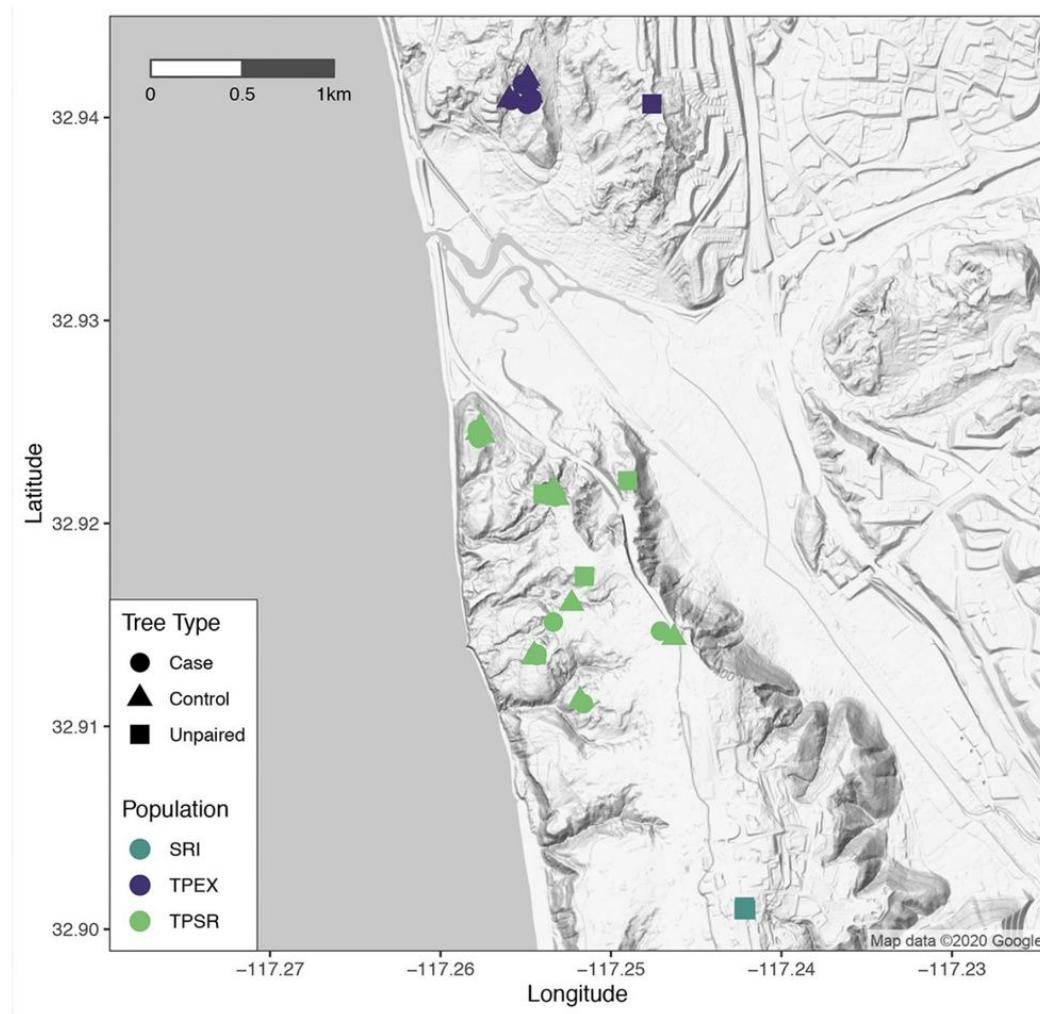


Figure 1

Map of sampling locations of Torrey pine trees (n = 40). Colors indicate population while shape reflects case/control status. 'Unpaired' trees indicate additional trees sampled to capture genetic diversity across the species but were not included in association tests. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2.

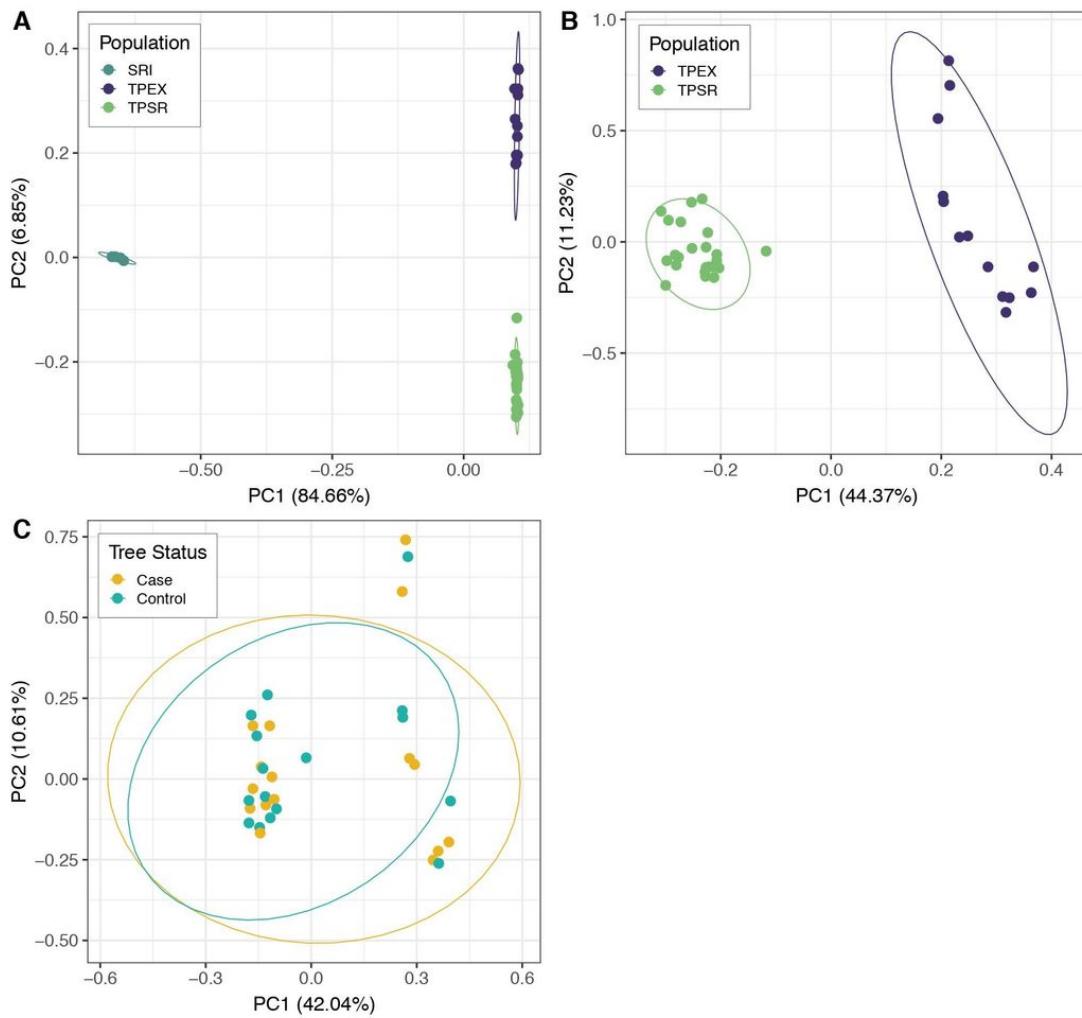


Figure 2

PC-AiR plots displaying the first two axes for LD-pruned SNPs for a) all individuals ($n = 40$), b) mainland individuals ($n = 36$), and c) paired case/control samples ($n = 32$). Points represent individuals and are color-coded by population or case/control status. The 95% confidence ellipse per group assuming a multivariate t-distribution is shown.

Figure 3.

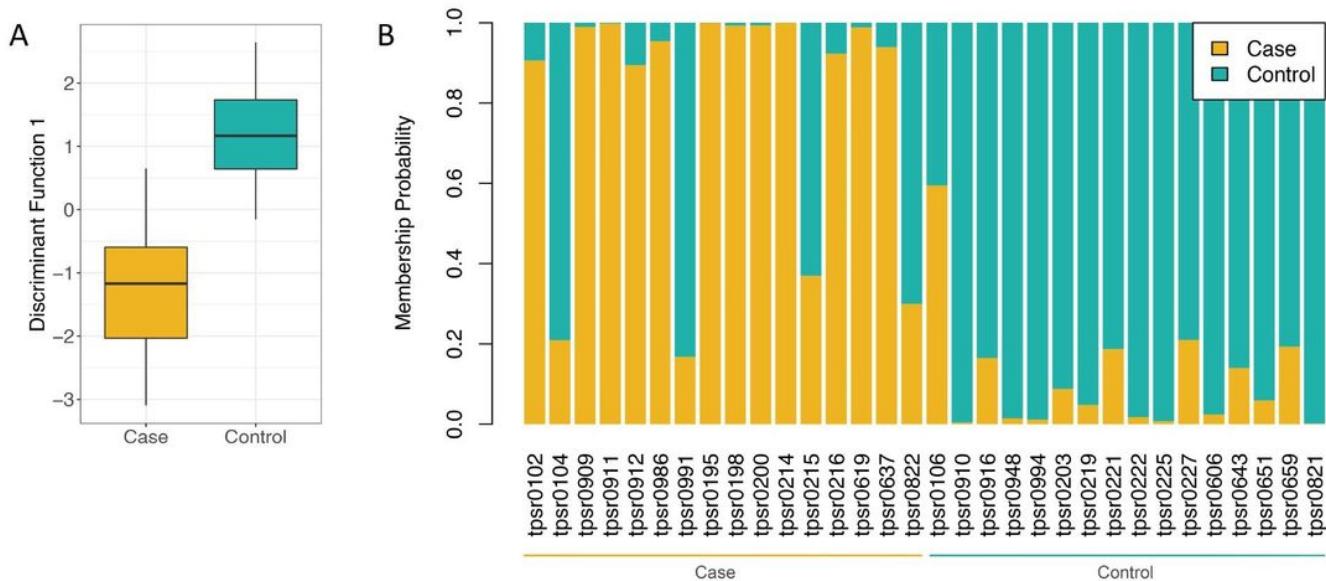


Figure 3

DAPC of the first discriminant function of 12 PCs summarizing 1560 biallelic SNPs ($n = 32$ case/control paired samples). a) Box plot of discriminant function 1 by case/control status. b) Component plot of membership probability of samples to case or control groups based on multivariate genotypes.

Figure 4.

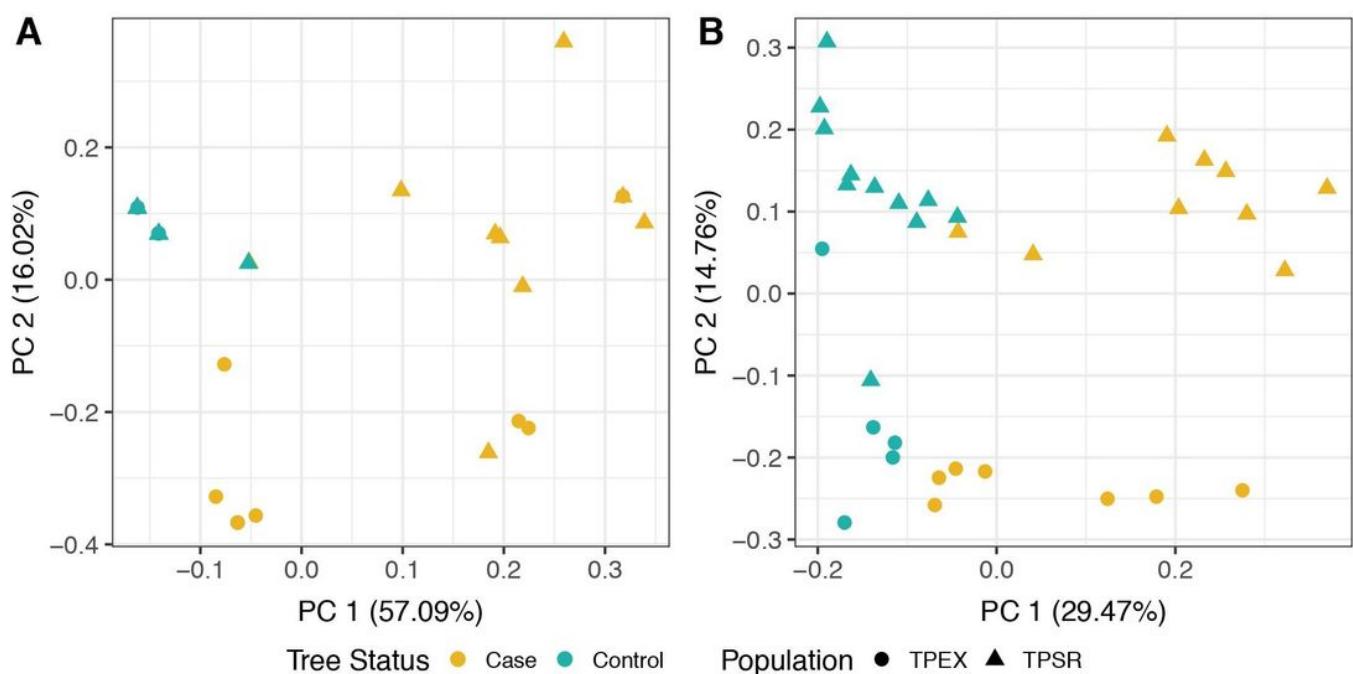


Figure 4

PCA plot of SNPs associated with case/control status in the GEMMA mixed model. Plots display a) SNPs in Table 3 and b) SNPs in the top 5% of the distribution. Only SNPs with the highest Q-value per gene are displayed.

Figure. 5.

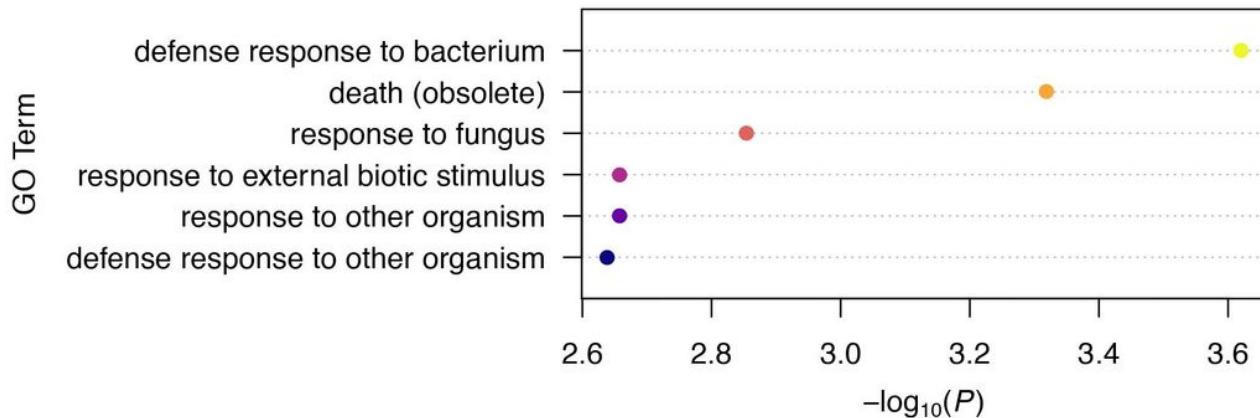


Figure 5

Biological process gene ontology terms significantly enriched among the top 153 SNPs (top 10%) relative to the background SNP set tested in GEMMA. The $-\log_{10}$ transformed P-values from AgriGO v2.0 are displayed, each of which met an FDR threshold of 0.05 after correction for multiple testing. EnTAP assigned biological process GO terms to 515 of 1526 SNPs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SteeleetalTorreypinesCoGeSupMtIsResubFinal.pdf](#)