Identifying genetic variation associated with environmental gradients and drought-tolerance phenotypes in ponderosa pine

Mengjun Shu[1] and Emily Moran[1]

[1]University of California Merced

April 29, 2023

Abstract

As climate changes, understanding the genetic basis of local adaptation in plants becomes an ever more pressing issue. Combining Genotype-Environment Association (GEA) with Genotype-Phenotype Association (GPA) analysis has an exciting potential to uncover the genetic basis of environmental responses. We use these approaches to identify genetic variants linked to local adaptation to drought in *Pinus ponderosa.* Over 4 million SNPs were identified using 223 individuals from across the Sierra Nevada of California. We found 1458 associated with five largely uncorrelated climate variables, with the largest number (1151) associated with April 1st snowpack. We also conducted a greenhouse study with various drought-tolerance traits measured in seedlings grown in control and drought treatments. 817 SNPs were associated with control-condition trait values, while 1154 were associated with responsiveness of these traits to drought. While no individual SNPs were associated with both the environmental variables and the measured traits, several annotated genes were associated with both, particularly those involved in cell wall formation, biotic and abiotic stress responses, and ubiquitination. However, the functions of many of the associated genes have not yet been determined due to the lack of gene annotation information for conifers. Future studies are needed to assess the developmental roles and ecological significance of these unknown genes.

Identifying genetic variation associated with environmental gradients and drought-tolerance phenotypes in ponderosa pine

Mengjun Shu[1], Emily V. Moran[1]

Author Affiliations:

1. Life and Environmental Sciences, University of California, Merced, CA, 5200 N Lake Rd, Merced CA, 95343, USA

Corresponding Author:

Mengjun Shu

Email: mshu@ucmerced.edu

ORCiD: 0000-0002-6323-2664

Abstract

As climate changes, understanding the genetic basis of local adaptation in plants becomes an ever more pressing issue. Combining Genotype-Environment Association (GEA) with Genotype-Phenotype Association (GPA) analysis has an exciting potential to uncover the genetic basis of environmental responses. We use these approaches to identify genetic variants linked to local adaptation to drought in*Pinus ponderosa* . Over 4 million SNPs were identified using 223 individuals from across the Sierra Nevada of California. We found 1458 associated with five largely uncorrelated climate variables, with the largest number (1151) associated with

April 1st snowpack. We also conducted a greenhouse study with various drought-tolerance traits measured in seedlings grown in control and drought treatments. 817 SNPs were associated with control-condition trait values, while 1154 were associated with responsiveness of these traits to drought. While no individual SNPs were associated with both the environmental variables and the measured traits, several annotated genes were associated with both, particularly those involved in cell wall formation, biotic and abiotic stress responses, and ubiquitination. However, the functions of many of the associated genes have not yet been determined due to the lack of gene annotation information for conifers. Future studies are needed to assess the developmental roles and ecological significance of these unknown genes.

Keywords: climate change, adaptive genetic variation, environmental association, phenotypic association, GBS, SNP

Introduction

Genomics promises exciting advances towards understanding adaptive genetic variation and evolutionary potential of plants under a rapidly changing and often increasingly variable environment (Hoffmann & Sgrò 2011; Savolainen *et al.* 2013; Harrisson *et al.* 2014). Intraspecific genetic variation represents the potential for adaptive change in response to new selective challenges, which is critical for local species persistence under environmental change (Rice & Emery 2003; Bell & Gonzalez 2009). Adaptation to local climate conditions has been considered typical for tree populations (Langlet 1971; Ying & Liang 1994; Kitzmiller 2005; Wright 2007), but organisms with such long generation times and a sessile lifestyle can become maladapted if environmental shifts rapidly occur (Aitken *et al.* 2008; Anderson*et al.* 2012; Alberto *et al.* 2013). Plants also exhibit plastic changes in their growth form and physiology in response to stress, and the level of plasticity can itself be heritable (Van Kleunen & Fischer 2005; Auld *et al.* 2010) and may be under the selection (Zettlemoyer & Peterson 2021). Understanding the distribution of genetic variation related to environmental responses may help us better predict changes and manage forests in a shifting climate (Neale & Kremer 2011; Oney *et al.* 2013). This includes selecting seed sources for restoration or breeding that have desirable characteristics such as drought tolerance (Beaulieu *et al.* 2014; Isik 2014).

Landscape genomics offers enormous potential to discover genes responsible for local adaptation by investigating the statistical association between genetic variation at individual loci and the causative environmental factors (Eckert *et al.* 2010, 2015; Sork*et al.* 2013; Lu *et al.* 2019). This approach is sometimes known as Genotype-Environment Association (GEA) analysis. Prior studies in *Arabidopsis* – the primary plant model organism - have found that environmentally-associated SNPs can predict performance in common gardens (Hancock *et al.* 2011). A *Pinus pinaster* study suggests this could be true in trees as well, even when only a modest number of the genetic variants involved have been identified (Jaramillo-Correa *et al.* 2015). However, GEA studies don't by themselves reveal why specific alleles are more prevalent in particular environments – for example, are they responsible for selectively favored traits? Genotype-Phenotype Association (GPA) analysis identifies loci linked to a specific phenotype (Eckert *et al.* 2009; Holliday*et al.* 2010). In plant GPA studies, individuals are typically grown in a common environment to eliminate the effects of environmental variation on phenotypes. However, this approach does not reveal whether a trait variant would be favored in the field. GEA and GPA association are thus complementary, and combining them might better identify the loci and traits that are selectively favored in particular conditions than either could alone (Eckert *et al.* 2015; Mahony *et al.*2020).

The large genome size of conifer trees (>19 GBP) represents a challenge for analysis. Most association studies in conifers have focused on SNPs within a few hundred genes (Eckert *et al.* 2009, 2015; Holliday *et al.* 2010; Hamilton *et al.* 2013; Dillon*et al.* 2014; Housset *et al.* 2018), or fewer than 2,000 genome-wide SNPs (Uchiyama *et al.* 2013). One notable exception is a recent study on lodgepole pine that used a sequence capture dataset created by mapping the *Pinus contorta* transcriptome to the*P. taeda* genome sequence (Mahony *et al.* 2020). A genome-wide SNP climate-association study was also recently completed for *P. lambertiana* , one of the few other pines species with a full genome sequence (Weiss *et al.* 2022). Still, most conifers have neither a published genome sequence nor a complete transcriptome. Though targeted sequencing is efficient, candidate gene approaches may miss other vital genes with previously unsuspected

roles in local adaptation, and focusing solely on variants within genes may miss significant variants within regulatory regions.

Several approaches to identifying more genetic variants for genome-wide association studies (GWAS) utilizing next-generation sequencing (NGS) have been proposed in recent years (Davey *et al.* 2011; Poland & Rife 2012). Genotyping-by-Sequencing (GBS), which can generate tens of thousands of SNP markers (Single Nucleotide Polymorphisms) without the need for a reference genome or whole transcriptome, has emerged as a cost-effective strategy (Elshire *et al.* 2011; Andrews *et al.* 2016). By combining the power of multiplexed NGS with restriction-enzyme-based genome complexity reduction, GBS can genotype large populations of individuals for thousands of SNPs in an increasingly rapid and inexpensive way (Poland *et al.* 2012; Poland & Rife 2012).

Despite the high economic and ecological importance of ponderosa pine (*Pinus ponderosa*) in the western United States (Graham & Jain 2005), no previous study has attempted to identify the relationship between gene sequence variation and drought tolerance in this species. Some studies have investigated *P. ponderosa's* evolutionary history and phylogeography using mitochondrial DNA markers; these reflect the long-term biogeographical process contributing to the modern distribution of the species but have limited adaptive significance in themselves (Johansen & Latta 2003; Potter *et al.* 2013). Other studies have emphasized the importance of intraspecific variation of*P. ponderosa* in environmental responses but focus on the phenotypic variation within and among populations without identifying the underlying genetic variation (Kolb *et al.* 2016; Maguire*et al.* 2018). California's historic 2012–2016 drought may represent an increasingly common condition as climate changes (Griffin & Anchukaitis 2014; Berg & Hall 2015). Such "hot droughts" can lead to mass tree mortality, even in relatively drought tolerant species like ponderosa pine, negatively impacting the sustainability of conifer forests (Fettig *et al.* 2019). A deep understanding of the genetic basis of adaptation in ponderosa pine and other western conifers is critical for successful reforestation and conservation programs.

In this study, we conducted a GEA analysis on 223 ponderosa pine genotypes from a range of climates across the central Sierra Nevada mountains of California. We then planted seeds collected from a subset of these trees in the greenhouse. The resulting seedlings provided the basis of a GPA analysis of putative drought-response traits. We ran gene annotation to ascribe biological function to the genes that the associated SNPs were in or adjacent to. Then we assessed overlap in SNP identity or gene functions among GEA and GPA association analysis that might indicate particular importance for local adaptation.

Materials and Methods

Sampling and DNA sequencing

In the 1970s, the Forest Service's Pacific Southwest Regional Genetic Resources Program planted clones of 302 wild ponderosa pines from diverse climate conditions in the central portion of the Sierra Nevada mountains in an orchard located in Chico, California. We chose 223 individual *P. ponderosa* genotypes from the orchard for the GEA analysis whose collection locations span the full climatic range included in the collection. The source locations for these genotypes (Fig. 1) fall within just one of the several genetic subdivisions previously identified in ponderosa pine (Conkle & Critchfield 1988; Williams 2009; Potter *et al.* 2015). Fresh needles were collected from these individuals and placed in labeled tea bags over silica gel to dry them and quickly preserve the DNA for extraction.

DNA was extracted from the dried needles using a modified Qiagen plant kits protocol by adding proteinase K and quantified using an Eppendorf BioSpectrometer (Eppendorf, AG, Germany). Samples were frozen and sent to the UC Davis Genome Center for library construction. Four 48-plex GBS libraries consisting of 47 DNA samples and negative control (no DNA) and one 36-plex GBS library composed of 35 DNA samples and negative control were prepared. The pool was quantified via qPCR using the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, MA, USA) for Illumina sequencing platforms, with 0.9X bead cleanup to remove small fragments (<250 bp). Additional DNA purification using the Zymo DNA Clean & Concentrator kit (Zymo Research, Irvine, CA) was performed to increase the purity of the extracted DNA. The libraries were then sequenced (single-end read 90 bp or 100 bp) using an Illumina HiSeq 4000 (Illumina, San Diego,

CA), one library per lane.

SNP calling and filtering

No reference genome is available for ponderosa pine (*Pinus ponderosa*), but one does exist for loblolly pine (*Pinus taeda)*(Neale *et al.* 2014; Zimin *et al.* 2014). Of the conifers that have been sequenced to date, *P. taeda* is the most closely related to *P. ponderosa* (Gernandt *et al.* 2009; Willyard*et al.* 2009). Furthermore, the *P. taeda* reference genome was successfully used to design probes for sequence capture in *P. contorta* (Suren *et al.* 2016; Yeaman *et al.* 2016), a distant relative. Based on preliminary analyses, we selected the Stacks v.2.2 pipeline (Rochette & Catchen 2017) with this reference genome (https://treegenesdb.org/FTP/Genomes/Pita/) for SNP calling (Shu 2020). Each step in the Stacks reference pipeline is performed internally in Stacks algorithms except alignment with BWA v.0.7.17 (Li & Durbin 2009) and the Samtools v.1.9 (Li 2011) step used to get read position. Default settings were used in Stacks, BWA and Samtools.

After calling the SNPs, we ran SnpEff (Cingolani *et al.* 2012) to identify the location of the gene containing each SNP. We used the database of annotated genome and the reference genome of loblolly pine v.2.01 in TreeGenes (http://treegenesdb.org/FTP/Genomes/Pita/v2.01/). The location of each SNP is listed in the output file of SnpEff as one of six primary location categories, including intragenic variants, intergenic variants, upstream SNPs, downstream SNPs, synonymous, and missense variants in the gene coding sequence. In Snp Eff, "intragenic" refers to SNPs in introns, while "missense" refers to any non-synonymous mutation in the transcribed region.

Many SNPs identified by GBS fall between genes and regulatory regions (in the intergenic regions) and likely have no direct effect on gene expression or function. In addition, because of the low amount of linkage disequilibrium in conifers (Namroud *et al.* 2008; Isik*et al.* 2016), any associations identified between such intergenic SNPs and a phenotype or environment of interest are likely false positives rather than reflecting linkage between the SNP and a causal variant. Therefore, we first filtered out the intergenic SNPs before running the association analysis using a Python script (https://github.com/shumengjun/LFMM).

Climate data

We obtained 30-year (1921–1950) averages of climate data for each genotype source location from the 270 m resolution California Basin Characterization Model (BCM) (Flint *et al.* 2013). These mid-20$^{\text{th}}$-century values were used instead of more recent climate data because they more closely resemble the conditions when the genotypes were establishing as seedlings.

For the GEA, we chose to focus on raw environmental variables rather than environmental PCA axes, as several previous studies have done (Eckert *et al.* 2010, 2015). This is because PCA associations can be challenging to interpret if, for example, the axes include both temperature and moisture variables. Instead, we used PCA (Fig. S4) to select five environmental variables that have low correlation with one another across tree source locations: mean climatic water deficit (CWD, a measure of evaporative demand exceeding soil moisture); mean minimum winter (December-February) temperature (TMIN); mean maximum summer (June - August) temperature of summer (TMAX); mean monthly winter precipitation (PPTW); and mean April 1$^{\text{st}}$ snowpack (PCK4). Other climate variables considered but not included in the analysis were actual evapotranspiration (AET), potential evapotranspiration (PET), excess water, recharge, runoff, snowfall, snowmelt, soil water storage, and snow sublimation.

Genotype-Environment association analysis

We used latent factor mixed model 2 (LFMM2) for GEA association, which has been shown to outperform similar approaches with several orders-of-magnitude faster computing (Caye *et al.* 2019), which also controls for the effects of demographic processes and population structure (Wang *et al.* 2017). This approach is robust to high amounts of missing data, such as GBS sequencing tends to produce, when sample sizes are >100 (Xuereb *et al.* 2017).

LFMM2 regression models combine fixed and latent effects with the following equation:

$$\mathbf{Y} = \mathbf{XB}^{T} + \mathbf{W} + \mathbf{E} .$$

$\mathbf{Y}$ is a matrix of genetic information measured from $p$genetic markers for n individuals, and $\mathbf{X}$ is a matrix of$d$ environmental variables measured for n individuals. The fixed effect sizes are recorded in the B matrix, which has dimension$p * d$. The $\mathbf{E}$ matrix represents residual errors with the same dimensions as the response matrix. The matrix $\mathbf{W}$is a matrix of rank K, defined by K latent factors where model choice procedures can determine K. The K factors represent unobserved confounders - usually geographical structure in the genotypes of the samples – defined as an n *K matrix, $\mathbf{U} . \mathbf{V}$ is a$p \times$ K matrix of loadings. The matrix U is obtained from the matrix's singular value decomposition (SVD):

$$\mathbf{W} = \mathbf{UV}^{T}$$

We used the two approaches implemented in the LEA v.2.6.0 R package to determine K: principal component analysis (PCA) and admixture analysis (Frichot *et al.* 2013; Frichot & François 2015). First, we ran the LEA function PCA to select the number of significant PCA components by computing Tracy-Widom tests with the LEA function Tracy.widom (Patterson *et al.* 2006). Second, we ran the LEA function snmf for K values between 1 and 5 with ten repetitions each. The most likely K value was identified by minimizing the cross-validation error evaluated in the 10-fold cross-validation procedure (Frichot & Francois, 2014). We then chose significant associations based on $p(<10^{-5})$ value.

Greenhouse experiment and phenotype measurements

Seedling traits that may be related to drought tolerance include height growth, root length, dry shoot weight, dry root weight, root-to-shoot dry mass ratio, and specific root length. Allocation to roots versus shoots can affect a seedling's ability to take up water (Brunner & Godbold 2007; Markesteijn & Poorter 2009; Moran *et al.* 2017a), while the number and size of needles affect the area over which they may lose water through the stomata (Parker 1949). Two other traits related to this latter point are stomata density and the number of rows on the adaxial versus abaxial sides of the needles. However, it should be noted that lower stomatal conductance can be achieved with either fewer stomata, smaller stomata, or by closing the stomata more often (Irvine*et al.* 1998; Ryan 2011).

Seeds were collected from a subset of 50 genotyped parent trees in the summer of 2018. We placed 2-3 mature cones from each mother tree into paper bags and put them in a warm, dry place until seeds were released. We aimed to have ten seedlings from each of the fifty maternal families in wet and dry treatments, 1000 seedlings in total. Forty-eight of the 50 families had enough seeds in their cones to be included in the experiment (Fig. S1). During winter 2018, the seeds were stratified to break dormancy by placing them in aerated water for 48 hours, then surface-drying them and storing them in plastic bags in the refrigerator (˜1.7°C) for six weeks. Because pines are wind-pollinated and outcrossing (Williams 2009), seeds from the same tree are mostly half-siblings, occasionally full-sibs.

Because the maximum first-year seedling root length observed in a pilot experiment was more than 110 cm, we used plastic tubes with an 8 cm width and 120 cm depth for planting. The bottom of each tube was capped with mesh to prevent the soil from falling out while allowing drainage. The lightweight clear tubes were wrapped in black plastic to keep roots in the dark. The planting soil was a mixture of 70% sand, 20% vermiculite, and 10% organic-rich potting mix to mimic the coarse texture of the soil of many Sierra Nevada conifer forests (Bales*et al.* 2011). To keep tubes upright, we used PVC pipes to build 10 frames that could each hold 100 tubes. Two seeds from each family were planted in each tube in February 2019, and two tubes from each family were randomly placed within each frame. In April 2019, we replanted more stratified seeds of the correct family in tubes without seedlings. All the tubes were watered every other day during the germination and seedling establishment period (February through June).

At the end of June 2019, all but one seedling per tube was removed, and alternating frames were assigned to the wet treatment and the dry treatment (5 frames containing up to 500 seedlings per treatment) (Fig. S1). The wet treatment group was watered twice weekly, and the drought treatment group was watered once every three weeks until mid-October (3.5 months). While wild ponderosa pine seedlings would receive little to

no precipitation during the summer, this occasional watering was necessary for the greenhouse environment. Temperatures inside the greenhouse in the low-elevation environment of Merced, CA, reached as high as 37°C on the hottest days and the soil volume of the tubes was limited, with no access to groundwater, both of which made evaporation and drought stress more intense than the no-precipitation condition in the wild.

Multiple phenotypic traits were measured during and after the greenhouse experiment. Only 42 out of 48 mother trees had enough germination to carry out these measurements across both treatments. We calculated shoot growth as final height minus height at the initiation of the treatments. The length of fresh roots was measured from the soil surface to the taproot tip immediately after the harvesting to avoid shrinkage. Following harvest, needles, fresh stems, and fresh roots of all the seedlings were separately put into paper bags and dried at 75 °C for 48 hours. We measured root dry mass (RW) and shoot weight (SW, total of stem and needles). We then calculated the root-shoot ratio (R2S) as RW/SW. Specific Root length (SRL) was calculated as root length/root weight.

Before harvest, we also collected 3-4 fresh needles from living seedlings to calculate stomatal density. In pines, stomata are arranged into longitudinal rows. We put each needle on a slide and photographed it at 100x magnification using a Leica DME compound microscope equipped with a Leica DFC290 digital camera. All counts were conducted near the middle of the needle to avoid variation that might occur at the base and the tip. Approximately 1.96 mm lengths of the needle were surveyed for the number of stomata and stomatal rows on their adaxial (upper) and abaxial (lower) surfaces. Needle width was measured in magnified images using the line measure tool in the Leica software. Then we calculated the stomata density on each side as the number of stomata divided by 1.96*width of needle. An average density and number of rows were calculated for each individual across sampled needles.

Genotype-Phenotype association analysis

We used the SNPs identified in the 42 mother trees for the GPA association analysis, focusing on the traits significantly associated with drought treatments. For the wet treatment traits, we use the average trait value across all members of each family in the wet treatment to run GPA analysis. For the drought response traits, we deduct the average trait value for a given family in the wet treatment from the value for each family's offspring in the drought treatment and then use the mean difference as the input for GPA. We used LFMM 2 (Caye et al. 2019) to run the GPA association analysis and then identified associations based on p ($<10^{-5}$) value.

Gene annotation

After identifying the significantly associated SNPs in GEA and GPA, we aligned the gene sequences for these regions against the nonredundant protein sequences database using UniProt to identify the gene and protein with the implemented Blastx (2.9.0+, E $<$ 1e^{-10}). The Gene Ontology Annotation Database ("UniProt" 2015; Bateman *et al.* 2017) was used to identify the potential functions of the genes further. If a SNP is in the intragenic region, we performed a search by querying the flanking sequence 400 bp from the beginning position of the gene. This had to be done separately because the "start" and "end" positions for the genes containing the introns were too far apart; Blastx could obtain no hits.

Results

Genetic diversity and population structure

A total of 4,155,896 SNPs were identified from GBS data of the 223 genotypes after initial filtering. With these SNPs, we ran both principal component analysis (PCA) and admixture analysis to determine the number of populations (K) represented by these individuals. Two principal components best explained the genetic variation between our samples, but nearly all individuals clustered together (Table S1). According to the admixture analysis result, the best K value was one (Fig. S2). We also plotted the admixture of each tree. We found that the identified "populations" when K=2 completely overlapped geographically (Fig. 1 B and Fig. S3). Thus, we concluded that the sampled genotypes belong to one interbreeding population and used K = 1 for the association analysis.

Environmental associations at individual loci

After filtering out the intergenic SNPs that might result in false positives, we were left with 927,740 (22.3%) SNPs in or near genes. These were then used for the association analyses. This is similar to the approach used by Jordan et al. (2017) for *Eucalyptus* . After the running of LFMM2 ($p < 10^{-5}$) for GEA, we found 1,458 significant associations with the five selected environmental variables (Table 1). PCK4 (April 1stsnowpack) had the most associations, with TMIN (minimum winter temperature) having the following highest number. Few SNPs were associated with more than one climatic variable, with the highest degree of overlap between PCK4 and TMIN (64 SNPs) and between CWD and TMIN (17 SNPs) (Fig. 2).

For PCK4 and TMIN, there were roughly similar numbers of associated SNPs in upstream and downstream regions versus the gene itself, with 14% of associated SNPs being missense (non-synonymous) mutations (Table 2). SNPs associated with CWD were also roughly evenly split between flanking regions and the main gene sequence, but only 3% were missense mutations. A higher proportion of SNPs associated with TMAX (maximum summer temperature) were within the gene, with 22% being missense mutations, while PPTW (winter precipitation) showed the opposite pattern, with 69% of SNPs being in the flanking regions.

Phenotypic associations at individual loci

Although 50 maternal families were initially selected for the greenhouse experiment, only 42 had sufficient germination for measurements to be included in analyses. Six out of the eight measured phenotypic traits were significantly different in the drought treatment versus the wet treatment. Height growth (GR) and shoot weight (SW) decreased, while root length (RL), the root-shoot dry mass ratio (R2S), stomata density on the adaxial side (SD_AD), and the number of stomatal rows on the abaxial side (NR_AB) increased. We therefore focused on these traits for the G2P association. We measure the association of SNPs to either the average measurement of control treatment family for each trait or the average change in the trait from wet to dry conditions (drought responsiveness). Heritabilities of trait responses to drought ranged from 0.15 to 0.65, and are discussed further in (Wu *et al.* 2023), with variation in shoot growth in response to drought being particularly variable.

More SNPs were associated with the trait drought responses (1,154) than with the control traits (817). While control R2S had the most associations and SW the least (Table 2), the opposite was the case for drought responsiveness (Table 3). The number of SNPs associated with more than one trait was low in both G2P analyses. The highest degree of overlap was in control traits of RL and R2S (12 SNPs) and of R2S and NR_AB (9 SNPs) (Fig. 3). The proportion of associated upstream SNPs was similar across control traits (32-40%), but proportions of other categories varied widely, with the proportion of missense SNPs ranging from 8-25%. For drought response, the distribution of SNPs in all categories differed, with the proportion of upstream being 19-34% and the proportion of missense being 7-16% for traits other than R2S. R2S was only associated with 6 SNPs, five upstream and one downstream.

Gene annotation for the significantly associated SNPs

Of the 1458 SNPs associated with environmental gradients, functions could be assigned for 788 (54%), while the rest had no matches in available gene ontology databases. We found that 283 SNPs with identifiable functions belonged to protein types that may be directly related to drought tolerance or other environmental responses (Fig. 4). We categorized these genes into five main functional groups: (a) the ubiquitination pathway, (b) seed, pollen and ovule formation, (c) cell wall formation, (d) stress responses, and (e) cell division and growth. Other associated SNPs with known functions were in or near transcription factors and genes with expression-regulating functions.

Many of the SNPs associated with TMAX, TMIN, CWD, and PCK4 were in or near genes in the protein ubiquitination pathway or the jasmonic acid synthesis response pathways (Fig. 4 and Table S2), both of which are involved in responses to biotic or abiotic stress (Creelman & Mullet 1995; Lyzenga & Stone 2012; Stone 2014). CWD and PCK4 were also associated with SNPs in or near genes involved in seed dormancy, cell wall organization, and the abscisic acid (ABA) signaling pathway, which have been previously linked to

drought responses in trees (Moran*et al.* 2017b). Genes involved in reproduction, including pollen and ovule formation, were associated with TMAX, TMIN, and PCK4. Genes involved in vascular tissue formation, growth regulation, and stress responses were associated with TMAX and PCK4. Genes involved in stomatal regulation and pathogen responses were associated with TMIN and PCK4. Further biotic and abiotic stress response genes were associated with PCK4, as were genes involved in nutrient transport, photosynthesis, respiration, sugar synthesis, and light responses (Table S2).

Of the 817 SNPs associated with seedling control (wet treatment) trait values and 1,154 SNPs associated with trait drought responsiveness, 43% and 51% could be assigned functions by gene ontology (Additional file 2: Table S3 and Table S4). Many of the same functional categories of genes associated with the environment were also related to measured phenotypes. This includes ubiquitination, seed development, cell wall organization, stress response, cell division (Fig. 4, 5, 6), and transcription factors. However, there was no overlap in specific SNPs identified.

The control treatment levels of the two stomatal traits were associated with genes involved in ubiquitination, cell wall organization or modification, growth and development, and ABA response. Control root-to-shoot ratio was associated with genes involved in biotic & abiotic stress responses, cell wall organization or modification, cell division or differentiation, lateral root formation, and ubiquitination. Control height growth had no associated SNPs, and root length was only associated with one SNP located in a gene involved in ubiquitination (Fig. 5). However, drought responsiveness of height growth, shoot weight, and root length were associated with all five functional categories (Fig. 6). Drought responsiveness of the two stomatal traits was associated with genes involved in stress responses, cell wall formation/organization, cell division/differentiation, and root formation.

Besides the five main functional groups of genes with SNPs associated with climatic, phenotypic, and drought response variables, several other functional groups were identified in the GEA and GPA annotation results (Table S2, S3, and S4). For example, 111 (14%) of the environmentally associated SNPs, 53 (6%) of SNPs associated with control traits, and 121 (12%) of the SNPs associated with trait drought responses were in genes relating to ATP binding or protein kinases. It was also fairly common for associated SNPs to be in genes associated with RNA/DNA binding, metal ion binding, translation, and protein transport.

Overlapping annotated genes in GEA and GPA

While, as noted in the section above, there was no overlap in the exact SNPs identified by GEA and GPA analyses, a few of the associated SNPs were found to be in the same genes. There were 14 genes identified in both the GPA for control traits and the GEA (Table 4). One of these is a ubiquitin-binding gene. Peptidyl-prolyl cis-trans isomerase, involved in protein folding, is known to be heat-induced in wheat (Kurek *et al.* 1999). Two genes are involved in glycerophospholipid synthesis or metabolism, suggesting some role related to cell membranes. Aspartyl proteases, like the one linked to winter precipitation and the number of stomatal rows, have been linked to the wood formation and to plant growth and development more generally (Cao *et al.* 2019). Butanoate–CoA ligases are often involved in the secondary compound synthesis (Beuerle & Pichersky 2002) and so could be involved in defenses against biotic antagonists or other stress responses. There were 15 genes identified in both the GPA for trait drought responsiveness and the GEA (Table 5). Most share the same functions as those in Table 4. Moreover, two overlapping genes are directly related to the stress response. Gene wsc1 is involved in cell wall biosynthesis under conditions of stress (Zu *et al.* 2001; Maddi *et al.*2012). Gene PAT14 is involved in leaf senescence in response to stresses (Lai *et al.* 2015; Zeng *et al.* 2018). However, several of the overlapping genes in each table have unknown functions, and most of these do not match any sequence in the database.

Discussion

In the GEA analysis, over half of the SNPs were associated with April 1st snowpack (PCK4). In this Mediterranean climate region, almost all of the annual precipitation occurs during the winter, and the melting of winter snow accumulation at high elevations feeds spring and summer streamflow (Serreze *et al.* 1999). Lack of snow can limit seedling establishment (Andrus *et al.* 2018). A "blanket" of snow can also

insulate seedlings from extremely cold temperatures, but may also delay the start of their growing season (Ettinger & HilleRisLambers 2013; Renard *et al.* 2016). Consistent with this latter possibility, one of the associated SNPs was in a gene involved in light responses. Winter minimum temperature (TMIN), which has frequently been found to limit growth in tree-ring studies (Harvey *et al.* 2020), shows the next highest number of associations. The number of SNPs associated with more than one climatic variable was low (Fig. 2), which may indicate that we successfully selected semi-independent climatic variables that require different genetic adaptations. The highest overlap was between PCK4 and TMIN (64 SNPs) and between CWD and TMIN (17 SNPs). The former SNP set may be related to adaptation to cold and snow depth, while the latter SNP set may be related to how quickly the site warms up in spring, drying out the soil. A similar GEA we conducted for the co-occurring species *Pinus lambertiana* also identified April snowpack as a key environmental variable that may have shaped local adaptation, and found low overlap in loci associated with different climate variables (Moran *et al.* In review).

In the GPA analysis, most SNPs associated with control phenotypic traits were linked with root-to-shoot ratio (R2S) and the number of abaxial stomatal rows (NR_AB). In contrast, most SNPs associated with phenotypic responses to drought were linked with shoot weight (SW), root length (RL), and R2S. Drought-stressed ponderosa pine seedlings allocated more to their root system, with longer root length, higher root-to-shoot dry mass ratio, less dry shoot mass, and less height growth. Other studies in pines have found similar patterns (Seiler & Johnson 1988; Irvine *et al.* 1998; Cregg & Zhang 2001; Taeger *et al.* 2015). This may indicate investment in greater water harvesting capacity at the cost of the overall low growth of aboveground structures – though low shoot growth can have the benefit of further reducing transpirational water loss (Moran *et al.* 2017b). We found that dry treatment root-to-shoot ratio was positively associated with survival in that treatment (Wu et al. 2023). Many of the SNPs associated with phenotypic drought responses were in genes associated with cell division & differentiation and with root growth, both of which make sense in light of the observed changes in allocation to root vs. shoot growth. The number of SNPs associated with more than one trait was low in both GPA analyses. The highest degree of overlap was in drought responsiveness of RL and R2S and of R2S and NR_AB (Fig. 6).

Non-synonymous (AKA missense) variants that may directly affect phenotype by changing protein form and function included 195 of the climate-associated, 93 of the control environment phenotype-associated, and 140 of the phenotype drought-response-associated SNPs (Tables 1, 2, & 3). Intragenic or synonymous variants are assumed to be neutral with respect to fitness but might be in linkage disequilibrium with a nearby causal variant. While linkage disequilibrium is usually low in conifers (Neale & Savolainen 2004), the GBS sequence fragments were relatively short (90-100 bp or less) and were trimmed further before SNP calling, so a linked non-synonymous variant could have been missed. We also found quite a few upstream and downstream SNPs in both GEA and GPA analysis that might directly affect gene expression or be linked to a protein-altering variant.

While we found no overlaps in specific SNPs between our GEA and GPA, identified several SNP-containing genes that were the same across the analyses (Tables 4 & 5). Most of these genes have been linked to stress responses in other studies. For example, gene wsc1 is involved in cell wall biosynthesis and gene PAT14 involved in leaf senescence, both in response to stress (Zu *et al.* 2001; Maddi *et al.* 2012; Lai *et al.* 2015; Zeng *et al.* 2018). Moreover, there was substantial overlap in functional categories found to be directly related to drought tolerance or other environmental responses in previous studies (Fig. 3, 4, 5). The prevalence of genetic associations related to abscisic acid (ABA)-signaling pathways and ubiquitination in GEA and GPA analyses is consistent with prior observations (Moran *et al.* 2017b) and with results of the *P. lambertiana analysis* (Moran *et al.* In review). Increasing ABA concentrations are used as a signal to keep stomata closed during dry conditions, reducing water loss (Brodribb *et al.* 2014). In addition, ABA signaling can also affect shoot growth and water uptake (Buckley 2005; Hamanishi & Campbell 2011). Ubiquitination is involved in drought responses in model species by playing a role in ABA-mediated dehydration stress responses (Ryu *et al.* 2010; Kim *et al.* 2012) or through the downregulation of plasma membrane aquaporin levels (Lee *et al.* 2009). Understanding of the role of ubiquitin in conifer drought response is still somewhat limited. A study in black spruce (*Picea mariana*) identified 16 candidate genes correlated with precipitation, including the

genes in the ubiquitin protein handling pathway (Prunier *et al.* 2011). The association between ubiquitin protein and roots and stomatal density may indicate previously unidentified roles in drought response.

Moreover, genes associated with seeds and seed dormancy can also be directly involved in drought tolerance; for instance, dehydrins can protect proteins from desiccation in both seeds and other plant tissues (Moran *et al.* 2017b). However, reproduction-related genes might also show associations with environmental gradients if they are involved in reproductive timing. Genes involved in xylem & phloem differentiation or cell wall formation could shape the hydraulic safety of water-transporting cells, which can be quite plastic in pines (Lauder et al. 2019). Other than these functions directly related to drought tolerance or different environmental responses, the other overlapping functions among GEA and GPA analysis are involved in gene expression (RNA or DNA binding, transcription factors, helicase activity, ribosome components, methylation) or ATP binding (motifs found in membrane transporters, microtubule subunits, enzymes, and other cell components that require energy). Our findings suggest the efficiency of combining GEA and GPA analysis with GBS to uncover potentially important adaptive genetic variation.

In conclusion, by investigating adaptive genetic variation in ponderosa pine with GEA and GPA association analysis, our study found thousands of genomic variants associated with response to climate and physiological traits. Some of these have previously-identified functions associated with drought responses, but for others, the gene function – or how that function is relevant for environmental responses – is still unknown. Molecular tools based on the associated genetic markers could be developed to assist breeders and land managers speed up selection for drought tolerance or selecting appropriate seed sources for a changing climate. In addition, our results should open new opportunities for functional studies to determine the molecular roles of the genes underlying these associated genetic makers in influencing trees' adaptation.

The two environmental variables with the most genetic associations – snowpack and winter temperatures – are among those that have already undergone significant shifts in recent decades, with further substantial shifts being projected due to anthropogenic climate change (Rapacciuolo*et al.* 2014; Fyfe *et al.* 2017). This suggests that tree populations in the Western US will be under rapidly shifting selective pressures, making exploring the potential of genomic selection for seed selection of pressing concern. We found considerable heritable variation in drought-responsive traits (Wu *et al.* 2023), suggesting adaptive potential exists if change is not too rapid. We are also following up on this study by testing the ability of the SNP associations detected here to predict performance in post-fire restoration plantings.

References

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T. & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* , 1, 95–111.

Alberto, F.J., Aitken, S.N., Alia, R., Gonzalez-Martinez, S.C., Hanninen, H., Kremer, A., *et al.* (2013). Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology* , 19, 1645–1661.

Anderson, J.T., Panetta, A.M. & Mitchell-Olds, T. (2012). Evolutionary and Ecological Responses to Anthropogenic Climate Change: Update on Anthropogenic Climate Change. *Plant Physiology* , 160, 1728–1740.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* , 17, 81–92.

Andrus, R.A., Harvey, B.J., Rodman, K.C., Hart, S.J. & Veblen, T.T. (2018). Moisture availability limits subalpine tree establishment.*Ecology* , 99, 567–575.

Auld, J.R., Agrawal, A.A. & Relyea, R.A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B: Biological Sciences* , 277, 503–511.

Bales, R.C., Hopmans, J.W., O'Geen, A.T., Meadows, M., Hartsough, P.C., Kirchner, P., *et al.* (2011). Soil Moisture Response to Snowmelt and Rainfall in a Sierra Nevada Mixed-Conifer Forest. *Vadose Zone Journal*

, 10, 786–799.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., *et al.* (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res* , 45, D158–D169.

Beaulieu, J., Doerksen, T., Clement, S., MacKay, J. & Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* , 113, 343–352.

Bell, G. & Gonzalez, A. (2009). Evolutionary rescue can prevent extinction following environmental change. *Ecology Letters* , 12, 942–948.

Berg, N. & Hall, A. (2015). Increased Interannual Precipitation Extremes over California under Climate Change. *Journal of Climate* , 28, 6324–6334.

Beuerle, T. & Pichersky, E. (2002). Purification and characterization of benzoate:coenzyme A ligase from Clarkia breweri. *Arch Biochem Biophys* , 400, 258–264.

Brodribb, T.J., McAdam, S.A.M., Jordan, G.J. & Martins, S.C.V. (2014). Conifer species adapt to low-rainfall climates by following one of two divergent pathways. *PNAS* , 111, 14489–14493.

Brunner, I. & Godbold, D.L. (2007). Tree roots in a changing world.*Journal of Forest Research* , 12, 78–82.

Buckley, T.N. (2005). The control of stomata by water balance. *New Phytologist* , 168, 275–292.

Cao, S., Guo, M., Wang, C., Xu, W., Shi, T., Tong, G., *et al.*(2019). Genome-wide characterization of aspartic protease (AP) gene family in Populus trichocarpa and identification of the potential PtAPs involved in wood formation. *BMC Plant Biology* , 19, 276.

Caye, K., Jumentier, B., Lepeule, J. & Francois, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol Biol Evol* , 36, 852–860.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L.,*et al.* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* , 6, 80–92.

Conkle, M.T. & Critchfield, W.B. (1988). Genetic variation and hybridization of ponderosa pine. *In: Ponderosa Pine: the species and its management, Washington State University Cooperative Extension, 1988: p. 27-43* .

Creelman, R.A. & Mullet, J.E. (1995). Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. *PNAS* , 92, 4114–4119.

Cregg, B.M. & Zhang, J.W. (2001). Physiology and morphology of Pinus sylvestris seedlings from diverse sources under cyclic drought stress.*Forest Ecology and Management* , 154, 131–139.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* , 12, 499–510.

Dillon, S., McEvoy, R., Baldwin, D.S., Rees, G.N., Parsons, Y. & Southerton, S. (2014). Characterisation of Adaptive Genetic Diversity in Environmentally Contrasted Populations of Eucalyptus camaldulensis Dehnh. (River Red Gum). *PLOS ONE* , 9, e103515.

Eckert, A.J., Bower, A.D., Wegrzyn, J.L., Pande, B., Jermstad, K.D., Krutovsky, K.V., *et al.* (2009). Association Genetics of Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* , 182, 1289–1302.

Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L., Nelson, C.D., Ross-Ibarra, J., Gonzalez-Martinez, S.C., *et al.* (2010). Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). *Genetics* , 185, 969–982.

Eckert, A.J., Maloney, P.E., Vogler, D.R., Jensen, C.E., Mix, A.D. & Neale, D.B. (2015). Local adaptation at fine spatial scales: an example from sugar pine (Pinus lambertiana, Pinaceae). *Tree Genetics & Genomes* , 11, 42.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., *et al.* (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species.*PLoS ONE* , 6, 1–10.

Ettinger, A.K. & HilleRisLambers, J. (2013). Climate isn't everything: Competitive interactions and variation by life stage will also affect range shifts in a warming world. *American Journal of Botany* , 100, 1344–1355.

Fettig, C.J., Mortenson, L.A., Bulaon, B.M. & Foulk, P.B. (2019). Tree mortality following drought in the central and southern Sierra Nevada, California, U.S. *Forest Ecology and Management* , 432, 164–178.

Flint, L.E., Flint, A.L., Thorne, J.H. & Boynton, R. (2013). Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. *Ecol Process* , 2, 1–21.

Frichot, E. & Francois, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* , 6, 925–929.

Frichot, E., Schoville, S.D., Bouchard, G. & Francois, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Mol Biol Evol* , 30, 1687–1699.

Fyfe, J.C., Derksen, C., Mudryk, L., Flato, G.M., Santer, B.D., Swart, N.C., *et al.* (2017). Large near-term projected snowpack loss over the western United States. *Nat Commun* , 8, 14996.

Gernandt, D.S., Hernandez-Leon, S., Salgado-Hernandez, E. & Perez de La Rosa, J.A. (2009). Phylogenetic relationships of Pinus subsection Ponderosae inferred from rapidly evolving cpDNA regions.*Systematic Botany* , 34, 481–491.

Graham, R.T. & Jain, T.B. (2005). Ponderosa pine ecosystems. *In: Ritchie, Martin W.; Maguire, Douglas A.; Youngblood, Andrew, tech. coordinators. Proceedings of the Symposium on Ponderosa Pine: Issues, Trends, and Management, 2004 October 18-21, Klamath Falls, OR. Gen. Tech. Rep PSW-GTR-198. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture: 1-32.* , 198, 1–32.

Griffin, D. & Anchukaitis, K.J. (2014). How unusual is the 2012–2014 California drought? *Geophysical Research Letters* , 41, 9017–9023.

Hamanishi, E.T. & Campbell, M.M. (2011). Genome-wide responses to drought in forest trees. *Forestry (Lond)* , 84, 273–283.

Hamilton, J.A., Lexer, C. & Aitken, S.N. (2013). Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis x P. glauca) hybrid zone. *New Phytologist* , 197, 927–938.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., *et al.* (2011). Adaptation to climate across the Arabidopsis thaliana genome. *Science* , 334, 83–86.

Harrisson, K.A., Pavlova, A., Telonis-Scott, M. & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* , 7, 1008–1025.

Harvey, J.E., Smiljanić, M., Scharnweber, T., Buras, A., Cedro, A., Cruz-García, R., *et al.* (2020). Tree growth influenced by warming winter climate and summer moisture availability in northern temperate forests. *Global Change Biology* , 26, 2505–2518.

Hoffmann, A.A. & Sgrò, C.M. (2011). Climate change and evolutionary adaptation. *Nature* , 470, 479–485.

Holliday, J.A., Ritland, K. & Aitken, S.N. (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). *New Phytologist* , 188, 501–514.

Housset, J.M., Nadeau, S., Isabel, N., Depardieu, C., Duchesne, I., Lenz, P., *et al.* (2018). Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. *New Phytologist* , 218, 630–645.

Irvine, J., Perks, M.P., Magnani, F. & Grace, J. (1998). The response of Pinus sylvestris to drought: stomatal control of transpiration and hydraulic conductance. *Tree Physiology* , 18, 393–402.

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* , 45, 379–401.

Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., *et al.* (2016). Genomic selection in maritime pine.*Plant Sci* , 242, 108–119.

Jaramillo-Correa, J.-P., Rodríguez-Quilón, I., Grivet, D., Lepoittevin, C., Sebastiani, F., Heuertz, M., *et al.* (2015). Molecular proxies for climate maladaptation in a long-lived tree (Pinus pinaster Aiton, Pinaceae). *Genetics* , 199, 793–807.

Johansen, A.D. & Latta, R.G. (2003). Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology* , 12, 293–298.

Kim, S.J., Ryu, M.Y. & Kim, W.T. (2012). Suppression of Arabidopsis RING-DUF1117 E3 ubiquitin ligases, AtRDUF1 and AtRDUF2, reduces tolerance to ABA-mediated drought stress. *Biochemical and Biophysical Research Communications* , 420, 141–147.

Kitzmiller, J.H. (2005). Provenance Trials of Ponderosa Pine in Northern California. *Forest Science* , 51, 595–607.

Kolb, T.E., Grady, K.C., McEttrick, M.P. & Herrero, A. (2016). Local-Scale Drought Adaptation of Ponderosa Pine Seedlings at Habitat Ecotones. *Forest Science* , 62, 641–651.

Kurek, I., Aviezer, K., Erel, N., Herman, E. & Breiman, A. (1999). The wheat peptidyl prolyl cis-trans-isomerase FKBP77 is heat induced and developmentally regulated. *Plant Physiol* , 119, 693–704.

Lai, J., Yu, B., Cao, Z., Chen, Y., Wu, Q., Huang, J., *et al.*(2015). Two homologous protein S-acyltransferases, PAT13 and PAT14, cooperatively regulate leaf senescence in Arabidopsis. *Journal of Experimental Botany* , 66, 6345–6353.

Langlet, O. (1971). Two Hundred Years Genecology. *Taxon* , 20, 653–721.

Lee, H.K., Cho, S.K., Son, O., Xu, Z., Hwang, I. & Kim, W.T. (2009). Drought Stress-Induced Rma1H1, a RING Membrane-Anchor E3 Ubiquitin Ligase Homolog, Regulates Aquaporin Levels via Ubiquitination in Transgenic Arabidopsis Plants. *The Plant Cell* , 21, 622–641.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* , 27, 2987–2993.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , 25, 1754–1760.

Lu, M., Loopstra, C.A. & Krutovsky, K.V. (2019). Detecting the genetic basis of local adaptation in loblolly pine (Pinus taeda L.) using whole exome-wide genotyping and an integrative landscape genomics analysis approach. *Ecology and Evolution* , 9, 6798–6809.

Lyzenga, W.J. & Stone, S.L. (2012). Abiotic stress tolerance mediated by protein ubiquitination. *Journal of Experimental Botany* , 63, 599–616.

Maddi, A., Dettman, A., Fu, C., Seiler, S. & Free, S.J. (2012). WSC-1 and HAM-7 Are MAK-1 MAP Kinase Pathway Sensors Required for Cell Wall Integrity and Hyphal Fusion in Neurospora crassa. *PLOS ONE* , 7, e42374.

Maguire, K.C., Shinneman, D.J., Potter, K.M. & Hipkins, V.D. (2018). Intraspecific Niche Models for Ponderosa Pine (Pinus ponderosa) Suggest Potential Variability in Population-Level Response to Climate Change.*Systematic Biology* , 67, 965–978.

Mahony, C.R., MacLachlan, I.R., Lind, B.M., Yoder, J.B., Wang, T. & Aitken, S.N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study.*Evolutionary Applications* , 13, 116–131.

Markesteijn, L. & Poorter, L. (2009). Seedling root morphology and biomass allocation of 62 tropical tree species in relation to drought- and shade-tolerance. *Journal of Ecology* , 97, 311–325.

Moran, E., Lauder, J., Musser, C., Stathos, A. & Shu, M. (2017a). The genetics of drought tolerance in conifers. *New Phytol* , 216, 1034–1048.

Moran, E.V., DeSilva, R., Canning, C. & Wright, J.W. (In review). Sugar pine association genetics and performance in a post-fire restoration planting. *Ecological Applications* .

Moran, E.V., Lauder, J., Musser, C., Stathos, A. & Shu, M.J. (2017b). The genetics of drought tolerance in conifers. *New Phytologist* , 216, 1034–1048.

Namroud, M.-C., Beaulieu, J., Juge, N., Laroche, J. & Bousquet, J. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce.*Mol Ecol* , 17, 3599–3613.

Neale, D.B. & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* , 12, 111–122.

Neale, D.B. & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science* , 9, 325–330.

Neale, D.B., Wegrzyn, J.L., Stevens, K.A., Zimin, A.V., Puiu, D., Crepeau, M.W., *et al.* (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies.*Genome Biol* , 15, R59.

Oney, B., Reineking, B., O'Neill, G. & Kreyling, J. (2013). Intraspecific variation buffers projected climate change impacts on Pinus contorta. *Ecology and Evolution* , 3, 437–449.

Parker, J. (1949). EFFECTS OF VARIATIONS IN THE ROOT-LEAF RATIO ON TRANSPIRATION RATE. *Plant Physiol* , 24, 739–743.

Patterson, N., Price, A.L. & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics* , 2, e190.

Poland, J.A., Brown, P.J., Sorrells, M.E. & Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* , 7, e32253.

Poland, J.A. & Rife, T.W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome* , 5, 92–102.

Potter, K.M., Hipkins, V.D., Mahalovich, M.F. & Means, R.E. (2013). Mitochondrial DNA haplotype distribution patterns in Pinus ponderosa (Pinaceae): range-wide evolutionary history and implications for conservation. *Am. J. Bot.* , 100, 1562–1579.

Potter, K.M., Hipkins, V.D., Mahalovich, M.F. & Means, R.E. (2015). Nuclear genetic variation across the range of ponderosa pine (Pinus ponderosa): Phylogeographic, taxonomic and conservation implications. *Tree Genetics & Genomes* , 11, 38.

Prunier, J., Laroche, J., Beaulieu, J. & Bousquet, J. (2011). Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce: SNPs and climate adaptation. *Molecular Ecology* , 20, 1702–1716.

Rapacciuolo, G., Maher, S.P., Schneider, A.C., Hammond, T.T., Jabis, M.D., Walsh, R.E., *et al.* (2014). Beyond a warming fingerprint: individualistic biogeographic responses to heterogeneous climate change in California. *Global Change Biology* , 20, 2841–2855.

Renard, S.M., McIntire, E.J.B. & Fajardo, A. (2016). Winter conditions – not summer temperature – influence establishment of seedlings at white spruce alpine treeline in Eastern Quebec. *Journal of Vegetation Science* , 27, 29–39.

Rice, K.J. & Emery, N.C. (2003). Managing microevolution: restoration in the face of global change. *Frontiers in Ecology and the Environment* , 1, 469–478.

Rochette, N.C. & Catchen, J.M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* , 12, 2640–2659.

Ryan, M.G. (2011). Tree responses to drought. *Tree Physiology* , 31, 237–239.

Ryu, M.Y., Cho, S.K. & Kim, W.T. (2010). The Arabidopsis C3H2C3-Type RING E3 Ubiquitin Ligase AtAIRP1 Is a Positive Regulator of an Abscisic Acid-Dependent Response to Drought Stress. *Plant Physiology* , 154, 1983–1997.

Savolainen, O., Lascoux, M. & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics* , 14, 807–820.

Seiler, J.R. & Johnson, J.D. (1988). Physiological and Morphological Responses of Three Half-Sib Families of Loblolly Pine to Water-Stress Conditioning. *Forest Science* , 34, 487–495.

Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A. & Pulwarty, R.S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNO) data. *Water Resources Research* , 35, 2145–2160.

Shu, M. (2020). Association genetics of drought tolerance in ponderosa pine (*Pinus ponderosa*). PhD. UC Merced, Merced, CA.

Sork, V.L., Aitken, S.N., Dyer, R.J., Eckert, A.J., Legendre, P. & Neale, D.B. (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* , 9, 901–911.

Stone, S.L. (2014). The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. *Frontiers in plant science* , 5, 135.

Suren, H., Hodgins, K.A., Yeaman, S., Nurkowski, K.A., Smets, P., Rieseberg, L.H., *et al.* (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources* , 16, 1136–1146.

Taeger, S., Sparks, T.H. & Menzel, A. (2015). Effects of temperature and drought manipulations on seedlings of Scots pine provenances.*Plant Biology* , 17, 361–372.

Uchiyama, K., Iwata, H., Moriguchi, Y., Ujino-Ihara, T., Ueno, S., Taguchi, Y., *et al.* (2013). Demonstration of Genome-Wide Association Studies for Identifying Markers for Wood Property and Male Strobili Traits in Cryptomeria japonica. *PLOS ONE* , 8, e79866.

UniProt: a hub for protein information. (2015). *Nucleic Acids Res* , 43, D204–D212.

Van Kleunen, M. & Fischer, M. (2005). Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist* , 166, 49–60.

Wang, J., Zhao, Q., Hastie, T. & Owen, A.B. (2017). CONFOUNDER ADJUSTMENT IN MULTIPLE HYPOTHESIS TESTING. *Ann Stat* , 45, 1863–1894.

Weiss, M., Sekhwal, M.K., Neale, D.B. & De La Torre, A.R. (2022). Genomics of Climate Adaptation in Pinus Lambertiana. In: *The Pine Genomes* , Compendium of Plant Genomes (ed. De La Torre, A.R.). Springer International Publishing, Cham, pp. 51–65.

Williams, C.G. (Ed.). (2009). The Dynamic Wind-Pollinated Mating System. In: *Conifer Reproductive Biology* . Springer Netherlands, Dordrecht, pp. 125–135.

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

Wright, J.W. (2007). Local adaptation to serpentine soils in Pinus ponderosa. *Plant Soil* , 293, 209–217.

Wu, D., Shu, M. & Moran, E.V. (2023). Heritability of plastic trait changes in drought-exposed ponderosa pine seedlings. *Ecosphere* , 14, e4454.

Xuereb, A., Stahlke, A., Bermingham, M., Brown, M., Nonaka, E., Razgour, O., *et al.* (2017). Effect of missing data and sample size on the performance of genotype-environment association methods.

Yeaman, S., Hodgins, K.A., Lotterhos, K.E., Suren, H., Nadeau, S., Degner, J.C., *et al.* (2016). Convergent local adaptation to climate in distantly related conifers. *Science* , 353, 1431–1433.

Ying, C.C. & Liang, Q. (1994). Geographic pattern of adaptive variation of lodgepole pine (Pinus contorta Dougl.) within the species' coastal range: field performance at age 20 years. *Forest Ecology and Management* , 67, 281–298.

Zeng, X., Xu, Y., Jiang, J., Zhang, F., Ma, L., Wu, D., *et al.*(2018). Identification of cold stress responsive microRNAs in two winter turnip rape (Brassica rapa L.) by high throughput sequencing. *BMC Plant Biol* , 18, 52.

Zettlemoyer, M.A. & Peterson, M.L. (2021). Does Phenological Plasticity Help or Hinder Range Shifts Under Climate Change? *Frontiers in Ecology and Evolution* , 9.

Zimin, A., Stevens, K.A., Crepeau, M.W., Holtz-Morris, A., Koriabine, M., Marcais, G., *et al.* (2014). Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics* , 196, 875–890.

Zu, T., Verna, J. & Ballester, R. (2001). Mutations in WSC genes for putative stress receptors result in sensitivity to multiple stress conditions and impairment of Rlm1-dependent gene expression in Saccharomyces cerevisiae. *Mol Gen Genomics* , 266, 142–155.

Acknowledgments

We thank the Forest Service's Pacific Southwest Regional Genetic Resources Program for allowing us to sample needles and collect seeds from their seed orchard, and XSEDE and UC Merced computer cluster for computational resources and support. We thank David Neale, Stephen Hart, Jason Sexton, Jeffrey Lauder, Dean Wu, and Rainbow DeSilva for their comments on the manuscript and the experimental design. Also, thank Vanessa Centeno and Shirley Calderon at UC Merced for greenhouse plant care. The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. For SNP identification, we used the MERCED computer cluster at UC Merced (funded by NSF Award ACI-1429783) and the Extreme Science and Engineering Discovery Environment (XSEDE, supported by NSF Award ACI-1548562). The funding organizations paid the experimental fees and computing resources for this research. Still, they did not play any role in the design of the study nor the collection analysis and interpretation of data, or the writing of the manuscript.

Data Accessibility Statement

Raw DNA sequencing data: available at National Center for Biotechnology Information under BioProject number PRJNA707049. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA707049. Individual tree and seedling SNP genotypes are available on Dryad. DOI: https://doi.org/10.6071/M3DQ1D. Greenhouse seedling data can be found as a supplement (Wu *et al.* 2023).

Competing Interests Statement

The authors declare that they have no competing interests.

Authors' contributions

MS: Research design, performed research, analyzed data, wrote the paper (corresponding author)

EM: Research design, edited the paper.

All authors have read and approved the manuscript.

Table 1. Number of environmentally associated SNPs located in different regions

Location of SNP	PCK4	TMIN	CWD	TMAX	PPTW
Upstream	335 (29%)	33 (23%)	11 (16%)	12 (24%)	16 (36%)
intragenic (intron)	336 (29%)	34 (23%)	24 (36%)	18 (36%)	7 (16%)
Synonymous	92 (8%)	22 (15%)	6 (9%)	5 (10%)	2 (4%)
Missense	157 (14%)	20 (14%)	2 (3%)	11 (22%)	5 (11%)
Downstream	229 (20%)	36 (25%)	24 (36%)	3 (6%)	15 (33%)
Other	2 (0.1%)	0	0	1 (2%)	0
Total	**1151**	**145**	**67**	**50**	**45**

Table 2. Number of SNPs associated with traits in control conditions.

Location of SNP	R2S	NR_AB	RL	GR	SD_AD	SW
upstream	166 (35%)	90 (32%)	12 (43%)	6 (40%)	4 (33%)	3 (33%)
intragenic (intron)	106 (23%)	79 (28%)	5 (18%)	2 (13%)	3 (25%)	1 (11%)
synonymous	40 (8%)	18 (6%)	1 (3%)	0 (0%)	2 (17%)	1 (11%)
missense	61 (13%)	21 (8%)	3 (11%)	3 (20%)	3 (25%)	2 (22%)
downstream	100 (21%)	72 (26%)	7 (25%)	4 (27%)	0 (0%)	1 (11%)
other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (11%)
Total	**473**	**280**	**28**	**15**	**12**	**9**

Table 3. Number of SNPs associated with drought responsiveness of traits.

Location of SNP	ΔP2Σ	ΔNP_AB	ΔPΛ	ΔΓP	ΔΣΔ_AΔ	ΔΣΩ
upstream	5 (83%)	43 (28%)	84 (22%)	48 (33%)	11 (19%)	138 (34%)
intragenic (intron)	0 (0%)	41 (26%)	115 (30%)	41 (27%)	33 (58%)	113 (28%)
synonymous	0 (0%)	10 (6%)	29 (8%)	11 (7%)	1 (2%)	43 (10%)
missense	0 (0%)	15 (10%)	60 (16%)	15 (10%)	4 (7%)	46 (11%)
downstream	1 (17%)	45 (29%)	85 (23%)	35 (23%)	8 (14%)	69 (17%)
other	0 (0%)	2 (1%)	3 (1%)	0 (0%)	0 (0%)	0 (0%)
Total	**6**	**156**	**376**	**150**	**57**	**409**

Table 4. Overlapping genes in GEA and the GPA for traits in control conditions

Climate variable	Phenotypic variable	Gene name	Gene function
PCK4	NR_AB	MARPO_0050s0076	Ubiquitin binding
PCK4	NR_AB	Unknown	Unknown
PCK4	NR_AB	Gotri_016876	Unknown
PCK4	NR_AB	Peptidyl-prolyl cis-trans isomerase	Protein folding, may be hea
PCK4	NR_AB	HAD-superfamily subfamily IIA hydrolase	Glycerophospholipid biosynt
PCK4 & TMIN	NR_AB	Unknown	Unknown
PCK4	NR_AB & R2S	Pyridoxal kinase	ATP/ADP conversion
PCK4	R2S	RNA pseudouridine synthase 4, mitochondrial	Synthesis of modified U in I
PCK4	R2S	Unknown	Unknown
PCK4	R2S	Glycerophosphodiester phosphodiesterase	Glycerophospholipid metabo
PCK4	R2S	MAP3K epsilon protein kinase 1	Control of cell division/expa
PPTW	NR_AB	Aspartyl protease	Protein breakdown, often in
PPTW	R2S	eukaryotic translation initiation factor 5B-like	Translation initiation
TMAX	R2S	Butanoate–CoA ligase	Secondary compound metab

Table 5. Overlapping genes in GEA and the GPA for trait drought responsivenes

Climate variable	Phenotype Variable	Gene name	Gene function
PCK4	ΔGR	CSUI_002384	ATP binding
PCK4	ΔGR	LOC109003013	DNA binding; regulation of translation
PCK4	ΔGR	EXO84A	exocytosis
PCK4	ΔNR_AB	EUGRSUZ_B03992	oxidoreductase activity
TMAX	ΔNR_AB	L195_g029008	nucleic acid binding
PCK4	ΔRL	T459_09847	RNA binding
PCK4	ΔRL	AMTR_s00007p00201600	Ubiquitin binding
CWD	ΔRL	NALOC109013111	RNA binding; regulation of translation
PCK4	ΔRL	MARPO_0181s0009	eoxyribonucleotide catabolic process
PCK4	ΔRL	PAT14	leaf senescence
PCK4	ΔSD_AD & ΔSW	Unknown	Unknown
PCK4	ΔSW	LOC109001250	peptidyl-prolyl cis-trans isomerase activity
PCK4	ΔSW	wsc1	regulation of cell wall organization or biogenesis
PCK4	ΔSW	CCAM_LOCUS30844	Unknown
CWD	ΔSW	Unknown	Unknown

Fig. 1 Location and the admixture analysis of the 223 ponderosa pine genotypes. Left: Original geographic distribution of the 223 ponderosa pine genotypes. Right: Proportion of each individual's genome allocated to "population 1" (green) and "population 2" (orange) by admixture analysis when K=2, illustrating lack of geographical isolation. Trees were subsequently treated as part of a single population.

Fig. 2 Venn diagram comparing overlap in environmentally associated SNPs. The number of overlapping SNPs that are associated with four climatic variables between April 1stsnowpack (PCK4), monthly winter precipitation (PPTW), climatic water deficit (CWD), and minimum winter temperature (TMIN).

Fig. 3 Venn diagram comparing overlap in phenotypically associated SNPs. Left: Overlap in SNPs significantly associated with control root length (RL), root-shoot ratio (R2S), and abaxial stomatal rows (NR_AB). SNPs associated with control height growth (15), adaxial stomatal density (12), and shoot weight (9) did not overlap with other categories. Right: Overlap in SNPs significantly associated with

drought responsiveness of shoot weight (ΔSW); root length (ΔRL); and the number of stomatal rows on abaxial side (ΔNR_AB). SNPs associated with drought responsiveness of height growth (150), adaxial stomatal density (57), and R2S (6) did not overlap with any other categories.

Fig. 4 Five types of annotated SNP functions associated with different climatic variables. The number of non-synonymous variants and other variants that are associated with the five climatic variables: Climatic water deficit (CWD); Minimum winter temperature (TMIN); Maximum summer temperature (TMAX); April 1st snowpack (PCK4), and Monthly winter precipitation (PPTW). Missense (non-synonymous) SNPs are shown in grey, and other types of SNP are in orange.

Fig. 5 Five types of annotated SNP functions associated with different traits in control conditions . The number of non-synonymous variants and other variants that are associated with four traits in control conditions: root length (RL), number of stomatal rows on abaxial surface (NR_AB), stomatal density on adaxial surface (SD_AD), and root-to-shoot ratio (R2S). No SNPs in these categories were associated with height growth or shoot weight. Missense (non-synonymous) SNPs are shown in grey, and other types of SNP are in orange.

Fig. 6 Five types of annotated SNP functions associated with drought responsiveness of different traits. The number of non-synonymous variants and other variants that are associated with drought responsiveness of five traits: changes in height growth (GR), root length (RL), shoot weight (SW), number of stomatal rows on abaxial surface (NR_AB), and stomatal density on adaxial surface (SD_AD). No SNPs in these categories were associated with root-to-shoot ratio (R2S). Missense (non-synonymous) SNPs are shown in grey, and other types of SNP are in orange.