

Identifying Genetic Variation Associated With Environmental Variation and Drought-tolerance Phenotypes in Ponderosa Pine

Mengjun Shu (✉ mshu@ucmerced.edu)

University of California Merced <https://orcid.org/0000-0002-6323-2664>

Emily V. Moran

UC Merced: University of California Merced <https://orcid.org/0000-0003-4624-1910>

Research article

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

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-689957/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Genotype-to-environment (G2E) association analysis coupled with genotype-to-phenotype (G2P) association analysis promises exciting advances towards discovering genes responsible for local adaptation. We combine G2E and G2P analysis with gene annotation in *Pinus ponderosa* (ponderosa pine), an ecologically and economically important conifer that lacks a sequenced genome, to identify genetic variants and gene functions that may be associated with local adaptation to drought.

Results

We identified SNP markers in 223 genotypes from across the Sierra Nevada by aligning GBS sequence fragments to the reference genome of *Pinus taeda* (loblolly pine). Focusing on SNPs in or near coding regions, we found 1458 associated with 5 largely-uncorrelated climate variables, with the largest number (1151) associated with April 1st snow pack. We also planted seeds from a subset of these trees in the greenhouse, subjected half of the seedlings to a drought treatment, and measured phenotypes thought to be associated with drought tolerance, including root length and stomatal density. 817 SNPs were associated with the control-condition values of six traits, while 1154 were associated with responsiveness of these traits to drought.

Conclusions

While no individual SNPs were associated with both the environmental variables and the measured traits, several categories of genes were associated with both, particularly those involved in cell wall formation, biotic and abiotic stress responses, and ubiquitination. However, functions of many of the associated genes have not yet been determined due to the lack of gene annotation information for trees and future studies are needed.

Introduction

Genomics promises exciting advances towards exploring adaptive genetic variation and evolutionary potential under rapidly changing and often unpredictable environment [1–3]. 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 [4, 5]. Adaptation to local climate conditions is common in tree populations [6–9]. However, tree populations with long life cycles may become maladapted if environmental shifts rapidly [10–12]. Understanding the distribution of genetic variation related to environmental responses may help us better predict changes and manage forests in a changing climate [13, 14]. This includes selecting seed sources for restoration or breeding that have desirable characteristics such as drought tolerance [15, 16].

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 [17–20]. This approach is sometimes known as genotype-to-environment (G2E) association analysis. Prior studies in *Arabidopsis* – the primary plant model organism - have found that environmentally-associated SNPs can predict performance in common gardens [21], and a study in *Pinus pinaster* suggests this could be true in trees as well [22]. However, G2E studies often don't by themselves reveal why certain allele are more prevalent or favored in particular environments – for example, are they responsible for selectively favored traits? Genotype-to-Phenotype (G2P) association identifies loci linked to a particular phenotype [23, 24]. To eliminate the effects of environment on phenotypes, traits must be measured in a common environment, such as a greenhouse. However, G2P association study does not reveal whether a trait variant would be favored in the field. G2E and G2P association are thus complementary, and by combining them one might identify both the loci and traits that are selectively favored in particular conditions [18, 25].

The large genome size of conifer trees (> 19 GB) represents a challenge for analysis. Most association studies in conifers have focused on SNPs within a few hundred genes [18, 23, 24, 26–28], or fewer than 2,000 genome-wide SNPs [29]. One notable exception is a recent study on lodgepole pine that made use of a sequence capture dataset created by mapping the *Pinus contorta* transcriptome to the *P. taeda* genome sequence [25], but most conifers have neither a published genome sequence nor a full transcriptome. Though targeted sequencing is efficient, candidate gene approaches may miss other important genes with previously unsuspected roles in local adaptation, while focusing on variants within genes may miss important 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 [30, 31]. Genotyping-by-Sequencing (GBS), which can generate tens of thousands of SNP markers (Single Nucleotide Polymorphisms) without the need for a reference genome or full transcriptome, has emerged as a cost-effective strategy [32, 33]. By combining the power of multiplexed NGS with restriction-enzyme-based genome complexity reduction, GBS is able to genotype large populations of individuals for many thousands of SNPs in an increasingly rapid and inexpensive way [31, 34].

Despite the high economic and ecological importance of ponderosa pine (*Pinus ponderosa*) in the western United States [35], 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* 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 little adaptive significance in themselves [36, 37]. 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 [38, 39]. California's historic 2012–2016 drought may represent an increasingly common condition as climate changes [40, 41]. Such "hot droughts" can lead to mass tree mortality, negatively impacting the sustainability of conifer forests

[42]. 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 the 1970's 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 that span the full climatic range included in the collection for the G2E analysis, and seedlings of a subset of 50 genotyped parent trees for a G2P analyses of putative drought-response traits based on a greenhouse experiment. We ran gene annotation to ascribe biological function to the genes that the associated SNPs were in or near. Then we assessed overlap in SNP identity or gene functions among G2E and G2P association analysis that might indicate particular importance for local adaptation.

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 principle components best explained the genetic variation between our samples, but nearly all individuals clustered together (Fig. 1 and Additional file 1: Table S1), and according to the admixture analysis result the best K value was one (Additional file 1: Fig. S1). We also plotted the admixture of each individual tree and found that “populations” completely overlapped geographically (Fig. 1 and Additional file 1: Fig. S2). Thus, we concluded that the sampled genotypes belong to one interbreeding population and used K = 1 for the association analysis.

Environmental and phenotypic associations at individual loci

Initial gene annotation revealed that many of the 4,155,896 SNPs fall between genes and regulatory regions (in the intergenic regions) – likely have no direct effect on gene expression or function. To eliminate false positives that might arise from this, we filtered out the intergenic SNPs, leaving 927,740 (22.3%) SNPs in or near genes for the association analyses. This is similar to the approach used in Jordan et al. for *Eucalyptus* [43]. We used latent factor mixed model 2 (LFMM2) for G2E and G2P association analysis with these 927,740 SNPs and K = 1.

The five 1921–1950 mean climate variables used in the G2E analysis – climatic water deficit (CWD); minimum winter temperature (TMIN); maximum summer temperature (TMAX); monthly winter precipitation (PPTW); and April 1st snow pack (PCK4) – had low-to-moderate correlations with one other (Additional file 1: Fig. S3). After the running of LFMM2 ($q < 0.05$) for G2E, we found 1,458 significant associations with the environmental variables out of the 927,740 filtered SNPs (Table 1). PCK4 had the most associations by far, with TMIN having the next highest number of associations. The number of

SNPs associated with more than one climatic variable was low, with the highest degree of overlap between PCK4 and TMIN (64 SNPs) and between CWD and TMIN (17 SNPs) (Fig. 2).

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

For both PCK4 and TMIN, there were roughly similar numbers of associated SNPs in upstream and downstream regions versus with 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 were within the gene, with 22% being missense mutations, while PPTW showed the opposite pattern, with 69% of SNPs being in the flanking regions.

Before running G2P analysis, seedlings from a subset of the 223 genotyped trees were grown in the greenhouse and subjected to wet (control) and drought treatments, with multiple phenotypic traits thought to be associated with drought response being measured for both groups. Fifty families were initially selected, although 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), shoot weight (SW), root length (RL), root-shoot dry mass ratio (R2S), stomata density on adaxial side (SD_AD), and number of stomatal rows on abaxial side (NR_AB). We therefore focused on these traits for the G2P association. We measure the association of SNPs to either the control treatment family breeding value (BV) for each trait, or the average change in the trait from wet to dry conditions (drought responsiveness).

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 proportion of upstream being 19–34% and proportion of missense being 7–16% for traits other than R2S. R2S was only associated with 6 SNPs, 5 upstream and 1 downstream.

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 responsive traits.

Location of SNP	Δ R2S	Δ NR_AB	Δ RL	Δ GR	Δ SD_AD	Δ SW
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

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 belonged to protein types that have functions 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.

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 Additional file 2: Table S2), both of which are involved in responses to biotic or abiotic stress [44–46]. CWD and PCK4 were also associated with SNPs in or near genes involved in seed dormancy and the abscisic acid (ABA) signaling pathway, both of which have been previously linked to drought responses in trees [47]. Genes involved in reproduction, including pollen and ovule formation, were associated with TMAX, TMIN, and PCK4. CWD and PCK4 were associated with genes involved in cell wall organization. Both TMAX and PCK4 were associated with genes involved in vascular tissue formation, growth regulation, and stress responses, while TMIN and PCK4 were associated with genes involved in stomatal regulation and pathogen responses. 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 (Additional file 2: Table S2),

Of the 817 SNPs associated with seedling control 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), respectively. Many of the same functional categories of genes that were found to be associated with the environment were also associated with measured phenotypes, though there was no overlap in specific SNPs identified. This includes ubiquitination, seed development, cell wall organization, stress response, and cell division (Fig. 4, 5, 6).

In the control treatment, 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 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 both the G2E and G2P annotation results (Additional file 2: Table S2, Table S3, and Table S4). For example, 111 (14%) of the environmentally-associated SNPs, 53 (6%) of SNPs associated control traits, and 121 (12%) of the SNPs associated with trait drought responses were in genes relating to ATP binding or protein kinases. Associated SNPs in genes associated with RNA/DNA binding, metal ion binding, translation, and protein transport were also fairly common.

Discussion

We identified 1458 SNPs associated with 5 climate variables, with April 1st snow-pack associated with most of the SNPs. We also identified 817 SNPs associated with the control-condition values of six phenotypic traits, while 1154 associated with responsiveness of these traits to drought. No individual SNPs overlapped between the genotype-to-environment (G2E) and genotype-to-phenotype (G2P) analyses. But the associated SNPs did share similar gene functional categories including (a) the ubiquitination pathway, (b) seed, pollen and ovule formation, (c) cell wall formation, (d) stress responses, and (e) cell division and growth. Other shared categories including ATP binding or protein kinases, RNA/DNA binding, metal ion binding, translation, and protein transport.

Different categories of SNPs may affect function in different ways. Non-synonymous (AKA missense) variants may directly affect phenotype by changing protein form and function; these included 195 of the climate-associated, 93 of the control environment phenotype-associated, and 140 of the phenotype drought-response-associated SNPs (Table 1, 2, 3). Intragenic or synonymous variants are assumed to be neutral with respect to fitness, but either might be in linkage disequilibrium with a nearby causal variant. While linkage disequilibrium is fairly low in conifers [48], the GBS sequence fragments were quite 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 G2E and G2P analysis that might either directly affect gene expression or be linked to a protein-altering variant.

For the G2E association, we used 1921–1950 average climate values to estimate the selective environment under which these genotypes established as seedlings and saplings prior to their collection in the 1970s. We chose to focus on raw environmental variables rather than environmental PCA axes, as a number of previous studies have done [17, 18], because PCA associations can be difficult to interpret if, for example, the axes include both temperature and moisture variables. We selected five climate variables that exhibit low correlation with one another across the collection area. The number of SNPs associated with more than one climatic variable was low (Fig. 2), which may indicate that we were successful in selecting semi-independent climatic variables which require different genetic adaptations. The highest degree of overlap was between PCK4 and TMIN (64 SNPs) and between CWD and TMIN (17 SNPs). The former SNP set might be related to adaptation to cold and/or snow depth, while the latter SNP set might be related to how quickly the site warms up in spring, drying out the soil.

In the G2E analysis, over half of the SNPs were associated only with April 1st snowpack (PCK4). Winter minimum temperatures (TMIN) — affecting the depth and duration of snowpack — shows the next highest number of associations. In this Mediterranean climate region, most of the annual precipitation occurs during the winter, and melting of winter snow accumulation at high elevations feeds spring and summer streamflow [49]. However, a heavy snowpack may also delay the start of the growing season for juvenile trees. Consistent with this, at least one of the associated SNPs was in a gene involved in light responses.

In the G2P analysis, most of the SNPs associated with control phenotypic traits were linked with root-to-shoot ratio (R2S) and number of abaxial stomatal rows (NR_AB,) while most of the SNPs associated with phenotypic responses to drought were linked with shoot weight (SW), root length (RL), and R2S ratio. 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 [50–53]. This may indicate acclimation to drought at the cost of overall low growth of aboveground structures. 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 G2P analyses. The highest degree of overlap was in drought responsiveness of RL and R2S and of R2S and NR_AB (Fig. 6).

Eckert et al. (2015) found two SNPs associated with both environmental PCAs and measured phenotypes out of 31 and 6, respectively – a low rate, but one which led us to think we might see more with a higher number of associations. We found no overlaps in specific SNPs between our G2E and G2P analyses, but there was substantial overlap in functional categories, which directly related to drought tolerance or other environmental responses in previous studies [44, 46, 47, 54]. The prevalence of genetic associations related to abscisic acid (ABA)-signaling pathways and ubiquitination in both G2E and G2P analyses is consistent with prior studies of drought response in conifers [47]. Increasing ABA concentrations are used as a signal to keep stomata closed during dry conditions, reducing water loss [55]. In addition, ABA signaling can also affect shoot growth and water uptake [56, 57]. Ubiquitination has been found to be involved in drought responses in model species by playing a role in ABA-mediated dehydration stress responses [58, 59], or through the downregulation of plasma membrane aquaporin levels [60]. The study of the role of ubiquitin in conifer drought response is still somewhat limited. A study in black spruce (*Picea mariana*) identified 16 out of 313 candidate genes correlated with precipitation, including the genes in the ubiquitin protein handling pathway [61]. 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 [47]. 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 play a role in shaping the hydraulic safety of water-transporting cells, something that can be quite plastic in pines (Lauder et al. 2019). Other than these functions directly related to drought tolerance or other environmental responses, the other overlapping functions among G2E and G2P 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 G2E and G2P analysis with GBS to uncover potentially important adaptive genetic variation.

For many of the other loci associated with environmental gradients, gene ontology results were too vague to draw many conclusions about their function or why the association might exist. However, some of these genes have been previously associated with stress, including Ras-related protein RABC1 to drought responses [62] and pentatricopeptide repeat-containing protein to cold stress [63]. Two of the SNPs associated with minimum temperature are found in the intragenic regions of CGS1 and RE2, genes found to be upregulated during cold stress [64] and heat stress [65], respectively.

Conclusions

In conclusion, by investigating adaptive genetic variation in ponderosa pine with G2E and G2P 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 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 in speeding up selection for drought tolerance or selecting appropriate seed sources for a changing climate. In addition, our results should open up new opportunities for functional studies to determine the molecular roles of the genes underlying these associated genetic markers in influencing trees adaptation.

Methods

Sampling

The original source locations for the 223 *P. ponderosa* genotypes in the Chico orchard (Fig. 1) fall within just one of the several genetic subdivisions previously identified in ponderosa pine [66–68]. Fresh needles were collected from these individuals and placed in labeled tea bags over silica gel to quickly dry them and preserve the DNA for extraction.

Seeds were collected from a subset of 50 genotyped parent trees in the summer of 2018. Because pines are wind-pollinated and outcrossing [67], seeds from the same tree are mostly half-siblings, occasionally full-sibs. We placed 2–3 mature cones from each mother tree into paper bags and placed them in a warm dry place until seeds were released. The seeds were stored in a refrigerator until the greenhouse experiment was carried out (see greenhouse experiment section below). All voucher specimens were maintained at Moran's Lab in University of California, Merced.

DNA sequencing

DNA was extracted from the dried needles using a modified Qiagen plant kits protocol and quantified using an Eppendorf BioSpectrometer (Eppendorf, AG, Germany). Samples were frozen and sent to the UC Davis Genome Center for processing. Four 48-plex GBS libraries consisting of 47 DNA samples and a negative control (no DNA) and one 36-plex GBS library consisting of 35 DNA samples and a 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

No reference genome is available for ponderosa pine (*Pinus ponderosa*), but one does exist for loblolly pine (*Pinus taeda*) [69, 70]. Of the conifers that have been sequenced to date, *P. taeda* is the most closely related to *P. ponderosa* [71, 72]. Furthermore, the *P. taeda* reference genome was used to successfully design probes for sequence capture in *P. contorta* [73, 74]. Based on preliminary analyses, we selected the Stack v.2.2 pipeline [75] with this reference genome (<https://treegenesdb.org/FTP/Genomes/Pita/>) for SNP calling (Shu & Moran, in review). Each step in the Stacks reference pipeline is performed internally in Stacks algorithms except alignment with BWA v.0.7.17 [76] and the Samtools v.1.9 [77] step used to get read position. Default settings were used in Stacks, BWA and Samtools.

SNP filtering

After calling the SNPs, we ran SnpEff [78] to identify the location of gene that the SNP locates. We built the data base with the 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 [79, 80], 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. We therefore filtered out the intergenic SNPs first 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) [81]. The five variables were 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 1st snow pack (PCK4).

Environmental associations

We used LFMM2, which was developed for G2E association and has been shown to outperform other similar approaches with several orders-of-magnitude faster computing [82], and which also controls for the effects of demographic processes and population structure on the distribution of genetic variation [83]. This approach is robust to high amounts of missing data, such as GBS sequencing tends to produce, when sample sizes are > 100 [84]. LFMM2 regression models combine fixed and latent effects with the following equation

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

where \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 \mathbf{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 K can be determined by model choice procedures. The K factors represent unobserved confounders - usually geographical structure in the genotypes of the samples - represented as an $n * K$ matrix, \mathbf{U} . \mathbf{V} is a $p \times K$ matrix of loadings. The matrix \mathbf{U} is obtained from a singular value decomposition (SVD) of the matrix.

$$\mathbf{W} = \mathbf{U}\mathbf{V}^T$$

To determine K , we used the two approaches implemented in the LEA v.2.6.0 R package: principal component analysis (PCA) and admixture analysis [85, 86]. 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` [87]. Second, we ran the LEA function `snmf` for K values between 1 and 5 with 10 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 a false rate of 5% ($q \leq 0.05$) using the R package `QVALUE` [88].

Greenhouse experiment

We used fifty half-sib families that span the climatic range of ponderosa pine in California for the greenhouse experiment and G2P analyses. We aimed to have 10 seedlings from each maternal family in both wet and dry treatments, 1000 seedlings in total. During winter 2018, the seeds were stratified to break dormancy by placing them in aerated water for 48 hours, then surface-drying them and placing them in plastic bags in the refrigerator ($\sim 1.7^\circ\text{C}$) for 6 weeks. Forty-eight of the 50 families had enough seeds in their cones to be included in the experiment (Additional file 1: Fig. S4).

Because maximum seedling root length in a pilot experiment conducted in 2017 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, and 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 [89]. 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 any 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) (Additional file 1: Fig. S4). The wet treatment group was watered twice every week 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 months, this occasional watering was necessary in the greenhouse environment to prevent complete mortality. 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 make 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. We calculated shoot growth as final height minus height at the initiation of the treatments. The length of fresh roots was measured from soil surface to taproot tip immediately after the harvesting, to avoid shrinkage. Following harvest, needles, fresh stem and fresh roots of all the seedlings were separately put into paper lunch bags and dried at 75°C for 48 hours. We measured root dry mass (RW) as well as shoot weight (SW, total of stem and needles). We then calculated 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 at the tip. Approximately 1.96 mm lengths of needle were surveyed for 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. Individual seedling means were calculated by averaging adaxial (AD) stomatal density and number of stomatal rows on both sides (AB & AD) across sampled needles. Only 42 out of 48 mother trees had enough germination to carry out these measurements across both treatments.

Genotype-phenotype association analysis

The 42 individual mother trees had already been genotyped, and we used these same SNPs for the G2P association analysis, focusing on the traits significantly associated with drought treatments. The

breeding value (BV) of a tree reflects the tendency of an individual to produce offspring with high values of that trait and is estimated by measuring relatives [16, 90]. For the wet treatment traits, we use the average trait value across all members of each family in the wet treatment as the BV. For the drought response traits, we deduct the average trait value for a given family in the wet treatment from the value for each offspring of that family in the drought treatment and then use the mean difference as the BV. We used LFMM 2 (Caye et al. 2019) to run the genotype to phenotype association analysis, and then identified associations based on $p (< 10^{-5})$ value.

Gene annotation

After identifying the significantly associated SNPs, 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 [91, 92] was used to further identify the potential functions of the genes. If a SNP is located 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 given for the genes containing the introns were too far apart; no hits could be obtained by Blastx.

Abbreviations

G2E: Genotype-to-environment

G2P: Genotype-to-phenotype

GBS: Genotyping-by-sequencing

SNP: Single nucleotide polymorphism

GWAS: Genome-wide association studies

NGS: Next generation sequencing

PCA: Principal component analysis

LFMM2: Latent factor mixed model 2

CWD: Climatic water deficit

TMIN: Minimum winter temperature

TMAX: Maximum summer temperature

PPTW: Winter precipitation

PCK4: April 1st snow pack

GR: Height growth

SW: Shoot weight

RL: Root length

R2S: Root-shoot dry mass ratio

SD_AD: Stomata density on adaxial side

NR_AB: Number of stomatal rows on abaxial side

BV: Breeding value

ABS: Abscisic acid

Declarations

Ethics approval and consent to participate

The permission of ponderosa pine needles and seeds collection for the experiment was obtained from Chico Orchard, CA. The authors declare that the experiments comply with the current laws of the country in which they were performed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Raw DNA sequencing data: available at National Center for Biotechnology Information under BioProject number PRJNA707049. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA707049>. Individual SNP genotypes: available on Dryad. DOI: <https://doi.org/10.6071/M3DQ1D>.

Funding

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 made use of the MERCED computer cluster at UC Merced (supported 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, but did not play any role in the design of the study nor in the collection analysis and interpretation of data, nor in the writing of the manuscript.

Authors' contributions

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

EM: Research design, edited paper.

All authors have read and approved the manuscript.

Acknowledgements

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. Also, special thanks to Vanessa Centeno and Shirley Calderon at UC Merced for greenhouse plant care.

References

1. Hoffmann AA, Sgrò CM. Climate change and evolutionary adaptation. *Nature*. 2011;470:479–85.
2. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nature Reviews Genetics*. 2013;14:807–20.
3. Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*. 2014;7:1008–25.
4. Rice KJ, Emery NC. Managing microevolution: restoration in the face of global change. *Frontiers in Ecology and the Environment*. 2003;1:469–78.
5. Bell G, Gonzalez A. Evolutionary rescue can prevent extinction following environmental change. *Ecology Letters*. 2009;12:942–8.
6. Langlet O. Two Hundred Years Genecology. *Taxon*. 1971;20:653–721.
7. Ying CC, Liang Q. 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*. 1994;67:281–98.
8. Kitzenmiller JH. Provenance Trials of Ponderosa Pine in Northern California. *Forest Science*. 2005;51:595–607.
9. Wright JW. Local adaptation to serpentine soils in *Pinus ponderosa*. *Plant Soil*. 2007;293:209–17.
10. Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*. 2008;1:95–111.
11. Anderson JT, Panetta AM, Mitchell-Olds T. Evolutionary and Ecological Responses to Anthropogenic Climate Change: Update on Anthropogenic Climate Change. *Plant Physiology*. 2012;160:1728–40.

12. Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, et al. Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology*. 2013;19:1645–61.
13. Neale DB, Kremer A. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics*. 2011;12:111–22.
14. Oney B, Reineking B, O'Neill G, Kreyling J. Intraspecific variation buffers projected climate change impacts on *Pinus contorta*. *Ecology and Evolution*. 2013;3:437–49.
15. Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity*. 2014;113:343–52.
16. Isik F. Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests*. 2014;45:379–401.
17. Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, et al. Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics*. 2010;185:969–82.
18. Eckert AJ, Maloney PE, Vogler DR, Jensen CE, Mix AD, Neale DB. Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics & Genomes*. 2015;11:42.
19. Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*. 2013;9:901–11.
20. Lu M, Loopstra CA, Krutovsky KV. 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*. 2019;9:6798–809.
21. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*. 2011;334:83–6.
22. Jaramillo-Correa J-P, Rodríguez-Quilón I, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, et al. Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*. 2015;199:793–807.
23. Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, et al. Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics*. 2009;182:1289–302.
24. Holliday JA, Ritland K, Aitken SN. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist*. 2010;188:501–14.
25. Mahony CR, MacLachlan IR, Lind BM, Yoder JB, Wang T, Aitken SN. Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications*. 2020;13:116–31.

26. Hamilton JA, Lexer C, Aitken SN. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* × *P. glauca*) hybrid zone. *New Phytologist*. 2013;197:927–38.
27. Dillon S, McEvoy R, Baldwin DS, Rees GN, Parsons Y, Southerton S. Characterisation of Adaptive Genetic Diversity in Environmentally Contrasted Populations of *Eucalyptus camaldulensis* Dehnh. (River Red Gum). *PLOS ONE*. 2014;9:e103515.
28. Housset JM, Nadeau S, Isabel N, Depardieu C, Duchesne I, Lenz P, et al. Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. *New Phytologist*. 2018;218:630–45.
29. Uchiyama K, Iwata H, Moriguchi Y, Ujino-Ihara T, Ueno S, Taguchi Y, et al. Demonstration of Genome-Wide Association Studies for Identifying Markers for Wood Property and Male Strobili Traits in *Cryptomeria japonica*. *PLOS ONE*. 2013;8:e79866.
30. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 2011;12:499–510.
31. Poland JA, Rife TW. Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*. 2012;5:92–102.
32. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17:81–92.
33. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*. 2011;6:1–10.
34. Poland JA, Brown PJ, Sorrells ME, Jannink J-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE*. 2012;7:e32253.
35. Graham RT, Jain TB. 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, US Department of Agriculture: 1-32. 2005;198:1–32.
36. Johansen AD, Latta RG. Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology*. 2003;12:293–8.
37. Potter KM, Hipkins VD, Mahalovich MF, Means RE. Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): range-wide evolutionary history and implications for conservation. *Am J Bot*. 2013;100:1562–79.
38. Kolb TE, Grady KC, McEtrick MP, Herrero A. Local-Scale Drought Adaptation of Ponderosa Pine Seedlings at Habitat Ecotones. *Forest Science*. 2016;62:641–51.
39. Maguire KC, Shinneman DJ, Potter KM, Hipkins VD. Intraspecific Niche Models for Ponderosa Pine (*Pinus ponderosa*) Suggest Potential Variability in Population-Level Response to Climate Change. *Systematic Biology*. 2018;67:965–78.

40. Griffin D, Anchukaitis KJ. How unusual is the 2012–2014 California drought? *Geophysical Research Letters*. 2014;41:9017–23.
41. Berg N, Hall A. Increased Interannual Precipitation Extremes over California under Climate Change. *Journal of Climate*. 2015;28:6324–34.
42. Fettig CJ, Mortenson LA, Bulaon BM, Foulk PB. Tree mortality following drought in the central and southern Sierra Nevada, California, U.S. *Forest Ecology and Management*. 2019;432:164–78.
43. Jordan R, Hoffmann AA, Dillon SK, Prober SM. Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Molecular Ecology*. 2017;26:6002–20.
44. Creelman RA, Mullet JE. Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. *PNAS*. 1995;92:4114–9.
45. Lyzenga WJ, Stone SL. Abiotic stress tolerance mediated by protein ubiquitination. *Journal of Experimental Botany*. 2012;63:599–616.
46. Stone SL. The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. *Frontiers in plant science*. 2014;5:135.
47. Moran EV, Lauder J, Musser C, Stathos A, Shu MJ. The genetics of drought tolerance in conifers. *New Phytologist*. 2017;216:1034–48.
48. Neale DB, Savolainen O. Association genetics of complex traits in conifers. *Trends in Plant Science*. 2004;9:325–30.
49. Serreze MC, Clark MP, Armstrong RL, McGinnis DA, Pulwarty RS. Characteristics of the western United States snowpack from snowpack telemetry (SNO_W) data. *Water Resources Research*. 1999;35:2145–60.
50. Seiler JR, Johnson JD. Physiological and Morphological Responses of Three Half-Sib Families of Loblolly Pine to Water-Stress Conditioning. *Forest Science*. 1988;34:487–95.
51. Irvine J, Perks MP, Magnani F, Grace J. The response of *Pinus sylvestris* to drought: stomatal control of transpiration and hydraulic conductance. *Tree Physiology*. 1998;18:393–402.
52. Cregg BM, Zhang JW. Physiology and morphology of *Pinus sylvestris* seedlings from diverse sources under cyclic drought stress. *Forest Ecology and Management*. 2001;154:131–9.
53. Taeger S, Sparks TH, Menzel A. Effects of temperature and drought manipulations on seedlings of Scots pine provenances. *Plant Biology*. 2015;17:361–72.
54. Houston K, Tucker MR, Chowdhury J, Shirley N, Little A. The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions. *Front Plant Sci*. 2016;7:984.
55. Brodribb TJ, McAdam SAM, Jordan GJ, Martins SCV. Conifer species adapt to low-rainfall climates by following one of two divergent pathways. *PNAS*. 2014;111:14489–93.
56. Buckley TN. The control of stomata by water balance. *New Phytologist*. 2005;168:275–92.

57. Hamanishi ET, Campbell MM. Genome-wide responses to drought in forest trees. *Forestry (Lond)*. 2011;84:273–83.
58. Ryu MY, Cho SK, Kim WT. The Arabidopsis C3H2C3-Type RING E3 Ubiquitin Ligase AtAIRP1 Is a Positive Regulator of an Abscisic Acid-Dependent Response to Drought Stress. *Plant Physiology*. 2010;154:1983–97.
59. Kim SJ, Ryu MY, Kim WT. Suppression of Arabidopsis RING-DUF1117 E3 ubiquitin ligases, AtRDUF1 and AtRDUF2, reduces tolerance to ABA-mediated drought stress. *Biochemical and Biophysical Research Communications*. 2012;420:141–7.
60. Lee HK, Cho SK, Son O, Xu Z, Hwang I, Kim WT. Drought Stress-Induced Rma1H1, a RING Membrane-Anchor E3 Ubiquitin Ligase Homolog, Regulates Aquaporin Levels via Ubiquitination in Transgenic Arabidopsis Plants. *The Plant Cell*. 2009;21:622–41.
61. Prunier J, Laroche J, Beaulieu J, Bousquet J. 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*. 2011;20:1702–16.
62. Khassanova G, Kurishbayev A, Jatayev S, Zhubatkanov A, Zhumalin A, Turbekova A, et al. Intracellular Vesicle Trafficking Genes, RabC-GTP, Are Highly Expressed Under Salinity and Rapid Dehydration but Down-Regulated by Drought in Leaves of Chickpea (*Cicer arietinum* L.). *Front Genet*. 2019;10:40.
63. Xing H, Fu X, Yang C, Tang X, Guo L, Li C, et al. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Scientific Reports*. 2018;8:1–9.
64. Dinari A, Niazi A, Afsharifar AR, Ramezani A. Identification of Upregulated Genes under Cold Stress in Cold-Tolerant Chickpea Using the cDNA-AFLP Approach. *PLOS ONE*. 2013;8:e52757.
65. Traylor-Knowles N, Rose NH, Sheets EA, Palumbi SR. Early Transcriptional Responses during Heat Stress in the Coral *Acropora hyacinthus*. *The Biological Bulletin*. 2017;232:91–100.
66. Conkle MT, Critchfield WB. Genetic variation and hybridization of ponderosa pine. In: *Ponderosa Pine: the species and its management*, Washington State University Cooperative Extension, 1988: p 27-43. 1988. <https://www.fs.usda.gov/treearch/pubs/32842>. Accessed 12 Mar 2018.
67. Williams CG, editor. The Dynamic Wind-Pollinated Mating System. In: *Conifer Reproductive Biology*. Dordrecht: Springer Netherlands; 2009. p. 125–35. doi:10.1007/978-1-4020-9602-0_8.
68. Potter KM, Hipkins VD, Mahalovich MF, Means RE. Nuclear genetic variation across the range of ponderosa pine (*Pinus ponderosa*): Phylogeographic, taxonomic and conservation implications. *Tree Genetics & Genomes*. 2015;11:38.
69. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol*. 2014;15:R59.
70. Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, et al. Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics*. 2014;196:875–90.

71. Willyard A, Cronn R, Liston A. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution*. 2009;52:498–511. doi:10.1016/j.ympev.2009.02.011.
72. Gernandt DS, Hernández-León S, Salgado-Hernández E, Pérez de La Rosa JA. Phylogenetic relationships of *Pinus* subsection *Ponderosae* inferred from rapidly evolving cpDNA regions. *Systematic Botany*. 2009;34:481–91.
73. Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, et al. Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*. 2016;16:1136–46.
74. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, et al. Convergent local adaptation to climate in distantly related conifers. *Science*. 2016;353:1431–3.
75. Rochette NC, Catchen JM. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*. 2017;12:2640–59.
76. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
77. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987–93.
78. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*. 2012;6:80–92.
79. Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol Ecol*. 2008;17:3599–613.
80. Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, et al. Genomic selection in maritime pine. *Plant Sci*. 2016;242:108–19.
81. Flint LE, Flint AL, Thorne JH, Boynton R. Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. *Ecol Process*. 2013;2:1–21.
82. Caye K, Jumentier B, Lepeule J, François O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol Biol Evol*. 2019;36:852–60.
83. Wang J, Zhao Q, Hastie T, Owen AB. CONFOUNDER ADJUSTMENT IN MULTIPLE HYPOTHESIS TESTING. *Ann Stat*. 2017;45:1863–94.
84. Xuereb A, Stahlke A, Bermingham M, Brown M, Nonaka E, Razgour O, et al. Effect of missing data and sample size on the performance of genotype-environment association methods. 2017.
85. Frichot E, Schoville SD, Bouchard G, François O. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Mol Biol Evol*. 2013;30:1687–99.
86. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*. 2015;6:925–9.
87. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLoS Genetics*. 2006;2:e190.

88. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *PNAS*. 2003;100:9440–5.
89. Bales RC, Hopmans JW, O’Geen AT, Meadows M, Hartsough PC, Kirchner P, et al. Soil Moisture Response to Snowmelt and Rainfall in a Sierra Nevada Mixed-Conifer Forest. *Vadose Zone Journal*. 2011;10:786–99.
90. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*. 2001;157:1819–29.
91. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43:D204–12.
92. Bateman A, Martin MJ, O’Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–69.

Figures

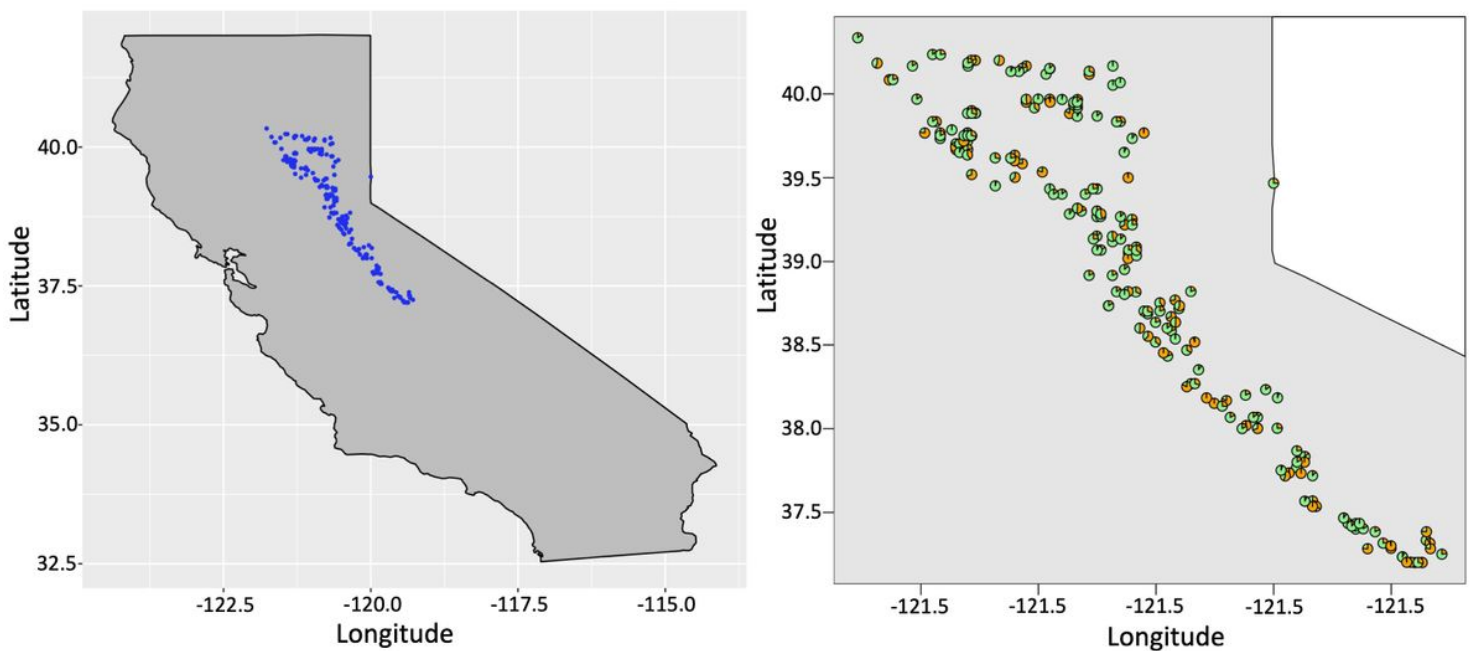


Figure 1

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.

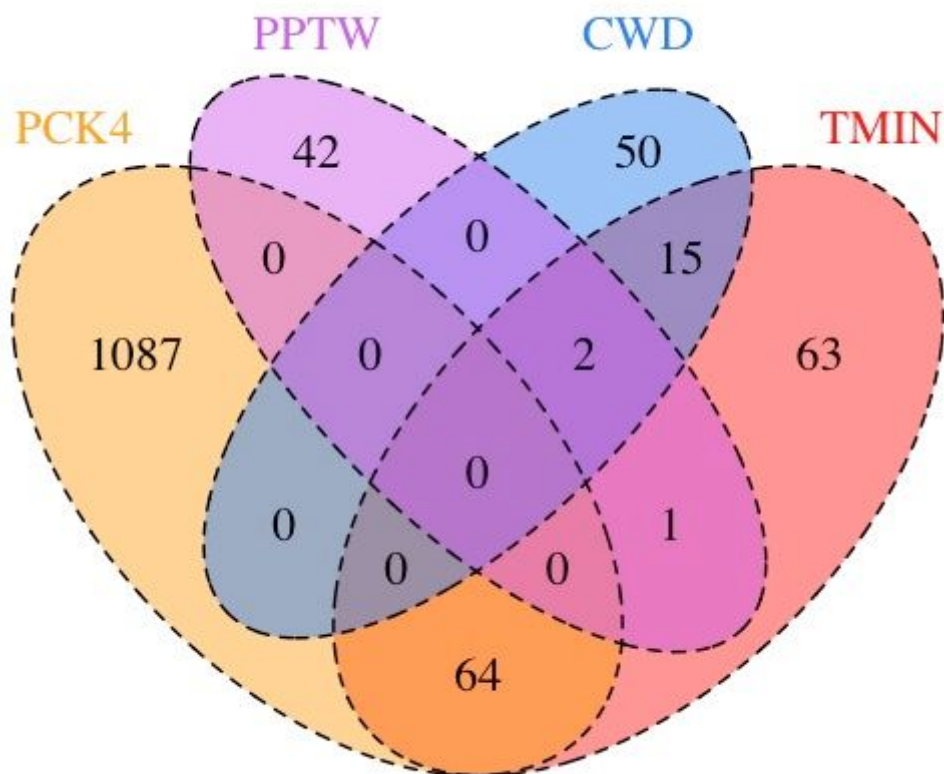


Figure 2

Venn diagram comparing overlap in environmentally associated SNPs. The number of overlapping SNPs that are associated with four climatic variables between April 1st snow pack (PCK4); mean monthly winter precipitation (PPTW); Mean climatic water deficit (CWD), and mean minimum winter temperature (TMIN). The 50 SNPS associated with mean maximum summer temperature (TMAX) did not overlap with the other four climatic variables, and so are not included in this figure.

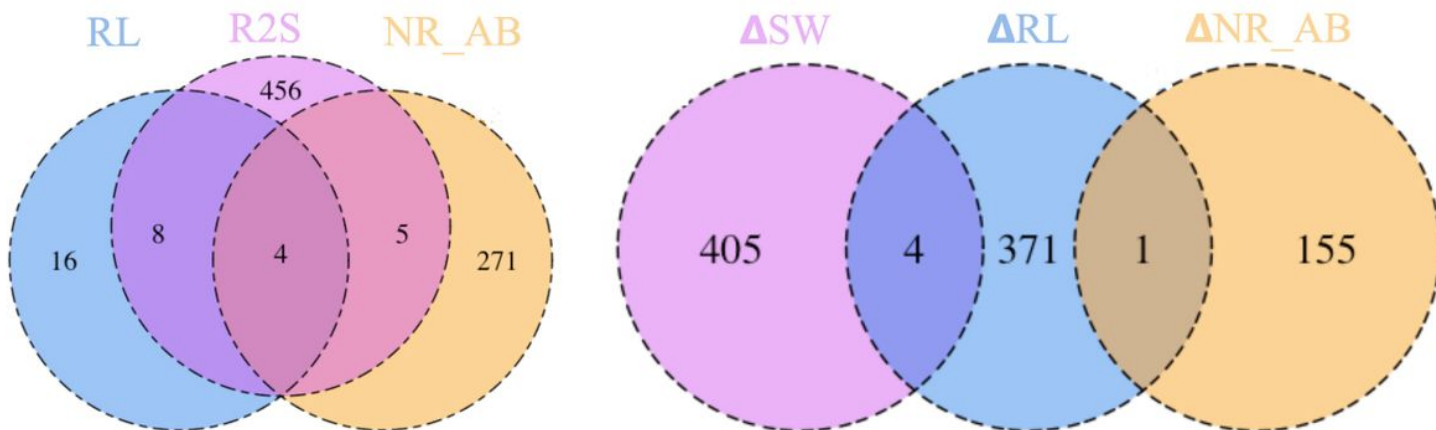


Figure 3

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 other categories. Right: Overlap in SNPs significantly associated with drought responsiveness of shoot weight (Δ SW); root length (Δ RL); and number of stomatal rows on abaxial side (Δ NR_AB). SNPs associated with drought responsiveness of height growth (150), adaxial stomatal density (57), and R2S (6) did not overlap with any other categories.

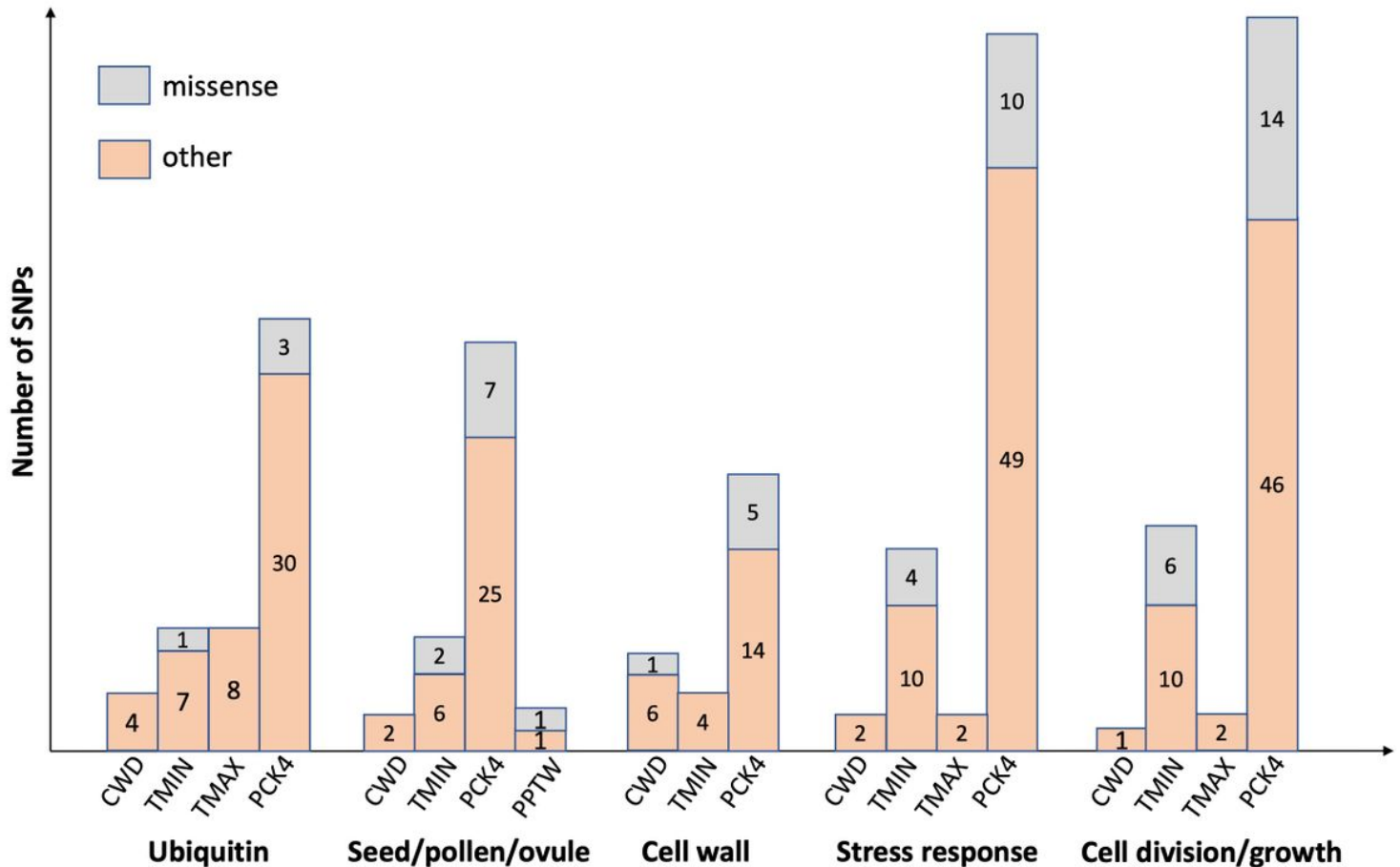


Figure 4

Five type 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 snow pack (PCK4), and Monthly winter precipitation (PPTW). Missense (non-synonymous) SNPs are shown in grey, other types of SNP in orange.

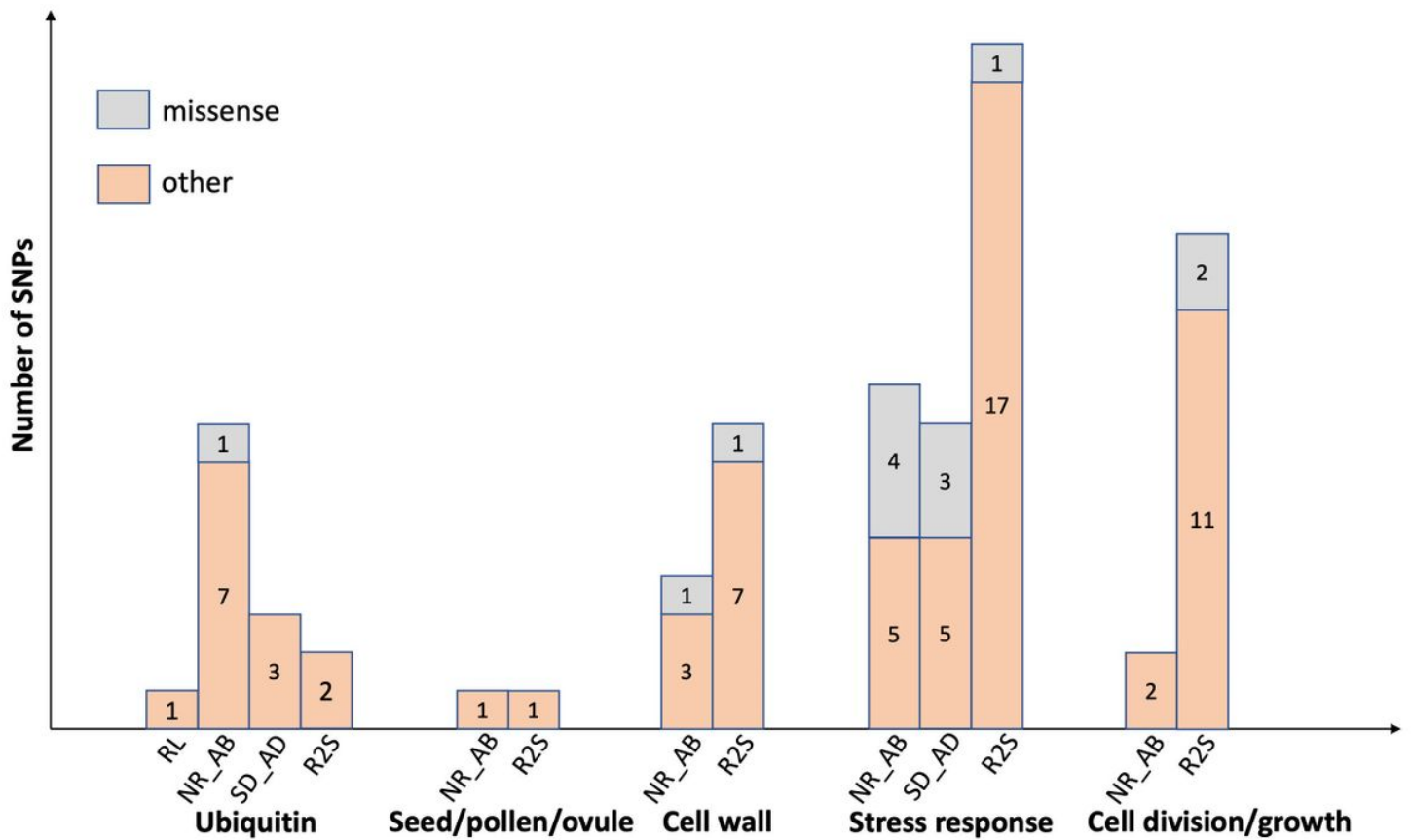


Figure 5

Five type 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, other types of SNP in orange.

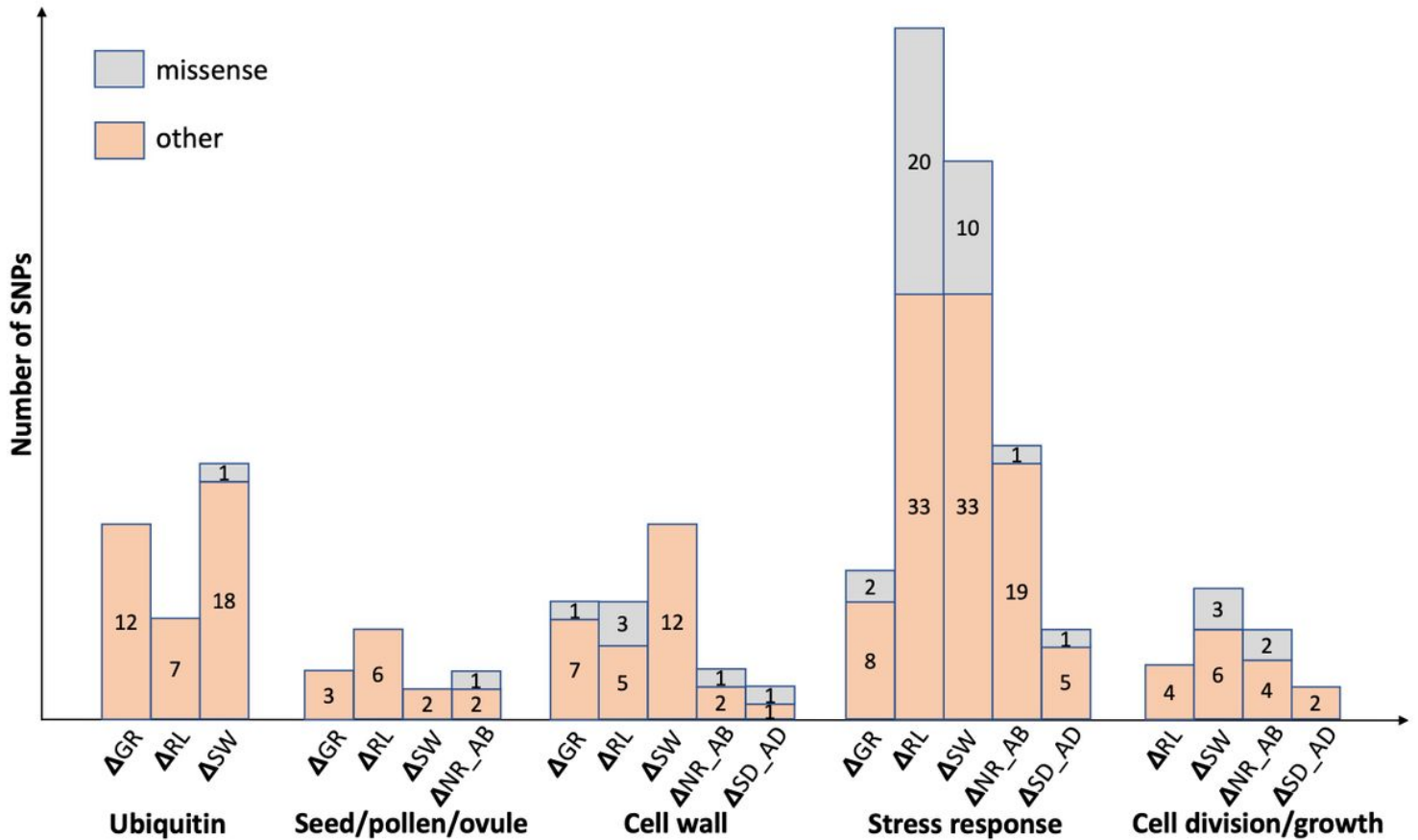


Figure 6

Five type 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, other types of SNP in orange.

Supplementary Files

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

- [SM1.docx](#)
- [SM2G2E.xlsx](#)
- [SM3G2Pcontrol.xlsx](#)
- [SM4G2Pdroughtresponse.xlsx](#)