Article

Selective Sweeps and Polygenic Adaptation Drive Local Adaptation along Moisture and Temperature Gradients in Natural Populations of Coast Redwood and Giant Sequoia

Amanda R. De La Torre [1],*, Manoj K. Sekhwal [1] and David B. Neale [2],*

1   School of Forestry, Northern Arizona University, 200 E. Pine Knoll, Flagstaff, AZ 86011, USA;
    manoj.kumar@nau.edu
2   Department of Plant Sciences, University of California-Davis, One Shields Avenue, Davis, CA 95616, USA
*   Correspondence: Amanda.de-la-torre@nau.edu (A.R.D.L.T.); dbneale@ucdavis.edu (D.B.N.)

Abstract: Dissecting the genomic basis of local adaptation is a major goal in evolutionary biology and conservation science. Rapid changes in the climate pose significant challenges to the survival of natural populations, and the genomic basis of long-generation plant species is still poorly understood. Here, we investigated genome-wide climate adaptation in giant sequoia and coast redwood, two iconic and ecologically important tree species. We used a combination of univariate and multivariate genotype–environment association methods and a selective sweep analysis using non-overlapping sliding windows. We identified genomic regions of potential adaptive importance, showing strong associations to moisture variables and mean annual temperature. Our results found a complex architecture of climate adaptation in the species, with genomic regions showing signatures of selective sweeps, polygenic adaptation, or a combination of both, suggesting recent or ongoing climate adaptation along moisture and temperature gradients in giant sequoia and coast redwood. The results of this study provide a first step toward identifying genomic regions of adaptive significance in the species and will provide information to guide management and conservation strategies that seek to maximize adaptive potential in the face of climate change.

Keywords: selective sweeps; polygenic adaptation; GEA; climate adaptation; *Sequoiadendron giganteum*; *Sequoia sempervirens*

1. Introduction

Understanding the genomic architecture of local adaptation is of great biological interest and paramount to predicting species' responses to present and future changes in the climate. Populations' response to shifts in trait optima because of changes in the climate will depend on the genetic basis of the trait and the population demography [1]. In traits controlled by many genes, natural selection will increase the frequency of advantageous alleles until the trait matches the new trait optimum [2]. Recent improvements in sequencing methods and genotyping approaches have allowed the detection of signatures of recent or ongoing positive selection from their molecular signature on neutral polymorphism in species genomes [2–4]. While most of this knowledge comes from model and domesticated species, the genomic architecture of local adaptation in natural populations of long-lived non-model species remains understudied [5].

Empirical and theoretical studies have suggested the presence of three different modes in the process of adaptation: hard sweeps, soft sweeps, and polygenic adaptation. In the classical hitchhiking model [6,7], new mutations spread rapidly to fixation due to natural selection, reducing genetic variation at neutral linked sites as they spread. These hard sweeps have received significant attention due to their distinctive signatures in the genome that are easily identified using genome-wide markers [8]. In this mutation-limited mode of adaptation, beneficial mutations are thought to be rare and therefore unlikely to be

present as standing genetic variation. More recent studies have proposed that selective sweeps arising from selection on standing variation or from recurrent mutation (soft sweeps) are more common than originally thought [9–11]. Both hard and soft sweeps reflect only recent or on-going adaptive events since signals such as an excess of rare alleles (picked up by Tajima's D), LD and haplotype statistics typically fade quickly (~0.01–0.1 Ne generations) [11]. In on-going adaptive events, the beneficial allele is still segregating in the population (partial sweep) until it reaches fixation and becomes a complete sweep [11]. In addition to sweeps at individual loci, relatively minor shifts in the allele frequencies of large numbers of small-effect loci (polygenic adaptation) can result in the rapid adaptation of phenotypic traits [12–14]. Selective sweeps and polygenic adaptation have rarely been studied together except for some theoretical [1,3,15–17] and a few empirical studies [18].

In contrast to genome scans that look for markers with high levels of population differentiation (Fst), multivariate genome-wide environment association (GEA) methods are more likely to identify the signals of subtle shifts in allele frequencies in weakly selected loci characteristic of polygenic adaptation [19–22]. Multivariate methods, which analyze many loci simultaneously, are better suited since they consider how sets of markers covary in response to the environment [20]. Constrained ordination procedures, such as principal component analysis (PCA) and redundancy analysis (RDA), combine the advantages of both multivariate methods and genotype-environment association procedures to detect adaptive variation based on genomic data [21,23,24]. Despite the high dimensional nature of genotype and environment, the latent factor mixed model (LFMM), which is a univariate genotype-environment association method, is also a frequently used GEA method [25].

Despite rich literature on the presence of local adaptation based on field and common garden experiments, the genomic basis of adaptation in conifer species lags behind other plant species. Non-model species attributes, such as long-generation times, large population sizes, and highly outcrossing rates, together with large genome sizes (10–30 Gb) have limited the development of genome-wide studies [26]. Genome-wide environmental association studies have been mostly based on pre-selected candidate genes in conifers [27–29], with just a few exceptions [19,30–33]. In all recent genome-wide studies in conifer species, adaptation to the environment occurred through many genes of small effect, characteristic of polygenic adaptation.

Giant sequoia (*Sequoiadendron giganteum*) and coast redwood (*Sequoia sempervirens*) are ancient species the fossil records of which date back to the Jurassic period, a time period in which they were more widely distributed across the northern hemisphere [34,35]. These iconic species are also some of the longest-lived trees on Earth, with records of 3200- and 2200-year-old trees, respectively [36]. Giant sequoia, the most "massive species" on Earth due to its enormous wood volume, of up to 1000 m^3 [37], is a diploid species with a genome size of 8.125 Gbp [38]. Generally considered as an outcrossing, sexually reproducing species, giant sequoias can also reproduce vegetatively up to 20 years of age. Giant sequoias are monoecious, meaning that pollen-producing and seed-bearing cones are borne of different branches of the same individual and develop serotinous cones, which require fire to release seeds [39]. The species grows in discrete groves on the western slope of the Sierra Nevada mountains at 830–2700 m of elevation. The elevational range includes a highly disjunct range consisting of approximately 75 groves and spanning around 420 km north to south [39]. Giant sequoias depend on the melting of the accumulated snow pack from the Sierra mountains and are the most moisture demanding species of the Sierra Nevada mixed conifer forests [40].

Coast redwood (*Sequoia sempervirens*) is considered the tallest tree species on Earth, reaching up to 115 m [41]. It is a hexaploid species with a genome size of 26.5 Gbp [42]. The species is monoecious and reproduces mostly asexually but can infrequently produce flowers and seeds every 10 or 20 years. It is characterized by a particular red-colored wood and an endemic narrow natural distribution range [43]. Once extensively distributed along the Pacific Coast in Oregon and California, populations of coast redwood were severely affected by intensive logging, leading to a reduction by more than 90% of their original

natural distribution by the late 19th century [44]. Both coast redwood and giant sequoia are currently listed as endangered species by the International Union for Conservation of Nature (IUCN) Red List of Threatened Species [45]. Despite their unique ecological and economic importance, the genomics of local adaptation in coast redwood and giant sequoia are insufficiently studied [38,40,46–50]. The objectives of this research were to dissect the genomic basis of local adaptation to climate in natural populations of giant sequoia and coast redwood by (i) locating genomic regions showing signals of natural selection that might provide information about the mode of adaptation, (ii) identifying the main environmental variables driving adaptation in the species, and (iii) understanding the main biological functions of genes associated with environmental variation.

2. Materials and Methods

2.1. DNA Extraction, Sequencing, and SNP Calling

Fresh needle tissue was collected from previously established common gardens spanning the natural distribution of the species (Figure 1). Coast redwood was collected from the hedge orchard growing in Russell Reserve (UC Field Station, Contra Costa County, CA, USA), and giant sequoia was collected from the Foresthill Divide Seed Orchard (Foresthill, CA, USA). A selected ramet was collected from each of the surviving 92 coast redwood (SESE) and 90 giant sequoia (SEGI) genets. In the lab, samples were flash-frozen in liquid nitrogen, stored in a −80 °C freezer for 48 h, and lyophilized. DNA was extracted with an E–Z 96 Plant DNA kit (Omega Biotek, Norcross, GA, USA). DNA was submitted to the UC Davis Genome Center for hybridization of baits for targeted sequencing. Exome capture was used to obtain loci from across the genome, from which SNPs were then identified and used in subsequent analyses. Exome capture reduces the complexity of large genomes by sequencing only coding regions and avoiding repeat regions of the genome. Details of the exome capture design can be found in [51]. The total capture region was 22.078 Mbp in SEGI and 37.529 Mbp in SESE. Libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform. Sequencing capture raw reads were aligned against the genome assemblies of coast redwood v2.1 (treegenesdb.org/FTP/Genomes/Sese) and giant sequoia v2.0 (treegenes.db.org/FTP/Genomes/Segi) using Bowtie2 v2.2.9 [52]. Alignments were sorted and later processed in parallel using SAMtools v1.3.1 and BEDtools v2.25.0 and SNPs were called using BCFtools with default parameters [53,54]. Haplotypes were called using Genome Analysis Toolkit (GATK v.4.1.7.0) HaplotypeCaller and GenotypeGVCF [55]. Functional annotations were obtained from the giant sequoia and coast redwood reference genome annotations (treegenesdb.org) by BLASTP [56] sequence alignment against the NCBI non-redundant protein sequences database (nr) using an *e*-value $< 1 \times 10^{-10}$. BCFtools was used to merge variant call format (VCF) files of individuals for further analysis [57].

2.2. Population Structure

The non-model-based principal component analysis (PCA) method was used to quantify the genetic structure in SEGI and SESE. Raw genotyping data containing high levels of missing data were filtered and imputed using TASSEL v.5.2.72 [58] with the following parameters: minor allele frequency (maf) = 0.05 and maximum allele frequency (max−maf) = 0.9. A minimum count (minimum number of samples in which the site must have been scored to be included in the filtered data set) of 50 was implemented for SEGI and 30 for SESE. The imputation was applied using the LD-KNNi method [59] at TASSEL. The PCA [60] was calculated using TASSEL with a merged VCF file containing 71 individuals of SEGI and 92 individuals of SESE. The ggplot2 R package (R version 4.0.4) was used for PCA visualization. Furthermore, the Python2.x fastStructure algorithm based on a variational framework for posterior inference of K clusters was used for population structure analysis [61]. To run fastStructure, *.bed*, *.bim*, and *.fam* files were generated using PLINK v.1.9 software with filtering parameters of maf = 0.05 and max−maf = 0.4 for SEGI and SESE.

The minimum count (mac) of 50 was implemented for SEGI and 30 for SESE using PLINK [62]. The chooseK.py script at fastStructure was used in order to choose the appropriate number of model components (K) that explain the genetic structure in the dataset. Models in fastStructure were replicated 10 times, with K values from 1 to 10. The pophelper R package was performed to visualize the ancestry bar plots generated from population assignments from the fastStructure results. All R packages were carried out on R version 4.0.4 at RStudio (https://www.rstudio.com/, accessed on 15 April 2021).
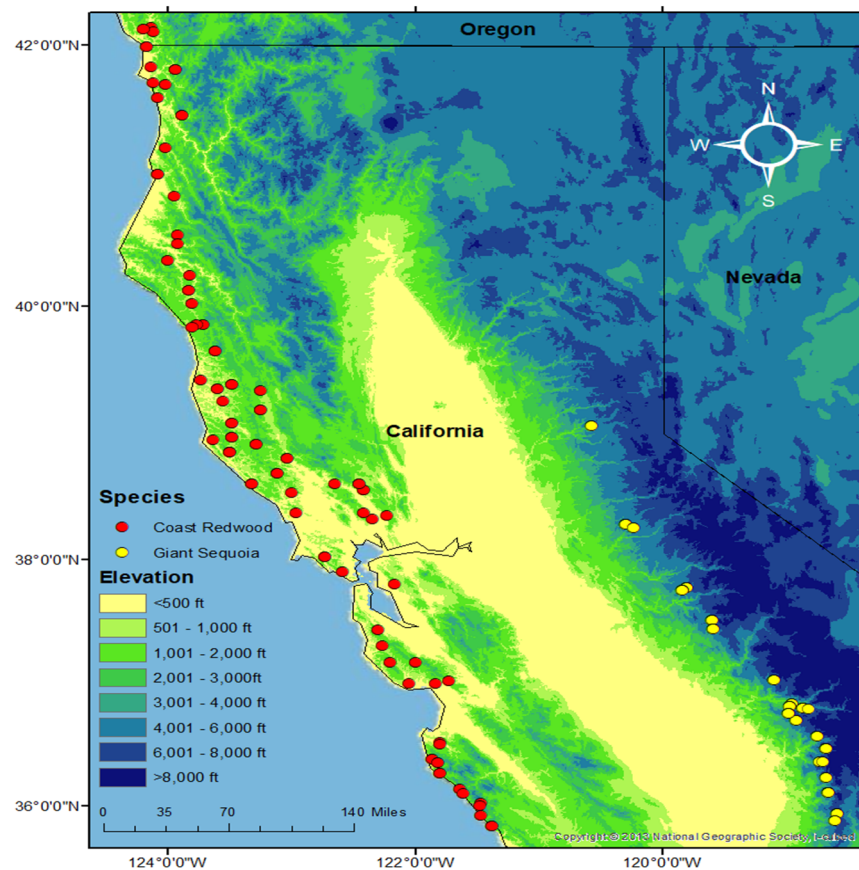

Figure 1. Sampling distribution of coast redwood and giant sequoia.

2.3. Genome-Wide Patterns of Diversity and Differentiation

Patterns of nucleotide diversity (pi) were estimated using non-overlapping sliding windows with a step of 1000 (window-pi-step option) and a window size of 10 kbp (window-pi) in VCFtools [57]. Levels of genetic differentiation among populations were estimated using the Fixation index (Fst) method from [63], with the -weir-fst-pop option in VCFtools [57]. Data were analyzed using sliding windows with a step of 1000 (–fst-window-step option) and a window size of 10 kbp (–fst-window-size option) in VCFtools. Fst was estimated using a priori populations (groves or provenances) based on geographic location. Other population genomics statistics estimated with VCFtools included a chi-squared test of the Hardy–Weinberg equilibrium per SNP (–hardy) and an estimation of the observed and expected number of homozygous sites and inbreeding coefficient (F) per individual (–het).

2.4. Signatures of Positive Selection in Outlier Regions

Genomic regions of exceptionally high or low differentiation were tested for evidence of natural selection in both species using non-overlapping sliding windows of 10 kbp in VCFtools. Signatures of positive selection were considered significant in genomic regions meeting three criteria: (1) increased levels of genomic differentiation accounted by Fst

values that are much larger than the average Fst across windows (after an initial analysis of overall Fst levels, an Fst > 0.1 was used as threshold), (2) reduced levels of nucleotide diversity (closer to zero), and (3) negative Tajima's D values, indicating a skewed allele frequency spectrum toward rare alleles.

2.5. Linkage Disequilibrium (LD) Analysis

Linkage disequilibrium (LD) was measured based on the concept of the square of the correlation coefficient between two loci (r^2), indicating the ability of the alleles present in a marker to predict the presence of alleles on a second marker located at a certain genetic distance measured in base pairs [64]. Squared correlation coefficients (r^2) among all pairs of SNPs within a distance of 500 kbp were calculated using PLINK v.1.9 [62]. To quantify LD, we used TASSEL-filtered dataset and ran PLINK with the following parameters for all r^2 values using –ld-window 100, –ld-window-kb 500 kb, and r^2 threshold 0. Non-linear regression of pairwise r^2 against the physical distance between sites (in base pairs) was performed to estimate the decay of LD in both SESE and SEGI. The decay line was calculated by nonlinear regression to the LD plot using the Hill and Weir equation [65].

2.6. Genome-Wide Environmental Association (GEA)

Coordinates (latitude, longitude, and elevation) representing the geographic origin of the sampled trees were collected for SESE individuals. For SEGI, only the geographic origin of the groves was known; therefore, the coordinates of the centroid of the grove polygon were employed as the geographic origin to obtain environmental data from public databases. Environmental data were obtained from WorldClim 2.0 [66] and ClimateNA [67]. All data were based on the averages for years 1962–1990 and included seasonal and annual variables. Environmental variables were selected based on previous ecological and physiological studies in the species [48,51]. Correlations among all geographic (latitude, longitude, and elevation) and environmental variables were tested in R v3.6.1. Associations among markers and variables were tested using univariate and multivariate methods. Latent factor mixed model (LFMM) version 2 [68] was used to identify univariate genome-wide environmental associations. The K values were determined using PCA, and the ridge approach was used at LFMM [68]. LFMM is computationally efficient and provides statistically optimal corrections, resulting in improved power and control for false discoveries. In the analysis, the lfmm ridge function was used in the lfmm method for estimating latent confounders (or factors). LFMM uses fitted latent factor mixed models to evaluate associations between a SNP genotype matrix and the environmental variable [69]. In addition, a redundancy analysis (RDA) multivariate GEA was implemented in the vegan R package [70]. Since RDA is a regression-based method, it can be subject to problems when using highly correlated predictors [71]; therefore, an $r > 0.7$ was used to exclude highly correlated environmental factors. The function "pairs panels" was used to visualize correlations among the environmental factors.

2.7. Functional Gene Annotations and Enrichment Analyses

The genomic positions of significant SNPs were used to identify the annotated genes by scanning the genomic VCF files of SEGI and SESE. Subsequently, the identified significant SNPs were annotated using annotation files downloaded from TreeGenes (https://treegenesdb.org/TripalContactProfile/588450, accessed on 1 March 2021). The annotation was confirmed using some other approaches, such as pfam [72], blastp [56] and BlastKOALA [73]. Pfam was obtained using the HMMER [74] at default parameters with *e*-value 1.0 to search for protein families. To search for similar hits, blastp ran at the expected threshold 0.05; matrix, BLOSUM 62; and database, non-redundant protein sequence (nr). BlastKOALA in KEGG [75] was performed for protein pathways and functional annotation. Identical matching genes were chosen for identifying annotation and KEGG pathways. Gene ontology (GO) enrichment analyses for biological process, molecular function, and

cellular component were calculated with Blast2GO [76]. False Discovery Rate (FDR) was used for correction for multiple testing.

3. Results

3.1. Sequence Capture and SNP Datasets

A total of 630,166 and 804,682 SNPs were called for 96 SEGI and 92 SESE individuals, respectively. From them, 52,987 (9.2%) SNPs from 71 SEGI individuals and 64,358 (8.0%) SNPs from 92 SESE individuals were retained after filtering using TASSEL. The missing SNPs statistics before and after imputation for each of the SEGI and SESE individuals are reported in Supplementary Tables S1 and S2, respectively. The filtered SNP dataset was retained for further analyses. The imputed dataset was only used for RDA.

3.2. Population Structure

To better understand the genetic structure in the datasets, a principal component (PCAs) and a fastStructure analyses were used to analyze the SNP data. The optimal number of inferred ancestral components (K) was estimated by the chooseK.py script. In giant sequoia, the output showed K = 2 as the model complexity that maximizes marginal likelihood and K = 8 as the number of model components that better explained the structure in dataset. In coast redwood, K = 1 maximized marginal likelihood and K = 3 explained the structure in dataset. Therefore, K = 2–8 was selected to describe the genetic structure of giant sequoia and K = 3 was selected for coast redwood (Figure 2). The PCA analysis showed consistent results to fastStructure, identifying eight genetic clusters in giant sequoia and three clusters in coast redwood (Supplementary Figure S1). These results were consistent with previous studies in the species using more individuals but fewer molecular markers [46,48].
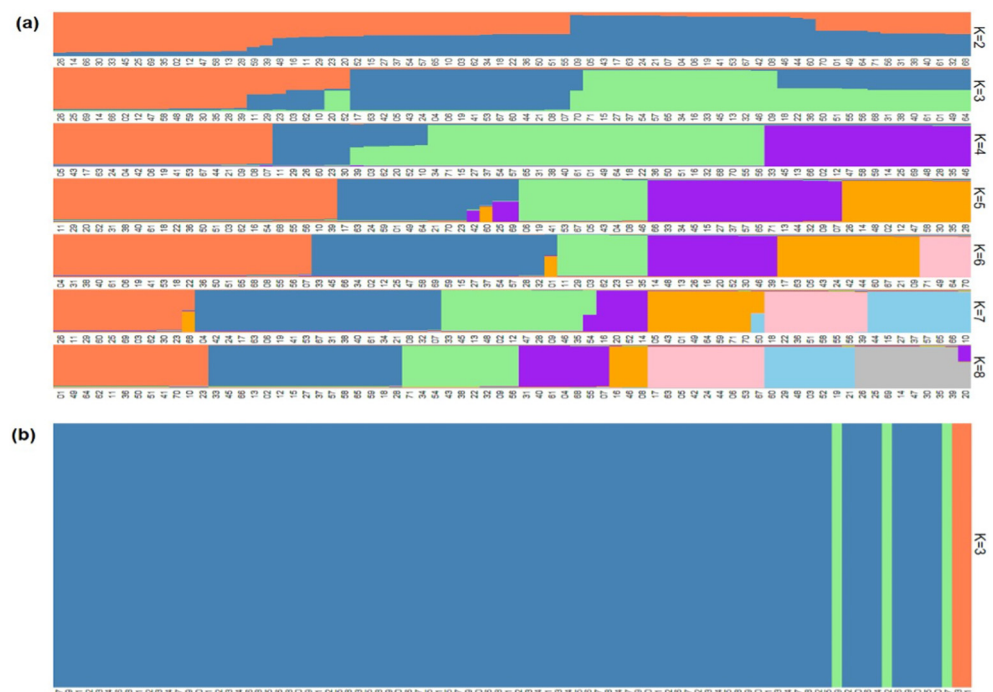

Figure 2. Ancestry plots of sampled individuals. Population structure plots were developed by fastStructure and R package pophelper in giant sequoia (**a**) and coast redwood (**b**).

3.3. Genome-Wide Patterns of Diversity and Differentiation

Nucleotide diversity (pi) estimated using non-overlapping sliding windows averaged 0.0001842 for SESE and 0.00008 for SEGI. Average Fst was estimated as 0.01127 for SESE and 0.01593 for SEGI (Supplementary Figure S2). A chi-squared test suggested that 86.5% of

SNPs are in the Hardy–Weinberg equilibrium in SESE and 46.4% in SEGI. Significant levels of inbreeding (F) as estimated based on the observed and expected number of homozygous sites per individual was found in seven SEGI individuals. No evidence of inbreeding was found in SESE individuals (Supplementary Figure S3).

3.4. Signatures of Positive Selection in Outlier Regions

Negative Tajima's D values were found in 1390 sliding windows, each containing between 182 and 411 SNPs, with a total of 3351 SNPs distributed across the 11 chromosomes of the SEGI genome. From those, 17 genomic regions across all 11 chromosomes (with the exception of chromosomes 6 and 9), each containing one or more consecutive windows, showed elevated Fst values, higher than 0.1, and reduced nucleotide diversity, suggesting signatures of positive selection (Figure 3; Supplementary Table S3). Similarly, 336 sliding windows containing 614 SNPs across 125 scaffolds in the SESE genome also harbored negative Tajima's D values. From these, 29 genomic regions across 21 scaffolds containing one or more sliding windows showed elevated Fst values and reduced nucleotide diversity (Supplementary Table S4).
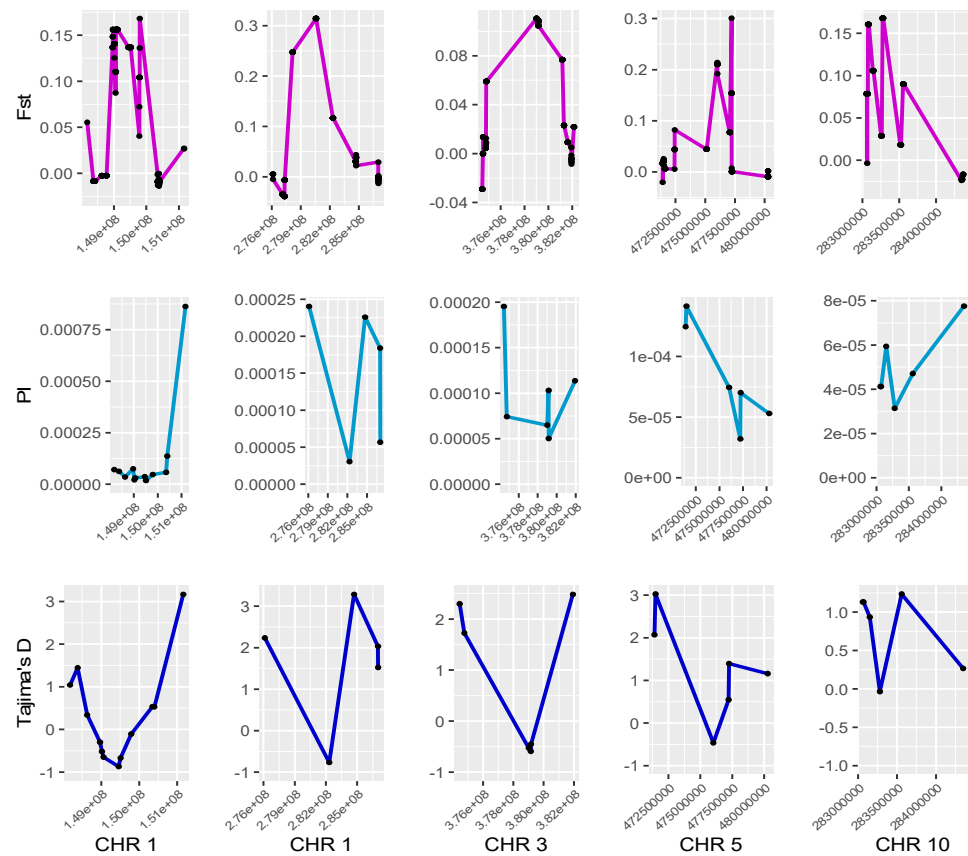

Figure 3. Selective sweeps in the giant sequoia genome. Patterns of nucleotide diversity (PI) and differentiation (Fst and Tajima's D) were evaluated using sliding windows to locate genomic regions showing signatures of positive selection. Five of the longest outlier genomic regions in chromosomes 1, 3, 5 and 10 are shown.

3.5. Genome-Wide Linkage Disequilibrium (LD)

The pattern of LD decay was quantified based on the filtered 52,987 SNPs genotyped in 71 individuals of SEGI and 64,358 SNPs genotyped in 92 individuals of SESE. The LD decay analysis of up to 500 kb of physical distance in SEGI and SESE is shown in Supplementary Figure S3. In SEGI, LD rapidly decayed between 0 and 100 kb and plateaued at ~300 kb. In contrast, LD decayed more rapidly in SESE, reaching r^2 levels of 0.02 at <100 k (Supplementary Figure S4).

3.6. Genome-Wide Environmental Association (GEA)

In SEGI, associations were tested among the SNP markers and six environmental variables: mean annual precipitation (MAP), mean summer precipitation (MSP), number of frost-free days (NFFD), frost-free period (FFP), precipitation as snow (PAS), and climate moisture deficit (CMD). In SESE, tested environmental variables included annual heat moisture index (AHM), mean annual temperature (MAT), MAP, and CMD. Results indicated strong correlations among all environmental factors (Supplementary Figure S5), suggesting the combined effect of groups of individual variables associated with an aspect of climate. Since RDA requires no missing data, an imputation approach was applied to the TASSEL-filtered dataset of total 52,987 SNPs for SEGI and 33,578 SNPs for SESE using the LD-KNNi method [59] at TASSEL. The multivariate RDA method identified significant associations between 292 SNP markers, matching 51 candidate genes, and six environmental variables for SEGI (Figure 4; Supplementary Table S5; Supplementary Figure S6). From these, we observed 68 SNP markers associated with CMD, 11 SNPs with FFP, 83 SNPs with MAP, 61 SNPs with MSP, 44 SNPs with NFFD, and 26 SNPs with PAS in SEGI. In SESE, RDA identified 1016 significant SNPs matching 297 candidate genes. From these, 281 SNPs were associated with AHM, 217 SNPs with CMD, 218 SNPs with MAP, and 300 SNPs with MAT, as measured by their loadings along the RDA axis (Supplementary Table S6). These results identified SNPs showing associations with multivariate aspects of California's Mediterranean climate.
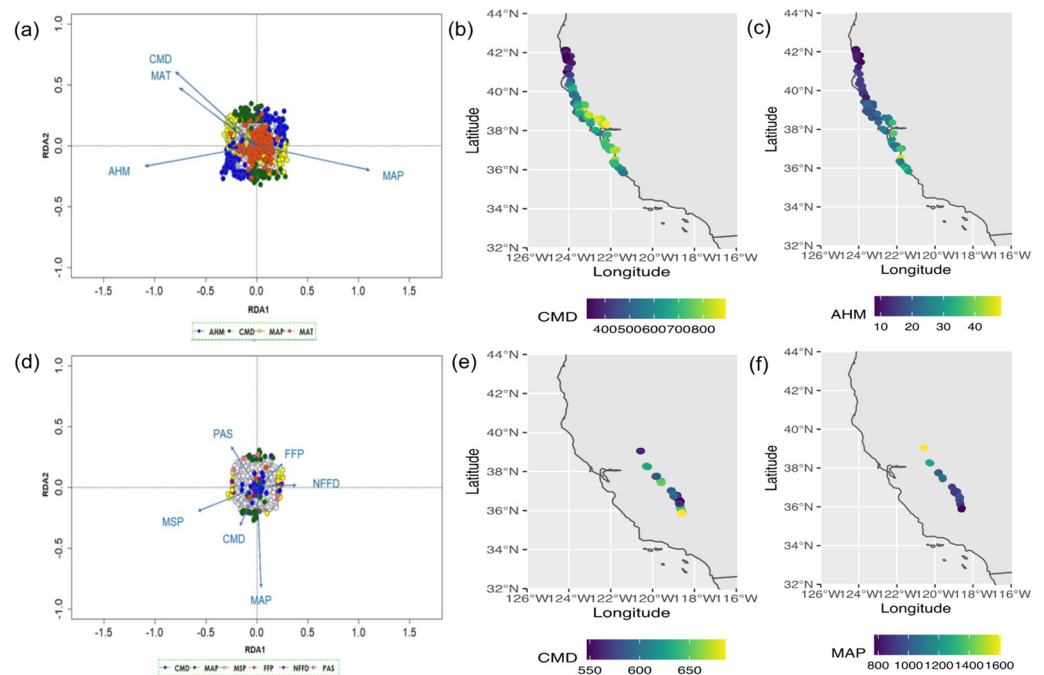

Figure 4. Results of multivariate RDA and distribution maps of environmental variables. Data triplots to highlight SNP loadings on RDA axes 1 and 2 (**a**) in SESE and (**d**) in SEGI. Candidate SNPs are shown as colored points with coding by most highly correlated environmental predictors. SNPs not identified as candidates (neutral SNPs) are shown in light gray. Blue vectors represent environmental predictors. Distribution maps of climate moisture deficit (CMD) and annual heat moisture index (AHM) across collected samples are displayed in (**b**) and (**c**), respectively, for SESE. CMD and mean annual precipitation (MAP) distribution maps are displayed in figures (**e**) and (**f**), respectively.

LFMM results using K = 8, with an FDR threshold of 0.10, detected only three candidate SNPs in response to the environmental predictors MAP, MSP, NFFD, FFP, PAS and CMD in giant sequoia. These three candidate SNPs were at chromosome 2 and gene SEGI 30997, which were annotated as AP-2 complex subunit alpha (Supplementary Table S5). In coast redwood, a total 71 candidate SNPs (eight different genes) were identified in

response to environmental predictors MAP, MAT, AHM, and CMD. Most of these candidate SNPs were located at genes SESE_067476 and SESE_121765 at scaffold 181730 and were annotated as cytochrome P450 family and respiratory burst oxidase, respectively (Supplementary Table S6). The genomic positions of significant SNPs are shown in the Manhattan plots in Supplementary Figure S7. A Venn diagram analyses between GEA methods identified 10 common SNP markers in both RDA and LFMM methods in SESE and no common markers in SEGI (Supplementary Table S7; Supplementary Figure S8).

Significant numbers of SNPs were involved in different metabolic and cellular processes in giant sequoia and coast redwood. For example, genes SEGI_09259 and SEGI_19226, associated with CMD environmental variable, were annotated as oxidoreductases and part of the ribosome biogenesis pathway system, respectively. Genes SEGI_01196, SEGI_17061, SEGI_02591, and SEGI_22391, associated with MAP, were involved in translation factors, MAPK signaling pathway, mRNA surveillance, and N-Glycan biosynthesis, respectively. MSP-associated genes were involved in linoleic acid metabolism, valine, leucine and isoleucine degradation, thermogenesis, and transfer RNA biogenesis. Some NFFD-associated genes were involved in N-Glycan biosynthesis and EGFR tyrosine kinase inhibitor resistance pathways (Supplementary Table S8). In SESE, major pathways of genes associated with annual heat moisture deficit were annotated as involved in glycolysis/gluconeogenesis, nitrogen metabolism, axon regeneration, glycine, serine and threonine metabolism, plant hormone signal transduction, flavonoid biosynthesis, MAPK signaling pathway, longevity regulating pathway, fatty acid degradation, and flavone and flavonol biosynthesis pathways. Genes associated with climate moisture deficit were involved in the pathways of DNA replication, Ubiquitin system, endocytosis, DNA replication, meiosis, zeatin biosynthesis, plant–pathogen interaction, circadian rhythm, glycosyltransferases, glycine, serine and threonine metabolism, MAPK signaling pathway, and monoterpenoid biosynthesis. Similarly, the major pathways of genes associated with mean annual precipitation were oxidoreductases, plant–pathogen interaction, RNA transport, amyotrophic lateral sclerosis, cyanoamino acid metabolism, propanoate metabolism, biosynthesis of secondary metabolites, calcium signaling pathway, cGMP-PKG signaling pathway, serine and threonine metabolism, pentose phosphate pathway, MMDE, C metabolism, and amino acid and nucleotide sugar metabolism (Supplementary Table S9).

3.7. Gene Enrichment Analyses

Results of Blast2GO suggested significant (*p*-value < 0.0001) over-representation of gene ontology in organic substance, primary, nitrogen compound, cellular, and macromolecule metabolic processes in giant sequoia (Supplementary Table S10; Supplementary Figures S9 and S10). In coast redwood, over-represented GO terms included metabolic processes (cellular, primary, nitrogen compound, organic substance, phosphorous, protein, organonitrogen, cellular aromatic compound, macromolecule, and phosphate-containing compound); biosynthetic processes (organic substance, cellular, and cellular macromolecule); binding (heterocyclic compound, small molecule, carbohydrate derivative, protein, organic cyclic compound, ion, nucleic acid, nucleoside phosphate, nucleotide, ribonucleotide, purine, adenyl, phosphorous-containing groups anion, and ATP); protein modification; and catalytic, oxidoreductase and transferase activities (Supplementary Table S11; Supplementary Figure S11).

4. Discussion

4.1. Selective Sweeps and Polygenic Adaptation Drive Genomic Architecture in the Species

The results of this study suggest a complex architecture of local adaptation to climate in coast redwood and giant sequoia shaped by on-going diversifying and stabilizing natural selection acting on a number of genes associated with moisture-related variables. Climate adaptation in the species is mostly driven by many genes of small effect widely distributed across the genomes of coast redwood and giant sequoia, suggesting widespread genomic changes consistent with polygenic adaptation. While recent genome-wide environmental association studies in conifer species have consistently reported polygenic

adaptation [19,30–33], our study suggests a more complex genomic architecture in coast redwood and giant sequoia, with regions of the genomes undergoing polygenic adaptation, other regions with sweep-like signatures, and a smaller third group of genomic regions showing signatures of both polygenic adaptation and selective sweeps.

Despite rich literature on the population genetics and quantitative genetics models of adaptation, the question of whether selective sweeps can occur at QTLs is still poorly understood [12,13]. Chevin and Hospital [3], based on a model with one major locus and infinitely many minor loci, predicted a very low probability of selective sweeps at QTLs. In contrast, other studies have found selective sweeps using simulations of various multi-locus models [16,17] and suggested that sweeps are present even for traits under weak selection where the genomic background explains most of the variation [1]. More recently, it has been suggested that a sudden change in the environment producing a change in the phenotype optimum might produce an initial phase in which directional selection introduces small allele frequency differences that are aligned or opposed to the shift and a final longer phase maintained by stabilizing selection [77].

In our study on coast redwood, we found that 79% of the scaffolds showing signatures of selective sweeps were also associated with regions identified by multivariate RDA (Supplementary Table S4). However, only three of these scaffolds showed signatures of sweeps and polygenic adaptation in the same regions or close genomic location within the scaffolds. Additional testing suggested these results are not likely to have happened by chance (Prob = 0 with 5000 and 10,000 bootstrap). An example of these genes is the transmembrane protein SESE_103346 gene (scaffold 142490), which showed sweep signatures, was associated with mean annual temperature and annual heat moisture index, and was in relatively close proximity to other genes associated with climate moisture deficit within the same scaffold (Supplementary Table S6). The WAT1-related protein SESE_031985 gene (scaffold 195192) showed sweep signatures, was associated with mean annual precipitation and annual heat moisture index, and was in relatively close proximity to other genes associated with climate moisture deficit and annual heat moisture index within the same scaffold (Supplementary Table S6). Finally, about 10 non-coding SNPs were co-located in a genomic region (within scaffold 353338) that showed sweep signatures and was associated with mean annual temperature.

In all cases, the presence of selective sweeps, based on negative Tajima's D; increased Fst; and reduced genetic variance, in which the advantageous alleles have not yet reach fixation (partial sweeps), suggests the presence of on-going directional or diversifying natural selection acting on genes involved in key metabolic processes, stress, transport, and reproduction in both coast redwood and giant sequoia (Supplementary Tables S3 and S4). This suggests on-going or recent genome-wide changes as a consequence of rapidly changing climate conditions, which might be beneficial for the survival of the species in the long term. This adaptive response is most likely aided by standing genetic variation, as expected in long-generation species [78], despite the low levels of nucleotide diversity found in both coast redwood and giant sequoia in this study. Considering the very slow mutation rates in conifers [79], the hard-sweep-like signatures observed in this study could also have originated from standing variation since hard sweeps might sometimes originate when a single copy from standing variation is the ancestor of all beneficial alleles [11]. Further evidence on this topic will require the identification of the ancestral alleles using an outgroup species in genealogical or coalescent history studies [11].

4.2. Patterns of Genomic Diversity and Divergence

Highly outcrossing conifers are expected to have a rapid LD decay. It has been reported that the r^2 decayed to less than 0.20 within approximately 1500 bp based on 19 candidate genes in loblolly pine [80]. In spruces, LD displayed diverse patterns among different genes or the same genes in different species, declining rapidly to half between a few base pairs and 2000 bp [81]. In Douglas-fir, LD decayed by greater than 50% over relatively short segments from $r^2 = 0.25$ to 0.10 within 2000 bp based on 18 genes [82]. In

this study, we found a rapid decay of LD in both coast redwood and giant sequoia, even though the former is mostly an asexual species. In spite of the genome-wide low levels of nucleotide diversity, coast redwood showed an elevated number of heterozygous sites per individual, based on the inbreeding coefficient F (Supplementary Figure S2). This could be due to an accumulation of alleles during mitotic divisions in asexual reproduction [83]. Alternatively, a mostly asexual species, such as coast redwood, could take advantage of non-additive genetic variation in the presence of a heterozygous advantage to achieve an equilibrium where most individuals are heterozygous [84].

4.3. Moisture-Related Variables Drive Adaptation in the Species

Our study indicates that climate adaptation in the species is driven by many genes involved in important biological functions related to metabolic processes, stress and signaling pathways, growth, plant defense, gene expression regulation, and other mechanisms that strongly depend on variation in moisture-related environmental variables (Supplementary Tables S8 and S9). In our GEA analysis, we identified 292 SNPs associated with different environmental variables, such as climate moisture deficit (CMD), mean annual precipitation (MAP), mean summer precipitation (MSP), number of frost-free days (NFFD), frost-free period (FFP), and precipitation as snow (PAS) in giant sequoia, and 1016 significant SNPs associated with annual heat moisture index (AHM), mean annual temperature (MAT), MAP, and CMD in coast redwood (Figure 4), suggesting that the study of climate adaptation requires the analysis of groups instead of individual environmental variables. Previous studies have identified modular aspects of climate adaptation in species such *Pseudotsuga menziesii* [30], *Pinus contorta* [85], and *Pinus taeda* [19]. Other GEA studies were able to identify sets of loci associated with specific climate variables, such as relative humidity and vapor water deficit, in *Abies alba* [86]; maximum temperature and annual precipitation in *Pinus cembra* and *Pinus mugo* [87]; aridity index in *Eucalyptus* [88]; mean annual temperature and precipitation in *Populus trichocarpa* [89]; and mean coldest-month temperature, extreme minimum temperature over a 30-year period, and mean annual precipitation in *P. trichocarpa* [90].

4.4. Functional Annotation of Genes Associated with Environmental Variables

Trees respond to changes in the environment in numerous ways, reflected as physiological, genetic, cellular, and morphological changes [91]. In this study, we identified genes associated with different environmental factors involved in many important biological processes, such as secondary metabolism (terpene, steroids, vitamins, mitochondrial, and chloroplastic), growth and reproductive development, transcription regulation, stress and signal transduction, disease resistance, and DNA processes.

Plant secondary metabolites play a crucial role in adaptation to climate change [92]. In coast redwood, we found secondary metabolism genes associated with several moisture and temperature variables. For example, SESE_068947 was associated with climate moisture deficit; SESE_021525, SESE_025033, and SESE_072616 were associated with mean annual precipitation; and SESE_038374 and SESE_075615 were associated with mean annual temperature. Other important genes involved in secondary metabolism (terpene biosynthesis) are the members of the cytochrome P450 supergene family [93]. We found cytochrome P450 genes SEGI_33526 associated with mean summer precipitation, SESE_100967 and SESE_120015 associated with annual heat moisture index, and SESE_067476 associated with climate moisture deficit and mean annual precipitation. Members of the cytochrome P450 (CYP450) superfamily have been shown to play essential roles in regulating secondary metabolite biosynthesis in *Aralia elata* (Miq.) [94]. A comprehensive genome-wide comparative expression analysis of CYP93 genes in 60 green plants revealed that CYP93 genes in dicots and monocots are preferentially expressed in the roots and tend to be induced by biotic and/or abiotic stresses [95].

Plants employ complex signaling pathways to regulate the expression of genes that allow resistance to environmental stress [96] The ubiquitin-proteasome system controls

many cellular processes by degrading specific proteins and regulates key biological processes, such as hormonal signaling, growth, embryogenesis, senescence and environmental stress, and DNA repair [97,98]. Protein degradation starts when an E1 enzyme joins a Ubiquitin protein in a three-step conjugation cascade (E1 > E2 > E3) that detects specific ubiquitination signals [99]. Our study identified gene SESE_030219 annotated as E3 ubiquitin ligase, SESE_123278 as ubiquitin fusion degradation, and SESE_004990 as ubiquitin carboxyl-terminal hydrolase. E3 Ubiquitin-ligases act as central regulators of many key cellular and physiological processes, including responses to biotic and abiotic stresses in plant species [100]. Recent functional genomics studies have revealed that about 5% of the Arabidopsis genome codes for proteins are involved in the ubiquitination pathway [101].

It has been reported that calcium-dependent signaling and mitogen-activating protein kinases (MAPKs) act during abiotic stress. For instance, in *Populus euphratica*, calcium-dependent protein kinase 10 (CPK10) is expressed under drought and frost and activates both drought- and frost-responsive genes to induce stress tolerance [102]. In *Populus trichocarpa*, MAPK cascades are involved in the promotion of antioxidant stress responses and the expression of drought-related genes [103,104]. In our study, we identified genes SEGI_01196 and SEGI_29292, which were associated with mean annual precipitation and were involved in the MAPK signaling pathway. In coast redwood, MAPK signaling pathway genes included SESE_052680, associated with annual heat moisture index; SESE_027642 and SESE_105796, associated with climate moisture deficit; SESE_121765, SESE_062974, and SESE_121765, associated with mean annual precipitation; and SESE_032016, SESE_035122, and SESE_085808, associated with mean annual temperature.

Stress-inducible dehydrins belonging to the late-embryogenesis abundant (LEA) protein family are regulated by transcription factors (TFs) binding to specific responsive elements, such as ABRE, CRT/DRE/LTRE, MYB, and MYC [105]. Dehydrins are a major group of versatile proteins that play a role in many oxidative stress responses. For example, they participate in the protection of membrane integrity [106]. In our study, we identified genes SESE_085808 (transcription factor, MYC2) and SESE_066024 (MYB-like protein) (Supplementary Table S9).

The most-known R proteins in plants are the nucleotide-binding (NB) site leucine-rich repeat (LRR), which play important roles in plant defense responses to various pathogens. Plant NB-LRR proteins often have, at the N terminus, a Toll/Interleukin-1 receptor (TIR) or coiled coil (CC) domain. In plants, the functions of two additional TIR-containing protein families, TIR-NB site (TN) and TIR-unknown/random (TX), have been investigated in *Arabidopsis thaliana*, suggesting that TN proteins might act in guard complexes monitoring pathogen effectors [107–109]. The study of Zhang et al. [110] provided a set of 167 *NBS-LRR* genes for *Dioscorea rotundata*, which may serve as a primary resource for functional *NBS-LRR* genes against various pathogens [110]. Xu et al. [111] found that the *NBS-LRR* gene (*ZmNBS25*) in maize enhances disease resistance in rice and Arabidopsis [111]. In our study, we have identified SESE_037263, SESE_107044, and SESE_026090 as TIR/NBS/LRR disease resistance proteins that were associated with the annual heat moisture index in coast redwood.

Finally, we annotated several other genes with transporter functions, such as SESE_111042 (NRT1/PTR_family), SESE_115496 (detoxification), SESE_039889 (sugar transporter), SESE_090540 (folate-biopterin transporter), SESE_050983 and SESE_026109 (amino acid transporter), SESE_079497 (nucleobase-ascorbate transporter), and SESE_094426 (NPF family transporter) (Supplementary Table S9). The NRT1/PTR family protein NPF7.3/NRT1.5 reported in Arabidopsis is an indole-3-butyric acid transporter that is involved in root gravitropism [112]. Amino acid transporters have been identified in several model crop species, such as Arabidopsis [113], tomato [114], barley [115], maize [116] and rice [117].

5. Conclusions

This study provides a step toward our understanding of the genomics of adaptation to changing climates in giant sequoia and coast redwood, by identifying genomic regions and

genes of adaptive significance and by reporting the geographical locations of populations or groves of importance for conservation.

References

1. Stetter, M.G.; Thornton, K.; Ross-Ibarra, J. Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. *PLoS Genet.* **2018**, *14*, e1007794. [CrossRef]
2. Pritchard, J.K.; Pickrell, J.K.; Coop, G. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **2010**, *20*, R208–R215. [CrossRef]
3. Chevin, L.M.; Hospital, F. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **2008**, *180*, 1645–1660. [CrossRef]
4. Nielsen, R.; Williamson, S.; Kim, Y.; Hubisz, M.; Clark, A.; Bustamante, C.D. Genomic scans for selective sweeps using SNP data. *Genome Res.* **2005**, *15*, 1566–1575. [CrossRef] [PubMed]
5. Anderson, J.T.; Lee, C.R.; Rushworth, C.A.; Colautti, R.I.; Mitchell-Olds, T. Genetic tradeoff and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **2013**, *22*, 699–708. [CrossRef]
6. Smith, J.M.; Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **1974**, *23*, 23–35. [CrossRef] [PubMed]
7. Kaplan, N.L.; Hudson, R.R.; Langley, C.H. The 'hitchhiking effect' revisited. *Genetics* **1989**, *123*, 887–899. [CrossRef]

8. Sabeti, P.C.; Schaffner, S.F.; Fry, B.; Lohmueller, J.; Varilly, P.; Shamovsky, O.; Palma, A.; Mikkelsen, T.S.; Altshuler, D.; Lander, E.S. Positive natural selection in the human lineage. *Science* **2006**, *312*, 1614–1620. [CrossRef]

9. Hermisson, J.; Pennings, P.S. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **2005**, *169*, 2335–2352. [CrossRef] [PubMed]

10. Przeworski, M.; Coop, G.; Wall, J.D. The signature of positive selection on standing genetic variation. *Evolution* **2005**, *59*, 2312–2323. [CrossRef] [PubMed]

11. Hermisson, J.; Pennings, P.S. Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **2017**, *8*, 700–716. [CrossRef]

12. Jain, K.; Stephan, W. Modes of rapid polygenic adaptation. *Mol. Biol. Evol.* **2017**, *34*, 3169–3175. [CrossRef]

13. Jain, K.; Stephan, W. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics* **2017**, *206*, 389–406. [CrossRef] [PubMed]

14. Pritchard, J.K.; Di Rienzo, A. Adaptation–not by sweeps alone. *Nat. Rev. Genet.* **2010**, *11*, 665–667. [CrossRef]

15. Barghi, N.; Schlotterer, C. Distinct patterns of selective sweep and polygenic adaptation in evolve and resequence studies. *Genome Biol. Evol.* **2020**, *12*, 890–904. [CrossRef] [PubMed]

16. Wollstein, A.; Stephan, W. Adaptive fixation in two-locus models of stabilizing selection and genetic drift. *Genetics* **2014**, *198*, 685–697. [CrossRef]

17. Pavlidis, P.; Jensen, J.D.; Stephan, W.; Stamatakis, A. A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.* **2012**, *29*, 3237–3248. [CrossRef]

18. Sanjak, J.S.; Sidorenko, J.; Robinson, M.R.; Thornton, K.R.; Visscher, P.M. Contemporary directional and stabilizing selection. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 151–156. [CrossRef] [PubMed]

19. De La Torre, A.R.; Wilhite, B.; Neale, D.B. Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. *Genome Biol. Evol.* **2019**, *11*, 2976–2989. [CrossRef] [PubMed]

20. Forester, B.R.; Lasky, J.R.; Wagner, H.H.; Urban, D.L. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* **2018**, *27*, 2215–2233. [CrossRef]

21. De Kort, H.; Vandepitte, K.; Bruun, H.H.; Closset-Kopp, D.; Honnay, O.; Mergeay, J. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Mol. Ecol.* **2014**, *23*, 4709–4721. [CrossRef] [PubMed]

22. Hancock, A.M.; Alkorta-Aranburu, G.; Witonsky, D.B.; DiRienzo, A. Adaptations to new environments in humans: The role of subtle allele frequency shifts. *Philos. Trans. R. Soc. B* **2010**, *365*, 2459–2468. [CrossRef] [PubMed]

23. Lasky, J.R.; Des Marais, D.L.; McKAY, J.K.; Richards, J.H.; Juenger, T.E.; Keitt, T.H. Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Mol. Ecol.* **2012**, *21*, 5512–5529. [CrossRef] [PubMed]

24. Steane, D.A.; Potts, B.M.; McLean, E.; Prober, S.M.; Stock, W.D.; Vaillancourt, R.E.; Byrne, M. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Mol. Ecol.* **2014**, *23*, 2500–2513. [CrossRef]

25. Mitton, J.B.; Linhart, Y.B.; Hamrick, J.L.; Beckman, J.S. Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range. *Theor. Appl. Genet.* **1977**, *51*, 5–13. [CrossRef]

26. De La Torre, A.R.; Birol, I.; Bousquet, J.; Ingvarsson, P.K.; Jansson, S.; Jones, S.J.; Keeling, C.I.; MacKay, J.; Nilsson, O.; Ritland, K.; et al. Insights into Conifer Giga-genomes. *Plant Physiol.* **2014**, *166*, 1724–1732. [CrossRef]

27. Beaulieu, J.; Doerksen, T.; Boyle, B.; Clément, S.; Deslauriers, M.; Beauseigle, S.; Blais, S.; Poulin, P.; Lenz, P.; Caron, S.; et al. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics* **2011**, *188*, 197–214. [CrossRef]

28. Cumbie, W.P.; Eckert, A.J.; Wegrzyn, J.; Whetten, R.W.; Neale, D.B.; Goldfarb, B. Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. *Heredity* **2011**, *107*, 105–114. [CrossRef]

29. Gonzalez-Martinez, S.C.; Huber, D.; Ersoz, E.; Davis, J.M.; Neale, D.B. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **2008**, *101*, 19–26. [CrossRef]

30. De La Torre, A.R.; Wilhite, B.; Puiu, D.; St Clair, J.B.; Crepeau, M.W.; Salzberg, S.L.; Langley, C.H.; Allen, B.; Neale, D.B. Dissecting the polygenic basis of cold adaptation using genome-wide association of traits and environmental data in Douglas-fir. *Genes* **2021**, *12*, 110. [CrossRef]

31. Lu, M.; Krutovsky, K.; Nelson, C.D.; West, J.B.; Reilly, N.A.; Loopstra, C.A. Association genetics of growth and adaptive traits in loblolly pine (*Pinus taeda* L.) using whole-exome-discovered polymorphisms. *Tree Genet. Genomes* **2017**, *13*, 57–75. [CrossRef]

32. Lu, M.; Seeve, C.M.; Loopstra, C.A.; Krutovsky, K.V. Exploring the genetic basis of gene transcript abundance and metabolite levels in loblolly pine (*Pinus taeda* L.) using association mapping and network construction. *BMC Genet.* **2018**, *19*, 100–113. [CrossRef]

33. Yeaman, S.; Hodgins, K.A.; Lotterhos, K.E.; Suren, H.; Nadeau, S.; Degner, J.C.; Nurkowski, K.A.; Smets, P.; Wang, T.; Gray, L.K.; et al. Convergent local adaptation to climate in distantly related conifers. *Science* **2016**, *353*, 1431–1433. [CrossRef] [PubMed]

34. Endo, S. A record of Sequoia from the Jurassic of Manchuria. *Bot. Gaz.* **1951**, *113*, 228–230.

35. Miller, C.N. Mesozoic conifers. *Bot. Rev.* **1977**, *43*, 217–280. [CrossRef]

36. Lanner, R.M. *Conifers of California*; Cachuma Press: Los Olivos, CA, USA, 1999; ISBN 0-9628505-9628503-9628505.

37. Sillett, S.C.; Van Pelt, R.; Carroll, A.L.; Kramer, R.D.; Ambrose, A.R.; Trask, D. How do tree structure and old age affect growth potential of California redwoods? *Ecol. Monogr.* **2015**, *85*, 181–212. [CrossRef]

38. Scott, A.D.; Zimin, A.V.; Puiu, D.; Workman, R.; Britton, M.; Zaman, S.; Caballero, M.; Read, A.C.; Bogdanove, A.J.; Burns, E.; et al. A reference genome sequence for Giant Sequoia. *G3 (Bethesda)* **2020**, *10*, 3907–3919. [CrossRef] [PubMed]

39. Weatherspoon, C.P. Sequoiadendron giganteum (Lindl.) Buchholz Giant Sequoia. 1990. Available online: https://dendro.cnre.vt.edu/dendrology/USDAFSSilvics/136.pdf (accessed on 1 March 2021).

40. Dodd, R.S.; DeSilva, R. Long-term demographic decline and late glacial divergence in a Californian paleoendemic: *Sequoiadendron giganteum* (giant sequoia). *Ecol. Evol.* **2016**, *6*, 3342–3355. [CrossRef] [PubMed]

41. Ishii, H.R.; Azuma, W.; Kuroda, K.; Sillett, S.C. Pushing the limits to tree height: Could foliar water storage compensate for hydraulic constraints in *Sequoia sempervirens*? *Funct. Ecol.* **2014**, *28*, 1087–1093. [CrossRef]

42. Neale, D.B.; Zimin, A.V.; Zaman, S.; Scott, A.D.; Shrestha, B.; Workman, R.E.; Puiu, D.; Allen, B.J.; Sekhwal, M.K.; De La Torre, A.R.; et al. Assembled and annotated 26.5 Gbp coast redwood genome: A resource for estimating evolutionary adaptive potential and investigating hexaploidy origin. *G3 Genes Genomes Genet.* **2021**, in press.

43. Ahuja, M.R. Genetic constitution and diversity in four narrow endemic redwoods from the family Cupressaceae. In *Biodiversity and Conservation of Woody Plants*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; Volume 165, pp. 5–19.

44. Burns, E.E.; Campbell, R.; Cowan, P.D. *State of Redwoods Conservation Report: A Tale of Two Forests, Coast Redwoods, Giant Sequoia*; Save the Redwoods League: San Francisco, CA, USA, 2018.

45. Farjon, A.; Schmid, R. Sequoia sempervirens. The IUCN red list of threatened species 2013, e.T34051A2841558. Available online: https://www.iucnredlist.org/species/34051/2841558 (accessed on 1 April 2021).

46. Breidenbach, N.; Gailing, O.; Krutovsky, K.V. Genetic structure of coast redwood (Sequoia sempervirens [D. Don] Endl.) populations in and outside of the natural distribution range based on nuclear and chloroplast microsatellite markers. *PLoS ONE* **2020**, *15*, e0243556. [CrossRef] [PubMed]

47. Breidenbach, N.; Sharov, V.V.; Gailing, O.; Krutovsky, K.V. De novo transcriptome assembly of cold stressed clones of the hexaploid *Sequoia sempervirens* (D. Don) Endl. *Sci. Data* **2020**, *7*, 239–247. [CrossRef]

48. DeSilva, R.; Dodd, R.S. Association of genetic and climatic variability in giant sequoia, *Sequoiadendron giganteum*, reveals signatures of local adaptation along moisture-related gradients. *Ecol. Evol.* **2020**, *10*, 10619–10632. [CrossRef]

49. Klápště, J.; Meason, D.; Dungey, H.S.; Telfer, E.J.; Silcock, P.; Rapley, S. Genotype-by-environment interaction in coast redwood outside natural distribution search for environmental cues. *BMC Genet.* **2020**, *21*, 15–29. [CrossRef] [PubMed]

50. Scott, A.D.; Stenz, N.W.; Ingvarsson, P.K.; Baum, D.A. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. *New Phytol.* **2016**, *211*, 186–193. [CrossRef]

51. De La Torre, A.R.; Sekhwal, M.K.; Puiu, D.; Salzberg, S.L.; Scott, A.D.; Allen, B.; Neale, D.B.; Chin, A.R.O.; Buckley, T.N. Genome-wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. *bioRxiv* **2021**. [CrossRef]

52. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

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

54. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

55. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]

56. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T.L. NCBI BLAST: A better web interface. *Nucleic Acids Res.* **2008**, *36*, W5–W9. [CrossRef] [PubMed]

57. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

58. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

59. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.-Y.; Myles, S. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3 (Bethesda)* **2015**, *5*, 2383–2390. [CrossRef] [PubMed]

60. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef]

61. Raj, A.; Stephens, M.; Pritchard, J.K. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **2014**, *197*, 573–589. [CrossRef]

62. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.D.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]

63. Weir, B.S.; Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **1984**, *38*, 1358–1370.

64. Ke, X.; Hunt, S.; Tapper, W.; Lawrence, R.; Stavrides, G.; Ghori, J.; Whittaker, P.; Collins, A.; Morris, A.P.; Bentley, D.; et al. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum. Mol. Genet.* **2004**, *13*, 577–588. [CrossRef]

65. Hill, W.G.; Weir, B.S. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet. Res.* **2011**, *93*, 47–64. [CrossRef]

66. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]

67. Wang, T.L.; Hamann, A.; Spittlehouse, D.; Carroll, C. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE* **2016**, *11*, e0156720. [CrossRef]

68. Caye, K.; Jumentier, B.; Lepeule, J.; Francois, O. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* **2019**, *36*, 852–860. [CrossRef]

69. Frichot, E.; Schoville, S.D.; Bouchard, G.; Francois, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **2013**, *7*, 1687–1699. [CrossRef] [PubMed]

70. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R. *Vegan: Community Ecology Package*; CRAN: Helsinki, Finland, 2020. Available online: https://cran.r-project.org/web/packages/vegan/vegan.pdf (accessed on 2 February 2021).

71. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [CrossRef]

72. Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]

73. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef]

74. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef] [PubMed]

75. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [CrossRef]

76. Götz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **2008**, *36*, 3420–3435. [CrossRef]

77. Hayward, L.K.; Sella, G. Polygenic adaptation after a sudden change in environment. *bioRxiv* **2021**, 792952. [CrossRef]

78. Chhatre, V.E.; Fetter, K.C.; Gougherty, A.V.; Fitzpatrick, M.C.; Soolanayakanahally, R.Y.; Zalensy, R.S. Climatic niche predicts the landscape structure of locally adaptive standing genetic variation. *bioRxiv* **2019**. [CrossRef]

79. De La Torre, A.R.; Li, Z.; Van de Peer, Y.; Ingvarsson, P.K. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Biol. Evol.* **2017**, *34*, 1363–1377. [CrossRef]

80. Neale, D.B.; Savolainen, O. Association genetics of complex traits in conifers. *Trends Plant. Sci.* **2004**, *9*, 325–330. [CrossRef]

81. Namroud, M.C.; Guillet-Claude, C.; Mackay, J.; Isabel, N.; Bousquet, J. Molecular evolution of regulatory genes in spruces from different species and continents: Heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. *J. Mol. Evol.* **2010**, *70*, 371–386. [CrossRef] [PubMed]

82. Krutovsky, K.V.; Neale, D.B. Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* **2005**, *171*, 2029–2041. [CrossRef] [PubMed]

83. Birky, C.W. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **1996**, *144*, 427–437. [CrossRef]

84. Peck, J.R.; Waxman, D. What's wrong with a little sex? *J. Evol. Biol.* **2000**, *13*, 63–69. [CrossRef]

85. Lotterhos, K.E.; Yeaman, S.; Degner, J.; Aitken, S.; Hodgins, K.A. Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biol.* **2018**, *19*, 157–181. [CrossRef]

86. Roschanski, A.M.; Csilléry, K.; Liepelt, S.; Oddou-Muratorio, S.; Ziegenhagen, B.; Huard, F.; Ullrich, K.K.; Postolache, D.; Vendramin, G.G.; Fady, B. Evidence of divergent selection for drought and cold tolerance at landscape and local scales in Abies alba Mill. in the French Mediterranean Alps. *Mol. Ecol.* **2016**, *25*, 776–794. [CrossRef]

87. Mosca, E.; Gugerli, F.; Eckert, A.J.; Neale, D.B. Signatures of natural selection on *Pinus cembra* and *P.* mugo along elevational gradients in the Alps. *Tree Genet. Genomes* **2016**, *12*, 9–24. [CrossRef]

88. Steane, D.A.; Potts, B.; McLean, E.H.; Collins, L.; Holland, B.; Prober, S.; Stock, W.D.; Vaillancourt, R.; Byrne, M. Genomic Scans across Three Eucalypts Suggest that Adaptation to Aridity is a Genome-Wide Phenomenon. *Genome Biol. Evol.* **2017**, *9*, 253–265. [CrossRef]

89. Evans, L.M.; Slavov, G.T.; Rodgers-Melnick, E.; Martin, J.; Ranjan, P.; Muchero, W.; Brunner, A.M.; Schackwitz, W.; Gunter, L.; Chen, J.-G.; et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat. Genet.* **2014**, *46*, 1089–1096. [CrossRef] [PubMed]

90. Geraldes, A.J.M.; Farzaneh, N.; Grassa, C.; McKown, A.; Guy, R.D.; Mansfield, S.D.; Douglas, C.J.; Cronk, Q.C.B. Landscape genomics of Populus trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* **2014**, *68*, 3260–3280. [CrossRef]

91. Kijowska-Oberc, J.; Staszak, A.M.; Kaminski, J.; Ratajczak, E. Adaptation of forest trees to rapidly changing climate. *Forests* **2020**, *11*, 123. [CrossRef]

92. Ramakrishna, A.; Ravishankar, G.A. Influence of abiotic stress signals on secondary metabolites in plants. *Plant. Signal. Behav.* **2011**, *6*, 1720–1731.

93. Bathe, U.; Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **2019**, *161*, 149–162. [CrossRef]

94.	Cheng, Y.; Liu, H.; Tong, X.; Liu, Z.; Zhang, X.; Li, D.; Jiang, X.; Yu, X. Identification and analysis of CYP450 and UGT supergene family members from the transcriptome of *Aralia elata* (Miq.) seem reveal candidate genes for triterpenoid saponin biosynthesis. *BMC Plant Biol.* **2020**, *20*, 1–38. [CrossRef] [PubMed]

95.	Du, H.; Ran, F.; Dong, H.-L.; Wen, J.; Li, J.-N.; Liang, Z. Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. *PLoS ONE* **2016**, *11*, e0165020.

96.	Overmyer, K.; Vuorinen, K.; Brosche, M. Interaction points in plant stress signaling pathways. *Physiol. Plant.* **2018**, *162*, 191–204. [CrossRef]

97.	Sharma, B.; Joshi, D.; Yadav, P.K.; Gupta, A.K.; Bhatt, T.K. Role of ubiquitin-mediated degradation system in plant biology. *Front. Plant Sci.* **2016**, *7*, 806–814. [CrossRef] [PubMed]

98.	Xu, F.Q.; Xue, H.W. The ubiquitin-proteasome system in plant responses to environments. *Plant. Cell Environ.* **2019**, *42*, 2931–2944. [PubMed]

99.	Scheffner, M.; Nuber, U.; Huibregtse, J.M. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **1995**, *373*, 81–83. [CrossRef]

100.	Serrano, I.; Campos, L.; Rivas, S. Roles of E3 ubiquitin-ligases in nuclear protein homeostasis during plant stress responses. *Front. Plant Sci.* **2018**, *9*, 139–146. [CrossRef]

101.	Mazzucotelli, E.; Belloni, S.; Marone, D.; De Leonardis, A.; Guerra, D.; Di Fonzo, N.; Cattivelli, L.; Mastrangelo, A. The e3 ubiquitin ligase gene family in plants: Regulation by degradation. *Curr. Genom.* **2006**, *7*, 509–522. [CrossRef]

102.	Chen, J.H.; Xue, B.; Xia, X.L.; Yin, W.L. A novel calcium-dependent protein kinase gene from *Populus euphratica*, confers both drought and cold stress tolerance. *Biochem. Biophys. Res. Commun.* **2013**, *441*, 630–636. [CrossRef]

103.	Hamel, L.P.; Miles, G.P.; Samuel, M.A.; Ellis, B.E.; Séguin, A.; Beaudoin, N. Activation of stress-responsive mitogen-activated protein kinase pathways in hybrid poplar (*Populus trichocarpa* x *Populus deltoides*). *Tree Physiol.* **2005**, *25*, 277–288. [CrossRef]

104.	Wang, L.; Su, H.; Han, L.; Wang, C.; Sun, Y.; Liu, F. Differential expression profiles of poplar MAP kinase kinases in response to abiotic stresses and plant hormones, and overexpression of PtMKK4 improves the drought tolerance of poplar. *Gene* **2014**, *545*, 141–148. [CrossRef] [PubMed]

105.	Hanin, M.; Brini, F.; Ebel, C.; Toda, Y.; Takeda, S.; Masmoudi, K. Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signal. Behav.* **2011**, *6*, 1503–1509. [CrossRef]

106.	Yu, Z.; Wang, X.; Zhang, L. Structural and functional dynamics of dehydrins: A plant protector protein under abiotic stress. *Int. J. Mol. Sci.* **2018**, *19*, 3420. [CrossRef]

107.	Meyers, B.C.; Kozik, A.; Griego, A.; Kuang, H.; Michelmore, R.W. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell* **2003**, *15*, 809–834. [CrossRef]

108.	Meyers, B.C.; Morgante, M.; Michelmore, R.W. TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. *Plant J.* **2002**, *32*, 77–92. [CrossRef] [PubMed]

109.	Nandety, R.S.; Caplan, J.L.; Cavanaugh, K.; Perroud, B.; Wroblewski, T.; Michelmore, R.W.; Meyers, B.C. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* **2013**, *162*, 1459–1472. [CrossRef]

110.	Zhang, Y.-M.; Chen, M.; Sun, L.; Wang, Y.; Yin, J.; Liu, J.; Sun, X.-Q.; Hang, Y.-Y. Genome-wide identification and evolutionary analysis of NBS-LRR genes from dioscorea rotundata. *Front. Genet.* **2020**, *11*, 484–495. [CrossRef] [PubMed]

111.	Xu, Y.; Liu, F.; Zhu, S.; Li, X. The maize NBS-LRR gene ZmNBS25 enhances disease resistance in rice and Arabidopsis. *Front. Plant Sci.* **2018**, *9*, 1033–1046. [CrossRef] [PubMed]

112.	Watanabe, S.; Takahashi, N.; Kanno, Y.; Suzuki, H.; Aoi, Y.; Takeda-Kamiya, N.; Toyooka, K.; Kasahara, H.; Hayashi, K.; Umeda, M.; et al. The Arabidopsis NRT1/PTR FAMILY protein NPF7.3/NRT1.5 is an indole-3-butyric acid transporter involved in root gravitropism. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 31500–31509. [CrossRef]

113.	Liu, X.; Bush, D.R. Expression and transcriptional regulation of amino acid transporters in plants. *Amino Acids* **2006**, *30*, 113–120. [CrossRef]

114.	Schwacke, R.; Grallath, S.; Breitkreuz, K.E.; Stransky, E.; Stransky, H.; Frommer, W.B.; Rentsch, D. LeProT1, a transporter for proline, glycine betaine, and gamma-amino butyric acid in tomato pollen. *Plant Cell* **1999**, *11*, 377–392.

115.	Kohl, S.; Hollmann, J.; Blattner, F.R.; Radchuk, V.; Andersch, F.; Steuernagel, B.; Schmutzer, T.; Scholz, U.; Krupinska, K.; Weber, H.; et al. A putative role for amino acid permeases in sink-source communication of barley tissues uncovered by RNA-seq. *BMC Plant Biol.* **2012**, *12*, 154–172. [CrossRef] [PubMed]

116.	Pan, X.; Hasan, M.; Li, Y.; Liao, C.; Zheng, H.; Liu, R.; Li, X. Asymmetric transcriptomic signatures between the cob and florets in the maize ear under optimal- and low-nitrogen conditions at silking, and functional characterization of amino acid transporters ZmAAP4 and ZmVAAT3. *J. Exp. Bot.* **2015**, *66*, 6149–6166. [CrossRef] [PubMed]

117.	Zhao, H.M.; Ma, H.L.; Yu, L.; Wang, X.; Zhao, J. Genome-wide survey and expression analysis of amino acid transporter gene family in rice (*Oryza sativa* L.). *PLoS ONE* **2012**, *7*, e49210. [CrossRef] [PubMed]