

10-2022

## **Meta-Analysis Reveals Challenges and Gaps for Genome-to-Phenome Research Underpinning Plant Drought Response**

Anthony E. Melton  
*Boise State University*

Stephanie J. Galla  
*Boise State University*

Carlos Dave C. Dumaguit  
*Boise State University*

John M.A. Wojahn  
*Boise State University*

Stephen Novak  
*Boise State University*

*See next page for additional authors*

---

**Authors**

Anthony E. Melton, Stephanie J. Galla, Carlos Dave C. Dumaguit, John M.A. Wojahn, Stephen Novak, Marcelo Serpe, Peggy Martinez, and Sven Buerki



Article

# Meta-Analysis Reveals Challenges and Gaps for Genome-to-Phenome Research Underpinning Plant Drought Response

Anthony E. Melton <sup>\*,†</sup> , Stephanie J. Galla <sup>†</sup> , Carlos Dave C. Dumaguit , John M. A. Wojahn , Stephen Novak , Marcelo Serpe, Peggy Martinez and Sven Buerki <sup>\*</sup>

Department of Biological Sciences, Boise State University, Boise, ID 83725, USA

\* Correspondence: anthonymelton@boisestate.edu (A.E.M.); svenbuerki@boisestate.edu (S.B.)

† These authors contributed equally to this work.

**Abstract:** Severe drought conditions and extreme weather events are increasing worldwide with climate change, threatening the persistence of native plant communities and ecosystems. Many studies have investigated the genomic basis of plant responses to drought. However, the extent of this research throughout the plant kingdom is unclear, particularly among species critical for the sustainability of natural ecosystems. This study aimed to broaden our understanding of genome-to-phenome (G2P) connections in drought-stressed plants and identify focal taxa for future research. Bioinformatics pipelines were developed to mine and link information from databases and abstracts from 7730 publications. This approach identified 1634 genes involved in drought responses among 497 plant taxa. Most (83.30%) of these species have been classified for human use, and most G2P interactions have been described within model organisms or crop species. Our analysis identifies several gaps in G2P research literature and database connectivity, with 21% of abstracts being linked to gene and taxonomy data in NCBI. Abstract text mining was more successful at identifying potential G2P pathways, with 34% of abstracts containing gene, taxa, and phenotype information. Expanding G2P studies to include non-model plants, especially those that are adapted to drought stress, will help advance our understanding of drought responsive G2P pathways.

**Keywords:** arid habitats; climate change; drought stress; green plants; genome-to-phenome; text mining



**Citation:** Melton, A.E.; Galla, S.J.; Dumaguit, C.D.C.; Wojahn, J.M.A.; Novak, S.; Serpe, M.; Martinez, P.; Buerki, S. Meta-Analysis Reveals Challenges and Gaps for Genome-to-Phenome Research Underpinning Plant Drought Response. *Int. J. Mol. Sci.* **2022**, *23*, 12297. <https://doi.org/10.3390/ijms232012297>

Academic Editors: Isabel Marques and Honghong Hu

Received: 21 September 2022

Accepted: 12 October 2022

Published: 14 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Anthropogenically driven climate change is expected to exacerbate drought conditions in landscapes around the world, currently affecting up to 85% of the world's population [1–3], reviewed in [4]. Climate change is causing an increase in mean annual temperatures, which could potentially increase by 2 °C globally by 2050, and nearly 6 °C by 2100 [5]. Changes to precipitation patterns are also expected, with some regions predicted to experience severe flooding or drought with greater frequency [6]. For example, a recent study by Williams et al. [6] revealed that southwestern North America has recently experienced the largest megadrought in the past 1500 years, which is largely attributable to anthropogenic climate change magnifying natural climatic variability [7]. The expected increases in drought due to higher temperatures and reduced precipitation are significant threats to agriculture, and natural communities and ecosystems [8]. The capacity of crop plants to adapt to new drought conditions is essential for providing food security as human population growth continues [9,10] (more text on this topic and literature review is provided in the Supplemental Text S1 given in supplementary materials).

Beyond agricultural plant species, there is great interest in understanding how natural plant species and communities respond to intensifying drought stressors to preserve stable and functioning ecosystems in a changing world reviewed in [11]. For many plant species, the ability to naturally disperse towards areas with optimal niche space is limited by

factors such as short seed-dispersal distances and habitat loss and fragmentation. As a result of these or other limitations, geographic shifts may not keep pace with new and altered habitats produced by climate change [12,13]; reviewed in [11]. This is especially true when other anthropogenic threats (e.g., habitat loss, fragmentation, and reductions in population size) occur jointly with climate change [14,15]. Therefore, identifying the mechanisms allowing plants to persist as climate change, and especially drought, proceeds and intensifies is paramount for evaluating adaptive capacity of these populations and designing and implementing best management practices for a sustainable future [16].

There is an urgent need to identify mechanisms of response to the stresses caused by climate change, particularly drought. Reductions in the cost of high-throughput nucleotide sequencing over the past two decades [17] have made it possible to investigate the genomic underpinnings of these responses in non-model species at a much greater scale and rate. While the genomic basis of drought adaptation has been explored in well-resourced crop plants (e.g., corn, *Zea mays*; rice, *Oryza* spp.; wheat, *Triticum* spp.) [18–20], plant model systems (e.g., *Arabidopsis thaliana*) [21,22], and forestry species reviewed in [23], fewer non-model or natural plant systems have been explored to the same extent. This disparity is reflected in the number of genomic resources that are readily available for plant species. For example, of the 243 unique and chromosome-level plant reference genomes available on NCBI (as of December 2020), only 20% are from species of non-human use, as classified by the Kew Gardens [24]. We anticipate this disparity in genomic resources may be due to several barriers to conducting conservation genomic research (i.e., the ‘conservation genomics gap’) [25,26], including the cost of sequencing and assembling the large and complex genomes that many plants possess [27,28]. Therefore, a cost-effective and targeted approach to finding genes associated with drought response is required to fill this gap for natural populations.

Reviewing the state of research focused on G2P pathways will allow for focused hypothesis-testing research on non-model plant species for which little genomic data is available. Mining genomic and phenomic data across a wide phylogenetic breadth will allow for a better understanding of the diverse and complex plant responses to drought and provide novel opportunities for crop improvement. While mining methods applied to publicly available datasets and publications can advance our understanding of G2P processes, there is a stark lack of infrastructure available to perform such searches and analyses in a thorough fashion. While many genetics and genomics journals now require that sequence data be publicly available in a repository (e.g., PLoS ONE enacted a “Data Availability Statement” requirement in March of 2014 [29]), this has not always been the case and many studies lack connectivity to an available dataset [30]. This creates an urgent need to increase database connectivity to aid in increasing reproducibility and allowing for meta-analyses.

To make efficient and rapid gains in understanding the genomic-basis of plant response towards drought, we developed a literature and database search pipeline and performed a meta-analysis to identify focal taxa of research on genes underpinning drought-stress response across the primary literature. Abstracts from these articles were mined using the R package ‘G2PMineR’ [31] to produce networks connecting genes to phenotypic traits associated with drought-stress responses. Identifying key gene networks and drought-adapted species for which genomic resources may be available for will provide potential routes forward in drought G2P research. Databases containing genetic and taxonomic data related to these publications were queried to identify patterns of genome-to-phenome research and to assess connectivity of these publications and databases. This methodology enables investigations into how genes are associated with physiological processes and drought-stress response strategies across a broad range of taxa. This study gives a broader understanding of drought G2P research in plants. It has generated a list of genes and associated phenotypes to provide a foundation for studying the genomic and phenotypic basis of drought tolerance of less-studied natural plant species worldwide.

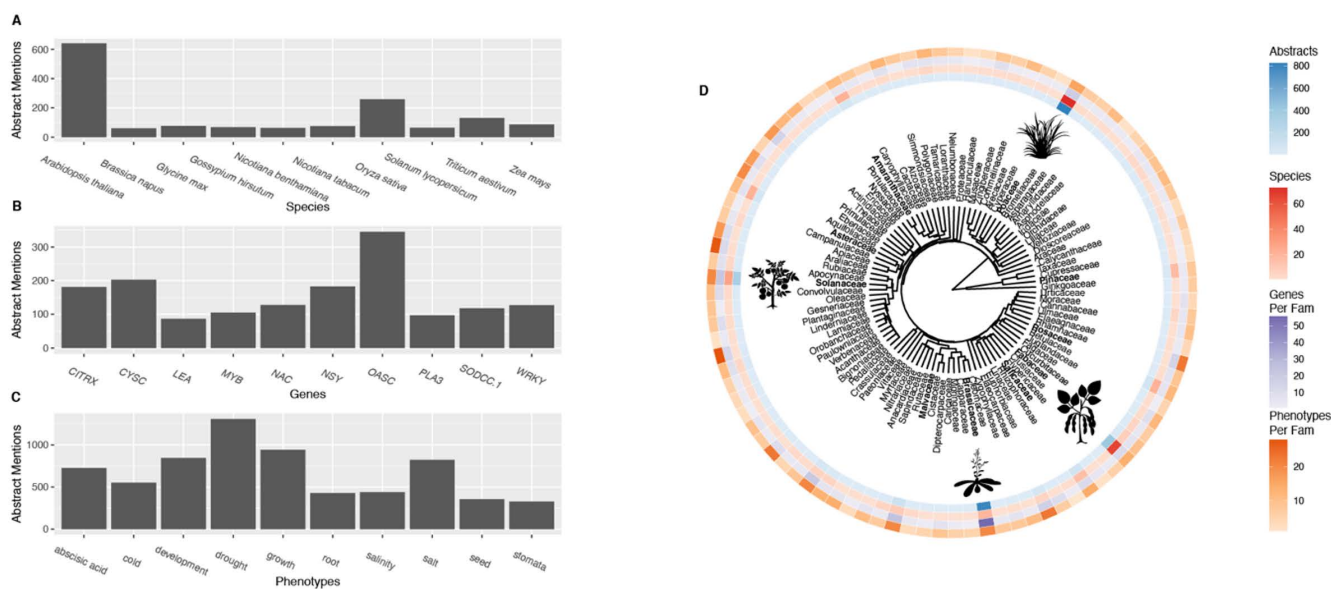
## 2. Results

### 2.1. Literature Mining

The PubMed search query used in this study resulted in the acquisition of 7730 abstracts from papers published from January 1998–December 2020. Abstract mining using ‘G2PMineR’ [31] produced a total of 2642 consensus abstracts having data for taxa, genes, and phenotypes (Table S1).

### 2.2. Taxa

Within the ‘G2PMineR’ consensus abstracts, 497 species or subspecies, representing 296 genera and 98 families (91 angiosperm families, four gymnosperm families, two fern families, and one lycophyte family; Table S2) were identified. The top reported species is *Arabidopsis thaliana* L., with 641 mentions in the consensus abstracts (Figure 1A). Other highly studied species include *Oryza sativa* L. (Poaceae; 259 mentions), *Triticum aestivum* L. (Poaceae; 131 mentions), *Zea mays* L. (Poaceae; 87 mentions), and *Glycine max* (L.) Merr. (Fabaceae; 77 consensus mentions). The most studied genes are reported in a wide breadth of taxa, indicating that they play a role in drought response across diverse plant lineages (Figure 1D). Poaceae had the highest representation, with 73 taxa mentioned in consensus abstracts. Other highly represented families included Fabaceae (66 taxa) and Solanaceae (29 taxa). Poaceae species were the most mentioned in consensus abstracts (828 mentions), followed by Brassicaceae (800 mentions) and Fabaceae (417 mentions). Within the consensus abstracts, 414 (83.30% of species mentioned) are classified for human use (Table S2). Of the plants with human uses according to the Kew Gardens Checklist of Useful Plants [24], the most common uses were medicinal (318 taxa), environmental (251 taxa), food (229 taxa), materials (218 taxa), and gene sources (208 taxa), with these categories not being mutually exclusive. Of the plants identified in consensus abstracts, 103 (20.7% of consensus abstract taxa) were listed as having chromosome level assemblies per NCBI database (Table S2).



**Figure 1.** Most highly referenced (A) species, (B) genes, and (C) phenotype words within the consensus abstracts. (D) Phylogeny of families represented in the consensus abstracts shows that families with model or crop plants (e.g., Brassicaceae, Fabaceae, and Poaceae) were the most highly represented, both in terms of species included in studies and number of abstracts in which they were mentioned.

Locality data were successfully downloaded for 106 species mentioned in the total abstracts lacking human use according to the Kew Gardens Checklist of Useful Plants [24]. Of these, 31% (33 of 106 species) were found to occur in hyper-arid environments based on

aridity index values ( $AI < 0.05$ ), though none of these species were included in the consensus abstracts (Figure S1, Table S3). One occurrence for *Tillandsia ionantha* Planchon was found in northern Africa and was excluded as it is native to Central America and Mexico and the accuracy of this occurrence could not be confirmed [32]. The most mentioned plant species lacking from hyper-arid environments lacking human use per Kew list were: *Cleistogenes songorica* (Roshev.) Packer (Poaceae; eight abstracts), *Caragana korshinskii* Kom. (Fabaceae; seven abstracts), *Malus prunifolia* (Willd.) Borkh. (Rosaceae; seven abstracts), *Aeluropus littoralis* (Gouan) Parl. (Poaceae; six abstracts), *Malus sieversii* (Ledeb.) M.Roem. (Rosaceae; six abstracts), *Boea hygrometrica* (Bunge) R. Brown (Gesneriaceae; five abstracts), *Reaumuria soongorica* (Pall.) Maxim. (Tamaricaceae; five abstracts), *Alhagi sparsifolia* Shap. (Fabaceae; four abstracts), *Rosa hybrida* (Rosaceae; four abstracts), and *Chorispora bungeana* Fisch. & C.A. Mey (Brassicaceae; three abstracts).

### 2.3. Genes

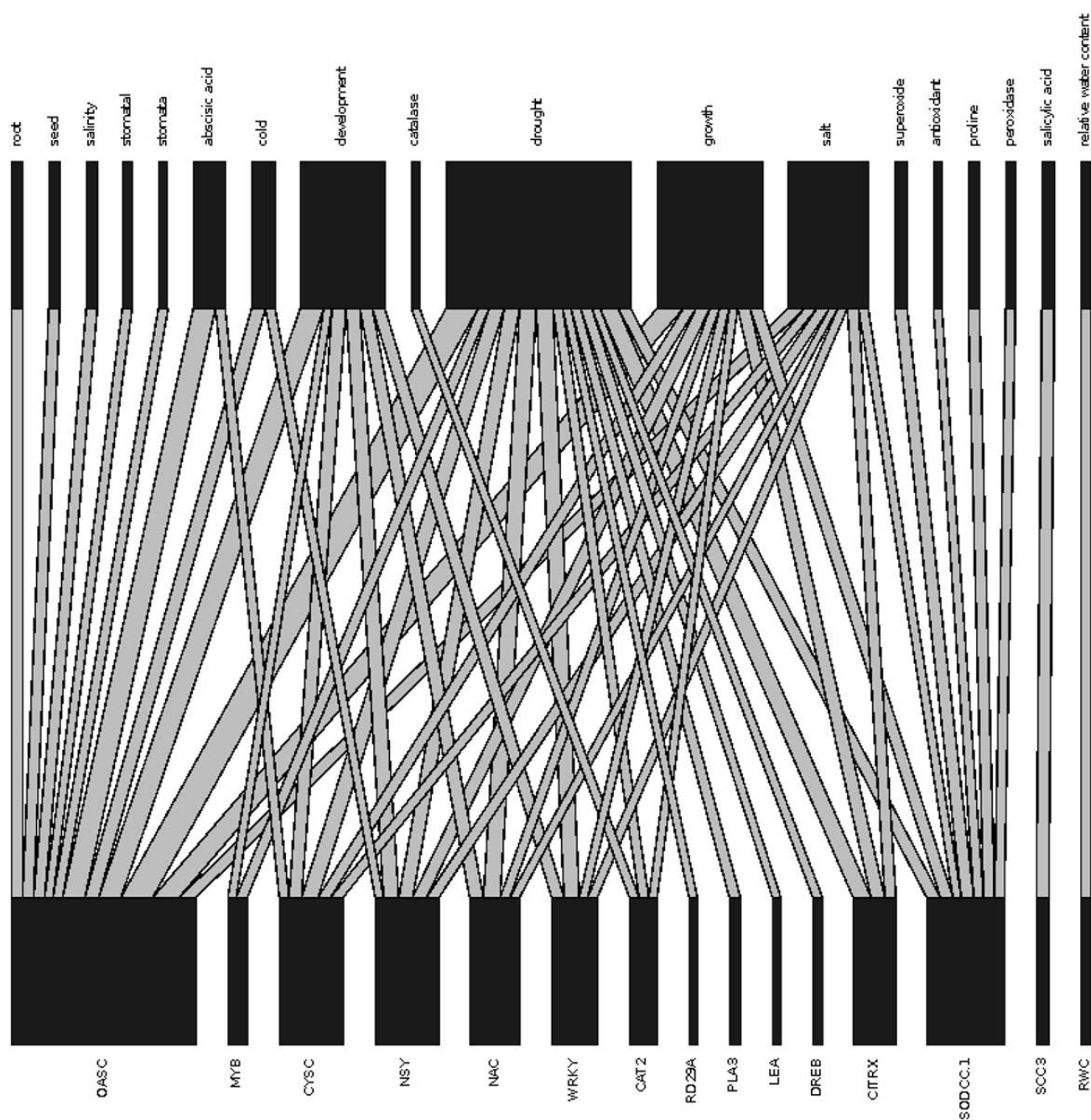
A total of 1634 genes, including single-copy genes and gene groups (e.g., *WRKY*) associated with drought response phenotypes were found in the 'G2PMineR' consensus abstracts (Table S4). The top-10 cited genes included *oasC* (O-acetylserin (thiol) lyase isoform C; 345 mentions), *cysC* (cysteine synthase C1; 203 mentions), *NSY* (chloroplasmic neoxanthin synthase; 183 mentions), *CITRX* (Cf-9-interacting thioredoxin; 181 mentions), *NAC* (NAM, ATAF, and CUC transcription factors; 128 mentions), *WRKY* (*WRKY*-domain containing transcription factors; 127 mentions), *SODCC.1* (superoxide dismutase [Cu-Zn] 1; 118 mentions), *MYB* (MYB DNA-binding domain-containing transcription factors; 105 mentions), *PLA3* (probable glutamate carboxypeptidase; 97 mentions), and *LEA* (late embryogenesis abundant proteins; 87 mentions; Figure 1B).

### 2.4. Phenotypes

A total of 118 phenotypes from our curated drought phenotype lexicon were identified in the consensus abstracts (Table S5). The top 10 phenotype words within the consensus abstracts were "drought" (1307 mentions), "growth" (940 mentions), "development" (845 mentions), "salt" (823), "abscisic acid" (726 mentions), "cold" (551 mentions), "salinity" (439 mentions), "root" (430), "seed" (357 mentions), and "stomata" (329 mentions) (Figure 1C). The top drought strategy for which phenotypes were mentioned was the "avoidance" strategy, with 6383 mentions. For other drought strategies, the total mentions for phenotypes of a given group was: 1751 for "detection", 2945 for "recovery", 3570 for "general stress", 3703 for "escape", and 5765 for "tolerance".

### 2.5. Gene to Phenotype Interactions

Of the top gene and phenotypes, *oasC* and general phenotype "drought" had the greatest connectivity (Figure 2). *oasC* also had high connectivity with "abscisic acid", "development", and "growth". *CITRX*, *WRKY*, *NAC*, *NSY*, and *cysC* genes were also highly connected to the phenotype "drought". Most genes were associated with multiple drought-response strategies: a total of 759 genes (~48% of genes identified in consensus abstracts) were associated with phenotypes of all major drought-response strategies, 366 (21%) were associated with "avoidance", "escape", "recovery", and "tolerance", and 120 (8%) of consensus genes were associated with "avoidance", "detection", and "tolerance" (Figure 3). These genes were associated with a variety of gene ontology (GO) categories, though the most common GO categories were abscisic acid response (GO:0009737, 73 genes), water deprivation response (GO:0009414, 66 genes), and abscisic acid activated signalling pathways (GO:0009738, 62 genes).



**Figure 2.** Bipartite plot showing the relationship between most referenced phenotypes (top row) and the most referenced genes with which they are associated (bottom row). The width of the block indicates the strength of the association for a given phenotype or gene and wider connections indicate more abstracts demonstrating that connection. The most highly mentioned phenotype word in the consensus abstracts was “drought”, which was highly associated with all top ten most highly mentioned genes, as well as *CAT2*, *RD29A*, and *DREB* genes.

## 2.6. Assessing Connectivity of Sequence Data and Taxonomy Databases

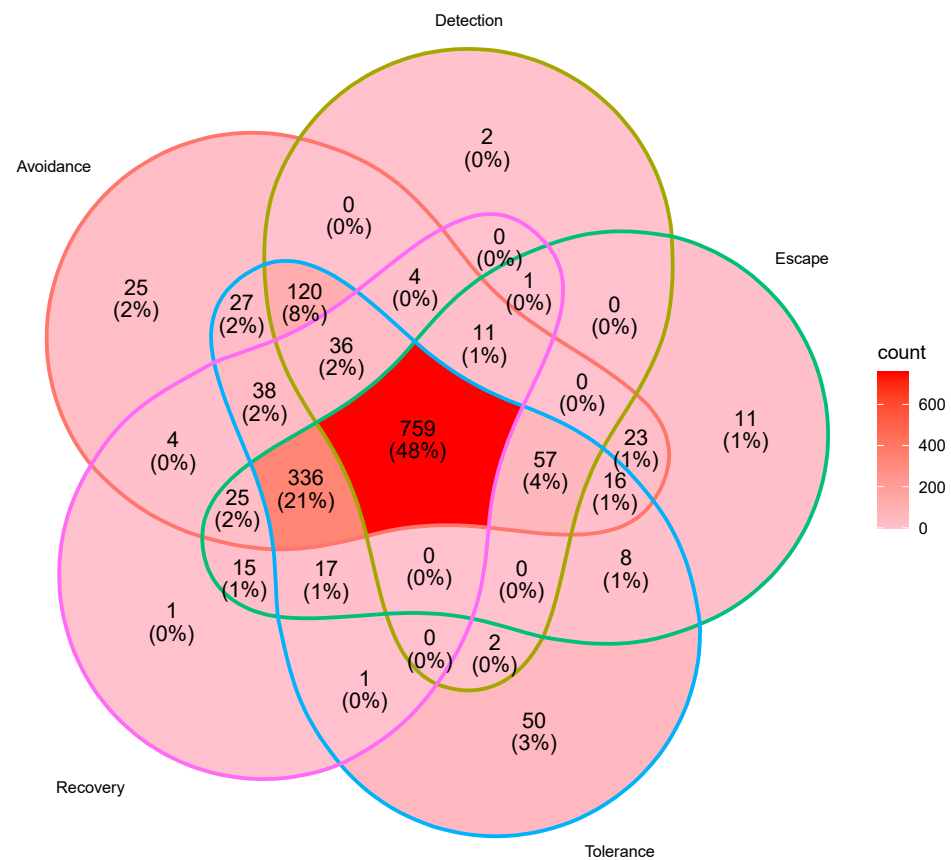
In total, 1712 (22.15%) of the 7730 abstracts were associated with sequence data and 1865 (24.13%) were associated with taxonomic data on NCBI. Only 1624 (21.01%) abstracts had connectivity to GenBank databases, for both taxonomy and gene data, for each publication, per assessment using ‘rentrez’ [33].

A total of 583 taxa were found to be associated with the PubMed IDs using ‘rentrez’. Of these, five taxa were at the familial level, 25 taxa were at the generic level, and 15 taxa

were identified as non-plants. The publications included one to 152 species/taxa, with all but six focusing on less than 10 taxa. A total of 192 taxa were shared with the taxa identified in the abstract mining search.

A total of 423 genes with the GenBank database were associated with the PubMed IDs. Few studies were associated with many nucleotide sequences (e.g., 37 studies were associated with greater than 1000 sequence accessions). The most common genes from the studies with sequence accessions in GenBank included *oasC* (274 accessions), *NAC* (37 accessions), and *WRKY* (35 accessions). A total of 234 genes identified from sequences were shared with those identified in the abstract mining search.

Due to a lack of phenotypic data in NCBI accessions, phenotype data had to be mined from abstracts. This produced a total of 114 phenotypes identified from abstracts that had associated sequence and taxonomy data, all of which were included in the abstract mining search.



**Figure 3.** Venn diagram of the five major drought response strategies (avoidance, detection, escape, tolerance, and recovery) showing the overlap of genes associated with each category.

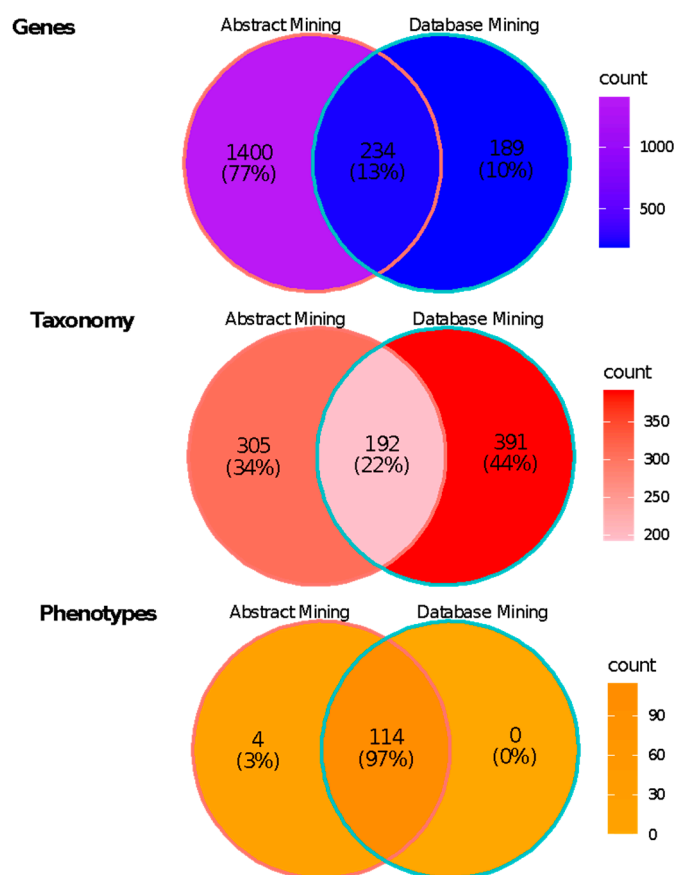
### 3. Discussion

#### 3.1. Trends and Gaps in Plant Drought Research

Only 21% of studies (1624 of 7730 abstracts) had connectivity to NCBI databases for gene sequences or taxa. Of these, 1533 studies focused on one taxon, with the greatest number of taxa studied in any one study being 152. Most studies also focused on only one sequence, while some were associated with over 100,000 sequences. Presumably, the studies with greater numbers of sequences but only one taxon associated with them in databases represent expression characterization studies with many replicates of the taxon (e.g., [34,35]). The most studied species from papers with both sequence and taxon database connectivity was *Arabidopsis thaliana*, with 730 studies on this species being linked to both sequence and taxon data. While these sequence data were associated with plant-oriented studies, several taxa associated with these data were not plants. For example, a study on calcineurin B-like



proteins in *Arabidopsis* is also associated with sequence data from *Rattus norvegicus* (brown rat), which was used to test for interactions between calcineurin B-like proteins [36]. Some bacteria species were identified, which were mostly used in gene transformation or cloning experiments (e.g., [37]). While the 'G2PMineR' workflow is naïve to whether a given gene mentioned in abstracts is characterized from an expression study or from genomic sequence, this does not affect the interpretation of the workflow results. The 'G2PMineR' abstract mining workflow was more successful at making G2P connections, with 34% of abstracts containing gene, taxa, and phenotype data. Consensus abstracts contained all requisite information for making associations between taxa, genes, and phenotypes, which is apparently not possible to any meaningful level using database associations. Characterization of genes, species, and phenotypes involved in drought tolerance was more successful using the G2PMineR pipeline, compared to simply using manual searches. For example, 1634 genes were identified using 'G2PMineR' while only 423 genes associated with the PubMed abstract set were found within GenBank. Interestingly, only 234 genes were shared within the two datasets (Figure 4).



**Figure 4.** Overlap of genes, taxa, and phenotypes found using abstract mining with 'G2PMineR' and database mining with 'rentrez'. Little overlap was found for gene and taxonomy data mined abstracts and databases (13% and 22%, respectively). Almost all phenotypes (97%) identified in the consensus abstracts were able to be identified in abstracts that were associated with gene and taxonomy data, as well.

The 'G2PMineR' workflow illustrates the current state of drought response research among plant species regarding different plant species under investigation. Approximately 83% of plant species identified in plant drought abstracts represent either model species, crop species, or species of human use. The top species identified was the model plant species *Arabidopsis thaliana* L., which is not surprising because *Arabidopsis* is the primary model plant for genetic research and has a well-assembled and annotated genome with

vast genetic resources available for G2P research [38]. In addition to *Arabidopsis thaliana*, crop species such as *Oryza sativa* L. (rice) or *Zea mays* L. (corn) have been the primary focus of drought research, which is expected given that climate change, especially droughts, poses a threat to agriculture [39]. A large phylogenetic breadth of species was found in our search across the plant kingdom. The taxa identified in the consensus abstracts represented 98 green plant families, with 95 of the represented plant families being seed-plant families. While species of many families were included in the consensus abstracts, these species were generally biased towards families with model plant species and economically important species, such as Poaceae, Fabaceae, and Solanaceae (Figure 3). Outside of seed plants, only two fern (Hymenophyllaceae and Pteridaceae) and one lycophyte (Selaginellaceae) families were represented. Because of this bias, researchers may have only a narrow understanding of how plants adapt and respond to drought stress. The use of genetic model plants and crop plants in any genomic or adaptive capacity research is of great importance and may be broadly applicable across a wide taxonomic breadth (e.g., [40]). However, non-model plants, plants of arid ecosystems, and plants native to threatened ecosystems need broader inclusion in G2P drought research. Studying the G2P drought response pathways of plants that are well-adapted to dry environments, such as the taxa highlighted in Figure S1, will contribute to our understanding of the various metabolic and physiological pathways that increase drought tolerance, and could potentially be used for crop improvement. While a similar number of taxa were identified using 'rentrez' and 'G2PMineR' (391 versus 305, respectively), there was little overlap, with only 192 (~28% of all taxa identified from both abstract and database mining) being found in both abstract sets.

Relatively few plant species found in consensus abstracts (103 of 497; 20.7% of consensus abstract taxa) had high quality genome assemblies (i.e., assembled to chromosome level) available, with a bias towards model organisms. With decreasing costs, increasing quality, and the ease of generating whole genome assemblies for non-model organisms [17], there are more opportunities to include non-model and natural plant species in G2P research. We anticipate that the efforts of individual research labs and genomic consortiums (e.g., 10,000 Plant Genomes Project [41]) will help bridge this gap in the coming years.

Few of the species targeted in G2P research thus far occur in arid or drought-prone areas. Only 16.70% of species and subspecies identified in the consensus abstracts lacked human use per the Kew Gardens checklist. Of the non-model or non-frequently used plant species identified across all abstracts, roughly 33% occur in arid regions (Figure S1). While the study of model and crop plant species is important, as these species have the financial resources and infrastructure available for in-depth genomic research, more effort needs to be directed to study species in ecosystems in which drought-adapted species have evolved. Among these, the species identified from all abstracts native to hyper-arid habitats and well-adapted to severe droughts belong primarily to families for which genomic resources are readily available (e.g., Brassicaceae, Fabaceae, Poaceae, and Rosaceae). Consequently, these species are good candidates for G2P research, potentially revealing new mechanisms to improve drought tolerance in con-familial crop species. Given that no species in the consensus abstracts occur in areas where plants must be highly adapted to drought-stress, this is an area that needs more attention by researchers.

### 3.2. Drought Response Strategies

In the consensus abstracts, the most-studied responses fell under the drought avoidance strategy, which includes the ability of plants to reduce drought stress by minimizing water loss and increasing water uptake [42]. This observation is consistent with the high occurrence of keywords such as abscisic acid (ABA) and osmotic adjustment. ABA is well known to induce stomatal closure and increase the root to shoot ratio, while osmotic adjustment allows plants to extract more water from drying soil profiles [43–45]. Many of the top genes mentioned in consensus abstracts comprise transcription factors (e.g., *NAC*, *MYB*, *WRKY*) that affect stomatal activity [46,47], which agrees with the high occurrence of avoidance strategy words in consensus abstracts. In addition to avoidance, strategies

that reduce drought-induced cellular damage due to accumulation of toxic molecules, particularly “synthesis of osmoprotectants” and “antioxidant production”, were also often mentioned in consensus abstracts. Relatively fewer studies were associated with mechanisms of drought detection or escape and recovery from drought, suggesting a need for more research in these areas. While drought was the main focus of this G2P investigation, other stressors like cold and salt were commonly mentioned. This result reflects the fact that plants exhibit specific, as well as shared, responses to different abiotic stresses [44]. Some of the damage caused by various abiotic stresses (e.g., increased ROS production and dehydration) can be similar [48,49]. With shared responses, plants can use many of the same genes to minimize damage and cope with multiple stresses. We anticipate that as plant G2P research continues, findings across stress types (e.g., temperature, light, salt, heavy metals) will be applicable to one another and overall enhance our understanding of stress response.

### 3.3. Candidate Genes for Future G2P Research

Highly reported genes signal broad function in stress response pathways, particularly those in osmoprotectant and antioxidant pathways, and were associated with taxa of many of the plant families represented in this study. Of the top ten most referenced genes, *oasC*, *cysC*, *LEA*, *NAC*, and *SODCC.1* have functions associated with, but not limited to, osmoprotectant and oxidative pathways [50–52]. Other commonly referenced genes function in carotenoid metabolic pathways involved in synthesizing ABA precursors or antioxidants such as  $\beta$ -Carotene [53,54]. Many of the genes in the consensus abstracts include transcription factors, such as *WRKY*, *NAC*, and *MYB*, which have widespread functions in stress response pathways, including stomatal closures [46,47]. These transcription factors act like ‘switches’ to regulate expression of many genes and are predicted to be one of the most important mechanisms for plant drought response and increase our capacity to engineer more resilient plant species [55,56].

Most of the top occurring genes were found to be highly connected to the phenotype “drought” in the consensus abstract, which is to be expected given our initial search query. However, our top reporting genes (e.g. *SODCC.1*, *oasC*) showed high connectivity with many phenotypes of different stresses (e.g., “salt”, “cold”), response strategies (e.g., “growth”, “development”), and phytochemical pathways (e.g., “abscisic acid”, “proline”, “salicylic acid”). This indicates how genes with more targeted functions (e.g., osmoprotectants and antioxidants) can play general roles in stress response. Indeed, DroughtDB by Alter et al. (2015) [57] shows how some genes can play multiple roles across drought response. As G2P plant drought research continues to expand, we expect the functions of genes will become better characterized. In the meantime, this meta-analysis provides a list of candidate genes identified across taxa that can be broadly explored in species of interest using various methods, such as the mining of draft genomes (e.g., [58]).

While there is a large focus on phenotypes associated with the drought-avoidance and -tolerance strategies in the consensus abstracts, genes contributing to phenotypes associated with all of the drought response strategies were highly referenced in the consensus abstracts (Figure 3). These include all the most studied genes identified in this meta-analysis. This indicates that while there are some gaps in focal phenotypes and drought response strategies, many genes underpinning these phenotypes and response strategies have likely been well characterized and could readily be transferred to and used in studies of many other drought response strategies if they are not already in use.

### 3.4. Limitations of Literature and Database Mining

While we have been able to identify numerous genes and species used in the study of drought stress response, there are limitations to the text mining methods employed here. One primary limitation is that the text mining methods outlined here mine only published abstracts. Given publisher constraints, such as word count, abstracts often give broad summaries and may not list specific genes or species which were studied. For

example, studies on expression patterns and functional pathways for gene families are quite common and these publications may only provide the gene family or group being studied (e.g., *WRKY*) and not specific members of these gene families. While measures were taken to account for such studies (e.g., including the gene family/group name in the searches, as well as limiting the final consensus abstract genes to and grouping by SwissProt families), conclusions cannot be drawn on which specific genes within the gene families confer improved drought-stress responses without a more in-depth reading of Results and Discussions sections of published literature.

Furthermore, there is a disconnect between publications and the databases in which sequence data are stored: only 21% of abstracts were associated with sequence and taxonomy data in NCBI databases. Some of this may be due to the timing of publication (e.g., older publications are less likely to have sequence data deposited into GenBank (Figure S2)), but as noted by searching GenBank, even recent publications are not always linked to the genes, genomes, or proteins within databases. G2P research and meta-analyses/reviews such as this one will benefit from greater connectivity between sequence databases and publications (e.g., linking a sequence in GenBank to a publication by including an identification number such as a PubMed ID).

#### 4. Conclusions

Gaps have been identified in the plant taxa used for G2P drought research. Most effort in this area has focused on model plants and not on plants naturally occurring in drought-prone ecosystems. Several species of hyper-arid ecosystems, with close relatives whose genomes have been sequenced, were identified providing species for future research on drought response G2P pathways. Many commonly studied genes contribute to multiple drought response strategies, even though the research focused on a subset of these strategies. These genes provide excellent starting points for G2P research on non-model plants with few genomic resources. Given the gaps in focal taxa and the availability of genomic and genetic resources, especially for commonly studied genes and gene families, this presents a clear path for rapidly advancing drought response research in the future.

#### 5. Materials and Methods

##### 5.1. PubMed Search

The PubMed (<https://pubmed.ncbi.nlm.nih.gov/> (accessed on 3 January 2021)) database was chosen for our literature search, as it includes all genetic and genomic studies hosted through NCBI's GenBank and is open source. We developed search query strings in PubMed to find all publications relevant to literature pertaining to genes underpinning plant drought. We manually evaluated results of several query strings for content and verified targeted searches using the *AbstractClusterMakeR* and *MembershipInvestigateR* functions in the R package 'G2PMineR' ([https://buerkilabteam.github.io/G2PMineR\\_Web/](https://buerkilabteam.github.io/G2PMineR_Web/) [31]). The results of the final query, completed in December 2020, included publications between the years 1998 and 2020 with the terms "plant AND drought AND (resistance OR tolerance OR response OR avoidance OR resilience OR susceptibility OR escape) AND gene". Figure 5 shows the workflow for analyses performed on the downloaded abstract data.

##### 5.2. Literature Mining and Analyses

All abstracts were mined for species studied, genes referenced, and phenotypes expressed using the R package 'G2PMineR' [31]. We used the function *SpeciesLookeR* to mine abstracts for all species characterized as 'green plants' (Class Viridiplantae) within the Global Biodiversity Information Facility (GBIF; GBIF.org) database. To ensure that our assessment included up-to-date taxonomies, we used the 'taxize' R package [59] to identify synonyms and the primary accepted scientific name for each taxon used in the abstracts. We further determined how many plants are assigned as being useful to humans (i.e., plants for foods, poisons, medicines, environmental uses, social uses, fuels, or materials) per a list generated by the Kew Gardens [24], to assess bias towards non-model organisms. NCBI

(<https://www.ncbi.nlm.nih.gov/genome/> (accessed on 5 January 2021)) was queried for species with chromosome level genome assemblies to determine how many of the species mentioned in abstracts included here have high-level genome assemblies. To determine how species of interest identified from these abstracts are geographically distributed, GBIF (see [60] for the GBIF dataset) and iDigBio (<https://www.idigbio.org/> (accessed on 7 April 2021)) were queried for all species lacking human use per the Kew Gardens Checklist of Useful Plants [24]. This search was restricted to these non-model and/or crop plants, as species of agriculture and horticulture are often logged in public databases through citizen scientists, due to biases in observations being made in or near areas of human activity, and many of these occurrences do not represent natural occurrence [61]. Occurrence data were cleaned to remove data lacking complete coordinates, having unlikely coordinates (e.g., 0,0), and lacking environmental data using the ‘scrubr’ R package [62] using aridity index (AI) as a reference (<https://cgiaresi.community/data/global-aridity-and-pet-database/>; (accessed on 14 February 2021) [63]). Occurrence data were used to extract AI values, which were then used to identify species that are adapted to low-water stress, particularly those found in hyper-arid (AI < 0.05) environments (all occurrence data used to identify species of hyper-arid environments are available at DOI:10.5281/zenodo.7153278).

To understand which genes are being described in each study, the ‘G2PMineR’ function *GenesLookeR* was used to search abstracts for 40,435 genes, their putative gene families, and ontologies as listed in SwissProt for Viridiplantae, as of July 2020. To capture genes and broad family groups (e.g., *WRKY* as opposed to *WRKY1*) that were cited in abstracts, but not always captured in SwissProt, we generated an additional lexicon of genes and gene family names from the DroughtDB [57], the Plant Transcription Factor Database (<http://plantfdb.gao-lab.org/> (accessed on 31 August 2020)), and those reported in previously published plant drought gene reviews (e.g., [9,64,65]). The ‘G2PMineR’ function *SynonymReplaceR* is used to replace gene synonyms with accepted gene names from SwissProt. See Table S4 for a complete list of gene and gene family names identified in extracted abstracts.

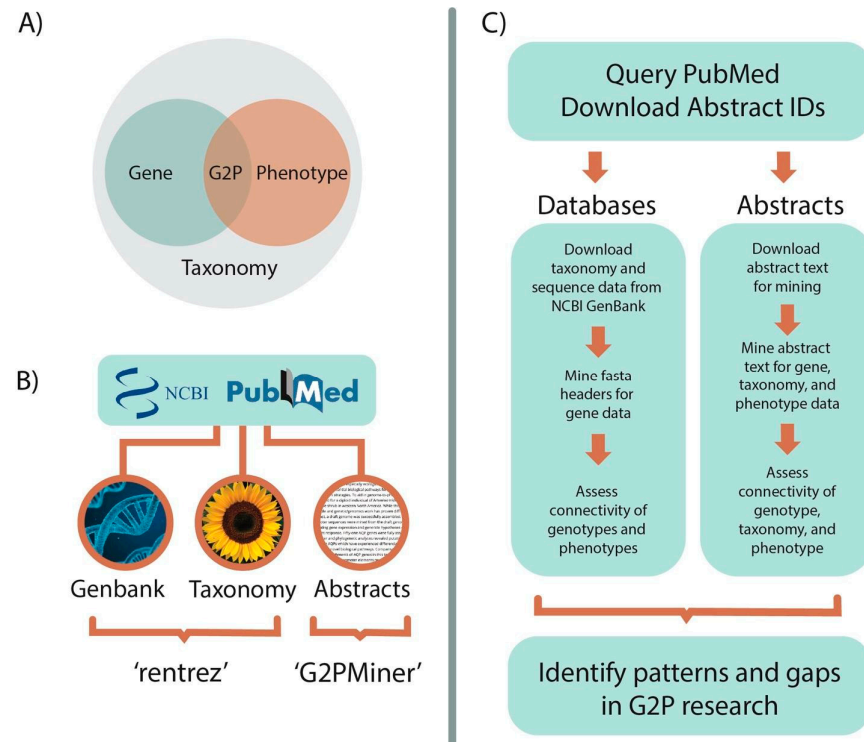
Finally, we used the ‘G2PMineR’ function *PhenotypeLookeR* to search abstracts for phenotypic words of interest, as defined by a curated glossary from the Missouri Botanical Garden [66]. To make a lexicon that is specific to drought, we curated a drought phenotypic lexicon using terminology from Farooq et al. (2012) [67], De Micco & Aronne (2012) [68], and Vilagrosa et al. (2012) [69]. To test whether genes were associated with phenotypes that represented different drought response categories (Figure S3), all phenotypes were manually categorized as being relevant to drought detection, resistance (i.e., escape, avoidance, and tolerance), and/or recovery. See Table S6 for a full list of phenotypes and categories used.

All downstream analyses for species, genes, and phenotypes were conducted for abstracts that had all three search items represented using the *ConsensusInferreR* function within ‘G2PmineR’. In order to elucidate the link between genotypes and phenotypes, networks were developed using the ‘G2PMineR’ function *PairwiseDistanceInferreR*, which generates a pairwise distance matrix between each gene and phenotype using the number of shared abstract matches. To place taxa represented in the consensus abstracts into a phylogenetic context, families of all represented taxa were placed into a family level flowering plant phylogeny [70]. Family placement for each taxon was acquired from NCBI using the ‘taxize’ R package [59]. Consensus abstract mentions, taxa, genes, and phenotypes represented were then tabulated for each family.

### 5.3. Assessing Connectivity to Databases

The R package ‘rentrez’ [33] was used to mine gene and taxonomy information for publications that have both GenBank sequence and taxonomy data linked to the respective PubMed ID. To ensure that taxonomy was standardized across comparisons, the R package ‘taxize’ [59] was used to evaluate taxa and replace any synonyms with the accepted taxon name. Due to a lack of standardized formatting of data within fasta headers, ‘G2PMineR’ [31] was used to mine gene names from headers using ‘G2PMineR’ databases.

Phenotype data for abstracts with both gene and taxonomy data database associations were then subset from the consensus abstract phenotype data for comparisons to abstract mining results.



**Figure 5.** Workflow of abstract and database mining analyses. (A) The goal of these pipelines was to assess the connectivity of gene data to phenotype data all within a taxonomic framework, focusing on Viridiplantae. (B) NCBI data bases were chosen as the hub of this research as all data within these databases can be associated with an NCBI PubMed abstract ID. (C) Once abstracts of interest are identified, the PubMed IDs and abstract text can be downloaded for abstract and database mining analyses.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232012297/s1>, Supplemental Literature Cited [71–100].

**Author Contributions:** A.E.M., S.J.G. and S.B. designed research; A.E.M., S.B., S.J.G., C.D.C.D. and J.M.A.W. performed research. A.E.M. and S.J.G. wrote the paper, with edits and contributions by A.E.M., S.J.G., C.D.C.D., J.M.A.W., S.N., M.S., P.M. and S.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** We gratefully acknowledge the National Science Foundation EPSCoR Track I grant GEM3 (Genes by Environment: Modeling, Mechanisms, Mechanisms) under award number OIA-1757324 for funding this work. We thank additional support from the NSF Track 2 EPSCoR Program GUTT (Genomes Underlying Toxin Tolerance) under award number OIA-1826801.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The ‘G2PMineR’ R package is available at <https://github.com/BuerkiLabTeam/G2PMineR>. Supplemental R scripts for post-G2PMineR analyses are available via GitHub [https://github.com/aemelon/G2PMineR\\_Supplemental\\_Scripts](https://github.com/aemelon/G2PMineR_Supplemental_Scripts). Raw and quality-controlled occurrence data are available at DOI:10.5281/zenodo.7153278.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mpelasoka, F.; Hennessy, K.; Jones, R.; Bates, B. Comparison of suitable drought indices for climate change impacts assessment over Australia towards resource management. *Int. J. Climatol.* **2008**, *28*, 1283–1292. [[CrossRef](#)]
2. Zhang, L.; Zhou, T. Drought over East Asia: A review. *J. Clim.* **2015**, *28*, 3375–3399. [[CrossRef](#)]
3. Callaghan, M.; Schleussner, C.-F.; Nath, S.; Lejeune, Q.; Knutson, T.R.; Reichstein, M.; Hansen, G.; Theokritoff, E.; Andrijevic, M.; Brecha, R.J.; et al. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. Clim. Chang.* **2021**, *11*, 966–972. [[CrossRef](#)]
4. Trenberth, K.E.; Dai, A.; Van Der Schrier, G.; Jones, P.D.; Barichivich, J.; Briffa, K.R.; Sheffield, J. Global warming and changes in drought. *Nat. Clim. Chang.* **2014**, *4*, 17–22. [[CrossRef](#)]
5. IPCC. Climate Change 2021: The Physical Science Basis. In *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V.P., Zhai, A., Pirani, S.L., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M.L., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2021. [[CrossRef](#)]
6. Williams, A.P.; Cook, E.R.; Smerdon, J.E.; Cook, B.I.; Abatzoglou, J.T.; Bolles, K.; Baek, S.H.; Badger, A.M.; Livneh, B. Large contribution from anthropogenic warming to an emerging North American megadrought. *Science* **2020**, *370*, 314–318. [[CrossRef](#)] [[PubMed](#)]
7. Stahle, D.W. Anthropogenic megadrought. *Science* **2020**, *368*, 238–239. [[CrossRef](#)]
8. Ault, T.R. On the essentials of drought in a changing climate. *Science* **2020**, *368*, 256–260. [[CrossRef](#)] [[PubMed](#)]
9. Gupta, A.; Rico-Medina, A.; Caño-Delgado, A.I. The physiology of plant responses to drought. *Science* **2020**, *368*, 266–269. [[CrossRef](#)]
10. Voosen, P. The hunger forecast. *Science* **2020**, *368*, 226–229. [[CrossRef](#)] [[PubMed](#)]
11. Brodribb, T.J.; Powers, J.; Cochard, H.; Choat, B. Hanging by a thread? Forests and drought. *Science* **2020**, *368*, 261–266. [[CrossRef](#)] [[PubMed](#)]
12. Gray, L.K.; Hamann, A. Tracking suitable habitat for tree populations under climate change in western North America. *Clim. Chang.* **2013**, *117*, 289–303. [[CrossRef](#)]
13. Urban, M.C.; Zarnetske, P.; Skelly, D.K. Moving forward: Dispersal and species interactions determine biotic responses to climate change. *Ann. N. Y. Acad. Sci.* **2013**, *1297*, 44–60. [[CrossRef](#)] [[PubMed](#)]
14. Opdam, P.; Wascher, D. Climate change meets habitat fragmentation: Linking landscape and biogeographical scale levels in research and conservation. *Biol. Conserv.* **2004**, *117*, 285–297. [[CrossRef](#)]
15. Pearson, R.G.; Dawson, T.P. Long-distance plant dispersal and habitat fragmentation: Identifying conservation targets for spatial landscape planning under climate change. *Biol. Conserv.* **2005**, *123*, 389–401. [[CrossRef](#)]
16. Nicotra, A.B.; Beever, E.A.; Robertson, A.L.; Hofmann, G.E.; O’Leary, J. Assessing the components of adaptive capacity to improve conservation and management efforts under global change. *Conserv. Biol.* **2015**, *29*, 1268–1278. [[CrossRef](#)] [[PubMed](#)]
17. Hayden, E. Technology: The \$1,000 genome. *Nature* **2014**, *507*, 294–295. [[CrossRef](#)] [[PubMed](#)]
18. Liu, S.; Wang, X.; Wang, H.; Xin, H.; Yang, X.; Yan, J.; Li, J.; Tran, L.-S.P.; Shinozaki, K.; Yamaguchi-Shinozaki, K.; et al. Genome-Wide Analysis of ZmDREB Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of *Zea mays* L. *PLoS Genet.* **2013**, *9*, e1003790. [[CrossRef](#)] [[PubMed](#)]
19. Hong, Y.; Zhang, H.; Huang, L.; Li, D.; Song, F. Overexpression of a stress-responsive NAC transcription factor gene ONAC022 improves drought and salt tolerance in rice. *Front. Plant Sci.* **2016**, *7*, 4. [[CrossRef](#)] [[PubMed](#)]
20. Condorelli, G.E.; Maccaferri, M.; Newcomb, M.; Andrade-Sanchez, P.; White, J.W.; French, A.N.; Sciara, G.; Ward, R.; Tuberosa, R. Comparative aerial and ground based high throughput phenotyping for the genetic dissection of NDVI as a proxy for drought adaptive traits in durum wheat. *Front. Plant Sci.* **2018**, *9*, 893. [[CrossRef](#)]
21. Abe, H.; Yamaguchi-Shinozaki, K.; Urao, T.; Iwasaki, T.; Hosokawa, D.; Shinozaki, K. Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* **1997**, *9*, 1859–1868.
22. He, G.-H.; Xu, J.-Y.; Wang, Y.-X.; Liu, J.-M.; Li, P.-S.; Chen, M.; Ma, Y.-Z.; Xu, Z.-S. Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in Arabidopsis. *BMC Plant Biol.* **2016**, *16*, 116. [[CrossRef](#)] [[PubMed](#)]
23. Hamanishi, E.T.; Campbell, M.M. Genome-wide responses to drought in forest trees. *Forestry* **2011**, *84*, 273–283. [[CrossRef](#)]
24. Diazgranados, M.; Allkin, B.; Black, N.; Cámara-Leret, R.; Canteiro, C.; Carretero, J.; Ulian, T. World Checklist of Useful Plant Species. In *Knowledge Network Biocomplexity*; KNB Data Repository: Richmond, UK, 2020.
25. Shafer, A.B.; Wolf, J.B.; Alves, P.C.; Bergström, L.; Bruford, M.W.; Brännström, I.; Colling, G.; Dalén, L.; De Meester, L.; Ekblom, R.; et al. Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **2015**, *30*, 78–87. [[CrossRef](#)] [[PubMed](#)]
26. Galla, S.J.; Buckley, T.R.; Elshire, R.; Hale, M.L.; Knapp, M.; McCallum, J.; Moraga, R.; Anna, W.S.; Phillip, W.; Steeves, T.E. Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. *Mol. Ecol.* **2016**, *25*, 5267–5281. [[CrossRef](#)]
27. Li, F.-W.; Harkess, A. A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **2018**, *6*, e1030. [[CrossRef](#)]
28. Michael, T.P.; VanBuren, R. Building near-complete plant genomes. *Curr. Opin. Plant Biol.* **2020**, *54*, 26–33. [[CrossRef](#)]
29. Federer, L.M.; Belter, C.; Joubert, D.J.; Livinski, A.; Lu, Y.-L.; Snyders, L.N.; Thompson, H. Data sharing in *PLoS ONE*: An analysis of Data Availability Statements. *PLoS ONE* **2018**, *13*, e0194768. [[CrossRef](#)]

30. Tedersoo, L.; Küngas, R.; Oras, E.; Köster, K.; Eenmaa, H.; Leijen, Ä.; Pedaste, M.; Raju, M.; Astapova, A.; Lukner, H.; et al. Data sharing practices and data availability upon request differ across scientific disciplines. *Sci. Data* **2021**, *8*, 192. [[CrossRef](#)]
31. Wojahn, J.; Galla, S.; Melton, A.; Buerki, S. G2PMineR: A Genome to Phenome Literature Review Approach. *Genes* **2021**, *12*, 293. [[CrossRef](#)]
32. Ancona, J.J.; Pinzón, J.P.; Díaz, J.J.O.; Morillo, I.R.; Tun-Garrido, J.; Palma-Silva, C.; Till, W. Botanical history and typification in the *Tillandsia ionantha* complex. *Taxon* **2021**, *70*, 1317–1326. [[CrossRef](#)]
33. Winter, D.J. rentrez: An R package for the NCBI eUtils API. *R J.* **2017**, *9*, 520–526. [[CrossRef](#)]
34. Mochida, K.; Kawaura, K.; Shimosaka, E.; Kawakami, N.; Shin-I, T.; Kohara, Y.; Yamazaki, Y.; Ogihara, Y. Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. *Mol. Genet. Genom.* **2006**, *276*, 304–312. [[CrossRef](#)] [[PubMed](#)]
35. Xiao, L.; Yang, G.; Zhang, L.; Yang, X.; Zhao, S.; Ji, Z.; Zhou, Q.; Hu, M.; Wang, Y.; Chen, M.; et al. The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5833–5837. [[CrossRef](#)] [[PubMed](#)]
36. Kudla, J.; Xu, Q.; Harter, K.; Gruissem, W.; Luan, S. Genes for calcineurin B-like proteins in Arabidopsis are differentially regulated by stress signals. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4718–4723. [[CrossRef](#)] [[PubMed](#)]
37. Yu, T.-F.; Xu, Z.-S.; Guo, J.-K.; Wang, Y.-X.; Abernathy, B.; Fu, J.-D.; Chen, X.; Zhou, Y.-B.; Chen, M.; Ye, X.-G.; et al. Improved drought tolerance in wheat plants overexpressing a synthetic bacterial cold shock protein gene *SeCspA*. *Sci. Rep.* **2017**, *7*, 44050. [[CrossRef](#)] [[PubMed](#)]
38. Bevan, M.; Walsh, S. The Arabidopsis genome: A foundation for plant research. *Genome Res.* **2005**, *15*, 1632–1642. [[CrossRef](#)]
39. Hammer, G.L.; McLean, G.; van Oosterom, E.; Chapman, S.; Zheng, B.; Wu, A.; Doherty, A.; Jordan, D. Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates. *Crop Sci.* **2020**, *60*, 605–621. [[CrossRef](#)]
40. Blümel, M.; Dally, N.; Jung, C. Flowering time regulation in crops—what did we learn from Arabidopsis? *Curr. Opin. Biotechnol.* **2015**, *32*, 121–129. [[CrossRef](#)]
41. Twyford, A.D. The road to 10,000 plant genomes. *Nat. Plants* **2018**, *4*, 312–313. [[CrossRef](#)]
42. Nicotra, A.; Atkin, O.; Bonser, S.; Davidson, A.; Finnegan, E.; Mathesius, U.; Poot, P.; Purugganan, M.; Richards, C.; Valladares, F.; et al. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* **2010**, *15*, 684–692. [[CrossRef](#)]
43. Jones, R.J.; Mansfield, T.A. Suppression of Stomatal Opening in Leaves Treated with Abscisic Acid. *J. Exp. Bot.* **1970**, *21*, 714–719. [[CrossRef](#)]
44. Xiong, L.; Wang, R.-G.; Mao, G.; Koczan, J.M. Identification of drought tolerance determinants by genetic analysis of root response to drought stress and abscisic acid. *Plant Physiol.* **2006**, *142*, 1065–1074. [[CrossRef](#)] [[PubMed](#)]
45. Blum, A. Osmotic adjustment is a prime drought stress adaptive engine in support of plant production. *Plant Cell Environ.* **2017**, *40*, 4–10. [[CrossRef](#)] [[PubMed](#)]
46. Gray, J. Guard Cells: Transcription Factors Regulate Stomatal Movements. *Curr. Biol.* **2005**, *15*, R593–R595. [[CrossRef](#)]
47. Cominelli, E.; Galbiati, M.; Tonelli, C. Transcription factors controlling stomatal movements and drought tolerance. *Transcription* **2010**, *1*, 41–45. [[CrossRef](#)]
48. Shinozaki, K.; Yamaguchi-Shinozaki, K. Molecular responses to drought and cold stress. *Curr. Opin. Biotechnol.* **1996**, *7*, 161–167. [[CrossRef](#)]
49. Golldack, D.; Li, C.; Mohan, H.; Probst, N. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front. Plant Sci.* **2014**, *5*, 151. [[CrossRef](#)]
50. Bian, S.; Jiang, Y. Reactive oxygen species, antioxidant enzyme activities and gene expression patterns in leaves and roots of Kentucky bluegrass in response to drought stress and recovery. *Sci. Hortic.* **2009**, *120*, 264–270. [[CrossRef](#)]
51. Akbudak, M.A.; Filiz, E.; Uylas, S. Identification of O-acetylserine(thiol)lyase (OASTL) genes in sorghum (*Sorghum bicolor*) and gene expression analysis under cadmium stress. *Mol. Biol. Rep.* **2019**, *46*, 343–354. [[CrossRef](#)]
52. Chen, Y.; Li, C.; Zhang, B.; Yi, J.; Yang, Y.; Kong, C.; Lei, C.; Gong, M. The role of the Late Embryogenesis-Abundant (LEA) protein family in development and the abiotic stress response: A comprehensive expression analysis of potato (*Solanum Tuberosum*). *Genes* **2019**, *10*, 148. [[CrossRef](#)]
53. Mibei, E.K.; Ambuko, J.; Giovannoni, J.J.; Onyango, A.N.; Owino, W.O. Carotenoid profiling of the leaves of selected African eggplant accessions subjected to drought stress. *Food Sci. Nutr.* **2017**, *5*, 113–122. [[CrossRef](#)] [[PubMed](#)]
54. Havaux, M. Carotenoid oxidation products as stress signals in plants. *Plant J.* **2014**, *79*, 597–606. [[CrossRef](#)] [[PubMed](#)]
55. Joshi, R.; Wani, S.H.; Singh, B.; Bohra, A.; Dar, Z.A.; Lone, A.A.; Pareek, A.; Singla-Pareek, S.L. Transcription factors and plants response to drought stress: Current understanding and future directions. *Front. Plant Sci.* **2016**, *7*, 1029. [[CrossRef](#)]
56. Manna, M.; Thakur, T.; Chirom, O.; Mandlik, R.; Deshmukh, R.; Salvi, P. Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. *Physiol. Plant.* **2021**, *172*, 847–868. [[CrossRef](#)]
57. Alter, S.; Bader, K.C.; Spannagl, M.; Wang, Y.; Bauer, E.; Schön, C.-C.; Mayer, K.F.X. DroughtDB: An expert-curated compilation of plant drought stress genes and their homologs in nine species. *Database* **2015**, *2015*, bav046. [[CrossRef](#)] [[PubMed](#)]
58. Melton, A.E.; Beck, J.; Galla, S.J.; Jenkins, J.; Handley, L.; Kim, M.; Grimwood, J.; Schmutz, J.; Richardson, B.A.; Serpe, M.; et al. A draft genome provides hypotheses on drought tolerance in a keystone plant species in Western North America threatened by climate change. *Ecol. Evol.* **2021**, *11*, 15417–15429. [[CrossRef](#)] [[PubMed](#)]
59. Chamberlain, S.A.; Szöcs, E. Taxize: Taxonomic search and retrieval in R. *F1000Research* **2013**, *2*, 191. [[CrossRef](#)] [[PubMed](#)]



60. Derived Dataset GBIF.org. Filtered Export of GBIF Occurrence Data. Available online: <https://www.gbif.org/derivedDataset/10.15468/dd.2u7f3z> (accessed on 6 October 2022).
61. Geldmann, J.; Heilmann-Clausen, J.; Holm, T.E.; Levinsky, I.; Markussen, B.; Olsen, K.; Rahbek, C.; Tøttrup, A.P. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers. Distrib.* **2016**, *22*, 1139–1149. [[CrossRef](#)]
62. Chamberlain, S.A. scrubr: Clean Biological Occurrence Records. 2021. Available online: <https://rdrr.io/github/ropensci/scrubr/> (accessed on 11 October 2022).
63. Trabucco, A.; Zomer, R. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. figshare. *Dataset* **2019**. [[CrossRef](#)]
64. Chaves, M.M.; Marôco, J.P.; Pereira, J.S. Understanding plant responses to drought-From genes to the whole plant. *Funct. Plant Biol.* **2003**, *30*, 239–264. [[CrossRef](#)]
65. Fang, Y.; Xiong, L. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell. Mol. Life Sci.* **2015**, *72*, 673–689. [[CrossRef](#)] [[PubMed](#)]
66. Missouri Botanical Gardens. 2020. Available online: [http://www.mobot.org/MOBOT/Research/APweb/top/glossarya\\_h.html](http://www.mobot.org/MOBOT/Research/APweb/top/glossarya_h.html) (accessed on 30 December 2020).
67. Farooq, M.; Hussain, M.; Wahid, A.; Siddique, K.H.M. Drought stress in plants: An overview. In *Plant Responses to Drought Stress: From Morphological to Molecular Features*; Springer Berlin: Heidelberg, Germany, 2012; pp. 1–33.
68. De Micco, V.; Aronne, G. Morpho-anatomical traits for plant adaptation to drought. In *Plant Responses to Drought Stress: From Morphological to Molecular Features*; Springer Berlin: Heidelberg, Germany, 2012; pp. 37–61.
69. Vilagrosa, A.; Chirino, E.; Peguero-Pina, J.J.; Barigah, T.S.; Cochard, H.; Gil-Pelegrin, E. Xylem cavitation and embolism in plants living in water-limited ecosystems. In *Plant Responses to Drought Stress: From Morphological to Molecular Features*; Springer Berlin: Heidelberg, Germany, 2012; pp. 63–109.
70. Magallón, S.; Gómez-Acevedo, S.; Sánchez-Reyes, L.L.; Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **2015**, *207*, 437–453. [[CrossRef](#)]
71. McDowell, N.; Pockman, W.T.; Allen, C.D.; Breshears, D.D.; Cobb, N.; Kolb, T.; Plaut, J.; Sperry, J.; West, A.; Williams, D.G.; et al. Mechanisms of Plant Survival and Mortality during Drought: Why Do Some Plants Survive While Others Succumb to Drought? *New Phytol.* **2008**, *178*, 719–739. [[CrossRef](#)] [[PubMed](#)]
72. Hacke, U.G.; Stiller, V.; Sperry, J.S.; Pittermann, J.; McCulloh, K.A. Cavitation Fatigue. Embolism and Refilling Cycles Can Weaken the Cavitation Resistance of Xylem. *Plant Physiol.* **2001**, *125*, 779–786. [[CrossRef](#)] [[PubMed](#)]
73. Li, Y.; Sperry, J.S.; Shao, M. Hydraulic Conductance and Vulnerability to Cavitation in Corn (*Zea mays* L.) Hybrids of Differing Drought Resistance. *Environ. Exp. Bot.* **2009**, *66*, 341–346. [[CrossRef](#)]
74. Vilagrosa, A.; Bellot, J.; Vallejo, V.R.; Gil-Pelegrin, E. Cavitation, Stomatal Conductance, and Leaf Dieback in Seedlings of Two Co-Occurring Mediterranean Shrubs during an Intense Drought. *J. Exp. Bot.* **2003**, *54*, 2015–2024. [[CrossRef](#)]
75. Choat, B.; Jansen, S.; Brodribb, T.J.; Cochard, H.; Delzon, S.; Bhaskar, R.; Bucci, S.J.; Feild, T.S.; Gleason, S.M.; Hacke, U.G.; et al. Global Convergence in the Vulnerability of Forests to Drought. *Nature* **2012**, *491*, 752–755. [[CrossRef](#)]
76. McDowell, N.G. Mechanisms Linking Drought, Hydraulics, Carbon Metabolism, and Vegetation Mortality. *Plant Physiol.* **2011**, *155*, 1051–1059. [[CrossRef](#)]
77. Sevanto, S.; McDowell, N.G.; Dickman, L.T.; Pangle, R.; Pockman, W.T. How Do Trees Die? A Test of the Hydraulic Failure and Carbon Starvation Hypotheses. *Plant, Cell Environ.* **2014**, *37*, 153–161. [[CrossRef](#)]
78. Jactel, H.; Petit, J.; Desprez-Loustau, M.L.; Delzon, S.; Piou, D.; Battisti, A.; Koricheva, J. Drought Effects on Damage by Forest Insects and Pathogens: A Meta-Analysis. *Glob. Chang. Biol.* **2012**, *18*, 267–276. [[CrossRef](#)]
79. Qiu, Y.; Yu, D. Over-Expression of the Stress-Induced OsWRKY45 Enhances Disease Resistance and Drought Tolerance in Arabidopsis. *Environ. Exp. Bot.* **2009**, *65*, 35–47. [[CrossRef](#)]
80. Ramegowda, V.; Senthil-Kumar, M. The Interactive Effects of Simultaneous Biotic and Abiotic Stresses on Plants: Mechanistic Understanding from Drought and Pathogen Combination. *J. Plant Physiol.* **2015**, *176*, 47–54. [[CrossRef](#)]
81. Szczepaniec, A.; Finke, D. Plant-Vector-Pathogen Interactions in the Context of Drought Stress. *Front. Ecol. Evol.* **2019**, *7*, 1–7. [[CrossRef](#)]
82. Kerr, K.L.; Meinzer, F.C.; McCulloh, K.A.; Woodruff, D.R.; Marias, D.E. Expression of Functional Traits during Seedling Establishment in Two Populations of *Pinus ponderosa* from Contrasting Climates. *Tree Physiol.* **2015**, *35*, 535–548. [[CrossRef](#)]
83. Adams, H.D.; Zeppel, M.J.B.; Anderegg, W.R.L.; Hartmann, H.; Landhäusser, S.M.; Tissue, D.T.; Huxman, T.E.; Hudson, P.J.; Franz, T.E.; Allen, C.D.; et al. A Multi-Species Synthesis of Physiological Mechanisms in Drought-Induced Tree Mortality. *Nat. Ecol. Evol.* **2017**, *1*, 1285–1291. [[CrossRef](#)]
84. Streck, N.A. Do We Know How Plants Sense a Drying Soil? *Ciência Rural* **2004**, *34*, 581–584. [[CrossRef](#)]
85. Robbins, N.E.; Dinneny, J.R. The Divining Root: Moisture-Driven Responses of Roots at the Micro- and Macro-Scale. *J. Exp. Bot.* **2015**, *66*, 2145–2154. [[CrossRef](#)]
86. Xiong, L.; Schumaker, K.S.; Zhu, J.-K. Cell Signaling during Cold, Drought, and Salt Stress. *Plant Cell* **2002**, *14*, S165–S183. [[CrossRef](#)] [[PubMed](#)]
87. Zhu, J.K. Salt and Drought Stress Signal Transduction in Plants. *Annu. Rev. Plant Biol.* **2002**, *53*, 247–273. [[CrossRef](#)] [[PubMed](#)]

88. Jakab, G.; Ton, J.; Flors, V.; Zimmerli, L.; Métraux, J.P.; Mauch-Mani, B. Enhancing Arabidopsis Salt and Drought Stress Tolerance by Chemical Priming for Its Abscisic Acid Responses. *Plant Physiol.* **2005**, *139*, 267–274. [[CrossRef](#)]
89. Zhang, J.; Jia, W.; Yang, J.; Ismail, A.M. Role of ABA in Integrating Plant Responses to Drought and Salt Stresses. *F. Crop. Res.* **2006**, *97*, 111–119. [[CrossRef](#)]
90. Moran, E.; Lauder, J.; Musser, C.; Stathos, A.; Shu, M. The Genetics of Drought Tolerance in Conifers. *New Phytol.* **2017**, *216*, 1034–1048. [[CrossRef](#)]
91. Cruz De Carvalho, M.H. Drought Stress and Reactive Oxygen Species: Production, Scavenging and Signaling. *Plant Signal. Behav.* **2008**, *3*, 156–165. [[CrossRef](#)] [[PubMed](#)]
92. Munné-Bosch, S.; Alegre, L. Die and Let Live: Leaf Senescence Contributes to Plant Survival under Drought Stress. *Funct. Plant Biol.* **2004**, *31*, 203–216. [[CrossRef](#)]
93. Blackman, C.J.; Brodribb, T.J.; Jordan, G.J. Leaf Hydraulics and Drought Stress: Response, Recovery and Survivorship in Four Woody Temperate Plant Species. *Plant, Cell Environ.* **2009**, *32*, 1584–1595. [[CrossRef](#)]
94. Chen, D.; Wang, S.; Cao, B.; Cao, D.; Leng, G.; Li, H.; Yin, L.; Shan, L.; Deng, X. Genotypic Variation in Growth and Physiological Response to Drought Stress and Re-Watering Reveals the Critical Role of Recovery in Drought Adaptation in Maize Seedlings. *Front. Plant Sci.* **2016**, *6*, 1–15. [[CrossRef](#)]
95. Goh, C.H.; Gil Nam, H.; Shin Park, Y. Stress Memory in Plants: A Negative Regulation of Stomatal Response and Transient Induction of Rd22 Gene to Light in Abscisic Acid-Entrained Arabidopsis Plants. *Plant J.* **2003**, *36*, 240–255. [[CrossRef](#)]
96. Virlovet, L.; Fromm, M. Physiological and Transcriptional Memory in Guard Cells during Repetitive Dehydration Stress. *New Phytol.* **2015**, *205*, 596–607. [[CrossRef](#)] [[PubMed](#)]
97. Jacobsen, A.L.; Pratt, R.B.; Davis, S.D.; Ewers, F.W. Cavitation Resistance and Seasonal Hydraulics Differ among Three Arid Californian Plant Communities. *Plant Cell Environ.* **2007**, *30*, 1599–1609. [[CrossRef](#)] [[PubMed](#)]
98. Luo, N.; Liu, J.; Yu, X.; Jiang, Y. Natural Variation of Drought Response in *Brachypodium distachyon*. *Physiol. Plant.* **2011**, *141*, 19–29. [[CrossRef](#)] [[PubMed](#)]
99. Gutzat, R.; Mittelsten Scheid, O. Epigenetic Responses to Stress: Triple Defense? *Curr. Opin. Plant Biol.* **2012**, *15*, 568–573. [[CrossRef](#)] [[PubMed](#)]
100. de Vries, F.T.; Griffiths, R.I.; Knight, C.G.; Nicolitch, O.; Williams, A. Harnessing Rhizosphere Microbiomes for Drought-Resilient Crop Production. *Science* **2020**, *368*, 270–274. [[CrossRef](#)] [[PubMed](#)]