

## Article

# Genome-Wide Identification of Potential mRNAs in Drought Response in Wheat (*Triticum aestivum* L.)

Muhammad Aqeel <sup>1</sup>, Wajya Ajmal <sup>1</sup>, Quratulain Mujahid <sup>2</sup>, Maryam Murtaza <sup>1</sup>, Mansour Almuqbil <sup>3</sup>, Shakira Ghazanfar <sup>1</sup>, Muhammad Uzair <sup>1,4</sup>, Ayesha Wadood <sup>5</sup>, Syed Mohammed Basheeruddin Asdaq <sup>6</sup>, Rameesha Abid <sup>1</sup>, Ghulam Muhammad Ali <sup>7,\*</sup> and Muhammad Ramzan Khan <sup>1,\*</sup>

- <sup>1</sup> National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Centre, Park Road, Islamabad 45500, Pakistan
- <sup>2</sup> Department of Botany, Lahore College for Women University Lahore, Lahore 54000, Pakistan
- <sup>3</sup> Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia
- <sup>4</sup> National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
- <sup>5</sup> Department of Biochemistry, Quaid-i-Azam University, Islamabad 45320, Pakistan
- <sup>6</sup> Department of Pharmacy Practice, College of Pharmacy, AlMaarefa University, Dariyah, Riyadh 13713, Saudi Arabia
- <sup>7</sup> Pakistan Agricultural Research Council, Islamabad 44000, Pakistan
- \* Correspondence: drgmali5@gmail.com (G.M.A.); mrkhan@parc.gov.pk (M.R.K.)

**Abstract:** Plant cell metabolism inevitably forms an important drought-responsive mechanism, which halts crop productivity. Globally, more than 30% of the total harvested area was affected by dehydration. RNA-seq technology has enabled biologists to identify stress-responsive genes in relatively quick times. However, one shortcoming of this technology is the inconsistent data generation compared to other parts of the world. So, we have tried, here, to generate a consensus by analyzing meta-transcriptomic data available in the public microarray database GEO NCBI. In this way, the aim was set, here, to identify stress genes commonly identified as differentially expressed ( $p < 0.05$ ) then followed by downstream analyses. The search term “Drought in wheat” resulted in 233 microarray experiments from the GEO NCBI database. After discarding empty datasets containing no expression data, the large-scale meta-transcriptome analytics and one sample proportional test were carried out (Bonferroni adjusted  $p < 0.05$ ) to reveal a set of 11 drought-responsive genes on a global scale. The annotation of these genes revealed that the transcription factor activity of RNA polymerase II and sequence-specific DNA-binding mechanism had a significant role during the drought response in wheat. Similarly, the primary root differentiation zone annotations, controlled by *TraesCS5A02G456300* and *TraesCS7B02G243600* genes, were found as top-enriched terms ( $p < 0.05$  and  $Q < 0.05$ ). The resultant standard drought genes, glycosyltransferase; Arabidopsis thaliana KNOTTED-like; bHLH family protein; Probable helicase MAGATAMA 3; SBP family protein; Cytochrome c oxidase subunit 2; Trihelix family protein; Mic1 domain-containing protein; ERF family protein; HD-ZIP I protein; and ERF family protein, are important in terms of their worldwide proved link with stress. From a future perspective, this study could be important in a breeding program contributing to increased crop yield. Moreover, the wheat varieties could be identified as drought-resistant/sensitive based on the nature of gene expression levels.

**Keywords:** drought; bread wheat; meta data; RNA seq; genomics



**Citation:** Aqeel, M.; Ajmal, W.; Mujahid, Q.; Murtaza, M.; Almuqbil, M.; Ghazanfar, S.; Uzair, M.; Wadood, A.; Asdaq, S.M.B.; Abid, R.; et al. Genome-Wide Identification of Potential mRNAs in Drought Response in Wheat (*Triticum aestivum* L.). *Genes* **2022**, *13*, 1906. <https://doi.org/10.3390/genes13101906>

Academic Editors: Asma Ayaz, Fazal Ullah and Saddam Saqib

Received: 24 September 2022

Accepted: 17 October 2022

Published: 20 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

More than two billion people worldwide, particularly in Asia, depend on wheat (*Triticum aestivum* L.) as their primary source of proteins, B vitamins, and dietary fibers [1,2]. Wheat is the largest produced commodity, reaching, approximately, 778.6 million metric tons [3]. A record 194 million tons of wheat will be traded globally in 2022, up 2.5 percent

(4.8 million tons) from the level in 2020/21. As a result of a better-than-anticipated demand, this represents a rise of 1.1 million tons from the previous report [4]. It is also predicted that in order to meet the demand and consumption in 2050, wheat production will need to rise by 50–60% from its current level.

Wheat biomass and yield are reduced by 25.0% and 27.5%, respectively, by drought [5]. To meet the problems of ensuring nutritional security, wheat grain quality must also increase in terms of its ability to withstand drought [6]. The frequency of rainfall, evaporation, and soil moisture all play a role in how severe or intense drought is, among other factors [7]. Drought stress not only affects more than one-third of all farmed land worldwide, but also facilitates insect infestation leading to low crop productivity. A total of 33% ( $9.9 \times 10^7$  hm<sup>2</sup>) of that territory is occupied by developing nations, 25% ( $6.0 \times 10^7$  hm<sup>2</sup>) by developed countries, and 42% ( $12.6 \times 10^7$  hm<sup>2</sup>) by underdeveloped states [8]. In Asia, there is only  $3.4 \times 10^7$  hm<sup>2</sup> of rainfed lowland and  $8.0 \times 10^6$  hm<sup>2</sup> of highland wheat that is under drought stress [9]. Wheat productivity can be increased in a sustainable and economically feasible way by breeding wheat types that are tolerant to drought stress [10]. Previous research has shown that drought has a negative influence on crop production. Given the inherent problems with drought stress for wheat breeders, a detailed understanding of the molecular basis of wheat drought stress can assist greatly in the identification of appropriate varieties. Numerous genes and microbial interactions are activated as a result of the drought-stress response in plants [1,11–14]. The wheat *NAC* gene (*TaNAC071-A*) is recently reported as showing one of the strongest associations with drought stress. They experimented with the knockout technique and found that the *TaNAC071-A* gene is a positive regulator of plant dehydration response [15,16]; Zinc finger protein (*ZFP*) consists of various member proteins constituting a family [17], apart from the RNA/DNA difference (RDD) during central dogma confer resistance to drought stress in plants [18].

The genes are included in selenium stress [19], response to auxin [20], electron transport chain, transcription regulated by the RNA polymerase II promoter, and different transcription factors. The primary transcription factors associated with stress include *TraesCS3D02G120200*, *TraesCS6A02G328700*, *Traes-CS2D02G000200*, *TraesCS6D02G086600*, *TraesCS6D02G260700*, *TraesCSU02G154600*, *TraesCS6B02G234100*, *TraesCS1D02G333000*, *TraesCS5A02G456300*, *TraesCS6A02G24-0400*, and *TraesCS7B02G243600*; these genes are crucial for controlling drought tolerance [21–24].

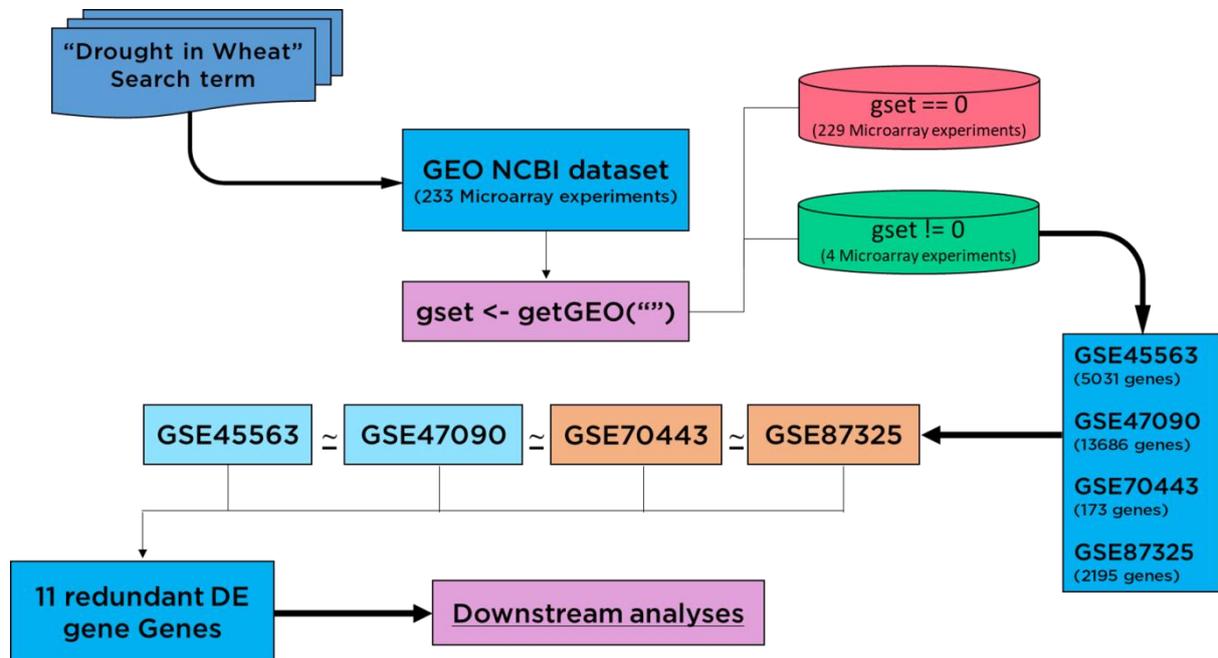
This research included a number of meta-analyses, summary studies, and model simulation results that only took drought into consideration, and discarded samples that had been subjected to stress other than drought. Under the conditions of climate- and water-availability constraints, it will be important to boost biomass production and economic yield. Before that can happen, it is necessary to comprehend the magnitude of the decline in wheat production and other agronomic features that are impacted by molecular parameters, such as drought proteins. Apart from generating knowledge of global drought genes, we are also interested to conduct an assessment of our methodology as a strategy of metatranscriptome analytics scale to address drought stress. In the future, this standardized methodology can serve as a road map of the potential for application in other crop species.

## 2. Materials and Methods

### 2.1. Meta-Analysis of Data from the Genome-Wide Transcriptome

The GEO Datasets [25] GSE45563 (without heat stress samples), GSE47090, GSE70443, and GSE87325 were employed as data sources. The expression profiles of drought-related genes (Figure 1) were targeted for *Triticum aestivum*, obtained in control and stress (water shortage) conditions [26]. The search term “Drought in wheat” was used to search for 233 microarray experiments from the GEO NCBI database. To identify changes in the expression patterns of genes, the expression profiles were downloaded using the *getGEO()* function. Out of 233 datasets, the empty experiments were discarded, containing no expression data by using the condition *gset* = 0. The large-scale meta-transcriptome analytics

and 1 sample proportional test (Bonferroni adjusted  $p < 0.05$ ) revealed a set of 11 drought-responsive genes on a global scale.



**Figure 1.** Flow chart diagram shown for screening of differentially expressed (DE) genes. Redundant genes were collected using expression datasets which were differentially expressed in each dataset.

## 2.2. Differential Expression Analyses

Researchers in bioinformatics have gained a lot of knowledge from studying microarray data. For instance, information sharing across all probes can increase the power to detect differential expressions and decrease false findings. One such method is limma [27], where the probe-wise variances were moderated using an empirical Bayes model. In the  $t$ - and  $F$ -statistic calculations, the moderated variances take the place of the probe-wise variances. EdgeR models count data using an over-dispersed Poisson model in a conceptually similar, but theoretically more challenging processes, and uses an empirical Bayes approach to control the degree of overdispersion among genes.

We anticipated that the information could be condensed into a table of counts, with rows denoting genes (or tags, exons, or transcripts) and columns denoting samples. These could be counts at the exon, transcript, or gene levels for RNA-seq research. The data are modelled using a negative binomial (NB) distribution,

$$Y_{ki} \sim NB(M_i p_{kj}, \phi_k) \quad (1)$$

for sample  $i$  and gene  $g$ . Here,  $M_i$  is the size of the library (the total number of reads),  $\phi_k$  denotes the dispersion, and  $p_{kj}$  denotes the relative abundance of the gene  $g$  in experimental group  $j$ , which includes sample  $i$ . We use the NB parameterization where the mean is  $\mu_{ki} = M_i p_{kj}$  and the variance is  $\mu_{ki} (1 + \mu_{ki} \phi_k)$ . For differential expression analysis, the parameters of interest are  $p_{kj}$ . The NB distribution was reduced to Poisson when  $\phi_k = 0$ . In some DGE applications, technical variation can be treated as Poisson. In general,  $\phi_k$  indicates the biological variation's coefficient of variance between the samples. Our approach is able to distinguish between biological and technical variance in this manner.

EdgeR calculates the gene-wise dispersion probabilities using conditional maximum likelihood, based on the total number of genes in that gene [28]. The dispersions were reduced using an empirical Bayes approach to a consensus value, efficiently utilizing information from other genes. Finally, the differential expression for each gene was evaluated using an exact test similar to Fisher's exact test but modified for highly distributed data.

R software (version 3.4.3; <https://cran.r-project.org/>) (accessed on 10 October 2022) was used to process drought-related microarray data acquired from the GEO database. The edgeR tool in R was used to identify differentially expressed genes (DEGs) across drought-stressed plants and control plants. Fold-change (FC) values were determined, and the following cutoff criteria were used to further select the DEGs:  $p < 0.05$ ,  $FDR < 0.05$  and  $\log |FC| > 2$ . The four datasets' overlapping DEGs were found using MS EXCEL manipulations (version 3.1.3; <http://www.funrich.org>) (accessed on 10 October 2022).

### 2.3. One-Sample Proportions Test for SDGs Screening

Only groups with finite numbers of gene presences and absences were used. The counts of presences and absences must be non-negative and, hence, not greater than the corresponding numbers of microarray experiments, which must be positive. All finite counts should be integers. In R, the `prop.test` function was utilized for  $>1$  gene presences in groups. Thus, for 2, 3, and 4 presences, the proportions 50 by 100, 75 by 100, and 100 by 100 were used, respectively. Later on, 2 presences in microarray experiments with the proportion 50 by 100 were discarded due to  $p = 1$  outcomes. The reproducibility can be checked using the R commands below:

```
> prop.test(50,100)$p.value
[1] 1
> prop.test(75,100)$p.value
[1] 9.583666e-07
> prop.test(100,100)$p.value
[1] 4.16275e-23
```

### 2.4. Bioinformatics Analyses

Sequence analysis and annotation predications were performed using the NCBI database containing BLAST tool. The default alignment parameter settings were as, max target sequences were restricted up to 100; threshold 100; Word size 28; Max matches in a query range 0; Match/Mismatch Scores 1, -2; Gap Costs were selected as linear; low-complexity region; mask for lookup table only during the BLAST search. Accession numbers against GI identifiers were identified, and then drought genes nucleotide and protein sequences were retrieved. PlantRegMap is a plant genomes database with advanced biocomputing tools (<http://plantregmap.gao-lab.org/>) (accessed on 10 October 2022). For transcription factor analysis, the protein sequences were aligned using the BLAST online tool in PlantTFDB v5.0 to mine for stress-related orthologs present in the database in the annotated form.

### 2.5. Functional and Pathway Enrichment Analyses of SDGs

An important resource for any functional investigation of the high-throughput gene-expression profiles is clusterProfiler v4.0.5, which we used for the functional annotation of GO and analysis of KEGG pathway enrichment [29]. The results from clusterProfiler were further imported to multiEnrich v17.900 in R v 4.1.1. The IGRAPH and TKPLOT were employed to illustrate the top 11 SDGs' functional enlargement. To thoroughly examine the SDGs connected to the GO terms and pathways, we conducted an integrative analysis using both the clusterProfiler v4.0.5 and IGRAPH packages. To obtain all Gene Ontological terms (GO) and trait-related pathways from the SDG dataset, the initial SDGs from the GEO2R tool were exposed to clusterProfiler v2.5.5/multiEnrich v17.900. clusterProfiler syndicates GO from the SDG dataset, delivering a fundamentally ordered functional network. Additionally, the enrichment of molecular/biological function GO analysis for SDGs was carried out, and  $p$ -values 0.05 were deemed significant.

### 2.6. Principal Component Analyses

Principal components analysis of the SDGs and the own gene expression dataset were calculated in R version 4.1.1 using the PCA function of the FactoMineR v2.4 package.

Subsequently, we compared the first and second PCs of both datasets in order to determine whether they span similar spaces.

### 2.7. Plant Ontology

A controlled vocabulary (ontology), known as Plant Ontology, explains the anatomy, morphology, and developmental stages of all plants. In order to enable meaningful cross-species queries on gene expression and phenotype data sets from plant genomics and genetics investigations, the PO aims to create a semantic framework.

The need to create new and better-adapted wheat varieties is brought on by the rise of the world's population and the changing climate. The Plant Trait Ontology (TO), the Crop Ontology (CO), and the GrainGenes database are three cohesive resources that give scientists and plant breeders connected resources and tools to make use of the vast amounts of genetic and genomic data that are available for plant genomics and crop development. The Planteome Project's reference-level ontology for plant characteristics is called the TO. More than 1500 plant traits are included in the TO of the most recent Planteome Release Version 4.0, which are arranged into nine higher-level categories: biochemical, biological, and plant growth and development processes, as well as plant shape, quality, vigour, stress, and yield. The TO is integrated with the Plant Ontology as part of the Planteome and is used to annotate or link to data objects for plant genomics and genetics (such as germplasm, QTLs, genes, and proteins) from a variety of plant taxa, including significant world crops and model plant species. There are around 165,000 data elements in the Planteome database that are linked by more than four million annotations in this release. Through the TO GitHub Issue Tracker, users are encouraged to request new TO terms or to leave comments.

## 3. Results

### 3.1. Meta-Analysis of Data from the Genome-Wide Transcriptome

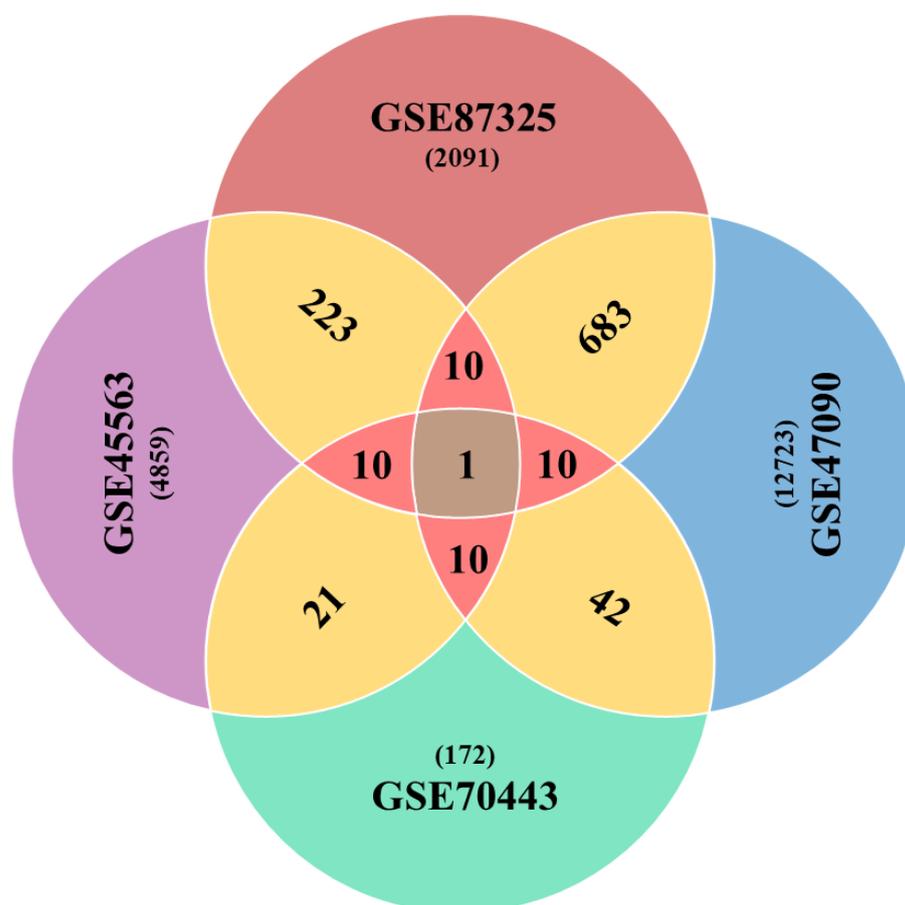
For bread wheat (*T. aestivum*), the GEO Datasets [25] were used as a source of data for the expression profiles of drought-related genes acquired in control and stress (normal water and water deficit) situations [26]. Initially, the samples consisted of 233 experiments for wheat, out of which four experiments were screened for available transcriptome data (GSE87325, GSE47090, GSE45563 and GSE70443), including information from microarray and NGS platforms (Supplementary Table S1). Other drought-stressed samples were excluded from the analyses, such as GSE45563 which had heat and analyzing 233 RNAseq IDs. Of the datasets GSE87325, GSE47090, GSE45563 and GSE70443, two originated from the USA, one dataset was from Germany and one from India, and were acquired from the GEO NCBI database. All datasets were checked for expression availability, whether the link "Analyze with GEO2R" is provided or not. Luckily, we got four out of 233 of which the expression data was available in GEO NCBI.

### 3.2. Global Drought Genes Screening

The differential expression analyses resulted in a variable number of genes in each case of the microarray experiment. These were 4859, 12,723, 172 and 2091 drought-related genes ( $p < 0.05$ ) from GSE45563, GSE47090, GSE70443 and GSE87325 microarray experiments, respectively. After a series of analytics, we came up with one gene with the GI accession 31369563 as a global factor towards drought stress in wheat as it was found as differentially expressed among all microarray experiments. Another ten genes passed the one sample proportion test ( $p < 0.05$ ) (Table 1). In total, we got a list of eleven Standard Drought Genes (SDGs). The pairwise comparison showed a larger number of DE genes, as shown in Figure 2. However, a comparison between three and four datasets was 10 and 1 DE genes for each combination, respectively.

**Table 1.** The SDGs were screened based on 1-sample proportions test with Bonferroni corrected  $p < 0.05$ , discarding other genes that did not become significant with similar criteria. The differential expression data included 4859, 12,723, 172 and 2091 genes from GSE45563, GSE47090, GSE70443 and GSE87325 microarray experiments, respectively.

| Sr # | GSE45563 | GSE47090 | GSE70443 | GSE87325 | Coutif | $p$ -Value             |
|------|----------|----------|----------|----------|--------|------------------------|
| 1    | 31369563 | 31369563 | 31369563 | 31369563 | 4      | $4.16 \times 10^{-23}$ |
| 2    | 25245708 | 25245708 | 25245708 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 3    | 31369545 | 31369545 | 31369545 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 4    | 25232126 | 25232126 | 25232126 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 5    | #N/A     | 19959095 | 19959095 | 19959095 | 3      | $9.58 \times 10^{-7}$  |
| 6    | 14320    | 14320    | 14320    | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 7    | 25550165 | 25550165 | 25550165 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 8    | 25270582 | 25270582 | 25270582 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 9    | 20313737 | 20313737 | 20313737 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 10   | 20334251 | 20334251 | 20334251 | #N/A     | 3      | $9.58 \times 10^{-7}$  |
| 11   | 25231635 | 25231635 | 25231635 | #N/A     | 3      | $9.58 \times 10^{-7}$  |



**Figure 2.** Venn diagram elaborating shared number of genes across the combinations of microarray datasets. Sharing of DE genes between two datasets are shown in gold color, while sharp pink color represents three combination outcomes, in terms of sharing the number of DE genes.

### 3.3. IWGSC Gene IDs Identification

Annotating genes is core to every biological experiment. We ran the BLAST search using ENSEMBL database, against wheat as the reference genome. The top hits accessed from aligned results are shown in Table 2. The high E-value, along with the alignment score and identity percentage for top hits homologs, elucidated that they had a strong reason to be considered as orthologs of the drought genes.

**Table 2.** Gene ID conversion provided the information about SDGs in the form of columns. First and second columns are the IDs that resulted from differential expression analyses. Similarly, third column was filled with their respective IWGSC IDs due to their high numbers in the form of alignment scores and residual identity in percentages with low E-values. Genomic locations were recorded by searching ENSEMBL database and provided as last three columns.

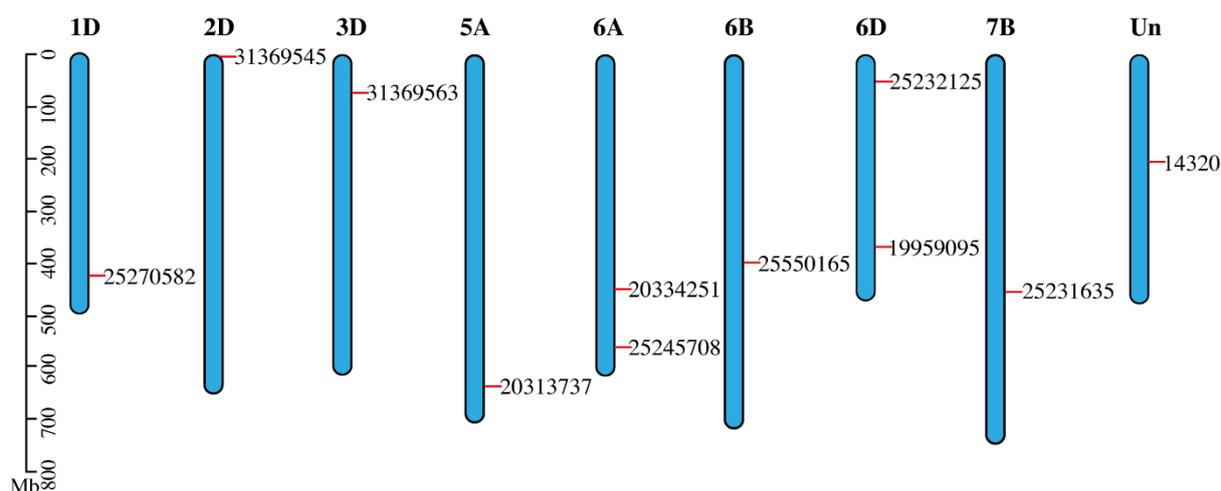
| Genes    | Accession Numbers | IWGSC Genes               | Score | E-Value                 | Identity (%) | Chromosome | Start     | End       |
|----------|-------------------|---------------------------|-------|-------------------------|--------------|------------|-----------|-----------|
| 31369563 | CD454935.1        | <i>TraesCS3D02G120200</i> | 484   | 0                       | 97.9         | 3D         | 75946541  | 75948546  |
| 25245708 | CA667105.1        | <i>TraesCS6A02G328700</i> | 146   | $1.50 \times 10^{-75}$  | 99.3         | 6A         | 562271850 | 562272738 |
| 31369545 | CD454917          | <i>TraesCS2D02G000200</i> | 424   | 0                       | 99.5         | 2D         | 39478     | 40878     |
| 25232126 | CA653601          | <i>TraesCS6D02G086600</i> | 248   | $2.50 \times 10^{-136}$ | 100          | 6D         | 52285210  | 52292603  |
| 19959095 | BJ220896          | <i>TraesCS6D02G260700</i> | 696   | 0                       | 99.9         | 6D         | 368305109 | 368306633 |
| 14320    | X52867            | <i>TraesCSU02G154600</i>  | 507   | 0                       | 98.3         | Un         | 206782100 | 206784099 |
| 25550165 | CA734567          | <i>TraesCS6B02G234100</i> | 105   | $9.40 \times 10^{-51}$  | 100          | 6B         | 393072749 | 393090605 |
| 25270582 | CA684029          | <i>TraesCS1D02G333000</i> | 118   | $1.30 \times 10^{-58}$  | 96.2         | 1D         | 423261565 | 423269992 |
| 20313737 | BQ168410          | <i>TraesCS5A02G456300</i> | 224   | $6.00 \times 10^{-122}$ | 100          | 5A         | 636429039 | 636430690 |
| 20334251 | BQ172428          | <i>TraesCS6A02G240400</i> | 308   | $5.40 \times 10^{-172}$ | 99.4         | 6A         | 451659872 | 451662196 |
| 25231635 | CA653110          | <i>TraesCS7B02G243600</i> | 235   | $3.20 \times 10^{-128}$ | 99.6         | 7B         | 452150743 | 452210823 |

### 3.4. Identification of Analyzed Genes Transcripts

This study analyzed genes encoding, glycosyltransferase, Auxin-responsive family protein, NAD(P)H-dependent oxidoreductase 1 (drought-responsive), proteins involved in gene expression regulations and auxin response. The EnsemblPlants (<https://plants.ensembl.org>) (accessed on 10 October 2022) database was used to find full-length cDNA sequences in the wheat genome (*Triticum aestivum* v. 2.2) (Table 2) 91 for genes producing antioxidant enzymes and enzymes involved in proline production. For identified sequences, Basic Local Alignment Search Tool (BLAST) for alignment resulted in accession IDs of the drought genes. It can be seen in the chromosome column that the D genome is at the top to have a larger number of drought genes, five out of eleven, *TraesCS3D02G120200*, *TraesCS2D02G000200*, *TraesCS6D02G086600*, *TraesCS6D02G260700* and *TraesCS1D02G333000* compared to any other chromosome. Genome A was depicted as second having SDGs the most after the D genome.

### 3.5. Mapping of SDGs on Wheat Genome

Chromosomal map explained how to alter new tracks by visualizing chromatin state transit—chromatin state in the genome has moved, for example, from one group of samples to the other. Variable methylation patterns are present in the genomic areas where chromatin states change, and these patterns may be an intriguing indicator of how chromatin states change. Most gar chromosomes have two counterparts in sterlet, according to an examination of conserved synteny between sterlet and gar. A total of 46 scaffolds were identified when sterlet homologous gene pairs were mapped against the genome in a paired manner. As expected from a WGD event, this finding shows homologous chromosomal fragments (Figure 3).



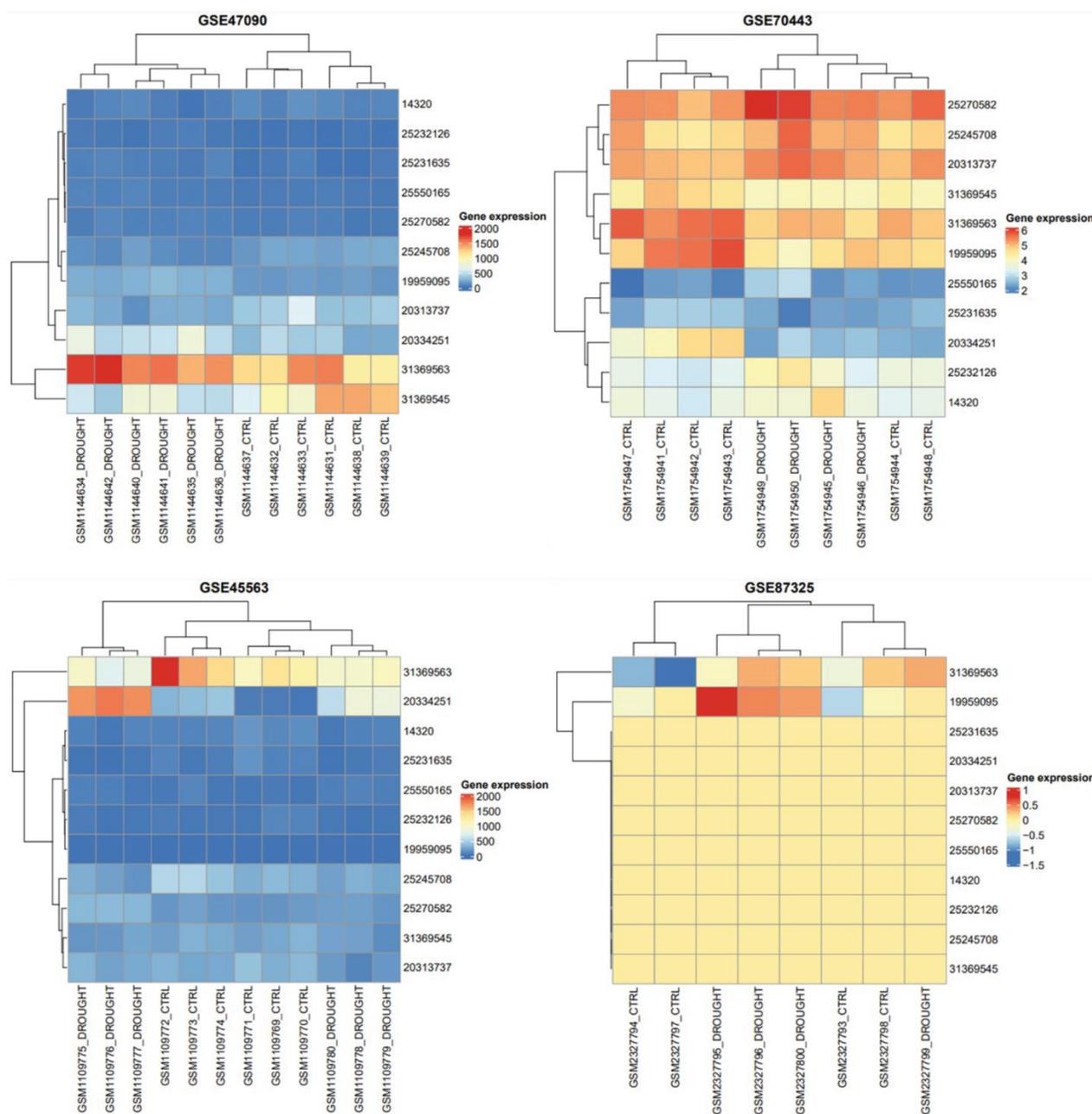
**Figure 3.** Chromosomal map depicted the SDGs distribution over their respective chromosomes. Out of all wheat chromosomes, all SDG-bearing chromosomes are included in the map. Gene places on chromosomes were made set according to their genomic locations. The chromosomes were drawn as blue bars.

### 3.6. Expression Analyses

An understanding of a gene's function in numerous biological processes can be gained by studying its expression profile. The expression studies of *T. aestivum* mRNAs were compared to drought and control phases in order to uncover the role of mRNA in distinct tissue developmental stages. Similar to the mRNAs, the majority of mRNA demonstrated differential expressions between the stages of drought and control (Figure 4). The results depicted their role in unrelated stages.

Physiological measures were utilized to generate a heatmap in order to pinpoint the critical variables for evaluating drought tolerance in wheat. As shown in Figure 2, for hierarchical (row) clustering, the morphological and physiological parameters of the 49 genotypes, which were grown either with drought treatment or with adequate water, were used. The 49 genotypes grouped together into group A when cultivated in well-watered conditions, while the identical 49 genotypes grouped together into group B when grown in drought-stricken conditions. By clear clustering, it can be seen that each switch-grass genotype's physiological and morphological traits are changed by drought-stress treatment in comparison to control circumstances. Surprisingly, in well-watered areas, the majority of the lowland genotypes tended to group together (group a in Figure 2, dot-highlighted); however, under drought-stress conditions, these genotypes were dispersed (group b in Figure 2 dot-highlighted).

A heatmap for SDGs is drawn from their expression pattern, observed in all microarray datasets. Heatmap of RNA-Seq transcriptome analysis for 11 chosen genes from the *Triticum aestivum* complex drought group and reference IWGSC genome. On the basis of a number of analysis parameters, about 11 genes are displayed. Based on average linkage and Euclidean distance of gene expression data, genes and samples were hierarchically clustered (dendrogram is shown for genes). It can be seen that drought and control samples showed variation in their protein secretion when we applied different watering conditions. The gene expression of each drought gene in terms of box colors provided the idea about whether it is under- or over-represented (Figure 4).

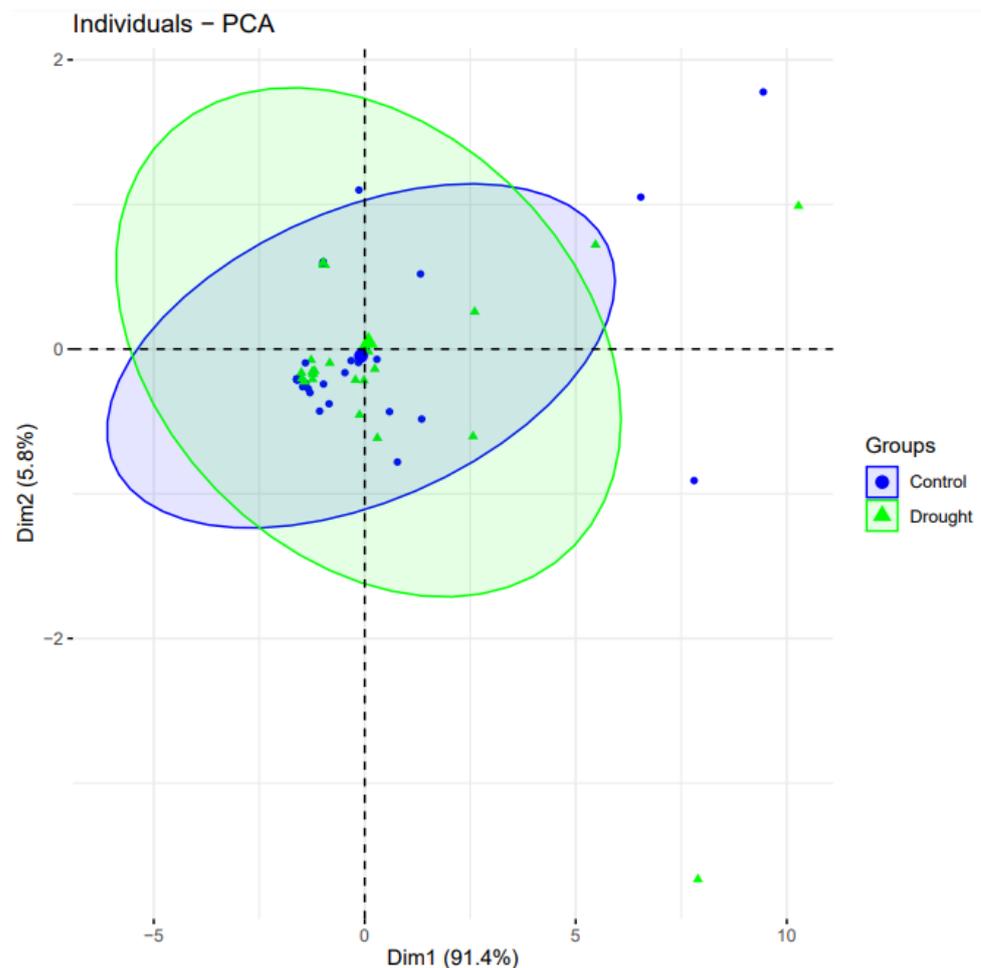


**Figure 4.** Heatmap and hierarchical clustering for gene expression matrices under well-watered and drought-stress conditions in 11 SDGs are shown as heatmaps. Color ranges are set between blue and red in order to visualize low to high gene expression, respectively. Clustering analysis of GSE47090, GSE70443, GSE45563 and GSE87325 showed two main horizontal groups of drought and controlled samples. Whereas, vertical scale is used to show all SDGs common in all gene expression datasets. Drought gene with GI accession 25550165 tends to be underexpressed in all experiments. Similarly, 313690563 gene in drought conditions had a trend towards overexpression in all gene expression datasets.

### 3.7. Statistical Analysis

Clustering and principal component analysis (PCA) was obtained using the “factoextra” program [30]. We used PCA to assess each drought gene’s contributions to the drought-treated and control wheat plants. The genes from the control case were more important in separating the groups than the parameters that had been subjected to drought. (Figure 5). Among the eleven drought genes, three genes from control condition samples

(blue-colored data points) covered an overall greater distance compared to the drought-treated groups, and thus, are considered positively correlated.



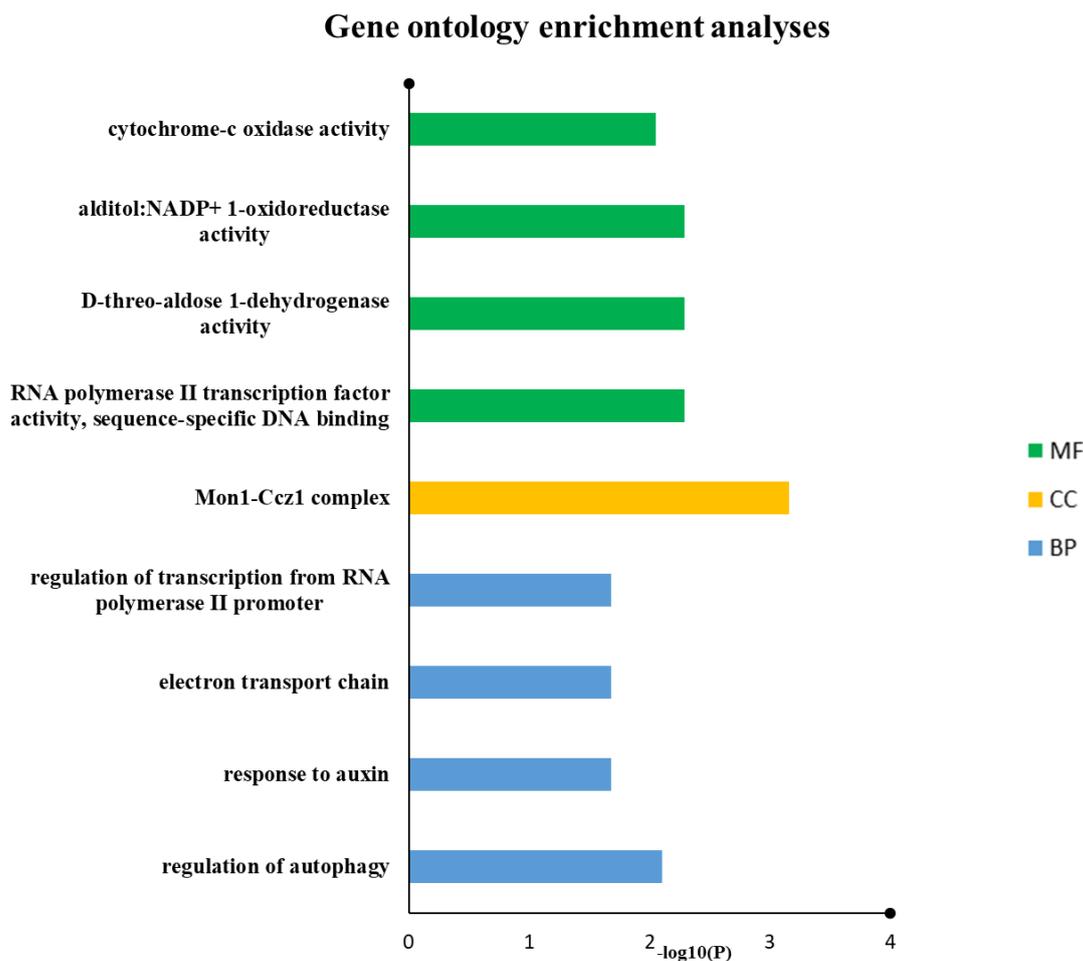
**Figure 5.** Principal component analysis biplot of gene expression of 11 genes studied in well-watered and drought conditions. PCA plot reveals the variation in the form of principal components (PC) 1 and 2. PC1 and PC2 drawn as Dim1 and Dim2 on horizontal and vertical axis, respectively, accounting for the variation up to 91.4% and 5.8%, respectively. Same genes are placed in plot based on their expression in drought and control conditions.

### 3.8. GO Categorization of SDGs

GO keywords were given to 10 out of 11 drought genes, according to the gene ontology analysis. In biological process classification, the drought genes were involved in the regulation of autophagy, response to auxin, electron transport chain and regulation of transcription from RNA polymerase II promoter. In molecular function categorization, RNA polymerase II transcription factor activity, sequence-specific DNA binding, D-threo-aldose 1-dehydrogenase activity, alditol: NADP+ 1-oxidoreductase activity and cytochrome-c oxidase activity were the most popular terms, especially in the field of cellular components, and Mon1-Ccz1 complex was the most dominant terms (Figure 6 and Table 3).

**Table 3.** SDGs ontology created using the reference annotation of IWGSC wheat genomes. For this purpose, drought protein sequences were annotated using R v3.4.1. Majority of genes ontology annotations were predicted for wheat sequences as given below.

| GO Type | Description   | Negative Log <sub>10</sub><br>( <i>p</i> Adjusted) | ID         | Gene Ratio | Bg. Ratio | <i>p</i> Value | P Adjust | Q Value  | Gene ID  | Count |
|---------|---|--|------------|------------|-----------|----------------|----------|----------|----------|-------|
| BP      | regulation of autophagy   | 2.107492695  | GO:0010506 | 1/4        | 3/7682    | 0.001561       | 0.007807 | 0.001644 | 25270582 | 1     |
|         | response to auxin   | 1.683984604  | GO:0009733 | 1/4        | 25/7682   | 0.012957       | 0.020702 | 0.004358 | 25245708 | 1     |
|         | electron transport chain  | 1.683984604  | GO:0022900 | 1/4        | 31/7682   | 0.016047       | 0.020702 | 0.004358 | 14320    | 1     |
|         | regulation of transcription<br>from RNA polymerase II<br>promoter                       | 1.683984604  | GO:0006357 | 1/4        | 32/7682   | 0.016562       | 0.020702 | 0.004358 | 20334251 | 1     |
| CC      | Mon1-Ccz1 complex   | 3.155214539  | GO:0035658 | 1/5        | 1/7148    | 0.000699       | 0.000699 | #N/A     | 25270582 | 1     |
| MF      | RNA polymerase II<br>transcription factor activity,<br>sequence-specific DNA<br>binding | 2.290980192  | GO:0000981 | 1/4        | 2/7814    | 0.001024       | 0.005117 | 0.001616 | 20334251 | 1     |
|         | D-threo-aldose<br>1-dehydrogenase activity  | 2.290980192  | GO:0047834 | 1/4        | 2/7814    | 0.001024       | 0.005117 | 0.001616 | 31369545 | 1     |
|         | alditol:NADP+<br>1-oxidoreductase activity  | 2.290980192  | GO:0004032 | 1/4        | 3/7814    | 0.001535       | 0.005117 | 0.001616 | 31369545 | 1     |
|         | cytochrome-c oxidase activity   | 2.048275717  | GO:0004129 | 1/4        | 7/7814    | 0.003579       | 0.008948 | 0.002826 | 14320    | 1     |



**Figure 6.** Drought proteins were annotated using reference genome of wheat from IWGSC. Predicted biological mechanisms indicated that auxin-related response, autophagy, electron transport chain and transcription regulations are prominent in this study.

### 3.9. Gene Annotation from PlantRegMap Database

The putative restrictive transcription factor was then found using our SDGs. We assumed that some of the linked genes may have known activities in the fruit dehiscence zone. To test this, we analyzed the smaller 11 gene set by first mapping each of the transcripts to the *Arabidopsis thaliana*, *Dichanthelium oligosanthes*, *Klebsormidium flaccidum*, *Medicago truncatula*, *Trifolium pratense*, *Utricularia gibba*, *Zea mays* and *Zoysia matrella* proteome, in order to find homologous gene loci, and those with E-values under 0.05 were chosen. Ten transcripts in all met this requirement, as shown in Table 4, and were then subjected to proxy analysis, using the most compatible *Arabidopsis thaliana*, *Dichanthelium oligosanthes*, *Klebsormidium flaccidum*, *Medicago truncatula*, *Trifolium pratense*, *Utricularia gibba*, *Zea mays* and *Zoysia matrella* gene locus IDs to conduct Plant Ontology (PO) and network analysis using PlantRegMap, to enable inference of putative gene function and regulatory networks. According to Table 5 linked PO keywords, 10 out of 11 drought gene homologues were found in the PlantRegMap database.

**Table 4.** Searching the homologs of SDGs in annotated genomes of *Arabidopsis thaliana*, *Dichanthelium oligosanthes*, *Klebsormidium flaccidum*, *Medicago truncatula*, *Trifolium pratense*, *Utricularia gibba*, *Zea mays* and *Zoysia matrella*.

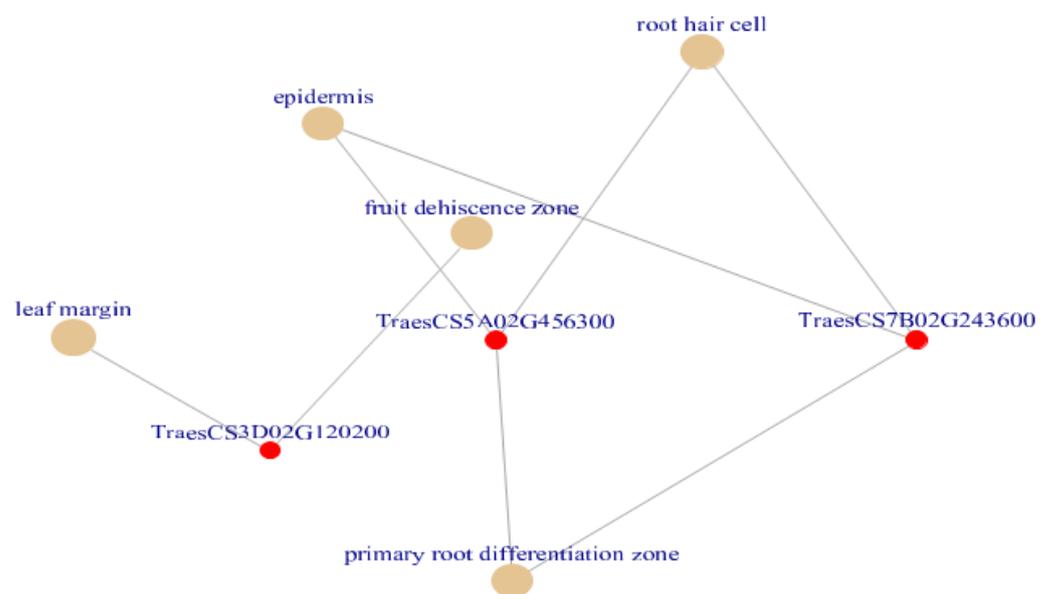
| Drought Genes             | Hit ID                    | Organism                          | Description   | Score | E-Value                |
|---------------------------|---------------------------|-----------------------------------|---|-------|------------------------|
| <i>TraesCS6D02G260700</i> | Zmw_sc01257.1.g00130.1    | <i>Zea mays</i>                   | <i>Zoysia matrella</i> SBP family protein                                   | 110   | $7.00 \times 10^{-6}$  |
| <i>TraesCS6D02G260700</i> | AT1G53160.1               | <i>Arabidopsis thaliana</i>       | <i>Arabidopsis thaliana</i> squamosa promoter-binding protein-like 4        | 278   | $3.00 \times 10^{-29}$ |
| <i>TraesCS6A02G240400</i> | AT2G46680.2               | <i>Arabidopsis thaliana</i>       | <i>Arabidopsis thaliana</i> homeobox 7                                      | 264   | $6.00 \times 10^{-27}$ |
| <i>TraesCS2D02G000200</i> | 678321642                 | <i>Utricularia gibba</i>          | <i>Utricularia gibba</i> bHLH family protein                                | 138   | $3.00 \times 10^{-8}$  |
| <i>TraesCS2D02G000200</i> | AT3G59060.4               | <i>Arabidopsis thaliana</i>       | phytochrome interacting factor 3-like 6                                     | -     | $3.00 \times 10^{-18}$ |
| <i>TraesCS6D02G086600</i> | kfl00432_0070             | <i>Klebsormidium flaccidum</i>    | <i>Klebsormidium flaccidum</i> C3H family protein                           | 709   | $2.00 \times 10^{-78}$ |
| <i>TraesCS6D02G086600</i> | AT1G29560.2               | <i>Arabidopsis thaliana</i>       | <i>Arabidopsis thaliana</i> C3H family protein                              | 101   | $1.00 \times 10^{-4}$  |
| <i>TraesCS6B02G234100</i> | Medtr2g092960.1           | <i>Medicago truncatula</i>        | <i>Medicago truncatula</i> Trihelix family protein                          | 104   | 0.001                  |
| <i>TraesCS6B02G234100</i> | AT3G58630.1               | <i>Arabidopsis thaliana</i>       | sequence-specific DNA binding transcription factors                         | -     | $2.00 \times 10^{-24}$ |
| <i>TraesCS6A02G328700</i> | AT1G70510.1               | <i>Arabidopsis thaliana</i>       | <i>Arabidopsis thaliana</i> KNOTTED-like from <i>Arabidopsis thaliana</i> 2 | 64    | 0.04                   |
| <i>TraesCS1D02G333000</i> | No hits found             |                                   |   |       |                        |
| <i>TraesCS5A02G456300</i> | Zmw_sc03344.1.g00030.1    | <i>Zoysia matrella</i>            | <i>Zoysia matrella</i> ERF family protein                                   | 160   | $1.00 \times 10^{-11}$ |
| <i>TraesCS5A02G456300</i> | AT5G19790.1               | <i>Arabidopsis thaliana</i>       | <i>Arabidopsis thaliana</i> related to AP2 11                               | 229   | $4.00 \times 10^{-22}$ |
| <i>TraesCS3D02G120200</i> | Tp57577_TGAC_v2_mRNA33215 | <i>Trifolium pratense</i>         | <i>Trifolium pratense</i> bHLH family protein                               | 325   | $5.00 \times 10^{-31}$ |
| <i>TraesCS3D02G120200</i> | AT2G22760.1               | <i>Arabidopsis thaliana</i>       | bHLH family protein   | -     | $1.00 \times 10^{-20}$ |
| <i>TraesCS7B02G243600</i> | Do013987.1                | <i>Dichanthelium oligosanthes</i> | <i>Dichanthelium oligosanthes</i> ERF family protein                        | 244   | $1.00 \times 10^{-20}$ |
| <i>TraesCS7B02G243600</i> | AT5G19790.1               | <i>Arabidopsis thaliana</i>       | related to AP2 11   | -     | $2.00 \times 10^{-18}$ |

**Table 5.** Plant ontology created using the reference annotation of *Arabidopsis thaliana* and *Zea mays* genomes. For this purpose, drought protein sequences were checked for homology across the available annotated sequences. Majority of homologs were found and plant ontology annotations were assigned to wheat sequences as given below:.

| ID         | Description                       | Gene Ratio | Bg Ratio  | p Value  | P Adjust | Q Value  | Gene ID  | Count |
|------------|-----------------------------------|------------|-----------|----------|----------|----------|--|-------|
| PO:0003015 | primary root differentiation zone | 2/9        | 42/24102  | 0.000106 | 0.001957 | 0.000309 | <i>TraesCS5A02G456300</i> /<br><i>TraesCS7B02G243600</i> | 2     |
| PO:0005679 | Epidermis                         | 2/9        | 57/24102  | 0.000196 | 0.001957 | 0.000309 | <i>TraesCS5A02G456300</i> /<br><i>TraesCS7B02G243600</i> | 2     |
| PO:0000256 | root hair cell                    | 2/9        | 208/24102 | 0.002564 | 0.015719 | 0.002482 | <i>TraesCS5A02G456300</i> /<br><i>TraesCS7B02G243600</i> | 2     |
| PO:0004707 | fruit dehiscence zone             | 1/9        | 11/24102  | 0.004101 | 0.015719 | 0.002482 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0020128 | leaf margin                       | 1/9        | 16/24102  | 0.00596  | 0.015719 | 0.002482 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0009064 | receptacle                        | 1/9        | 17/24102  | 0.006331 | 0.015719 | 0.002482 | <i>TraesCS6A02G328700</i>                                | 1     |
| PO:0000112 | shoot axis epidermis              | 1/9        | 18/24102  | 0.006702 | 0.015719 | 0.002482 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0000033 | fruit valve                       | 1/9        | 19/24102  | 0.007074 | 0.015719 | 0.002482 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0004006 | mesophyll cell                    | 1/9        | 19/24102  | 0.007074 | 0.015719 | 0.002482 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0006016 | leaf epidermis                    | 1/9        | 23/24102  | 0.008557 | 0.017114 | 0.002702 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0009053 | Peduncle                          | 1/9        | 31/24102  | 0.011518 | 0.019813 | 0.003128 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0020127 | primary root                      | 1/9        | 32/24102  | 0.011888 | 0.019813 | 0.003128 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0006036 | root epidermis                    | 1/9        | 62/24102  | 0.022919 | 0.035259 | 0.005567 | <i>TraesCS3D02G120200</i>                                | 1     |
| PO:0006504 | leaf trichome                     | 1/9        | 81/24102  | 0.029848 | 0.04264  | 0.006733 | <i>TraesCS3D02G120200</i>                                | 1     |

### 3.10. Plant Ontology Annotation Using Annotated Reference Genomes

The gene collection was collectively annotated for plant keywords, some of which we thought were compatible with BP-response to auxin (e.g., as for *TraesCS6A02G328700* with PO- cauline leaf, shoot apex, rib zone, inflorescence meristem, root, seed and plant embryo, vascular leaf, flower, sepal, petal, flower pedicel, gynoecium, receptacle, hypocotyl, collective leaf structure, plant embryo cotyledonary stage, mature plant embryo stage, plant embryo globular stage, plant embryo bilateral stage, petal differentiation and expansion stage and flowering stage); CC-cytosol; MF-oxidoreductase activity (e.g., as for *TraesCS2D02G000200* with PO-cauline leaf, shoot apex, inflorescence meristem, guard cell, leaf lamina base, shoot system, plant embryo, vascular leaf, stamen, carpel, sepal, petal, flower, stem, cotyledon, petiole, hypocotyl, leaf apex, collective leaf structure, vascular leaf senescent stage, plant embryo bilateral stage, LP.12 twelve leaves visible stage, LP.08 eight leaves visible stage, LP.02 two leaves visible stage, LP.10 ten leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, petal differentiation and expansion stage, and flowering stage); CC-integral component of membrane (e.g., as for *TraesCS5A02G456300* with PO- root hair cell, guard cell, primary root differentiation zone, and epidermis) BP-regulation of transcription, DNA-templated; CC-nucleus; MF-DNA-binding transcription factor activity, RNA polymerase II-specific (e.g., as for *TraesCS6A02G240400* with PO- cauline leaf, shoot apex, inflorescence meristem, guard cell, leaf lamina base, fruit, root, shoot system, plant embryo, portion of vascular tissue, vascular leaf, carpel, petal, stamen, flower, sepal, stem, flower pedicel, cotyledon, hypocotyl, collective leaf structure, pollen, vascular leaf senescent stage, plant embryo globular stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, petal differentiation and expansion stage and flowering stage) (Table 5). Full details of enriched GO and PO terms by inference from network neighbors for each of the 11 genes are given in (Figures 6 and 7, and Supplementary Table S2) which also shows the networking of three classes GO terms and PO terms along with SDGs ( $p < 0.05$ ) candidates from clusterProfiler.



**Figure 7.** Cnetplot created using plant ontology enriched terms (Adjusted  $p < 0.05$ ). Out of all enriched terms epidermis, root hair cell, leaf margin, fruit dehiscence zone and primary root differentiation zone terms are plotted by default which was associated with drought in wheat. The terms primary root differentiation zone, epidermis and root hair are governed by the DE genes are *TraesCS5A02G456300* and *TraesCS7B02G243600*.

#### 4. Discussion

Agricultural yields are impacted by drought, an abiotic stress that is a significant environmental barrier to global agriculture. By 2025, grain yields in regions at risk of drought will need to increase by 40% to accommodate nine billion people, the projected world population by 2050 [31]. Advanced molecular breeding or biotechnological methods can be used to grow crops that are more tolerant to environmental stresses [32]. However, the identification of drought-tolerant genes and the deciphering of drought-tolerant mechanisms are the prerequisite for the application of these strategies. In this direction, lots of efforts have been made, and one of the robust approaches is a meta-analysis of large microarray datasets [33].

A wheat reference transcriptome, associated with drought, is generated for a variety of agronomically valuable crops, and the development of global transcriptomics resources for droughts in wheat is missing in research to date. Currently, the information is related to drought stress in wheat impacted by multiple kinds of geographical environments. Here, the catalog of a wheat transcriptome from multiple RNASeq collections, reported by [21–24], has provided an important resource [34] to identify mRNA's potential roles in drought-stress conditions [35]. The analysis revealed eleven drought-responsive (11) genes, among which one (GI: 31369563, *TraesCS3D02G120200*) was common to all microarray datasets, and the other ten genes were common to three of the microarray datasets. The Ensembl Plant characterized the cluster of four genes out of them all (*TraesCS3D02G120200*, *TraesCS6D02G086600*, *TraesCSU02G154600* and *TraesCS6A02G240400*) as “Glycosyltransferase, Probable helicase MAGATAMA 3, Cytochrome c oxidase subunit 2 and D-ZIP I protein”, respectively. However, their homologous study of the remaining uncharacterized genes helped us to annotate them in other species (*Arabidopsis thaliana*, *Dichanthelium oligosanthes*, *Klebsormidium flaccidum*, *Medicago truncatula*, *Trifolium pratense*, *Utricularia gibba*, *Zea mays* and *Zoysia matrella*) and are given in Table 4. These genes have the capacity to control drought stress, so their downstream information can provide a more recent opportunity to agriculture biotechnologists.

Amongst all eleven genes, the one gene (*TraesCS3D02G120200*) found common in the four microarray datasets was stated as the standardized drought-related gene, involved in encoding glycosyltransferase. During abiotic stress conditions, secondary metabolites play an essential role in contending the stress environment because the plant's secondary metabolism is the result of the plant's interaction with its environment. UDP-glycosyltransferases (UGT) are enzymes that add sugars to the secondary metabolites, and thus, are believed to maintain secondary metabolites balance in plants [36]. A study revealed that UGT79B3 and UGT79B2 were involved in the glycosylation of anthocyanin and the lines with high expression of these genes assemble more anthocyanin to cope with drought stress [37]. Another study also confirmed the overexpression of glycosyltransferase in maize makes the plant drought-tolerant [38]. Overexpression of rice glycosyltransferase (UGT83A1) confers protection against salt, drought and cold stress in rice [39]. Thus, the expression level of glycosyltransferase imparts an essential part of drought tolerance in different crops.

We cannot neglect the rest of the ten genes, as they overlapped in three of the datasets used in the current study. Three (*TraesCS6D02G086600*, *TraesCSU02G154600* and *TraesCS6A02G240400*) of which contained gene descriptions in the Ensembl plant. *TraesCS6D02G086600* encodes Probable helicase MAGATAMA 3. RNA and DNA helicases plays important role in several cellular processes of protein protection and turnover. Many mechanisms involving helicases in conferring stress tolerance in plants have been put forth [40]. Overexpression of helicases alleviates abiotic stresses in chilli (*Capsicum annum* L.) [41], and *TraesCSU02G154600* encodes Cytochrome c oxidase subunit 2. In any abiotic stress, plants produce reactive oxygen species (ROS). Cytochrome c oxidase is an ideal antioxidant. None of the studies worked on cytochrome c oxidase in drought tolerance. So, this study provides a future avenue that cytochrome c oxidase might play a role in causing abiotic stress tolerance in plants. *TraesCS6A02G240400* encodes D-ZIP I protein.

Specific transcription factors called homeodomain-leucine zippers (HD-Zip) are crucial in a number of developmental processes and environmental variables [42]. Different studies revealed that the overexpression of HD-Zip 1 confers tolerance to stresses [43]. All the downstream protein products including Probable helicase MAGATAMA 3; Cytochrome c oxidase subunit 2 and D-ZIP I protein are, thus, suggested to be the main reservoirs of drought response in wheat.

SBP family protein was found as a homologue of *TraesCS6D02G260700* in *Zea mays* and *Arabidopsis thaliana*. Squamous promotor-binding protein (SBP)-box genes are exclusively present in plants and encode a family member protein of transcription factors. The study showed that the overexpression of SBP family proteins imparts abiotic tolerance to different plants [44]. bHLH family protein was identified as a homologue of *TraesCS2D02G000200* in *Utricularia gibba* and Phytochrome interacting factor 3-like 6 which is one of the bHLH family proteins was found as a homologue of the same gene in *Arabidopsis thaliana* [45]. Characterize the function of a bHLH transcription factor (AhHLH112) in drought conditions. The study revealed that the AhHLH112 transcription factor resides in the nucleus and was brought on by drought stress. Moreover, the high expression of this gene ameliorates drought stress in transgenic plants in both adult and seedling stages, many studies are also in line with the same results [46]. The trihelix family protein (a DNA-binding protein), the homologue of *TraesCS6B02G234100*, has been reported to impart a role in various developmental processes of plants. Over the last several years, this gene family also takes part in abiotic stress tolerance [47]. High expression of the trihelix gene transcription factor improves drought and salt stress in different plants [48]. KNOTTED-like, from *Arabidopsis thaliana* 2 homologue of *TraesCS6A02G328700* in *Arabidopsis thaliana*, mechanizes processes being a DNA-binding protein as itself. The study revealed that the water-deficit stress would lead to altering the expression of this DNA-binding protein in *Pandanus amaryllifolius* [49]. AP2 (an ERF family protein) was identified as a homologue of *TraesCS5A02G456300* and *TraesCS7B02G243600* in *Zoysia matrella*, *Dichanthelium oligosanthes* and *Arabidopsis thaliana*. PETALA2/ETHYLENE RESPONSIVE FACTOR (AP2/ERF) family transcription factors (AP2/ERFs) are involved in abiotic stresses and respond to multiple hormones [50]. Various AP2/ERFs mutants have identified alterations in abiotic stress response and hormonal sensitivity [51]. Thus, this transcription factor may perform roles in conferring abiotic stress tolerance to plants. This demonstrates that our results could provide further new horizons as a template to make predictions about traits in crops.

To cast light on the biological interpretation of SDGs derived from meta-analysis, the selected SDGs were subjected to GO and KEGG pathway-enrichment analysis and further to plant ontology. Results revealed that the RNA polymerase II transcription factor and primary root differentiation zone annotation controlled by *TraesCS5A02G456300* and *TraesCS7B02G243600* genes were found on top of the list of enriched terms ( $p < 0.05$  &  $Q < 0.05$ ) revealing drought impact on plant growth. Water availability and plant growth strongly correlate as water scarcity affects cell enlargement more than cell division. This results in the inhibition of growth which ultimately led to the reduction of cell wall extensibility and turgor [52]. Results of the present study in the form of drought genes are thereby considered potential targets for getting wheat varieties showing up- or downregulation depending upon the nature of fold changes. The detailed processing of the remaining unreported genes will elucidate the consensus between the drought and molecular basis of RNA polymerase II transcription factor-dependent rooting and leaf development-related mechanisms in wheat.

## 5. Conclusions

Here, we demonstrate that the complete transcriptome-level quantitative investigation of wheat gene expression is now possible without the use of a reference genome or well-established array platforms. Comprehensive bioinformatics research revealed transcriptional pathways that linked to various genotypes and biological interventions, opening up a wealth of new possibilities for fundamental researchers and wheat breeders. The

online repository, such as NCBI, is highly helpful for the construction of the predicted standard drought genes across the world, in addition to applications for resolving fundamental biological concerns, as stated in this work. We think that our strategy is broadly applicable to drought stress and will be perfect for the research of important agronomic qualities that will, undoubtedly, be influenced by the food security agenda in the future. This work will become more widely available as it has been shown to produce reliable reference drought gene groups. In fact, it might end up serving as a systems-based approach's main workhorse for drought response in wheat with scarce public genetic resources. This manuscript specifically establishes the groundwork for a precise molecular-level examination of root growth processes for the wheat crop.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101906/s1>, Table S1: Details of transcriptome data (GSE87325, GSE47090, GSE45563 and GSE70443), Table S2: Details related to GO and PO.

**Author Contributions:** Conceptualization, M.A. (Muhammad Aqeel); Data curation, M.A. (Muhammad Ali), M.M., and W.A.; Formal analysis, Q.M., M.U., A.W., M.M., R.A. and W.A.; Funding acquisition; Investigation, A.W.; Methodology, M.M.; Resources, G.M.A. and M.R.K.; Software, A.W.; Supervision, G.M.A. and M.R.K.; Validation, W.A.; Visualization, Q.M. and M.U.; Writing—original draft, M.A. (Muhammad Aqeel); Writing—review and editing, M.A. (Muhammad Aqeel), S.G., M.U., M.A. (Muhammad Ali), S.M.B.A. and M.R.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NIGAB out of SINO-PAK project and the Researchers Supporting Project (number RSP-2021/115) at King Saud University, Riyadh, Saudi Arabia.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the data is presented in the main text and Supplementary File.

**Acknowledgments:** The authors gratefully acknowledge the research facilities provided by the NIGAB, Pakistan. This research was funded by the Researchers Supporting Project (number RSP-2021/115) at King Saud University, Riyadh, Saudi Arabia. The authors are thankful to AlMaarefa University for their support.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Zulkiffal, M.; Ahsan, A.; Ahmed, J.; Musa, M.; Kanwal, A.; Saleem, M.; Anwar, J.; ur Rehman, A.; Ajmal, S.; Gulnaz, S.; et al. Heat and Drought Stresses in Wheat (*Triticum aestivum* L.): Substantial Yield Losses, Practical Achievements, Improvement Approaches, and Adaptive Mechanisms. *Plant Stress Physiol.* **2021**, *3*. [[CrossRef](#)]
2. Shewry, P.R.; Hey, S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* **2015**, *4*, 178–202. [[CrossRef](#)] [[PubMed](#)]
3. De Flaviis, R.; Mutarutwa, D.; Sacchetti, G.; Mastrocola, D. Could environmental effect overcome genetic? A chemometric study on wheat volatiles fingerprint. *Food Chem.* **2022**, *372*, 131236. [[CrossRef](#)] [[PubMed](#)]
4. Hashmi, H.A.; Belgacem, A.O.; Behnassi, M.; Javed, K.; Baig, M.B. Impacts of Climate Change on Livestock and Related Food Security Implications—Overview of the Situation in Pakistan and Policy Recommendations. In *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa*; Springer: Cham, Switzerland, 2021; pp. 197–239. [[CrossRef](#)]
5. Zhang, J.; Zhang, S.; Cheng, M.; Jiang, H.; Zhang, X.; Peng, C.; Lu, X.; Zhang, M.; Jin, J. Effect of drought on agronomic traits of rice and wheat: A meta-analysis. *Int. J. Environ. Res. Public Health* **2018**, *15*, 839. [[CrossRef](#)] [[PubMed](#)]
6. Bapela, T.; Shimelis, H.; Tsilo, T.J.; Mathew, I. Genetic Improvement of Wheat for Drought Tolerance: Progress, Challenges and Opportunities. *Plants* **2022**, *11*, 1331. [[CrossRef](#)] [[PubMed](#)]
7. Liu, Q.; Hao, Y.; Stebler, E.; Tanaka, N.; Zou, C.B. Impact of Plant Functional Types on Coherence Between Precipitation and Soil Moisture: A Wavelet Analysis. *Geophys. Res. Lett.* **2017**, *44*, 12,197–12,207. [[CrossRef](#)]
8. Panda, D.; Mishra, S.S.; Behera, P.K. Drought Tolerance in Rice: Focus on Recent Mechanisms and Approaches. *Rice Sci.* **2021**, *28*, 119–132. [[CrossRef](#)]

9. Singh, S.; Kumar, A.; Panda, D.; Modi, M.K.; Sen, P. Identification and characterization of drought responsive miRNAs from a drought tolerant rice genotype of Assam. *Plant Gene* **2020**, *21*, 100213. [[CrossRef](#)]
10. Pandey, V.; Shukla, A. Acclimation and tolerance strategies of rice under drought stress. *Rice Sci.* **2015**, *22*, 147–161. [[CrossRef](#)]
11. Naeem, M.; Waseem, M.; Zhu, Z.; Zhang, L. Downregulation of SIGRAS15 manipulates plant architecture in tomato (*Solanum lycopersicum*). *Dev. Genes Evol.* **2020**, *230*, 1–12. [[CrossRef](#)]
12. Sunera; Amna; Saqib, S.; Uddin, S.; Zaman, W.; Ullah, F.; Ayaz, A.; Asghar, M.; ur Rehman, S.; Munis, M.F.H.; et al. Characterization and phytostimulatory activity of bacteria isolated from tomato (*Lycopersicon esculentum* Mill.) rhizosphere. *Microb. Pathog.* **2020**, *140*, 103966.
13. Sadiqi, S.; Hamza, M.; Ali, F.; Alam, S.; Shakeela, Q.; Ahmed, S.; Ayaz, A.; Ali, S.; Saqib, S.; Ullah, F.; et al. Molecular Characterization of Bacterial Isolates from Soil Samples and Evaluation of their Antibacterial Potential against MDRS. *Molecules* **2022**, *27*, 6281. [[PubMed](#)]
14. Scarano, A.; Murmura, G.; Vantaggiato, G.; Lauritano, D.; Silvestre-Rangil, J.; Di Cerbo, A.; Lorusso, F. Delayed expansion of atrophic mandible (deam): A case report. *Oral Implantol.* **2017**, *10*, 190. [[CrossRef](#)] [[PubMed](#)]
15. Mao, H.; Li, S.; Chen, B.; Jian, C.; Mei, F.; Zhang, Y.; Li, F.; Chen, N.; Li, T.; Du, L.; et al. Variation in cis-regulation of a NAC transcription factor contributes to drought tolerance in wheat. *Mol. Plant* **2022**, *15*, 276–292. [[CrossRef](#)]
16. Ma, J.; Tang, X.; Sun, B.; Wei, J.; Ma, L.; Yuan, M.; Zhang, D.; Shao, Y.; Li, C.; Chen, K.-M.; et al. A NAC transcription factor, TaNAC5D-2, acts as a positive regulator of drought tolerance through regulating water loss in wheat (*Triticum aestivum* L.). *Environ. Exp. Bot.* **2022**, *196*, 104805. [[CrossRef](#)]
17. Wu, Z.; Shen, S.; Wang, Y.; Tao, W.; Zhao, Z.; Hu, X.; Yu, P. Genome-Wide Identification and Expression Analysis of the Zinc Finger Protein Gene Subfamilies under Drought Stress in *Triticum aestivum*. *Plants* **2022**, *11*, 2511. [[CrossRef](#)]
18. Pan, Y.; Li, M.; Huang, J.; Pan, W.; Shi, T.; Guo, Q.; Yang, G.; Nie, X. Genome-Wide Identification and Characterization of RNA/DNA Differences Associated with Drought Response in Wheat. *Int. J. Mol. Sci.* **2022**, *23*, 1405. [[CrossRef](#)]
19. Sharma, A.; Gupta, S.; Raina, M.; Kumar, D. Selenium transporters and their role in plant development and stress. In *Cation Transporters in Plants*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 307–336. ISBN 9780323857901.
20. Feng, X.; Ma, Q. Transcriptome and proteome profiling revealed molecular mechanism of selenium responses in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2021**, *21*, 584. [[CrossRef](#)]
21. Aprile, A.; Havlickova, L.; Panna, R.; Marè, C.; Borrelli, G.M.; Marone, D.; Perrotta, C.; Rampino, P.; De Bellis, L.; Curn, V.; et al. Different stress responsive strategies to drought and heat in two durum wheat cultivars with contrasting water use efficiency. *BMC Genom.* **2013**, *14*, 821. [[CrossRef](#)]
22. Peremarti, A.; Marè, C.; Aprile, A.; Roncaglia, E.; Cattivelli, L.; Villegas, D.; Royo, C. Transcriptomic and proteomic analyses of a pale-green durum wheat mutant shows variations in photosystem components and metabolic deficiencies under drought stress. *BMC Genom.* **2014**, *15*, 125. [[CrossRef](#)]
23. Begcy, K.; Walia, H. Drought stress delays endosperm development and misregulates genes associated with cytoskeleton organization and grain quality proteins in developing wheat seeds. *Plant Sci.* **2015**, *240*, 109–119. [[CrossRef](#)] [[PubMed](#)]
24. Kumar, J.; Gunapati, S.; Kianian, S.F.; Singh, S.P. Comparative analysis of transcriptome in two wheat genotypes with contrasting levels of drought tolerance. *Protoplasma* **2018**, *255*, 1487–1504. [[CrossRef](#)] [[PubMed](#)]
25. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)] [[PubMed](#)]
26. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)]
27. Smyth, G.K. Limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*; Springer: New York, NY, USA, 2005; pp. 397–420.
28. Smyth, G.K.; Verbyla, A.P. A Conditional Likelihood Approach to Residual Maximum Likelihood Estimation in Generalized Linear Models. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 565–572. [[CrossRef](#)]
29. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
30. Kassambara, A.; Mundt, F. others Factoextra: Extract and visualize the results of multivariate data analyses. *R Package Version* **2017**, *1*, 337–354.
31. Baldoni, E.; Frugis, G.; Martinelli, F.; Benny, J.; Paffetti, D.; Buti, M. A comparative transcriptomic meta-analysis revealed conserved key genes and regulatory networks involved in drought tolerance in cereal crops. *Int. J. Mol. Sci.* **2021**, *22*, 13062. [[CrossRef](#)]
32. Todaka, D.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front. Plant Sci.* **2015**, *6*, 84. [[CrossRef](#)]
33. Muthuramalingam, P.; Krishnan, S.R.; Pothiraj, R.; Ramesh, M. Global transcriptome analysis of combined abiotic stress signaling genes unravels key players in *Oryza sativa* L.: An in silico approach. *Front. Plant Sci.* **2017**, *8*, 759. [[CrossRef](#)]
34. Gallo, A.; Landi, R.; Rubino, V.; Di Cerbo, A.; Giovazzino, A.; Palatucci, A.T.; Centenaro, S.; Guidetti, G.; Canello, S.; Cortese, L.; et al. Oxytetracycline induces DNA damage and epigenetic changes: A possible risk for human and animal health? *PeerJ* **2017**, *5*, e3236. [[CrossRef](#)] [[PubMed](#)]

35. Buti, M.; Baldoni, E.; Formentin, E.; Milc, J.; Frugis, G.; Schiavo, F.L.; Genga, A.; Francia, E. A meta-analysis of comparative transcriptomic data reveals a set of key genes involved in the tolerance to abiotic stresses in rice. *Int. J. Mol. Sci.* **2019**, *20*, 5662. [[CrossRef](#)] [[PubMed](#)]
36. Lim, E.K.; Bowles, D.J. A class of plant glycosyltransferases involved in cellular homeostasis. *EMBO J.* **2004**, *23*, 2915–2922. [[CrossRef](#)] [[PubMed](#)]
37. Li, P.; Li, Y.-J.; Zhang, F.-J.; Zhang, G.-Z.; Jiang, X.-Y.; Yu, H.-M.; Hou, B.-K. The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *Plant J.* **2017**, *89*, 85–103. [[CrossRef](#)]
38. Li, Y.J.; Li, P.; Wang, T.; Zhang, F.J.; Huang, X.X.; Hou, B.K. The maize secondary metabolism glycosyltransferase UFGT2 modifies flavonols and contributes to plant acclimation to abiotic stresses. *Ann. Bot.* **2018**, *122*, 1203–1217. [[CrossRef](#)]
39. Dong, N.Q.; Sun, Y.; Guo, T.; Shi, C.L.; Zhang, Y.M.; Kan, Y.; Xiang, Y.H.; Zhang, H.; Yang, Y.B.; Li, Y.C.; et al. UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Commun.* **2020**, *11*, 2629. [[CrossRef](#)]
40. Pascuan, C.; Frare, R.; Alleva, K.; Ayub, N.D.; Soto, G. mRNA biogenesis-related helicase eIF4AIII from Arabidopsis thaliana is an important factor for abiotic stress adaptation. *Plant Cell Rep.* **2016**, *35*, 1205–1208. [[CrossRef](#)]
41. Shivakumara, T.N.; Sreevathsa, R.; Dash, P.K.; Sheshshayee, M.S.; Papolu, P.K.; Rao, U.; Tuteja, N.; UdayaKumar, M. Overexpression of Pea DNA Helicase 45 (PDH45) imparts tolerance to multiple abiotic stresses in chili (*Capsicum annuum* L.). *Sci. Rep.* **2017**, *7*, 2760. [[CrossRef](#)]
42. Harris, J.C.; Hrmova, M.; Lopato, S.; Langridge, P. Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytol.* **2011**, *190*, 823–837. [[CrossRef](#)]
43. Cabello, J.V.; Chan, R.L. The homologous homeodomain-leucine zipper transcription factors HaHB1 and AtHB13 confer tolerance to drought and salinity stresses via the induction of proteins that stabilize membranes. *Plant Biotechnol. J.* **2012**, *10*, 815–825. [[CrossRef](#)]
44. Wang, P.; Chen, D.; Zheng, Y.; Jin, S.; Yang, J.; Ye, N. Identification and expression analyses of SBP-box genes reveal their involvement in abiotic stress and hormone response in tea plant (*Camellia sinensis*). *Int. J. Mol. Sci.* **2018**, *19*, 3404. [[CrossRef](#)] [[PubMed](#)]
45. Li, C.; Yan, C.; Sun, Q.; Wang, J.; Yuan, C.; Mou, Y.; Shan, S.; Zhao, X. The bHLH transcription factor AhbHLH112 improves the drought tolerance of peanut. *BMC Plant Biol.* **2021**, *21*, 540. [[CrossRef](#)] [[PubMed](#)]
46. Zhao, Q.; Fan, Z.; Qiu, L.; Che, Q.; Wang, T.; Li, Y.; Wang, Y. MdbHLH130, an Apple bHLH Transcription Factor, Confers Water Stress Resistance by Regulating Stomatal Closure and ROS Homeostasis in Transgenic Tobacco. *Front. Plant Sci.* **2020**, *11*, 543696. [[CrossRef](#)] [[PubMed](#)]
47. Zheng, X.; Liu, H.; Ji, H.; Wang, Y.; Dong, B.; Qiao, Y.; Liu, M.; Li, X. The wheat GT factor TaGT2L1D negatively regulates drought tolerance and plant development. *Sci. Rep.* **2016**, *6*, 27042. [[CrossRef](#)]
48. Xi, J.; Qiu, Y.; Du, L.; Poovaiah, B.W. Plant-specific trihelix transcription factor AtGT2L interacts with calcium/calmodulin and responds to cold and salt stresses. *Plant Sci.* **2012**, *185*, 274–280. [[CrossRef](#)]
49. Amnan, M.A.M.; Aizat, W.M.; Khaidizar, F.D.; Tan, B.C. Drought Stress Induces Morpho-Physiological and Proteome Changes of *Pandanus amaryllifolius*. *Plants* **2022**, *11*, 221. [[CrossRef](#)]
50. Dietz, K.J.; Vogel, M.O.; Viehhauser, A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma* **2010**, *245*, 3–14. [[CrossRef](#)]
51. Xie, Z.; Nolan, T.M.; Jiang, H.; Yin, Y. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Front. Plant Sci.* **2019**, *10*, 228. [[CrossRef](#)]
52. Seleiman, M.F.; Refay, Y.; Al-Suhaibani, N.; Al-Ashkar, I.; El-Hendawy, S.; Hafez, E.M. Integrative effects of rice-straw biochar and silicon on oil and seed quality, yield and physiological traits of *Helianthus annuus* L. grown under water deficit stress. *Agronomy* **2019**, *9*, 637. [[CrossRef](#)]