bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Patterns of gene co-expression under water-deficit treatments and pan-genome occupancy in *Brachypodium distachyon*.

Authors: Rubén Sancho[1,2], Pilar Catalán[1,2,3], Bruno Contreras-Moreira[2,4,5,6], Thomas E. Juenger[7] David L. Des Marais*[8]

Affiliations:

[1] Department of Agricultural and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain

[2] Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Spain

[3] Tomsk State University, Tomsk, Russia

[4] Department of Genetics and Plant Breeding, Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas, Zaragoza, Spain

[5] Fundación ARAID, Zaragoza, Spain

[6] Current address: Ensembl Plants, European Bioinformatics Institute, EMBL-EBI, Hinxton, UK

[7] Department of Integrative Biology, The University of Texas at Austin, Austin, TX. USA

[8] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA. USA

* Corresponding author:

David L. Des Marais. 15 Vassar Street Room 48-325, Cambridge, MA, 02139 (USA). Phone: 617.258.6482. Email: dldesmar@mit.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Natural populations are characterized by abundant genetic diversity driven by a range of different types of mutation. The tractability of sequence complete genomes has allowed new insights into the variable composition of genomes, summarized as a species pan-genome, which demonstrate that many genes are absent from the reference genomes whose analysis has dominated the initial years of the genomic era. Our field now turns towards understanding the functional consequence of these highly variable genomes. Here, we analyzed weighted gene co-expression networks from leaf transcriptome data for drought response in the purple false brome *Brachypodium distachyon* and investigated network topology and differential expression of genes putatively involved in adaptation to this stressor. We specifically asked whether genes with variable "occupancy" in the pan-genome – genes which are either present in all studied genotypes or missing in some genotypes – show different distributions among co-expression modules. Co-expression analysis united drought genes expressed in drought-stressed plants into 9 modules covering 343 hub genes (440 hub isoforms), and genes expressed under controlled water conditions into 13 modules, covering 724 hub genes (911 hub isoforms). We find that low occupancy pan-genes are under-represented among several modules, while other modules are over-enriched for low-occupancy pan-genes. We also provide new insight into the regulation of drought response in *B. distachyon*, specifically identifying one module with an apparent role in primary metabolism that is strongly responsive to drought. Our work shows the power of integrating pan-genomic analysis with transcriptomic data using factorial experiments to understand the functional genomics of environmental response.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Soil water availability is a critical factor determining plant growth, development, and reproduction (Bohnert et al. 1995). Plants are able to cope with and acclimate to a range of soil water contents through the reprogramming of their physiology, growth, and development over time scales ranging from hours to seasons (Chaves et al. 2003); many of these acclimation strategies arise from altered transcriptional profiles (Fisher et al. 2016; Miao et al. 2017). Drought-responsive gene regulatory pathways have been investigated extensively in model plant systems such as *Arabidopsis*, maize, and rice (Hayano-Kanashiro et al. 2009; Nakashima et al. 2009; Nakashima et al. 2014; Janiak et al. 2015; Borah et al. 2017). A clear emerging theme, however, is that diverse species and varieties of plants exhibit diverse stress response mechanisms (Pinheiro and Chaves 2011; Des Marais et al. 2012; Juenger 2013), often controlled by complex regulatory networks. Understanding the genetic control of this phenotypic diversity is a priority for understanding the response of natural populations to climate change, and for designing resilient crop species (Benfey and Mitchell-Olds 2008).

Recent studies have brought attention to the remarkable variation in gene content among plant populations (Gordon et al. 2017; Gao et al. 2019; Alonge et al. 2020; Haberer et al. 2020), reflected in a species' pan-genome. A pan-genome refers to the genomic content of a species as a whole, rather than the composition of a single, reference, individual (Koonin and Wolf 2008). In practice, pan-genomes are estimated by deeply resequencing the genomes of a diversity panel of genotypes, often using a reference genome to aid in final assembly and annotation. In the diploid model grass *Brachypodium distachyon*, genomic analysis of 56 inbred natural "accessions" revealed that the total pan-genome of the species comprised nearly twice the number of genes in any single accession (Gordon et al. 2017). Remarkably, only 73% of genes in a given accession are found in at least 95% of the other accessions – so-called "core genes" (Koonin and Wolf 2008) – suggesting that a large number of genes are unique to subsets of accessions or even to individual accessions. The list of core genes in *B. distachyon* is enriched for annotations associated with essential cellular processes such as primary metabolism. Lower-occupancy genes, or "shell genes," are found in 5-94% of accessions and their annotations are enriched for many processes related to environmental response, including disease resistance genes. Similar patterns have been observed in the pan-genomes of *Arabidopsis thaliana*, barley, and sunflower (Contreras-Moreira et al. 2017; Hübner et al. 2019). The DNA sequence of core genes bear the hallmark of strong purifying selection and are typically expressed at a higher level and in more tissues as compared to shell genes. However, the vast majority of shell genes in *B. distachyon* appear to be functional, as homologs are found in other species' genomes (Gordon et al. 2017).

The preceding observations raise the intriguing possibility that shell genes may represent segregating variation that could be shaped by natural selection and thereby facilitate local adaptation or adaptive responses to a variable environment. Multiple studies in *Arabidopsis thaliana* demonstrate the role of segregating functional gene copies – effectively large-effect mutations – in shaping whole-plant response to the abiotic environment (Monroe et al. 2016; Monroe et al. 2018). The phenotypic effect size of a

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mutation can determine the likelihood that the mutant will become fixed in a population, with large-effect mutations more likely than not conferring deleterious phenotypes that are removed from populations by natural selection (Fisher 1930). The observation that two accessions of *B. distachyon* may vary in the presence or absence of hundreds of functional gene copies begs the question as to how potentially function changing gene deletions escaped the purging effects of purifying selection; pan-genomics requires that we reconceptualize how we interpret "gene loss" as we move beyond a reference-genome view of genome function. In principle, the pleiotropic effect of a mutation can be affected by the number of interacting genes (Jeong et al. 2001); if a gene has relatively few interacting partners then its presence or absence in a particular accession may have a small fitness effect and thus be maintained in populations. Similarly, the efficacy of selection to purge deleterious alleles may be reduced if a gene is only expressed in a subset of environments experienced by a species (Paaby and Rockman 2014). Here, we explore these ideas of functional gene turnover by testing the hypothesis that shell genes and core genes differ in their topological positions in gene co-expression networks.

Gene co-expression networks are now widely used to interpret functional genomic data by assessing patterns of correlation among genes via a threshold that assigns a connection weight to each gene pair (Zhang and Horvath 2005; Langfelder and Horvath 2008). Sets of genes, defined as nodes, with similar expression profiles are assigned to modules by applying graph clustering algorithms (Mao et al. 2009). Nodes in such networks show considerable variation in the extent to which their expression co-varies with other nodes; co-expression networks are generally considered "scale-free" (Guelzim et al. 2002). Modules are often comprised of genes with similar functions (Stuart et al. 2003; Wolfe et al. 2005). High connectivity "hub" nodes (genes) that show a high number of interactions with other genes within a weighted co-expression network are candidates for key players in regulating cellular processes (Albert et al. 2000; Carlson et al. 2006; Dong and Horvath 2007). A such, hub genes might be expected to more often than not be "essential" and thus show pleiotropic effects when mutated or deleted. By contrast, genes with fewer close co-expression relationships are often found to be situated on the periphery of networks and might, therefore, exhibit fewer pleiotropic effects when missing or mutated (Porth et al. 2014; Des Marais, Guerrero, et al. 2017; Masalia et al. 2017). In this context, we hypothesize that pan-genome core genes may be over-represented among co-expression network "hub-genes," as both appear to be involved in core cellular processes and may therefore show deleterious effects when deleted. Conversely, we predict that pan-genome "shell genes" – whose patterns of expression and thereby phenotypic effects are more restricted and condition-specific – will be enriched among lowly connected (non-hub) genes in gene co-expression networks.

Here, we explore the relationship between a plant's pan-genome and its gene co-expression network using *Brachypodium distachyon. Brachypodium* is a small genus of the subfamily Pooideae (Poaceae) that contains ~20 species distributed worldwide (Catalán et al. 2016; Scholthof et al. 2018). The annual diploid species *B. distachyon* is a model for temperate cereals and biofuel grasses (Vogel et al. 2010; Mur et al. 2011; Catalán et al. 2014; Scholthof et al. 2018); a reference genome for one *B. distachyon*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accession, Bd21 (IBI 2010) is now complemented by 54 deeply resequenced natural accessions (Gordon et al. 2017). Recent studies demonstrate the utility of *B. distachyon* and its close congeners for elucidating the evolution and ecology of plant-abiotic interactions, focusing especially on responses to soil drying, aridity, and water use strategy (Manzaneda et al. 2012; Verelst et al. 2013; Manzaneda et al. 2015; Fisher et al. 2016; Des Marais and Juenger 2016; Des Marais, Lasky, et al. 2017; Martínez et al. 2018; Handakumbura et al. 2019; Skalska et al. 2020). In the present study, we first identify and characterize gene co-expression modules associated with response to soil drying. We then test the hypothesis that the occupancy of pan-genes – whether they are part of the shell or core gene sets of the pan-genome – is associated with their connectivity in the *B. distachyon* gene co-expression network. Our work demonstrates the dynamic nature of plant genomes and sets up future work on the diversity and evolution of gene regulatory networks.

MATERIALS AND METHODS

Plant material, experimental design, total-RNA extraction and 3' cDNA tag libraries preparation

Sampling herein follows our earlier work documenting physiological and developmental response of 33 diploid natural accessions of *Brachypodium distachyon* (L.) P. Beauv. to soil drying (Table S1) (Des Marais, Lasky, et al. 2017). The sampled accessions were inbred for more than five generations (Vogel et al. 2006; Filiz et al. 2009; Vogel et al. 2009) and represent the geographic and ecological diversity of *B. distachyon* across the Mediterranean region. Whole genome resequencing data is available for all studied accessions (Gordon et al. 2017).

A total of 264 individual plants from the 33 accessions were grown under two greenhouse conditions, restriction of water (drought, D) and well-watered controls (water, W). We sampled four biological replicates per treatment-accession combination [33 accessions x 4 replicates x 2 treatments (D and W)]. Well-watered plants were watered to field-capacity every second day with fresh water, whereas drought plants were hand watered daily by pipette such that the soil water was reduced by 5% each day (Fig. 1; See Des Marais et al. (2017) for a full description of the growth and treatment conditions).

For each plant, the two youngest, fully expanded leaves of the tallest tiller were excised with a razor blade at the base of the lamina and flash-frozen on liquid nitrogen. Tissue was ground to a fine powder under liquid nitrogen using a Mixer Mill MM 300 (Retsch GmbH). RNA was extracted using the Sigma Spectrum Total Plant RNA kit, including on-column DNase treatment, following the manufacturer's protocol, and quantified using a NanoDrop (Thermo Scientific).

We used a RNA-Seq library protocol (3' cDNA tag libraries with fragment of 300-500 bp) for sequencing on the Illumina HiSeq platform adapted from Meyer et al. (2011). This Tag-Seq method yields only one

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequence per transcript, allowing for higher sequencing coverage per gene as a function of total sequencing effort (Tandonnet and Torres 2017).

Pre-processing of sequences, quantifying abundances of transcripts, normalizing and analysing of batch effects

Sequencing was carried out using an Illumina HiSeq2500 platform (100 bp Single-End (SE) sequencing). Quality control of SE reads was performed with FastQC software. Adapters and low quality reads were removed and filtered with Trimmomatic-0.32 (Bolger et al. 2014). Total numbers of raw and filtered SE reads for each accession and treatment are shown in Table S2.

Quantifying the abundances of transcripts from RNA-Seq data was done with Kallisto v0.43.1 (Bray et al. 2016). To accommodate the library preparation and sequencing protocols (3' tag from fragments of 300-500 bp), pseudoalignments of RNA-Seq data were carried out using as references 500 bp from the 3' tails of the *B. distachyon* 314 v3.1 transcriptome (IBI 2010; http://phytozome.jgi.doe.gov/). We applied an estimated average fragment lengths of 100 bp and standard deviations of fragment length of 20. Estimated numbers of transcripts per million (TPM) were recorded.

Exploratory analysis of the data set and the subsequent filtering and normalization of transcripts abundance steps between samples, and the *in silico* technical replicate step (bootstrap values computed with Kallisto), were conducted with the Sleuth package (Pimentel et al. 2017). A total of 16,386 targets (transcripts/isoforms) were recovered after the normalizing and filtering step using Sleuth package. This program was also used for batch-correction of data and of differentially expressed genes. To account for library preparation batch effects, date of library preparation was included as a covariate with condition variable in the full model.

Co-expression network analysis of normalized transcripts abundance

Co-expression networks for the Drought and Water (control) data sets were carried out using the transcripts per million (TPM) estimates and the R package WGCNA (Langfelder and Horvath 2008). We analysed 16,386 transcripts that were filtered and normalized for 127 and 124 Drought and Water individual samples (individual plants), respectively. After the removal of putative outliers, we ended with 121 Drought and 108 Water samples that were used for network construction.

The same parameters were fitted to the Drought and the Water data sets to construct their respective co-expression networks. The *BlockwiseModules* function was used to perform automatic network construction and module detection on the large expression data set of 16,386 transcripts. Parameters for co-expression network construction were fitted checking different values. We chose the Pearson correlation and unsigned network type, the soft thresholding power 6 (high scale free, R^2>0.85), a

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

minimum module size of 30, and a medium sensitivity (deepSplit = 2) for the cluster splitting. The topological overlap matrix (TOM) was generated using the TOMtype unsigned approach. Module clustering was performed with function *cutreeDynamic* and the Partitioning Around Medoids (PAM) option activated. Module merging was conducted with mergeCutHeight set to 0.30.

Isoform and gene counts were calculated. Isoform counts included all transcripts identified (e.g. Bradi1g1234.1; Bradi1g1234.2; Bradi1g1234.3) and gene counts only included different genes expressed, thus different isoforms from the same gene computed only once to gene counts (e.g. Bradi1g1234.1 and Bradi1g1234.2 count two isoforms but one gene, Bradi1g1234).

Analysis of network features in Drought and Water networks

Fundamental network features (Zhang and Horvath 2005; Dong and Horvath 2007; Horvath and Dong 2008) were assessed separately for the Drought and Water co-expression networks, based on an adjacency matrix calculated with the *fundamentalNetworkConcepts* function of the WGCNA package. We excluded the grey modules in these analyses. The features include Connectivity (for each gene, the sum of connection strengths with the other network genes, indicating how correlated a gene is with all other network genes), Scaled-Connectivity (K = Connectivity/max(Connectivity), which is used for computing the hub gene significance), Clustering-Coefficient (for each node, measures how 'cliquish' its neighbors are, which ranges from 0 to 1), Maximum Adjacency Ratio-MAR (to determine whether a node has high connectivity because of many weak connections or few strong connections), Density (the mean off-diagonal adjacency and is closely related to the mean connectivity), Centralization (1 for a network with star topology and 0 for a network where each node has the same connectivity), and Heterogeneity (the coefficient of variation of the connectivity distribution). Boxplot graphics and Wilcoxon tests of the network features were computed using ggplot2 (Wickham 2009) and ggsignif R packages, and the summary statistics using the summary function in R.

Detection of highly connected nodes (hub genes/ isoforms) within co-expression networks

Three representative descriptors of modules, module eigengene (ME), intramodular connectivity (k_{IM}) and eigengene-based connectivity (k_{ME}; or its equivalent module membership, MM) were calculated using the WGCNA package. Briefly, ME is defined as the first principal component of a given module and is often considered to represent the gene expression profiles within the module. k_{IM} measures how connected, or co-expressed, a given gene is with respect to the genes of a particular module. Thus, intra-modular connectivity is also the connectivity in the subnetwork defined by the module. MM is the correlation of gene expression profile with the module eigengene (ME) of a given module. MM values close to 1 or -1 indicate genes highly connected to the module. The sign of MM indicates a positive or a negative relationship between a gene and the eigengene of the module (Langfelder and Horvath 2010). Genes with

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

absolute MM value over 0.8 were considered "hub genes". Correlations between MM transformed by a power of $\beta = 6$ and k_{IM} were also calculated.

Pan-genome analyses: occupancy of clustered, hub and DE genes across accessions

Because the *B. distachyon* accessions studied herein comprise a subset of those included in the original pan-genome (Gordon et al. 2017) we re-ran the clustering procedures used in our earlier analysis with only the 33 accessions used here. We clustered CDS sequences from the annotated genomes of each of the studied accessions to define core, soft-core and shell genes with the software GET_HOMOLOGUES-EST v03012018 (Contreras-Moreira et al. 2017). This was performed with the OMCL algorithm (-M) and a stringent %-sequence identity threshold (–S 98). The resulting pan-genome matrix was interrogated to identify core genes observed in all 33 accessions, soft-core genes observed in 32 and 31 accessions, and shell genes observed in 30 or fewer. Occupancy was defined as the number of accessions that contain a particular gene model. We tested whether each module showed an excess or deficit of shell genes as compared to genome averages of pan-gene occupancy using a Fisher's Exact Test, as implemented in stats R package.

Enrichment analyses and GO/KEGG annotation of clustered genes

Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations for the *B. distachyon* 314 v.3.1 reference genome were retrieved (http://phytozome.jgi.doe.gov/; IBI 2010). Gene lists were tested for functional enrichments with the PANTHER (Protein ANalysis THrough Evolutionary Relationships) overrepresentation test (http://www.pantherdb.org). The original *B. distachyon* Bd21 v.3.1 gene ids were converted to v.3.0 with help from Ensembl Plants (Howe et al. 2020) to match those in PANTHER16.0 (Mi et al. 2021). Tests were conducted on all genes and on both conditions, drought, and water, applying the Fisher's Exact test with False Discovery Rate (FDR) multiple test correction. This analysis was applied on different data set: all genes, core, soft-core and shell genes for each co-expressed module.

Analysis of Drought versus Water modular structure preservation and comparison between consensus and set-specific modules

Permutation test was performed to check for preservation of the module topology in the Drought (discovery data) and the Water (test data) (Langfelder et al. 2011) networks by running 10,000 permutations using the modulePreservation function of the NetRep (Ritchie et al. 2016) R package with null="all" (include all nodes) for RNA-Seq data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

All test statistics (Module coherence, Average node contribution, Concordance of node contributions, Density of correlation structure, Concordance of correlation structure, Average edge weight and Concordance of weighted degree) were checked; therefore, a module was considered preserved if all the statistics have a permutation test P-value < 0.01. Searching for modules that could play a role in drought response, we focused on Drought modules that were unpreserved in the Water network (P-value > 0.01 in at least 1 of the seven statistics presented in Ritchie et al. (2016).

The Consensus modules (Cons) and the relation between Consensus and Drought (D) and Water (W) modules were performed as described in Langfelder and Horvath (2008).

Annotations and discovery of DNA motifs upstream of genes in the co-expressed modules

The co-expressed genes assigned to modules in the Drought and Water networks were further analyzed with the objective of discovering DNA motifs putatively involved in the co-expression of genes in each module. Motif analysis was carried out using a protocol based on RSAT::Plants (Ksouri et al. 2021). Briefly, this approach allowed us to discover DNA motifs enriched in promoters of co-expressed genes and match them to curated signatures of experimentally described transcription factors. First, -500 (upstream), +200 (downstream) sequences around transcription start sites of the genes in each module and 50 negative controls of equal size were extracted from the *B. distachyon* Bd21 v3.0 (Ensembl Plants version 46) reference genome. Then, peak-motifs (Thomas-Chollier et al. 2012) was used to discover enriched motifs with a 2nd order Markov genomic model, and, finally, GO enrichment was computed. The analyses generated a report with links to similar curated motifs in the database footprintDB as scored with normalized correlation (Ncor) (Sebastian and Contreras-Moreira 2014). For each module, a highly supported DNA motif was selected according to the Ncor, e-value and the number of sites (i.e. putative cis-regulatory elements, CREs) used to compile the motif. The matrix-scan tool (Turatsinze et al. 2008), using a weight threshold set to 70% of the motif length (60% for the small module D8), was used to scan the discovered motifs and identify individual genes within each module harboring putative CREs.

All the protein sequences for each co-expressed module were analyzed using iTAK (Plant Transcription factor & Protein Kinase Identifier and Classifier) online (v1.6) (Zheng et al. 2016) to annotate the transcription factors, transcriptional regulators and protein kinases (classification system by Lehti-Shiu and Shiu (2012)).

Analyses of differentially expressed (DE) isoforms/genes

In order to determine how many isoforms and genes were differentially expressed between the two treatments (D vs W), the two data sets were analyzed through the sleuth_result function (Pimentel et al. 2017). This function computes likelihood ratio tests (lrt) for null (no treatment effect) and alternative

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(treatment effect) models, attending to the full and reduced fitted models. A threshold of significance level of q-value ≤ 1E-6 was fixed to detect DE isoforms. The 50 most differentially expressed genes (25 up-regulated and 25 down-regulated) were classified based on fold-change of the average TPM values between the drought and water treatments.

Data availability

The *Brachypodium distachyon* filtered RNA-seq data were deposited in the ENA (European Nucleotide Archive; https://www.ebi.ac.uk/ena) with the consecutive accession numbers from ERR6133302 to ERR6133575. All supplementary files (Files S1, S2 and S3) and scripts executed to obtain our findings are available at https://github.com/Bioflora/Brachypodium_co_expression.

RESULTS

Modular distribution of the gene co-expression networks

Analysis of the Drought co-expression network identified 9 modules containing a total of 5,020 isoforms (min = 38, max = 2,477 isoforms per module), corresponding to 4,017 genes (min = 27, max = 1,986 genes per module). 11,366 isoforms (8,313 genes) were not clustered in any module (gray or "zero" module) (Fig. S1a; Table 1a). The Water (control) co-expression network showed 13 co-expression modules containing a total of 6,711 isoforms (min = 48, max = 1,866 isoforms per module), corresponding to 5,447 genes (min = 40, max = 1,439 genes per module). 9,675 isoforms (6,934 genes) were not clustered in any module (gray or "zero" module) (Fig. S1b; Table 1b). Hereafter, modules identified in the Drought RNA-Seq dataset will be labeled with the prefix "D" while those modules identified in the Water (control) dataset will have the "W" prefix.

The modular distribution of Drought and Water (control) co-expression networks showed differences both in the number and the size of the modules. In the Drought network, 69.4% of the isoforms (67.4% of the genes) were not clustered within any module (gray or "zero" module). The largest Drought module contained 15.1% of the isoforms (16.1% of the genes) whereas two modules clustered over 4-6% of the isoforms and genes, and the rest modules each clustered ≤ 2% of isoforms and genes (Fig. S1a; Table 1a). By contrast, in the Water network, 59.0% of the isoforms (56.0% of the genes) were not clustered within any module (grey or "zero" module), the largest module contained 11.4% of the isoforms (11.6% of the genes), six modules clustered over >2-6% of the isoforms and genes for each module, and other six modules clustered ≤ 2% of the isoforms and genes each (Fig. S1b; Table 1b).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hub nodes of the Drought and Water networks

Nodes with absolute MM values above 0.8 are inferred to be the most connected nodes or putative hub nodes for their respective co-expression modules. Hub nodes were detected in both the Drought (Table 1a) and the Water (Table 1b) networks. A total of 440 (2.7%) hub nodes from 343 (2.8%) hub genes in the Drought network, and 911 (5.6%) hub nodes from 724 (5.8%) hub genes in the Water network were detected. Roughly, twice the per-module fraction of hubs was detected in the Water network (5.6% hub nodes/5.8% hub genes; Table 1b) comparing to the Drought network (2.7% hub nodes/2.8% hub genes. Table 1a). The Drought and Water networks have been represented according to their hub nodes and connections (Fig. S1c, d).

Modular connectivity, correlation between modules and module membership in the co-expression networks

Relationships among modules within each network were established using module eigengene (ME) clustering (Fig. S2a, b) and their correlations (Table S3). Intra- and inter-modular connectivity was determined according to the difference between intra-modular connectivity (k_{IM}) and the connectivity out of each module (k_{out}), computed as the difference (k_{diff}) between total connectivity and intra-modular connectivity. k_{diff} values were calculated for all isoforms (nodes) within each module. The largest Drought modules showed a positive (D1, D2, D3, D4 and D5) or slightly negative (D6 and D7) mean k_{diff}, while the smallest modules had more negative values (id: 8 and 9) (Table 1a). Negative k_{diff} values for a node indicate that connectivity out of the module is higher than intra-modular connectivity. Similarly, the Water modules showed positive mean k_{diff} in the bigger modules (W1, W3, W5, W6, W7 and W9) or slightly negative (W4), with the exception of one large and one intermediate modules (W2 and W8) with negative values, similar to the smallest modules (Table 1b). High positive linear correlations between Module Membership (MM) transformed by a power of $\beta = 6$ and k_{IM} were recovered in both Drought (Fig. S3a) and Water (Fig. S3b) networks, thus validating the criterion of high MM (>0.8) for selecting hub genes. Collectively, these results suggest that modules of both Drought and Water networks were statistically well-supported.

Fundamental features of the Drought and Water networks

Fundamental network features were computed and compared between the Drought and the Water networks, with the exclusion of the gray modules, which comprise genes without clear co-expression relationships (Table 2; Fig. S4). While the Water network showed overall slightly higher though non-significant values of connectivity than the Drought network (Wilcoxon test, P-value > 0.05), the Drought network showed significantly higher values of scaled connectivity than the Water network (Table 2; Fig. S4). This difference appears to be due to the Water network exhibiting a very high maximum connectivity

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

value, more than twice the value observed for the Drought network. Moreover, the Water network had higher median and mean values, and lower maximum values of clustering coefficient than the Drought network, showing a significantly higher cohesiveness of the neighborhood of a node than the Drought network (Wilcoxon test, P-value < 0.001). The MAR parameter also showed significantly lower mean values in the Drought than in the Water network (Table 2; Fig. S4). The Water network had higher values of centrality and heterogeneity. By contrast, the lower density of the Water network compared to the Drought network could be a consequence of the lower correlation of the genes co-expressed in the Water network (Table 2). The differences between the features of the networks appear to be conditioned by the high number of hub genes present in the Water network compared to those detected in the Drought network. These hub genes present higher and more variable values of connectivity in the former than in the second network.

Pan-genome analyses: occupancy of all clustered and hub genes

We re-estimated the *B. distachyon* pan-genome following the methods of Gordon et al. (2017) using only the 33 accessions for which RNA-Seq data was generated herein. This revised pan-genome subset contains 34,310 pan-genome clusters (hereafter "pan-genes") which can be classified by the number of accessions with a gene model represented in a given cluster, or occupancy. We found 16,057 (46.8%) core gene clusters with at least one member in every accession. We analyzed these core genes along with the 5,642 (16.4%) clusters from the soft-core pan-genome (occupancy in 31 or 32 accessions) to account for gene annotation errors and uncertainty with orthology assignments. In contrast, there are 12,611 (36.8%) shell genes (genes found in fewer than 31 accessions). Of the 34,310 pan-genes, 12,137 (8,426 (69.4%) core, 1,869 (15.4%) soft-core and 1,842 (15.2%) shell genes) were represented by sequenced RNA tags, after filtering and normalizing steps, in our dataset. These results are consistent with prior studies which find that pan-genome core genes are frequently more highly expressed (Gordon et al. 2017; Gao et al. 2019; Tao et al. 2021) and therefore more likely to be sampled in RNA-Sequencing libraries.

The distribution of pan-gene occupancy varied considerably among modules in both Drought and Water networks (Table 3a,b; Fig. 2a,b). Most modules in both the Drought and Water co-expression networks showed ratios of shell genes consistent with the genome averages (Fisher's Test; Table 3a,b), with notable exceptions. Among Drought modules, module D8 shows a considerable excess of shell genes (63.2% of the total genes in this module are shell genes). Conversely, modules D2 (10.0% shell) and module 3 (11.2% shell) each show a deficit of shell genes. Water co-expression modules showed considerably more variation in pan-gene occupancy. Water modules W4 (27.8%), W9 (21.8%), W11 (25.5%), W12 (27.5%), and W13 (60.0%) all show an excess of shell genes relative to the genome averages, though W11, W12, and W13 are very small modules (fewer than 60 genes each). It is notable that most large modules in the water network, with the exception of W4, show significant deficits of shell genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We next investigated the proportion of putative hub genes (i.e., those nodes with module membership greater than 0.8) in each module that were members of the shell gene sets (Table 3c,d; Fig. 2c,d). Given the comparatively small number of these genes, greater departures from genome averages are required to reach statistical significance. Among the Drought modules, module D5 is exceptional in exhibiting no shell genes among its 25 putative hub nodes. Conversely, Drought module D7 (five of nine hub genes are in the shell) and module D8 (seven of eight hub genes are in the shell) show apparent excess of shell genes among their hub genes (Table 3c; Fig. 2c). Among putative hub genes found in the Water network, two modules stand out as having an exceptionally high proportion of shell genes. 33 of 42 predicted hub genes (78.6%) in module W4 and 6 of 9 hub genes (66.7%) in module W12 are shell genes. Shell genes are under-represented among the hub genes of several Water modules, including module W6 (3.6%) and module W7 (6.1%) (Table 3d; Fig. 2d).

Preservation and correspondence of Drought and Water networks

We tested the hypothesis that some gene co-expression modules are only observed under one of our two treatment conditions using two approaches. First, a permutation test was performed using NetRep (Ritchie et al. 2016) to test for the preservation of module topology in the Drought versus the Water networks, defined such as discovery and test data set respectively. This test computed seven topological statistics on each module and condition (drought vs water), quantifying the replication of the relationship structure between nodes composing each module under the null hypothesis that the module of interest is not preserved. All Drought and Water modules were topologically preserved according to the seven NetRep statistics (permutation test P-values < 0.01). We further tested for correspondence between Drought (D) and Water (W) set-specific and Drought-Water Consensus (Cons) co-expression modules using WGCNA (Fig.3a,b). Consensus modules are those shared by two or more networks (Langfelder and Horvath 2007). We found that several Drought specific modules were comprised of no or very few isoforms included in the Consensus modules (Fig. 3a). Thus, the Drought modules D5 (green), D7 (black) and D9 (magenta) did not show a clear correspondence with consensus modules using the protocol implemented in WGCNA. The seemingly conflicting results between NetRep – which identified module preservation between Drought and Water modules – and WCGNA – which identified three modules which were not preserved – reflects the different sensitivities of these two approaches to changing co-expression relationships within modules. One possible interpretation is that all modules are preserved in a broad sense between treatments, but that nodes within modules D5, D7, and D9 may have slightly different co-expression relationships than are observed in the corresponding modules in the Water treatment. We also found Water specific modules (Fig. 3b), W1 (turquoise), W6 (red), W8 (pink) and W12 (tan), did not overlap with the Consensus modules or only the gray module (non-co-expressed). The largest W1 module mainly overlapped with the gray consensus module, similar to W6. The smaller modules W8 and W12 did not overlap with any consensus module.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enrichment analysis of clustered genes with a pan-genomic perspective

Gene Ontology (GO) biological process enrichment was estimated for all co-expressed genes in modules found in each co-expression network. Additionally, GO enrichment was computed for core, soft-core, and shell genes within each module.

Seven of the nine modules detected in the Drought network and eleven of the thirteen modules in the Water network showed a significant GO term enrichment (Table 4a,b; Supplementary file S1). Both networks had modules enriched for biological processes such as nitrogen, amide and peptide metabolic and biosynthetic processes, photosynthesis, organonitrogen compound metabolic process, location/transport and phosphorylation, small molecule metabolic process, and nucleobase-containing compound and nucleic acid metabolic processes. We inferred the possible equivalent modules between the drought and water network attending to the common consensus module with which they overlap and comparing the GO enrichment (Table 4) of the Drought and Water modules (Fig. 3c). Thus, the D1 and W2 modules matched Cons1 and they are GO enriched in nitrogen, amide and peptide metabolic and biosynthetic processes. The D2 and W10, matched Cons4, are involved in the photosynthesis and the D3 and W3, matched Cons3, are GO enriched in processes of transport and locations of compounds. The D4 and W9 are involved in nucleic acid metabolic processes, and D9 and W13 in nitrogen, amide and peptide biosynthetic and metabolic processes. However, some modules showed a biological process unique to one of the two co-expression networks. For instance, module D5 is enriched in genes involved in protein folding, response to heat, temperature and abiotic stimulus. No modules in the water network showed these enrichments. Module W6 is enriched in genes predicted to be involved in cell wall organization or biogenesis; again, no modules in the drought network showed these enrichments.

As expected, core and soft-core genes were mostly enriched in the same GO terms as their respective modules (Table 4a,b). Similarly, shell genes were enriched with genes involved in the same biological process. However, the most significant GO term of the shell genes of the Drought module D1 was photosynthesis, which was not found to be significant among core genes. The shell genes of the D8 module were enriched in genes involved in photosynthesis, nitrogen, amide and peptide biosynthetic and metabolic processes, while core gene sets did not show any significant enrichments (Table 4a,b).

Regulatory motifs of genes in the Drought and Water modules

We detected statistically over-represented sequence motifs upstream of the genes in several modules; these motifs represent putative *cis*-regulatory elements (CREs) in proximal promoters of co-expressed genes in the Drought (Table 5a; Fig. S5a) and Water (Table 5b; Fig. S5b) networks. The results are summarized in Tables 5a,b support the hypothesis that genes in the same module share conserved regulatory architecture. Calmodulin-binding CREs are enriched in the D4 (associated with nucleobase-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

containing compound, heterocycle and nucleic acid metabolic processes GO terms) and W9 modules (also enriched for nucleobase-containing compound GO processes, among others). We also observed enriched motifs in treatment-specific modules. For example, the proximal promoters of 27.8% of the genes in module D5 contain CREs similar to those bound by transcription factor B-3 (HSFB3) (Table 5a) (Scharf et al. 2012), known to regulate heat shock responses in Arabidopsis (Nover et al. 2001; Bechtold et al. 2013; Guo et al. 2016).

A generally low but variable proportion of genes harboring putative CREs in their proximal promoters was detected in each module. The drought (Table 5a) and water (Table 5b) modules showed between 3.5-54.2% and 0.1-23% of genes with the predicted CREs, respectively. The occupancy and hub gene enrichments were analysed in the modules with at least 10 detected genes with putative CREs. These were predominantly core genes (>50%) in both networks, and between 7-22% in the Drought network, and 0-27.3% in the Water network, of all promoters with putative CREs were from shell genes. A variable proportion (9-23.4% in D network, and 9.3-41.4% in W network) of genes with putative CREs were hubs (Tables 5a,b).

Additionally, all the protein sequences for each module were analysed to annotate the transcription factors (TF), transcriptional regulators (TR) and kinases (Tables S4a,b). For example, only TR were annotated in the D5 module and they are involved, according to the GO-biological processes, in response to abscisic acid, response to heat stress, response to water deprivation, defense response, zinc ion binding, and chromatin binding and metal ion binding (Table S4a).

Functional analysis of differentially expressed isoforms

Gene expression analysis identified 3,489 genes and 4,941 isoforms as being differentially expressed (DE) between the water and drought treatments (Supplementary files S2; S3). Of these isoforms, 3,591 (72.7%) were upregulated and 1,350 (27.3%) were downregulated under drought comparing to water (control) condition (Table S5a,b; Supplementary file S3).

The 74.6%, 16.2% and 8.7% of upregulated isoforms in the drought condition were core, soft-core, and shell isoforms, respectively; 0.5% of upregulated DE genes did not match with the pan-genome (Table 6). Regarding the downregulated isoforms in the drought condition, 70.4%, 19.6% and 9.6% were core, soft-core, and shell isoforms, respectively; 0.4% of downregulated DE genes did not match with the pan-genome matrix (Table 6). 59.3% of DE isoforms (58.7% of DE genes) and 47.4% DE isoforms (44.6% DE genes) of Drought (Table S5a) and Water (Table S5b) networks, respectively, were not assigned to any modules (i.e., they are members of the gray or zero module). Among the DE isoforms, 58.9% and 41.7% of them corresponded to hub nodes of the Drought (Table S6a) and Water modules, respectively (Table S6b). Five of the nine Drought co-expression modules had a predominance (>50%) of upregulated DE isoforms except for the large modules D2 and D3 and the small modules D8 and D9 (Table S5a).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Similarly, among Water modules, only one large module (W4) and three small modules (W10, W11 and W13) had a predominance of downregulated DE isoforms in the drought condition compared to the water condition (Table S5b).

The 50 isoforms of the most differentially expressed genes (DEG), 25 most up-regulated and 25 most down-regulated, according to the fold-change difference between dry and wet, were analyzed separately. Interestingly, 21 of the 25 most strongly upregulated genes did not cluster with other genes in the drought co-expression network (i.e. were members of the grey D0 module; Table S7), while the majority of these strongly upregulated genes did cluster in a module in the water network. The 25 genes most strongly upregulated by drought show a range of predicted functions (Table S7) and include two predicted dehydrins, two ABA-associated proteins, and two lipid transfer proteins (LTPs). LTPs were among the mostly highly induced transcripts in an *Arabidopsis thaliana* experiment that imposed very similar soil drying conditions to those imposed in the present study (Des Marais et al. 2012). The 25 most strongly downregulated genes show markedly different patterns than the most strongly upregulated genes. Most of the genes cluster in the D2 drought module and the W10 module and several genes are hub genes in these two modules. Both of these modules are enriched for genes involved in photosynthesis (Table 4) and, indeed, the gene descriptions for many of these genes suggest associations with the light reactions of photosynthesis. Additionally, all the isoforms co-expressed in the putative drought-response module D5 were analyzed to check their relationships with the pan-genome, DEGs and GO enrichments (Table S8).

DISCUSSION

Large scale transcriptome data sets have been used to construct co-expression networks for gene and gene regulation discovery in model plant systems and crops (Aoki et al. 2007; Aoki et al. 2016; Masalia et al. 2017; Miao et al. 2017). The co-expression network approach further allows testing hypotheses on gene functions, from their connections with other functionally known genes classified in the same modules (Mochida et al. 2011), and on links between signalling pathways and phenotypic response to environmental stress (Des Marais et al. 2012). Gene networks operate in different biological contexts; an important proportion of the genetic interactions within a network have been demonstrated to be condition-specific (He and Maslov 2016). Our system-level approach allowed us to construct a drought-responsive gene co-expression network from leaf tissue transcriptome profiles of *B. distachyon* accessions and to identify modules of putatively co-regulated genes within it (Fig. S1). Drought response mechanisms consist of highly complex interactions of several metabolic processes, as observed previously in a range of grass species (e. g., barley, Mochida et al. 2011; rice, Yu et al. 2017; maize, Miao et al. 2017); however there is still a considerable gap in the knowledge of relationships between drought response genes and developmental signalling (Miao et al. 2017).

Regulatory control of *Brachypodium* response to soil drying

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B. distachyon is an annual species native to seasonally dry environments in the Mediterranean, where it has likely evolved mechanisms to tolerate short-term soil drying during the growing season as well as unpredictably timed end-of-season drought (López-Álvarez et al. 2015). Several past studies have identified mechanisms of response to soil drying comprising transcriptomic, metabolic, physiological, and developmental plasticity (Luo et al. 2011; Bertolini et al. 2013; Verelst et al. 2013; Gordon et al. 2014; Priest et al. 2014; Manzaneda et al. 2015; Chen et al. 2016; Fisher et al. 2016; Des Marais et al. 2016; Ruíz et al. 2016; Des Marais, Lasky, et al. 2017; Handakumbura et al. 2019), as well as considerable genetic diversity of response (GxE; Des Marais, Lasky, et al. (2017); Handakumbura et al. (2019)). Priest et al. (2014) provided the first transcriptomic assessment of response to drying, exposing the Bd21 accession to a simulated severe drying stress by removing plants from soil to desiccate on a lab benchtop. These authors observed a strong transcriptional signature of down-regulated photosynthesis, cell division, and cell growth. Subsequent work imposing a more gradual soil drying stress in Bd21 found the opposite pattern, directly observing sustained cell division and transcriptomic patterns of specific alterations to central metabolism, rather than outright downregulation (Verelst et al. 2013). Indeed, studies imposing moderate drying on diverse *B. distachyon* accessions revealed increased leaf mass per area and greater root biomass in several accessions in response to drying (Des Marais, Lasky, et al. 2017; Handakumbura et al. 2019), both of which require considerable investment of carbohydrates. We do, however, observe several strongly downregulated genes with annotated functions related to the light reactions of photosynthesis as well as RuBisCO assembly and function (Table S7).

How can we reconcile these transcriptional signatures of reduced photosynthesis with the observation that carbohydrate-intensive processes like root growth continue under drying? The effects soil drying on photosynthesis are complex, and that the reduction of internal leaf CO_2 (c_i) caused by stomatal closure can have strong effects on the redox status of cells (Pinheiro and Chaves 2011). As the Calvin Cycle reduces available CO_2, it is also a strong sink for energy captured by the photosystems. As this sink is lowered by decreased c_i, continued high irradiance can lead to increased expression of photoprotective mechanisms and decreased expression of photochemistry as cells try to protect themselves from excess energy (Demmig-Adams and Adams 1996). As a result, studies of soil drying responses often observe decreased activity of the photosystems and increased expression of, for example, photorespiration or other energy sinks (Wingler et al. 1999). While we have no direct measurements of photorespiration or the quantum yield of photosystem II in the current study, our observation of decreased expression of photosystem transcripts is consistent with these mechanisms.

In light of this past evidence for an important role of photosynthesis and primary metabolism in drying response, we focus here on Drought module 5 (D5). Drought module 5 showed a low correlation with the Consensus modules (Fig. 3a), consistent with the hypothesis that the genes in this module are involved in regulating plant response to drought stress. Its co-expressed genes, both core and soft-core, are involved in protein folding, response to heat, temperature and abiotic stimulus (Table 4a). All DE hub nodes of D5 were upregulated in the drought condition comparing to water condition (Table S8) and they were mainly

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

core, and soft-core to a lesser extent. The molecular chaperones, especially heat shock proteins (HSPs), were predominantly annotated in both co-expressed and DE hub genes, but also in the other upregulated DE and co-expressed genes of this module (Table S8). Related to the presence of chaperones, the annotated DNA motifs concurs with the GO enrichment of the protein folding and the Heat stress transcription factor B-3 (Table 5a).

Topological position of pan-genes

Pleiotropy can have a strong effect on the rate of evolution of genes and the roles that functional gene variants might play in evolutionary change. Pleiotropy is often correlated with the position of a gene or protein in biochemical and gene regulatory networks (Jeong et al. 2001; Erwin and Davidson 2009), which are now readily inferred from high dimensional datasets such as genome-wide gene co-expression networks. In the current study, we consider the case of potentially large-effect mutations – segregating gene copies identified from a grass pan-genome – and ask whether such genes are unevenly distributed in gene co-expression networks. Focusing on the Water (control) environment, we find that shell genes are under-represented among the lists of genes in five of the six largest (in terms of total number of genes) co-expression modules (Table 3b). These modules are enriched for GO terms comprising essential processes such as protein synthesis, primary metabolism, various processes related to phosphorus metabolism and signaling, and cell wall organization (Table 4b). Moreover, shell genes are lowly represented among module hub genes (diagnosed as those whose expression most highly correlated with the module as a whole, and thus possibly the most topologically connected among genes in a module) in these five Water modules (Tables 3c,d). Collectively, these results support the hypothesis that core pan-genome genes are centrally located in gene co-expression network and involved in biological processes likely to be under strong purifying selection.

Water module 4, comprising 590 genes, is the only large water module that is statistically enriched for shell genes (Table 3b). Shell genes in Water module 4 are enriched for a range of GO terms including processes related to photosynthesis (Table 4b). In general, lists of shell genes in modules do not tend to have strong GO enrichments, perhaps owing to the fairly small numbers of these gene lists. Interestingly, among the Drought modules, the only module for which shell genes do have an enrichment (D8; Table 3a) includes GO terms associated with photosynthesis (Table 4a). Shell genes represent genes found in some sampled accessions but missing in others, suggesting that *B. distachyon* may harbor genetic diversity in molecular pathways related to photosynthesis. Previously, we demonstrated significant genetic variation among these same *Brachypodium* accessions for leaf carbon content, leaf C:N ratios, and water use efficiency (WUE; Des Marais, Lasky, et al. (2017)). Among these, WUE was significantly associated with principal components summarizing climate diversity; we hypothesize that some of the segregating variation in photosynthesis gene presence/absence may be involved in local adaptation to climate.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

We thank R. Hopkins, E. Sukamtoh, J. Bonette, and B. Whitaker for assistance with data collection. This work was supported by the USDA (NIFA-2011-67012-30663) to D.L.D., NSF (IOS-0922457) to T.E.J., and the Spanish Ministry of Science and Innovation CGL2016-79790-P and PID2019-108195GB-I00 and University of Zaragoza UZ2016_TEC02 grant projects, and the Joint Genome Institute DE-AC02-05CH11231 contract to P.C. R.S. was funded by a Mineco FPI PhD fellowship, Mineco and Ibercaja-CAI mobility grants and Instituto de Estudios Altoaragoneses grant. B.C.M. was funded by Fundación ARAID. PC and RS were partially funded by a European Social Fund/Aragón Government Bioflora grant.

REFERENCES

Albert R, Jeong H, Barabási A-L. 2000. Error and Attack Tolerance of Complex Networks. Nature 406:378–382.

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell 182:145-161.e23.

Aoki K, Ogata Y, Shibata D. 2007. Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. Plant Cell Physiol. 48:381–390.

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T. 2016. ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. Plants Cell Physiol. 57:e5(1-9).

Bechtold U, Albihlal WS, Lawson T, Fryer MJ, Sparrow PAC, Richard F, Persad R, Bowden L, Hickman R, Martin C, et al. 2013. Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. J. Exp. Bot. 64:3467–3481.

Benfey PN, Mitchell-Olds T. 2008. From Genotype to Phenotype: Systems Biology Meets Natural Variation. Science (80-.). 320:495–497.

Bertolini E, Verelst W, Horner DS, Gianfranceschi L, Piccolo V, Inzé D, Pè ME, Mica E. 2013. Addressing the role of microRNAs in reprogramming leaf growth during drought stress in Brachypodium distachyon. Mol. Plant 6:423–443.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bohnert HJ, Nelson DE, Jensenayb RG. 1995. Adaptations to Environmental Stresses. Plant Cell 7:1099–1111.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borah P, Sharma E, Kaur A, Chandel G, Mohapatra T, Kapoor S, Khurana J. 2017. Analysis of drought-responsive signalling network in two contrasting rice cultivars using transcriptome-based approach. Sci. Rep. 42131:1–21.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34:525–527.

Carlson MRJ, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics 7:1–15.

Catalán P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LAJ, Pecchioni N, Rasmussen SK, Vogel JP, et al. 2014. Update on the genomics and basic biology of Brachypodium. Trends Plant Sci. 19:414–418.

Catalán P, López-Alvarez D, Díaz-Pérez A, Sancho R, López-Herranz ML. 2016. Phylogeny and Evolution of the Genus Brachypodium. In: Vogel JP, editor. Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models. Springer. p. 9–38.

Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought — from genes to whole plant. Funct. Plant Biol. 30:239–264.

Chen L, Han J, Deng X, Tan S, Li Lili, Li Lun, Zhou J, Peng H, Yang G, He G, et al. 2016. Expansion and stress responses of AP2/EREBP superfamily in Brachypodium Distachyon. Sci. Rep. 6:1–14.

Contreras-Moreira B, Cantalapiedra CP, García-Pereira MJ, Gordon SP, Vogel JP, Igartua E, Casas AM, Vinuesa P. 2017. Analysis of Plant Pan-Genomes and Transcriptomes with GET _ HOMOLOGUES-EST, a Clustering Solution for Sequences of the Same Species. Front. Genet. 8:1–16.

Demmig-Adams B, Adams WW. 1996. The role of xanthophyll cycle carotenoids in the protection of photosynthesis. Trends Plant Sci. 1:21–26.

Dong J, Horvath S. 2007. Understanding network concepts in modules. BMC Syst. Biol. 1:1–20.

Erwin DH, Davidson EH. 2009. The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. 10:141–148.

Filiz E, Ozdemir BS, Budak F, Vogel JP, Tuna M, Budak H. 2009. Molecular, morphological, and cytological analysis of diverse Brachypodium distachyon inbred lines. Genome 52:876–890.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fisher LH, Han J, Corke FM, Akinyemi A, Didion T, Nielsen KK, Doonan JH, Mur LA, Bosch M. 2016. Linking Dynamic Phenotyping with Metabolite Analysis to Study Natural Variation in Drought Responses of Brachypodium distachyon. Front. Plant Sci. 7:1–15.

Fisher RA. 1930. The genetical theory of natural selection. Oxford Clarendon Press

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, et al. 2019. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat. Genet. 51:1044–1051.

Gibson G. 2016. On the Evaluation of Module Preservation. Cell Syst. 3:17–19.

Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin A, Schackwitz W, Tyler L, et al. 2017. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nat. Commun. 8.

Gordon SP, Priest H, Des Marais DL, Schackwitz W, Figueroa M, Martin J, Bragg JN, Tyler L, Lee CR, Bryant D, et al. 2014. Genome diversity in Brachypodium distachyon: Deep sequencing of highly diverse inbred lines. Plant J. 79:361–374.

Guelzim N, Bottani S, Bourgine P, Képès F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. Nat. Genet. 31:60–63.

Guo M, Liu JH, Ma X, Luo DX, Gong ZH, Lu MH. 2016. The plant heat stress transcription factors (HSFS): Structure, regulation, and function in response to abiotic stresses. Front. Plant Sci. 7.

Haberer G, Kamal N, Bauer E, Gundlach H, Fischer I, Seidel MA, Spannagl M, Marcon C, Ruban A, Urbany C, et al. 2020. European maize genomes highlight intraspecies variation in repeat and gene content. Nat. Genet. 52:950–957.

Handakumbura PP, Stanfill B, Rivas-Ubach A, Fortin D, Vogel JP, Jansson C. 2019. Metabotyping as a Stopover in Genome-to-Phenome Mapping. Sci. Rep. 9:1–12.

Hayano-Kanashiro C, Calderón-Vázquez C, Ibarra-Laclette E, Herrera-Estrella L, Simpson J. 2009. Analysis of Gene Expression and Physiological Responses in Three Mexican Maize Landraces under Drought Stress and Recovery Irrigation. PLoS One 4:e7531.

He F, Maslov S. 2016. Pan- and core- network analysis of co-expression genes in a model plant. Sci. Rep. 6:1–11.

Horvath S, Dong J. 2008. Geometric Interpretation of Gene Coexpression Network Analysis. PloS Comput. Biol. 4:e1000117.

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, et al. 2020. Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res. 48:D689–D695.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, Lee JS, Baute GJ, Owens GL, Grassa CJ, et al. 2019. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nat. Plants 5:54–62.

IBI. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768.

Janiak A, Kwa M, Szarejko I. 2015. Gene expression regulation in roots under drought. J. Exp. Bot. 67:1003–1014.

Jeong H, Mason SP, Barabási AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. Nature 411:41–42.

Juenger TE. 2013. Natural variation and genetic constraints on drought tolerance. Curr. Opin. Plant Biol. 16:274–281.

Koonin E V., Wolf YI. 2008. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36:6688–6719.

Ksouri N, Castro-Mondragón JA, Montardit-Tarda F, van Helden J, Contreras-Moreira B, Gogorcena Y. 2021. Tuning promoter boundaries improves regulatory motif discovery in nonmodel plants: the peach example. Plant Physiol. 185:1242–1258.

Langfelder P, Horvath S. 2007. Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. 1:54.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:1–13.

Langfelder P, Horvath S. 2010. Overview of network terminology. :1–2. Available from: https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/ModulePreservation/Tutorials/glossaryTable.pdf

Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and reproducible? PLoS Comput. Biol. 7.

Lehti-Shiu MD, Shiu SH. 2012. Diversity, classification and function of the plant protein kinase superfamily. Philos. Trans. R. Soc. B Biol. Sci. 367:2619–2639.

López-Álvarez D, Manzaneda AJ, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, Mur L, Caicedo AL, Hazen SP, Breiman A, et al. 2015. Environmental niche variation and evolutionary diversification of the Brachypodium distachyon grass complex species in their native circum-Mediterranean range. Am. J. Bot. 102:1073–1088.

Luo N, Liu J, Yu X, Jiang Y. 2011. Natural variation of drought response in Brachypodium distachyon. Physiol. Plant. 141:19–29.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Manzaneda AJ, Rey PJ, Anderson JT, Raskin E, Weiss-Lehman C, Mitchell-Olds T. 2015. Natural variation, differentiation, and genetic trade-offs of ecophysiological traits in response to water limitation in Brachypodium distachyon and its descendent allotetraploid B. hybridum (Poaceae). Evolution (N. Y). 69:2689–2704.

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. 2012. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in Brachypodium distachyon (Poaceae). New Phytol. 193:797–805.

Mao L, Hemert JL Van, Dash S, Dickerson JA. 2009. Arabidopsis gene co-expression network and its functional modules. BMC Bioinformatics 10:1–24.

Des Marais DL, Guerrero RF, Lasky JR, Scarpino S V. 2017. Topological features of gene regulatory networks predict patterns of natural diversity in environmental response. Proc. R. Soc. B Biol. Sci. 284.

Des Marais DL, Juenger TE. 2016. Brachypodium and the Abiotic Environment. In: Vogel JP, editor. Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models, volume 18. Switzerland: Springer. p. 291–311.

Des Marais DL, Lasky JR, Verslues PE, Chang TZ, Juenger TE. 2017. Interactive effects of water limitation and elevated temperature on the physiology, development and fitness of diverse accessions of Brachypodium distachyon. New Phytol. 214:132–144.

Des Marais DL, Mckay JK, Richards JH, Sen S, Wayne T, Juenger TE. 2012. Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. Plant Cell 24:893–914.

Des Marais DL, Razzaque S, Hernandez KM, Garvin DF, Juenger TE. 2016. Quantitative trait loci associated with natural diversity in water-use efficiency and response to soil drying in Brachypodium distachyon. Plant Sci. 251:2–11.

Martínez LM, Fernández-ocaña A, Rey PJ, Salido T, Amil-ruiz F, Manzaneda AJ. 2018. Variation in functional responses to water stress and differentiation between natural allopolyploid populations in the Brachypodium distachyon species complex. Ann. Bot. 00:1–14.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PLoS One 12:e0182289.

Meyer E, Aglyamova G V., Matz M V. 2011. Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. Mol. Ecol. 20:3599–3616.

Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 49:D394–D403.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Miao Z, Han Z, Zhang T, Chen S, Ma C. 2017. A systems approach to a spatio- temporal understanding of the drought stress response in maize. Sci. Rep. 7:1–14.

Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K. 2011. Global Landscape of a Co-Expressed Gene Network in Barley and its Application to Gene Discovery in Triticeae Crops. Plants Cell Physiol. 52:785–803.

Monroe JG, McGovern C, Lasky JR, Grogan K, Beck J, McKay JK. 2016. Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana. Mol. Ecol. 25:3632–3644.

Monroe JG, Powell T, Price N, Mullen JL, Howard A, Evans K, Lovell JT, McKay JK. 2018. Drought adaptation in arabidopsis thaliana by extensive genetic loss-of-function. Elife 7:1–18.

Mur LAJ, Allainguillaume J, Catalán P, Hasterok R, Jenkins G, Lesniewska K, Thomas I, Vogel J. 2011. Exploiting the Brachypodium tool box in cereal and grass research. New Phytol. 191:334–347.

Nakashima K, Ito Y, Yamaguchi-Shinozaki K. 2009. Transcriptional Regulatory Networks in Response to Abiotic Stresses in Arabidopsis and Grasses. Plant Physiol. 149:88–95.

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 2014. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front. Plant Sci. 5:1–7.

Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf KD. 2001. Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? Cell Stress Chaperones 6:177–189.

Paaby AB, Rockman M V. 2014. Cryptic genetic variation: Evolution's hidden substrate. Nat. Rev. Genet. 15:247–258.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods 14:687–690.

Pinheiro C, Chaves MM. 2011. Photosynthesis and drought: Can we make metabolic connections from available data? J. Exp. Bot. 62:869–882.

Porth I, Klápště J, McKown AD, La Mantia J, Hamelin RC, Skyba O, Unda F, Friedmann MC, Cronk QCB, Ehlting J, et al. 2014. Extensive functional pleiotropy of REVOLUTA substantiated through forward genetics. Plant Physiol. 164:548–554.

Priest HD, Fox SE, Rowley ER, Murray JR, Michael TP, Mockler TC. 2014. Analysis of Global Gene Expression in Brachypodium distachyon Reveals Extensive Network Plasticity in Response to Abiotic Stress. PLoS One 9:e87499.

Ritchie SC, Watts S, Fearnley LG, Holt KE, Abraham G, Inouye M. 2016. A Scalable Permutation Approach Reveals Replication and Preservation Patterns of Network Modules in Large Datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cell Syst. 3:71–82.

Ruíz M, Quemada M, García RM, Carrillo JM, Benavente E. 2016. Use of thermographic imaging to screen for drought-tolerant genotypes in Brachypodium distachyon. Crop Pasture Sci. 67:99–108.

Scharf KD, Berberich T, Ebersberger I, Nover L. 2012. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. Biochim. Biophys. Acta - Gene Regul. Mech. 1819:104–119.

Scholthof K-BG, Irigoyen S, Catalán P, Mandadi KK. 2018. Brachypodium: A monocot grass model system for plant biology. Plant Cell.

Sebastian A, Contreras-Moreira B. 2014. footprintDB: a database of transcription factors with annotated cis elements and binding interfaces. Bioinformatics 30:258–265.

Skalska A, Stritt C, Wyler M, Williams HW, Vickers M, Han J, Tuna M, Tuna GS, Susek K, Swain M, et al. 2020. Genetic and methylome variation in Turkish Brachypodium distachyon accessions differentiate two geographically distinct subpopulations. Int. J. Mol. Sci. 21:1–17.

Stuart JM, Segal E, Koller D, Kim SK. 2003. A Gene Coexpression Network for Global Discovery of Conserved Genetic Modules. Science (80-.). 302:249–255.

Tandonnet S, Torres TT. 2017. Traditional versus 3′ RNA-seq in a non-model species. Genomics Data 11:9–16.

Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, Hathorn A, Wu X, Liu Y, Shatte T, et al. 2021. Extensive variation within the pan-genome of cultivated and wild sorghum. Nat. Plants 7:766–773.

Thomas-Chollier M, Herrmann C, Defrance M, Sand O, Thieffry D, Van Helden J. 2012. RSAT peak-motifs: Motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res. 40.

Turatsinze JV, Thomas-Chollier M, Defrance M, van Helden J. 2008. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat. Protoc. 3:1578–1588.

Verelst W, Bertolini E, De Bodt S, Vandepoele K, Demeulenaere M, Pè ME, Inzé D. 2013. Molecular and physiological analysis of growth-limiting drought stress in Brachypodium distachyon leaves. Mol. Plant 6:311–322.

Vogel JP, Garvin DF, Leong OM, Hayden DM. 2006. Agrobacterium-mediated transformation and inbred line development in the model grass Brachypodium distachyon. Plant Cell. Tissue Organ Cult. 84:199–211.

Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, et al. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vogel JP, Tuna M, Budak H, Huo N, Gu YQ, Steinwand MA. 2009. Development of SSR markers and analysis of diversity in Turkish populations of Brachypodium distachyon. BMC Plant Biol. 9:88.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag

Wingler A, Quick WP, Bungard RA, Bailey KJ, Lea PJ, Leegood RC. 1999. The role of photorespiration during drought stress: An analysis utilizing barley mutants with reduced activities of photorespiratory enzymes. Plant, Cell Environ. 22:361–373.

Wolfe CJ, Kohane IS, Butte AJ. 2005. Systematic survery reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics 6:1–10.

Yu H, Jiao B, Liang C. 2017. High-quality rice RNA-seq-based co-expression network for predicting gene function and regulation. bioRxiv.

Zhang B, Horvath S. 2005. A General Framework for Weighted Gene Co-expression Network Analysis. Stat. Appl. Genet. Mol. Biol. 4:Article 17.

Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, et al. 2016. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Mol. Plant 9:1667–1670.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Statistics of the number and percentage of isoforms (all and hub isoforms) and genes (all and hub genes), and mean k_{diff} (the difference between intra- and inter-modular connectivity) per module for each Drought **(a)** and Water **(b)** network (Note: the quantity of total counts of genes in drought and water networks is different because we counted the genes for each module independently. If there are several isoforms of the same gene that are clustered in different modules, this gene will be counted multiple times. The gene is not counted multiple times if this is clustered in the same module (e.g. The module D1 has three isoforms (Bradi1g1000.1, Bradi1g1000.2 and Bradi1g1000.3). In this case, we count only one gene (Bradi1g1000). However, if we have the isoforms in different modules, Bradi1g1000.1 in module D1, Bradi1g1000.2 in D2 and Bradi1g1000.3 in module D3, the gene is count three times in total, one in D1, one in D2 and one in D3). This only affects the total gene counts of the networks).

(a)

Drought modules		isoforms (nodes) per module		genes per module		mean Kdiff
		all	hub	all	hub	
D0	gray	11366 (69.4%)	nd	8313 (67.4%)	nd	nd
D1	turquoise	2477 (15.1%)	72 (2.9%)	1986 (16.1%)	63 (3.2%)	7.9
D2	blue	979 (6.0%)	100 (10.2%)	750 (6.1%)	71 (9.5%)	3.7
D3	brown	750 (4.6%)	114 (15.2%)	627 (5.1%)	99 (15.8%)	10.1
D4	yellow	318 (1.9%)	58 (18.2%)	258 (2.1%)	45 (17.4%)	3.7
D5	green	214 (1.3%)	34 (15.9%)	169 (1.4%)	25 (14.8%)	0.8
D6	red	111 (0.7%)	24 (21.6%)	95 (0.8%)	19 (20.0%)	-1.5
D7	black	68 (0.4%)	11 (16.2%)	48 (0.4%)	9 (18.8%)	-0.2
D8	pink	65 (0.4%)	16 (24.6%)	57 (0.5%)	8 (14.0%)	-5.9
D9	magenta	38 (0.2%)	11 (28.9%)	27 (0.2%)	4 (14.8%)	-8.8
Total counts		**16386**	**440 (2.7%)**	**12330**	**343 (2.8%)**	
Total unique counts		**16386**	**440 (2.7%)**	**12137**	**343 (2.8%)**	
Total unique counts (excluding gray module)		**5020**	**440 (8.8%)**	**4006**	**343 (8.6%)**	

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(b)

Water modules		isoforms (nodes) per module		genes per module		mean Kdiff
		all	hub	all	hub	
W0	gray	9675 (59.0%)	nd	6934 (56.0%)	nd	nd
W1	turquoise	1866 (11.4%)	321 (17.2%)	1439 (11.6%)	250 (17.4%)	26.8
W2	blue	870 (5.3%)	45 (5.1%)	727 (5.9%)	43 (5.9%)	-3.7
W3	brown	827 (5.0%)	67 (8.1%)	696 (5.6%)	49 (7.0%)	0.3
W4	yellow	701 (4.3%)	44 (6.3%)	590 (4.8%)	42 (7.1%)	-0.7
W5	green	607 (3.7%)	104 (17.1%)	509 (4.1%)	89 (17.5%)	3.1
W6	red	435 (2.7%)	65 (14.9%)	358 (2.9%)	55 (15.4%)	2.1
W7	black	390 (2.4%)	81 (20.8%)	313 (2.5%)	66 (21.4%)	4.6
W8	pink	288 (1.7%)	40 (13.9%)	245 (2.0%)	33 (13.5%)	-2.2
W9	magenta	282 (1.7%)	70 (24.8%)	239 (1.9%)	60 (25.1%)	5.8
W10	purple	246 (1.5%)	37 (15.0%)	185 (1.5%)	24 (13.0%)	-1.4
W11	greenyellow	89 (0.5%)	18 (20.2%)	55 (0.4%)	2 (3.6%)	-3.1
W12	tan	62 (0.4%)	9 (14.5%)	51 (0.4%)	9 (17.6%)	-0.9
W13	salmon	48 (0.3%)	10 (20.8%)	40 (0.3%)	2 (5.0%)	-1.6
Total counts		**16386**	**911 (5.6%)**	**12381**	**724 (5.8%)**	
Total unique counts		**16386**	**911 (5.6%)**	**12137**	**719 (5.9%)**	
Total unique counts (excluding gray module)		**6711**	**911 (13.6%)**	**5407**	**719 (13.3%)**	

Table 2. Statistics of topological features (Connectivity, Scaled-Connectivity, Clustering-Coefficient, Maximum adjacency ratio (MAR), Density, Centralization and Heterogeneity) for Drought (D) and Water (W) networks (excluding gray or "zero" module). Minimum (Min.) and Maximum (Max.) range, first (1st Qu.) and third (3rd Qu.) quartiles, median and mean values are detailed for connectivity, scaled connectivity, clustering coefficient and MAR variables in all cases.

	Connectivity		Scaled-Connectivity		Clustering Coefficient		MAR		Density	Centralization	Heterogeneity
D	**Min.**	0.1931	**Min.**	0.001884	**Min.**	0.007843	**Min.**	0.001215	**0.003128**	**0.01731**	**1.046**
	1st Qu.	4.73	**1st Qu.**	0.046137	**1st Qu.**	0.023497	**1st Qu.**	0.015039			
	Median	9.902	**Median**	0.096584	**Median**	0.03168	**Median**	0.050674			
	Mean	**15.6999**	**Mean**	**0.153137**	**Mean**	**0.056317**	**Mean**	**0.101807**			
	3rd Qu.	20.2376	**3rd Qu.**	0.197398	**3rd Qu.**	0.070095	**3rd Qu.**	0.139966			
	Max.	102.5221	**Max.**	1	**Max.**	0.640168	**Max.**	0.852346			
W	**Min.**	0.2478	**Min.**	0.001105	**Min.**	0.007156	**Min.**	0.001046	**0.002946**	**0.03048**	**1.519**
	1st Qu.	4.5082	**1st Qu.**	0.020103	**1st Qu.**	0.02305	**1st Qu.**	0.015927			
	Median	9.8379	**Median**	0.04387	**Median**	0.040211	**Median**	0.05922			
	Mean	**19.7662**	**Mean**	**0.088143**	**Mean**	**0.067956**	**Mean**	**0.111266**			
	3rd Qu.	21.279	**3rd Qu.**	0.094888	**3rd Qu.**	0.085417	**3rd Qu.**	0.162872			
	Max.	224.2527	**Max.**	1	**Max.**	0.472454	**Max.**	0.713113			

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Occupancies (core (33 accessions); soft-core (31-32 accessions) and shell (≤30 accessions)) of the co-expressed genes **(a; b)** and hub genes **(c; d)** for each Drought (D) and Water (W) networks. Asterisks note significant differences comparing shell and total genes between modules and network by Fisher test (*p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001).

(a)

			Drought genes occupancy		
Drought modules		**Genes**	**Core (33)**	**Soft-core (32 or 31)**	**Shell (<=30)**
D0	**gray**	8313	5764 (69.3%)	1269 (15.3%)	1280 (15.4%)
D1	**turquoise**	1986	1422 (71.6%)	265 (13.3%)	299 (15.1%)
D2	**blue**	750	549 (73.2%)	126 (16.8%)	75 (10.0%)***
D3	**brown**	627	440 (70.2%)	117 (18.7%)	70 (11.2%)*
D4	**yellow**	258	160 (62.0%)	49 (19.0%)	49 (19.0%)
D5	**green**	169	123 (72.8%)	27 (16.0%)	19 (11.2%)
D6	**red**	95	62 (65.3%)	18 (19.0%)	15 (15.8%)
D7	**black**	48	28 (58.3%)	9 (18.8%)	11 (22.9%)
D8	**pink**	57	17 (29.8%)	4 (7.0%)	36 (63.2%)***
D9	**magenta**	27	20 (74.1%)	6 (22.2%)	1 (3.7%)
Total counts		**12330**	**8585 (69.6%)**	**1890 (15.3%)**	**1855 (15.0%)**
Total unique counts		**12137**	**8426 (69.4)**	**1869 (15.4%)**	**1842 (15.2%)**
Total unique counts (excluding gray module)		**4006**	**2813 (70.2%)**	**618 (15.4%)**	**575 (14.4%)**

(b)

			Water genes occupancy		
Water modules		**Genes**	**Core (33)**	**Soft-core (32 or 31)**	**Shell (<=30)**
W0	**gray**	6934	4788 (69.1%)	1006 (14.5%)	1140 (16.4%)*
W1	**turquoise**	1439	1094 (76.0%)	188 (13.1%)	157 (10.9%)***
W2	**blue**	727	544 (74.8%)	111 (15.3%)	72 (9.9%)***
W3	**brown**	696	500 (71.8%)	128 (18.4%)	68 (9.8%)***
W4	**yellow**	590	353 (59.8%)	73 (12.4%)	164 (27.8%)***
W5	**green**	509	375 (73.7%)	101 (19.8%)	33 (6.5%)***
W6	**red**	358	258 (72.1%)	67 (18.7%)	33 (9.2%)**
W7	**black**	313	216 (69.0%)	59 (18.8%)	38 (12.1%)
W8	**pink**	245	162 (66.1%)	50 (20.4%)	33 (13.5%)
W9	**magenta**	239	124 (51.9%)	63 (26.4%)	52 (21.8%)*
W10	**purple**	185	127 (68.6%)	36 (19.5%)	22 (11.9%)
W11	**greenyellow**	55	33 (60.0%)	8 (14.5%)	14 (25.5%)
W12	**tan**	51	31 (60.8%)	6 (11.8%)	14 (27.5%)
W13	**salmon**	40	13 (32.5%)	3 (7.5%)	24 (60.0%)***
Total counts		**12381**	**8618 (69.6%)**	**1899 (15.3%)**	**1864 (15.1%)**
Total unique counts		**12137**	**8426 (69.4)**	**1869 (15.4%)**	**1842 (15.2%)**
Total unique counts (excluding gray module)		**5407**	**3795 (70.2%)**	**890 (16.5%)**	**722 (13.3%)**

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(c)

			Drought hub genes occupancy		
Drought modules		**Hub Genes**	**Core (33)**	**Soft-core (32 or 31)**	**Shell (<=30)**
D1	**turquoise**	63	38 (60.3%)	14 (22.2%)	11 (17.5%)
D2	**blue**	71	48 (67.6%)	16 (22.5%)	7 (9.9%)
D3	**brown**	99	59 (59.6%)	29 (29.3%)	11 (11.1%)
D4	**yellow**	45	30 (66.7%)	7 (15.6%)	8 (17.8%)
D5	**green**	25	22 (88.0%)	3 (12.0%)	0 (0%)
D6	**red**	19	12 (63.2%)	5 (26.3%)	2 (10.5%)
D7	**black**	9	3 (33.3%)	1 (11.1%)	5 (55.6%)*
D8	**pink**	8	1 (12.5%)	0 (0%)	7 (87.5%)**
D9	**magenta**	4	2 (50.0%)	2 (50.0%)	0 (0%)
Total counts		**343**	**215 (62.7%)**	**77 (22.4%)**	**51 (14.9%)**
Total unique counts		**343**	**215 (62.7%)**	**77 (22.4%)**	**51 (14.9%)**

(d)

			Water hub genes occupancy		
Water modules		**Hub Genes**	**Core (33)**	**Soft-core (32 or 31)**	**Shell (<=30)**
W1	**turquoise**	250	183 (73.2%)	30 (12.0%)	37 (14.8%)
W2	**blue**	43	29 (67.4%)	10 (23.3%)	4 (9.3%)
W3	**brown**	49	34 (69.4%)	11 (22.4%)	4 (8.2%)
W4	**yellow**	42	6 (14.3%)	3 (7.1%)	33 (78.6%)***
W5	**green**	89	55 (61.8%)	26 (29.2%)	8 (9.0%)
W6	**red**	55	45 (81.8%)	8 (14.5%)	2 (3.6%)*
W7	**black**	66	45 (68.2%)	17 (25.8%)	4 (6.1%)*
W8	**pink**	33	23 (69.7%)	9 (27.3%)	1 (3.0%)
W9	**magenta**	60	30 (50.0%)	15 (25.0%)	15 (25.0%)
W10	**purple**	24	15 (62.5%)	5 (20.8%)	4 (16.7%)
W11	**greenyellow**	2	2 (100%)	0 (0%)	0 (0%)
W12	**tan**	9	3 (33.3%)	0 (0%)	6 (66.7%)*
W13	**salmon**	2	1 (50.0%)	0 (0%)	1 (50.0%)
Total counts		**724**	**471 (65.1%)**	**134 (18.5%)**	**119 (16.4%)**
Total unique counts		**719**	**467 (65.0%)**	**134 (18.6%)**	**118 (16.4%)**

Table 4. Summary of the enrichment analysis according to the statistically significant GO biological process for the genes (all, core, soft-core and shell genes) clustered in the Drought (D) (**a**) and Water (W) (**b**) modules applying the statistical overrepresentation test of Panther (http://pantherdb.org/) tool. The biological processes summarized according to the lowest FDR values (see Supplementary file S1).

(a)

Modules		All genes	core genes	soft-core genes	Shell genes
D1	turquoise	nitrogen, amide and peptide metabolic and biosynthetic processes	nitrogen, amide and peptide metabolic and biosynthetic processes	nitrogen, amide and peptide metabolic and biosynthetic processes	photosynthesis, nitrogen, amide and peptide metabolic and biosynthetic processes
D2	blue	photosynthesis	photosynthesis	photosynthesis	*No statistically significant results*
D3	brown	organonitrogen compound metabolic process, location/transport and phosphorylation	organonitrogen compound metabolic process, location/transport and phosphorylation	*No statistically significant results*	*No statistically significant results*
D4	yellow	nucleobase-containing compound, heterocycle and nucleic acid metabolic processes	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*
D5	green	protein folding, response to heat, temperature and abiotic stimulus	protein folding, response to heat, temperature and abiotic stimulus	response to heat, temperature and protein folding	*No statistically significant results*
D6	red	lipid and small molecule metabolic process	lipid and small molecule metabolic process	*No statistically significant results*	*No statistically significant results*
D7	black	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*
D8	pink	photosynthesis, nitrogen, amide and peptide biosynthetic and metabolic processes	*No statistically significant results*	*No statistically significant results*	photosynthesis, nitrogen, amide and peptide biosynthetic and metabolic processes
D9	magenta	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*

(b)

	Modules	All genes	core genes	soft-core genes	Shell genes
W1	**turquoise**	nitrogen compound, organic substance, primary metabolic processes	nitrogen compound, organic substance, primary metabolic processes	protein ubiquitination, protein modification by small protein conjugation	*No statistically significant results*
W2	**blue**	peptide, amide and nitrogen biosynthetic and metabolic processes	peptide, amide and nitrogen biosynthetic and metabolic processes	peptide, amide and nitrogen biosynthetic and metabolic processes	peptide, amide and nitrogen biosynthetic and metabolic processes
W3	**brown**	protein, peptide, amide and ion transport/location	protein, peptide, amide and ion transport/location	protein and ion transport/location	*No statistically significant results*
W4	**yellow**	photosynthesis, generation of precursor metabolites and energy, cellular nitrogen compound and organic substance biosynthetic processes, and electron transport chain	organic substance, primary, nitrogen compound, nucleobase-containing compound metabolic processes	*No statistically significant results*	photosynthesis, generation of precursor metabolites and energy, cellular nitrogen compound and organic substance biosynthetic processes, and electron transport chain
W5	**green**	organonitrogen and phosphate-containing compounds, and phosphorus metabolic processes and protein phosphorylation	organonitrogen and phosphate-containing compounds, small molecule, protein and phosphorus metabolic processes and protein phosphorylation	*No statistically significant results*	*No statistically significant results*
W6	**red**	plant-type cell wall organization or biogenesis	plant-type cell wall organization or biogenesis	*No statistically significant results*	*No statistically significant results*
W7	**black**	glutathione, cellular amino acid, oxoacid, organic acid and sulfur compound metabolic processes	glutathione, cellular amino acid, oxoacid, organic acid and sulfur compound metabolic processes	*No statistically significant results*	*No statistically significant results*
W8	**pink**	small molecule, starch, proline and carbohydrate metabolic processes	small molecule, starch, proline and carbohydrate metabolic processes	*No statistically significant results*	*No statistically significant results*

W9	**magenta**	regulation of nucleic acid-templated transcription, and RNA, nucleobase-containing compound, macromolecule biosynthesis and metabolic processes	cellular protein modification process, protein dephosphorylation, protein modification process, macromolecule modification and dephosphorylation	regulation of nucleic acid-templated transcription, and RNA, nucleobase-containing compound, macromolecule biosynthesis and metabolic processes	*No statistically significant results*
W10	**purple**	photosynthesis	photosynthesis	photosynthesis	*No statistically significant results*
W11	**greenyellow**	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*
W12	**tan**	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*	*No statistically significant results*
W13	**salmon**	macromolecule, cellular nitrogen compound, peptide and amide biosynthetic processes	*No statistically significant results*	*No statistically significant results*	macromolecule, cellular nitrogen compound, peptide and amide biosynthetic processes

Table 5. DNA motifs and cis-regulatory elements (CREs) of co-expressed genes assigned to Drought (D) (**a**) and Water (W) (**b**) modules. Modules (numeric and color codes of the modules); Accession(footprintDB), Names (gene name); Consensus sequence of motif; Protein name (Swissprot); Ncor (normalized correlation score); e-value; sites (number of sites used to compile the DNA motif); Proportion of genes with CREs per module as detected by matrix-scan in -500 to +200 bp windows; Total genes with CREs [occupancy of core, soft-core and shell genes]; hub genes with CREs. The dashes indicate that no results were retrieved.

(a)

Modules		Accession (footprintDB)	Names	consensus	Protein (Swissprot)	Ncor	evalue	sites	Genes with putative CREs in promoter		
									Proportion of genes with CREs per module	Total genes with CREs [occupancy of core, soft-core and shell genes]	hub genes with CREs
D1	**turquoise**	MA1353.1 (JASPAR 2020), M0609 (Athaliana Cistrome v4_May2016)	AT1G72740, AT1G72740. DAP, T04493	ssmaAAACCCTAGcy	Telomere repeat-binding factor 5 (MYB transcription factor)	0.741	1.40E-56	225	9.0%	178 [139 (78.1%) 26 (14.6%) 13 (7.3%)]	16 (9.0%)
D2	**blue**	ABI4 (Athamap 20091028)	ABI4	scCACCaCCCrc	Ethylene-responsive transcription factor ABI4 (Protein ABSCISIC ACID INSENSITIVE 4)	0.697	5.80E-07	178	17.1%	128 [102 (79.7%) 17 (13.3%) 9 (7.0%)]	13 (10.2%)
D3	**brown**	M1684_1.02 (CISBP 1.02)	T153999_1.02, WRKY60	gtCGGTCAACgk	Probable WRKY transcription factor 60 (WRKY DNA-binding protein 60)	0.877	1.30E-09	165	11.6%	73 [54 (74.0%) 13 (17.8%) 6 (8.2%)]	12 (16.4%)
D4	**yellow**	MA1197.1 (JASPAR 2020), M0350 (Athaliana Cistrome v4_May2016)	CAMTA1, CAMTA1.DAP, T27097	bgCGACGCGCTks	Calmodulin-binding transcription activator 1, AtCAMTA1 (Ethylene-induced calmodulin-binding protein b,	0.717	7.80E-18	69	19.4%	50 [26 (52.0%) 13 (26.0%) 11 (22.0%)]	11 (22.0%)

					EICBP.b) (Signal-responsive protein 2, AtSR2)							
D5	**green**	M0455 (Athaliana Cistrome v4_May2016)	HSFB3.amp DAP, T08630	kkCTCTrGAAGgk	Heat stress transcription factor B-3, AtHsfB3 (AtHsf-05)	0.691	9.10E-13	89	27.8%	47 [36 (76.6%) 5 (10.6%) 6 (12.8%)]	11 (23.4%)
D6	**red**	-	-	-	-	-	-	-	-	-	-
D7	**black**	AHL25_3 (Arabidopsis PBM 20140210)	AHL25	wcAAAATAAATak	AT-hook motif nuclear-localized protein 25 (AT-hook protein of GA feedback 1)	0.643	0.46	35	54.2%	26 [16 (61.6%) 5 (19.2%) 5 (19.2%)]	5 (19.2%)
D8	**pink**	MA1274.1 (JASPAR 2020), M0256 (Athaliana Cistrome v4_May2016)	OBP3, OBP3.DAP, T16682	mrArGAATArrrAARrdh	Probable plastid-lipid-associated protein 14, chloroplastic, AtPap14 (Fibrillin-11) (OBP3-responsive protein 1)	0.623	1.50E-12	14	3.5%	2	-
D9	**magenta**	-	-	-	-	-	-	-	-	-	-

(b)

	Modules	Accession (footprintDB)	Names	consensus	Protein (Swissprot)	Ncor	evalue	sites	Genes with putative CREs in promoter		
									Proportion of genes with CREs per module	Total genes with CREs [occupancy of core, soft-core and shell genes]	hub genes with CREs
W1	**turquoise**	M0587_1.02 (CISBP 1.02)	PK22848.1, T073801_1.02	wkwcTAAAATTTTAGwww	-	0.59	6.80E-29	32	0.1%	2	-
W2	**blue**	At2g20350:M05132:TRANSFAC	At2g20350/M05132/TRANSFAC	mmAGGCCCATCwv	Ethylene-responsive transcription factor ERF120	0.59	1.90E-97	244	22.3%	162 [115 (71.0%) 25 (15.4%) 22 (13.6%)]	15 (9.3%)
W3	**brown**	-	-	-	-	-	-	-	-	-	-
W4	**yellow**	MA1379.1 (JASPAR 2020), M0357 (Athaliana Cistrome v4_May2016)	SOL1, SOL1.DAP, T15201	wwTTAAwwwddAAAwr	Carboxypeptidase SOL1, EC 3.4.17.- (Protein SUPPRESSOR OF LLP1 1)	0.662	1.10E-09	458	0.3%	2	-
W5	**green**	UP00581A_2 (UniPROBE 20160601)	vascular plant one zinc finger protein 2, VOZ2	sgAGTCAACGtcgv	Transcription factor VOZ2 (Protein VASCULAR PLANT ONE-ZINC FINGER 2, AtVOZ2)	0.88	1.90E-10	126	5.7%	29 [20 (69.0%) 4 (13.8%) 5 (17.2%)]	12 (41.4%)
W6	**red**	M0523 (Athaliana Cistrome v4_May2016)	MYB58.DAP, T03869	ytgCyACCAACCAvm	Transcription factor MYB58	0.837	1.10E-11	45	2.5%	9	-

W7	**black**	UP00582A_2 (UniPROBE 20160601)	ARABIDOPSIS THALIANA WRKY DNA-BINDING PROTEIN 18, At4g31800, AtWRKY18, F11C18.16, F28M20.10, WRKY transcription factor 1, WRKY transcription factor 18, WRKY1	mmsGGTCA AACGyr	WRKY transcription factor 18	0.698	7.40E-07	27	6.1%	19 [17 (89.5%) 2 (10.5%) 0 (0.0%)]	4 (21.1%)
W8	**pink**	-	-	-	-	-	-	-	-	-	-
W9	**magenta**	MA1197.1 (JASPAR 2020), M0350 (Athaliana Cistrome v4_May2016)	CAMTA1, CAMTA1.DAP, T27097	ssaCCGCGTc ss	Calmodulin-binding transcription activator 1, AtCAMTA1 (Ethylene-induced calmodulin-binding protein b, EICBP.b) (Signal-responsive protein 2, AtSR2)	0.755	4.50E-19	95	23.0%	55 [28 (50.9%) 12 (21.8%) 15 (27.3%)]	15 (27.3%)
W10	**purple**	-	-	-	-	-	-	-	-	-	-
W11	**greenyellow**	-	-	-	-	-	-	-	-	-	-
W12	**tan**	-	-	-	-	-	-	-	-	-	-
W13	**salmon**	-	-	trAAATymw wwTTCAht	-	0.504	0.00017	35	20.0%	8	-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

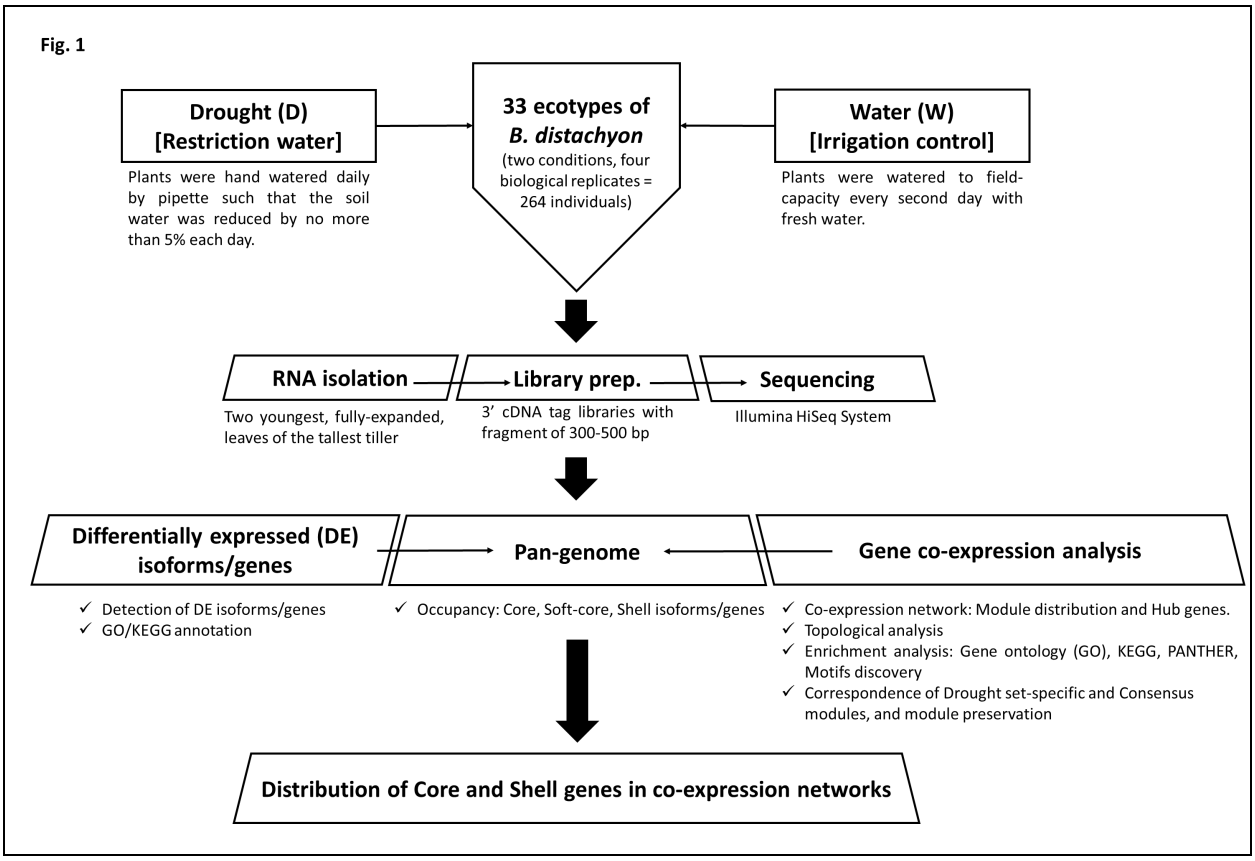
Table 6. Pan-genome analysis of the differentially expressed (DE) isoforms according to their occupancy: core (33 accessions), soft-core (31-32 accessions) and shell (≤30 accessions). Non-matched indicates genes not located in the pan-genome reference.

	core	soft-core	shell	non-matched	TOTAL
upregulated	2678 (74.6%)	583 (16.2%)	312 (8.7%)	18 (0.5%)	3591
downregulated	950 (70.4%)	264 (19.6%)	130 (9.6%)	6 (0.4%)	1350
TOTAL	**3628**	**847**	**442**	**24**	**4941**

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

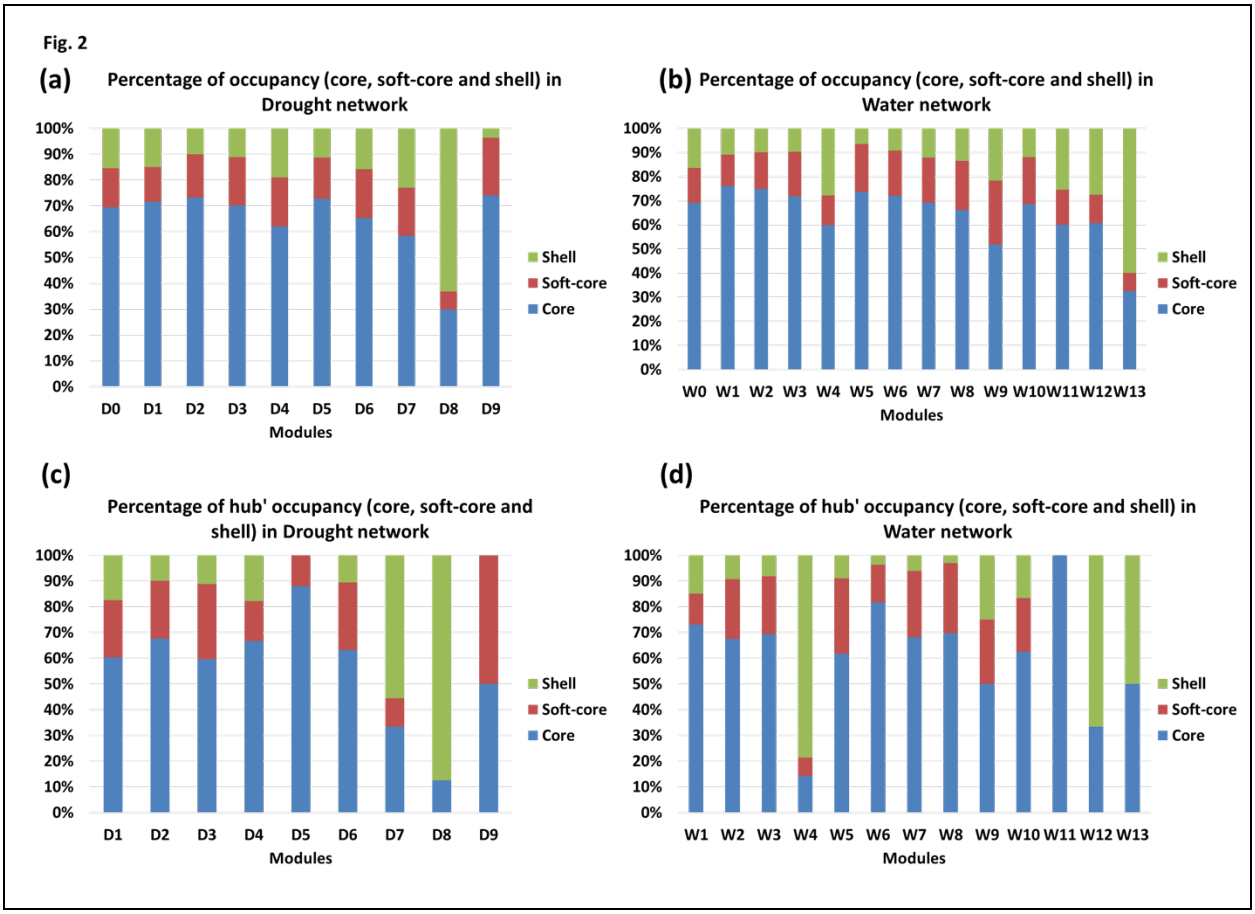
Figures:

Figure 1. Summary of the experimental design and analyses performed in the 33 accessions of the model grass *Brachypodium distachyon* under drought (D) and water (control) (W) conditions.



bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. Proportion (%) of occupancies (core (33); soft-core (31-32) and shell (≤30)) of the co-expressed genes (**a; b**) and hub genes (**c; d**) for each Drought (D) and Water (W) network.



bioRxiv preprint doi: https://doi.org/10.1101/2021.07.21.453242; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. Correspondence (number of nodes) of Drought (D) (**a**) and Water (W) (**b**) set-specific and Drought-Water consensus (Cons) modules. Each row of the table corresponds to one Drought/Water set-specific module, and each column corresponds to one consensus module. Numbers in the table indicate node counts in the intersection of the corresponding modules. Coloring of the table encodes − log(p), with p being the Fisher's exact test p-value for the overlap of the two modules. The stronger the red color, the more significant the overlap is. (**c**) Comparison of the GO enrichments for each D and W module, according to the correlation with the common consensus module.