



Aberystwyth University

*Differential gene expression and gene ontologies associated with increasing water-stress in leaf and root transcriptomes of perennial ryegrass (*Lolium perenne*)*

Fradera-Sola, Albert; Thomas, Ann; Gasior, Dagmara; Harper, John; Hegarty, Matthew; Armstead, Ian; Fernandez Fuentes, Narcis

Published in:

PLoS One

DOI:

[10.1371/journal.pone.0220518](https://doi.org/10.1371/journal.pone.0220518)

Publication date:

2019

Citation for published version (APA):

Fradera-Sola, A., Thomas, A., Gasior, D., Harper, J., Hegarty, M., Armstead, I., & Fernandez Fuentes, N. (2019). Differential gene expression and gene ontologies associated with increasing water-stress in leaf and root transcriptomes of perennial ryegrass (*Lolium perenne*). *PLoS One*, *14*(7), [e0220518]. <https://doi.org/10.1371/journal.pone.0220518>

Document License

CC BY

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400

email: is@aber.ac.uk

RESEARCH ARTICLE

Differential gene expression and gene ontologies associated with increasing water-stress in leaf and root transcriptomes of perennial ryegrass (*Lolium perenne*)

Albert Fradera-Sola¹, Ann Thomas², Dagmara Gasior², John Harper², Matthew Hegarty², Ian Armstead², Narcis Fernandez-Fuentes^{2*}

1 Quantitative Proteomics, Institute of Molecular Biology (IMB), Mainz, Germany, **2** Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom

* naf4@aber.ac.uk



OPEN ACCESS

Citation: Fradera-Sola A, Thomas A, Gasior D, Harper J, Hegarty M, Armstead I, et al. (2019) Differential gene expression and gene ontologies associated with increasing water-stress in leaf and root transcriptomes of perennial ryegrass (*Lolium perenne*). PLoS ONE 14(7): e0220518. <https://doi.org/10.1371/journal.pone.0220518>

Editor: Niranjana Baisakh, Louisiana State University, UNITED STATES

Received: April 4, 2019

Accepted: July 17, 2019

Published: July 30, 2019

Copyright: © 2019 Fradera-Sola et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from ENA repository accession number PRJEB31812.

Funding: This work has been funded by the Biotechnology and Biological Science Research Council (BBSRC) through the following institutional awards: BB/J004405/1 (ISPG) and BB/CSP1730/1 (GS). The funders had no role in study design, data

Abstract

Perennial ryegrass (*Lolium perenne*) is a forage and amenity grass species widely cultivated in temperate regions worldwide. As such, perennial ryegrass populations are exposed to a range of environmental conditions and stresses on a seasonal basis and from year to year. One source of potential stress is limitation on water availability. The ability of these perennial grasses to be able to withstand and recover after periods of water limitation or drought can be a key component of grassland performance. Thus, we were interested in looking at changes in patterns of gene expression associated with increasing water stress. Clones of a single genotype of perennial ryegrass were grown under non-flowering growth room conditions in vermiculite supplemented with nutrient solution. Leaf and root tissue was sampled at 4 times in quadruplicate relating to estimated water contents of 35%, 15%, 5% and 1%. RNA was extracted and RNAseq used to generate transcriptome profiles at each sampling point. Transcriptomes were assembled using the published reference genome sequence and differential gene expression analysed using 3 different programmes, DESeq2, edgeR and limma (with the voom transformation), individually and in combination, deriving Early, Middle and Late stage comparisons. Identified differentially expressed genes were then associated with enriched GO terms using BLAST2GO. For the leaf, up-regulated differentially expressed genes were strongly associated with GO terms only during the Early stage and the majority of GO terms were associated with only down-regulated genes at the Middle or Late stages. For the roots, few differentially expressed genes were identified at either Early or Middle stages. Only one replicate at 1% estimated water content produced high quality data for the root, however, this indicated a high level of differential expression. Again the majority of enriched GO terms were associated with down-regulated genes. The performance of the different analysis programmes and the annotations associated with identified differentially expressed genes is discussed.

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Perennial ryegrass (*Lolium perenne*) is a widely cultivated perennial forage and amenity grass in temperate areas worldwide and, as such, is either the major component or is a key constituent of many perennial pasture types. As a consequence of this wide distribution, perennial ryegrasses are exposed to, and have to respond to, a variety of environmental stresses, some seasonal, but others with increasingly unpredictable onsets as a consequence of changing climates. Reduced water availability, as a result of low rainfall or irrigation, is one such environmental stress which can have significant effects on the establishment and persistence of grasslands. This, in turn, can influence both agricultural productivity and the economics of farming as well as having broader impacts on the environment and grassland sustainability [1, 2]. Thus, there is a driver to understand perennial ryegrasses responses to water deficit at both whole plant and molecular levels and to integrate this information into the process of developing new perennial ryegrasses which have the ability to withstand and recover from periods of drought.

Numerous studies have reported relative changes in transcriptome profiles of model and crop plants in response to water-stress, often with a focus on the identification of differentially expressed genes (DEGs) distinguishing drought tolerant from susceptible genotypes. Such studies frequently indicate the likely roles and interplay of phytohormone signalling, specific transcription factors, receptor kinases, transporter and stress-response proteins (among others) in implementing the metabolic and physiological programmes required to deal with the onset of water-stress (e.g., [3–17]). An aspect of the majority of these studies is that, for obvious practical reasons, they have isolated, analysed and compared transcriptomes from the aerial parts of the plant. While this is experimentally more tractable, this has led to a relative lack of knowledge of the gene expression responses of roots to the progression of water stress. To address this, experimental systems have been developed which allow access to growing roots during the imposition of water stress, both for transcriptomic and physiological analyses. These include hydroponic systems, with osmotic stress being imposed using high molecular weight polyethylene glycol (PEG) [18–21] and the use of other suitable soil, sand [4, 13, 22, 23] and non-soil based hydrated media such as vermiculite [24–28], from which roots can be isolated relatively rapidly while minimising mechanical damage, at least for simpler root systems.

In addition to developing suitable experimental systems, there are also a number of contrasting analytical methods, implemented in different computational approaches, which can be taken to the identification of differential gene expression. A number of studies have compared the performance of different packages for identifying differentially expressed genes (DEGs) from RNAseq datasets [29–32] and made recommendations, taking into account factors such the structure of the overall analysis pipeline and the numbers of replicates. For instance, Schurch et al. [30] identified 5 programs that show a high level of inter-software consistency (DESeq [33], DESeq2 [34], EBSec [35], edgeR [36], limma [37]) and recommended the use of DESeq2 for experiments with fewer replicates per condition. Costa-Silva et al. [29] identified that combining different packages can refine sets of DEGs to make them more concordant with qRT-PCR data. The identification of numbers of DEGs within experiments is often given a broader biological relevance through the use of Gene Ontologies (GO). These allow the results of individual experiments to be contextualised within the body of existing experimental evidence to identify consistent patterns of gene expression associated with defined biological processes, molecular functions and cellular compartments. To enable this, a number of open-access and other resources have become available which allow for the assignment of functions or putative functions to genes or gene models through the use of sequence similarity searches in combination with direct or indirect experimental evidences (e.g., BLAST2GO [38]; The Gene Ontology Consortium [39] QuickGO [40])

Our overall aim in this study has been to develop a greater understanding of changes in gene expression patterns in leaves and roots of perennial ryegrass and, particularly, how these reflect underlying biological processes as indicated through GO terms. Additionally, we were interested in whether derived conclusions might be influenced by the use of different softwares. To approach this we have: A) generated leaf and root transcriptomes from a single genotype of perennial ryegrass at four points during a time course reflecting increasing water stress, B) compared changes in DEG expression patterns using 3 commonly used programmes, DESeq2, limma with voom transformation (hereafter referred to as limma-voom) and edgeR and C) identified enriched GO terms associated with groups of up- and down-regulated DEGs at the different time points, comparing DEGs identified by: i) the 3 programmes individually, ii) all the programmes, and iii) any of the programmes. The results are discussed in the contexts of the different analytical approaches and the biological responses to increasing water-stress.

Materials and methods

Plant material and drought treatments

A single, largely homozygous, genotype of perennial ryegrass (p226/135/16), derived from an inbred line was used for the RNAseq experiments. This is the same genotype used as the subject of the first published perennial ryegrass reference genome assembly [41]. All replicates were derived by clonal propagation (tillering) from this single genotype.

Throughout the experiment, plants were maintained in a 20°C growth room with an 8 hour photoperiod. Perennial ryegrass is a long-day plant and these conditions maintained the replicates in the vegetative growth stage (i.e., no induction to flowering).

Sixteen single tillers were taken from compost grown clonal replicates of p226/135/16, rinsed of compost, and transferred to containers of water until they showed new root growth, after about 6 days on average. At this point they were transferred to 90 mm pots containing vermiculite (graded for horticultural use, 2–5 mm) to establish and were watered with Hoagland's solution [42] twice a week. Once established, between 15 and 21 days after tillering, watering was stopped and water content was estimated (estimated water content; EWC) using a moisture meter HH2 Delta-T meter (AT Delta-T devices, Cambridge, UK). At each estimation 3 different moisture readings were taken and averaged. Leaves and roots were sampled at 35% (full watering, day 0) and when 15% (day 5 –no visible wilting), 5% (day 12 –visibly starting to wilt) and 1% (day 20 –marked wilting) EWC levels were reached (see S1 Fig for images of the replicates at the different EWC stages). Leaf samples were cut and flash frozen in liquid nitrogen and stored at -80°C. The roots were briefly washed with distilled water to remove the growing medium and then blotted dry prior to storage at -80°C. Four different clones were sampled at each EWC-point, deriving 4 biological replicates for each stage.

RNA extraction and sequencing

Total RNA was extracted from leaf material using the Trizol method (Sigma-Aldrich, Poole, UK) and quantified using Qubit fluorescence spectrophotometry. 1 µg of total RNA was used per sample for library construction according to the Illumina TruSeq Stranded mRNA Library Preparation Kit protocol. Samples were indexed such that 24 samples could be multiplexed per lane of a HiSeq2500 platform (2x126bp format). Samples were run across two high-output flowcells and reads demultiplexed using the bcl2fastq script.

RNAseq processing and quality control and mapping

Prior to mapping, raw reads were processed using Trimmomatic v.0.33 [43] to remove adapters using the following parameters (optimized after several run tests): ILLUMINACLIP:TruSeq3-PE-2.fa LEADING:15 SLIDINGWINDOW:4:15 MINLEN:30 HEADCROP:12, and the quality of resulting trimmed and cleaned reads was assessed using FastQC v.0.11 [44]. Reads were then mapped to perennial ryegrass (p226/135/16) reference genome assembly [41] using the splice-aware mapper hisat2 v2.0.0 [45] with default parameters to generate the bam files.

Pre-processing and quantification of transcripts. Prior to analysing the count matrices a pre-processing filtering was performed to remove potential artefacts and assess the quality of the replicates as described elsewhere [46, 47]. Count matrices were derived from bam files above using the *GenomicFeatures* and *GenomicAlignments* R libraries. Transcripts with a count lower than in any samples were discarded. We applied the regularized logarithm transformation (rlog) as implemented in the DESeq2 package to decrease the variance among gene expression values as proposed by Love et al. [34] and then calculated a distance matrix between samples and performed a principle component analysis (PCA) to quantify experimental covariates and batch effects among samples and replicates [48]. The R scripts implementing the pre-processing and generation of count matrices are available in [S1 Methods](#).

Estimating the completeness of transcriptomes. The transcriptome in each sample was assessed for its completeness as a measure of quality of the sequencing. Clean reads were mapped to the reference genome [41] using the splice-aware mapper Hisat2 v.2.0.0 [45] with default parameters and special options:—dta (required for downstream processing with StringTie)—phred33 (required to handle Illumina reads). Subsequently transcripts were assembled and merged using StringTie v1.1.0 [49] using default parameters. The completeness of each transcriptome was assessed using BUSCO [50] on the early_release plantdb set, composed of 1440 core genes.

Identification of DEGs using DESeq2, edgeR and limma-voom and deriving Jury 1 and Jury 3 DEG categories

Three different, count-based, methods were used to identify DEGs: DESeq2 [34], edgeR [36], and limma [51] with voom transformation [37]—limma-voom. In the case of DESeq2 the counts were transformed using the Relative Log Expression function [33][30] while in the case of edgeR and limma-voom the trimmed mean M-values were used [52]. Two scenarios were considered for the identification of DEGs referred to as *against reference* (AR) and *time-course* (TC). In the case of AR, the RNA samples taken at 35% EWC were considered the reference or control, hence samples at 15%, 5% and 1% EWC were compared against the 35% sample to generate the Early, Middle and Late comparisons, respectively. In the case of TC analyses the comparisons performed were: 35 vs 15% (Early), 15 vs 5% (Middle) and 5 vs 1% (Late), i.e., each sampling point was compared to the previous sampling point, the latter being considered to be the reference. Only genes with a 2-fold change in expression level and a 5% false discovery rate (FDR) were considered as significant (R scripts code are available in [S2–S4 Methods](#) for DESeq2, edgeR and limma-voom, respectively).

For each of the AR and TC categories, we further defined the Jury 1 (J1) and Jury 3 (J3) classes. For J1 we considered any gene as differentially expressed if identified as significant by any of the methods used, i.e., a gene with ≥ 2 -fold change in expression level and $\leq 5\%$ FDR in any of the following: DESeq2, edgeR or limma-voom. In the case of J3, only those genes identified as significant by all three softwares were considered.

Functional annotation of DEGs. The reference genome was functionally re-annotated using Blast2GO 5.25 (Pro) [38] as a prior step before computing GO term enrichments. The

functional annotation was done as follows: BLAST searches were performed on the nr database (release Jan 2017) using BLASTx command from ncbi-blast-2.2.28+ release [53] at an e-value cut-off of 0.000001 and selecting the top 20 hits. InterPro searches were performed using InterProScan v.5.18–57 [54] on TIGRFAM [55], PFAM [56], SMART [57], PANTHER [58], Gene3d [59] and PIRSF [60] databases. The annotation table is available in [S1 Table](#).

Partitioning of DEGs between expression categories. RNAseq was performed on samples extracted at the four EWC stages, 35%, 15%, 5% and 1% and DEGs were identified using DESeq2, edgeR and limma-voom and the combination categories J1 and J3 (as described above). DEGs identified in the Early, Middle and Late AR and TC comparisons were classified as to whether they were up- or down-regulated relative to the 35% EWC stage (AR) or the previous EWC stage (TC). This meant that every gene model could be assigned to one of 27 expression categories for AR and TC comparisons, describing whether it was significantly up-regulated, down-regulated or showed no significant change in expression levels at Early, Middle and Late comparison stages. For example, a gene model identified as up-regulated at the Early AR comparison, down-regulated at the Middle AR comparison and showing no significant change at the Late AR comparison is described as belonging to an expression category using the notation *AR-up_down_ns*.

Identification of enriched GO terms

After placing each gene model showing significant differential expression in at least one of the Early, Middle or Late comparisons in one of the AR and/or TC expression categories, significantly enriched GO terms were identified for each expression category using the Enrichment Analysis (Fisher's Exact Test) module in BLAST2GO with the FDR set at 0.05, and the GO terms associated with the annotated perennial ryegrass genome as the reference set. Subsequently, the gene models contributing to the significant enrichment of the individual GO terms associated with each expression category were extracted from the BLAST2GO output files.

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis

Enzyme activity codes were assigned to subsets of gene models using the BLAST2GO KEGG [61] interface. Where the numbers of up- and down-regulated DEGs assigned to particular enzyme codes within certain expression categories were compared, significance was estimated using Fisher's Exact Test.

Additional functional information on proteins

As described in the Results and Discussion in relation to particular DEGs, additional functional information was inferred from annotations using the Uniprot database[62].

Results

Pre-processing, mapping, and quality of sequencing and replicates

The RNA libraries were sequenced in a single batch yielding an average of around 15 and 10M reads per sample for shoot and root tissue respectively (Tables 1 and S2). The number of reads across replicates was fairly consistent with the exception of root tissues at 5 and 1% EWC where there were large differences among replicates. As a consequence root sample replicate 3 from 5% EWC and replicates 1, 3 and 4 from 1% EWC were not considered when computing summary statistics or DEGs (see below). Thus, the 1% EWC root sample was unreplicated. After removing adaptors, low quality and non-paired reads, the number of reads dropped by

Table 1. Sequencing and mapping statistics. The total number and number of paired reads shown as the aggregate value across all four replicates at different estimated water contents. Alignment rates are shown as the range among replicates. The mapped reads (bam files) were submitted to the European Nucleotide Archive (ENA) repository under study accession number PRJEB31812.

	Estimated water content			
	35%	15%	5%	1%
Leaf				
Total reads	71830102	61024525	86519927	67957287
Paired Reads	52433026	45412817	60453862	46278550
Alignment rate (%)	95–96	94–95	95–95	87–95
Root				
Total reads	31810877	54311559	27220469	11123223
Paired Reads	26154592	43984322	21061835	8504146
Alignment rate (%)	77–91	91–95	95–96	92

¹Average/range of 3 biological replicates

²One biological replicate

<https://doi.org/10.1371/journal.pone.0220518.t001>

around 13% on average (Tables 1 and S2). The completeness of the transcriptome was estimated using BUSCO (Table 2); the average number of core genes represented in the transcriptome, including complete and partial genes, was c. 93% for the leaf samples and 88% for the root samples (excluding the low quality root replicates).

The variability among sampling point replicates are shown in Figs 1 and S2 and S3. For the shoot samples (S2 Fig), the distributions of rlog values were homogeneous across the 16 samples (four EWC points and four replicates) indicating consistent biological replication. As referred to above, there was more variation among the root samples at 5% and 1% EWC (S3 Fig), leading to the omission of some of the replicates from the DEG analysis. The analysis of variability among replicates in the form of PCA and heat maps using the count matrices showed that shoot replicates within each EWC sampling point clustered together and the first two components of the PCA accounts for up to 69% of the overall variability (Fig 1). Thus, change in EWC as the experiment progressed was the main driver explaining the variability between EWC sampling points. This analysis also indicated that samples taken at 1% EWC were the most dissimilar, and that 35% and 15% EWC sampling points were the most similar. The PCA and heat map plots generated from the root data, by contrast, indicated a higher degree of variability between replicates and a greater degree of overlap between the 35% and

Table 2. Transcriptome completeness. The % completeness of the different transcriptomes at the different estimated water contents (EWCs) as estimated using the BUSCO core gene set.

Tissue	EWC (%)	BUSCO		
		Complete	Partial	Missing
Leaf	35	84.41	7.11	8.47
	15	84.41	10.67	7.43
	5	84.73	6.90	8.37
	1	85.04	7.32	7.64
Root	35	81.90	8.05	10.04
	15	84.52	7.32	8.16
	5 ¹	79.81	7.64	12.55
	1 ²	71.15	9.31	19.54

¹Average of 3 biological replicates

²One biological replicate

<https://doi.org/10.1371/journal.pone.0220518.t002>

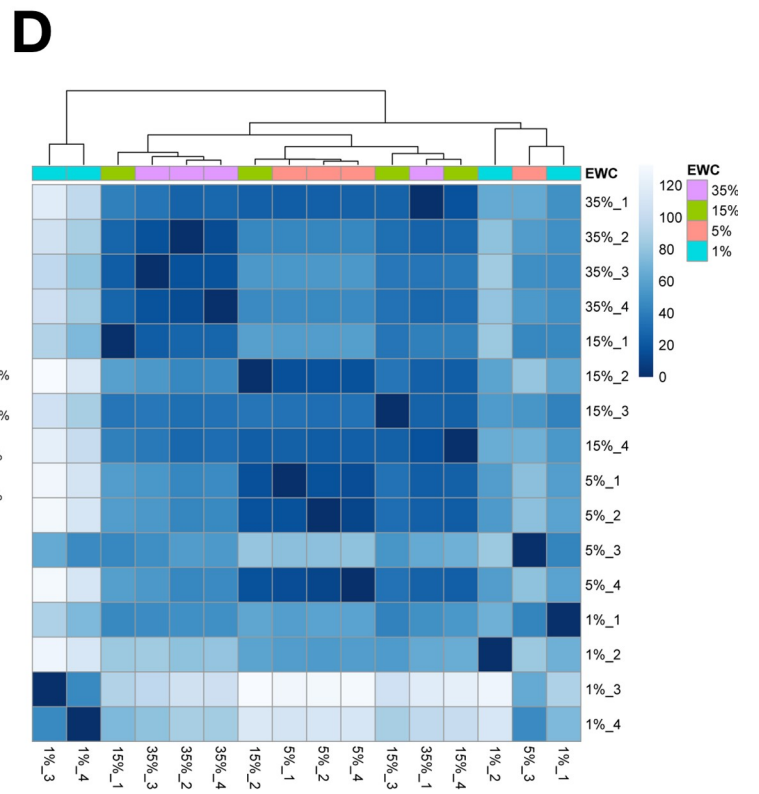
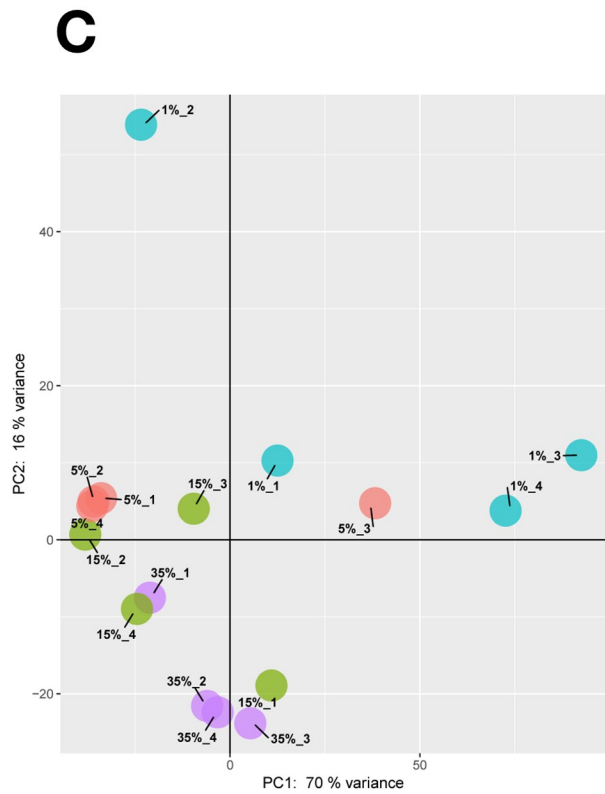
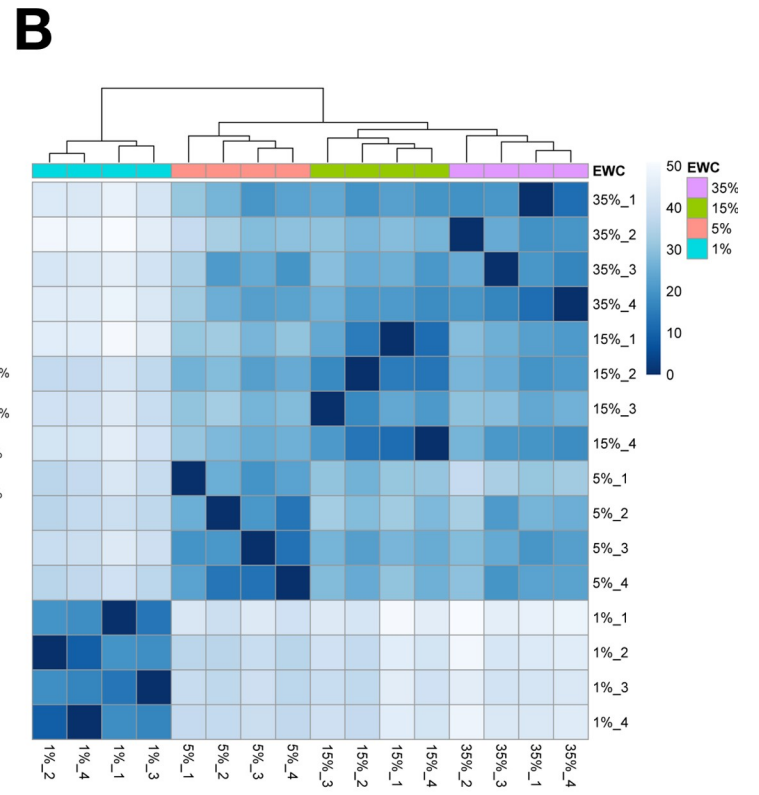
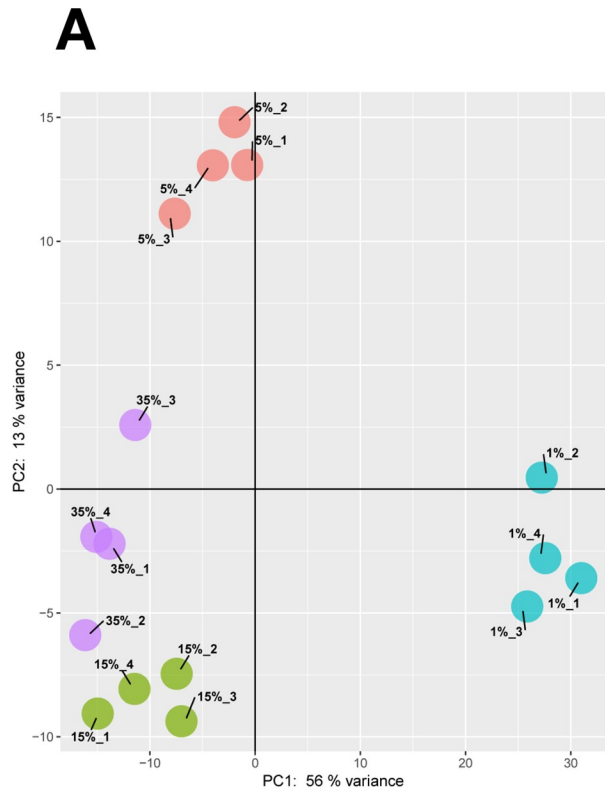


Fig 1. Variability among replicates from leaf and root transcriptomes. Panels A and C show PCA and heat map plots for shoot and root tissues respectively and Panels B and D portrays the heat maps computed with Euclidian distance within samples for shoot and root tissues respectively. Samples at EWC 35, 15, 5, and 1% are coloured in purple, green, red and cyan respectively.

<https://doi.org/10.1371/journal.pone.0220518.g001>

15% EWC sampling points (Fig 1). Three of the 4 replicates at 5% EWC were very consistent, though the fourth was an outlier that was not used. The single replicate retained from the 1% EWC sampling point was distinctly separated from the other retained sampling point replicates, particularly on the basis of the second principle component.

Differentially expressed gene models

A summary of the numbers of DEGs detected at the different comparison stages and by the 3 analysis programmes is given in Table 3. As a trend, the number of DEGs detected at each stage comparison increased as the experiment proceeded in both AR and TC comparisons, varying between 642 DEGs (J3, Early AR/TC) to 3762 (J1, Late, AR). The pattern of DEG identification in the root transcriptomes contrasted markedly with the leaf transcriptomes particularly in reference to the performance of the different analysis programmes. Only 5 DEGs were called in the Early comparison stage, all by DESeq2; in the Middle AR comparison stage a total of 844 DEGs were called, but again only by DESeq2 (with the exception of 1 DEG also called by limma-voom). For the Late comparison stage, as described earlier, 3 of the 4 replicates generated low or very low alignments with the reference genome and only 1 of the replicates was useable. However, from this one sample, between 3380 and 3673 DEGs were identified by DESeq2, edgeR and limma-voom which indicates substantial differential expression as the roots approached this end stage. Thus, major changes in differential expression in the leaf transcriptome occurred over the duration of the experiment, whereas major changes in differential expression in the root transcriptome did not occur in the Early and Middle comparison stages and were detected only during the most severe water stress (bearing in mind this conclusion is based on an unreplicated result for the Late stage comparison).

Comparison of individual DEGs identified at different stages and through different analysis methods

Table 4 indicates the frequencies with which individual DEGs were called across the different combinations of methods. For the leaf transcriptome, between 61% and 80% of the DEGs included within J1 were called by all of the methods (J3). Between 3 and 5% of the DEGs

Table 3. The number of significant differentially expressed genes detected by each analysis method for each expression comparison.

	Comparison	Analysis method				
		J1	J3	DESeq2	edgeR	limma-voom
Leaf	AR/TC Early	1050	642	934	754	774
	AR Middle	1838	1195	1564	1450	1494
	TC Middle	1860	1290	1622	1502	1553
	AR Late	3762	3024	3311	3352	3497
	TC Late	2701	2065	2397	2334	2403
Root	AR/TC Early	5	0	5	0	0
	AR Middle	844	0	844	0	1
	TC Middle	186	1	186	4	1
	AR Late	3932	2903	3673	3344	3380
	TC Late	2757	1787	2622	2196	2189

<https://doi.org/10.1371/journal.pone.0220518.t003>

Table 4. The total number of differentially expressed genes (J1) identified at each comparison stage and the percentage of these identified by all possible combinations of the analysis methods DESeq2 (D), edgeR (E) and limma-voom (L).

	Comparison	Combinations of analysis methods							
		J1 n	D,E,L (J3) %J1	D,E %J1	D,L %J1	E,L %J1	D %J1	E %J1	L %J1
Leaf	AR/TC Early	1050	61	8	2	2	18	1	8
	AR Middle	1838	65	9	3	3	8	2	10
	AR Late	3762	80	4	2	4	2	1	7
	TC Middle	1860	69	9	2	2	7	1	10
	TC Late	2701	76	7	2	3	4	1	8
Root	AR/TC Early	5	0	0	0	0	100	0	0
	AR Middle	844	0	0	0.1	0	100	0	0
	AR Late	3932	74	7	3	3	10	1	3
	TC Middle	186	1	2	0	0	98	0	0
	TC Late	2757	65	10	4	2	17	1	2

<https://doi.org/10.1371/journal.pone.0220518.t004>

included in J1 were called by combinations of just 2 of the programmes and between 3 and 9% were called by only one of the programmes—primarily DESeq2 or limma-voom. The root transcriptome again contrasted with the leaf transcriptome. As described earlier, few DEGs were called by any of the programmes in the Early and Middle stages and by far the majority of those that were, were called uniquely by DESeq2. For the single sample Late stage comparisons, in which a total of 3932 and 2757 DEGs were included in J1 for the AR and TC comparisons, respectively, 74% (AR) and 65% (TC) were called by all of the programmes (J3). These levels are comparable to those obtained for the leaf transcriptomes.

DEG expression patterns across early, middle and late AR and TC stage comparisons and contribution to GO category enrichment

For the shoot transcriptomes, when the Early, Middle and Late comparisons were considered in sequence and in combination with whether a particular gene was significantly up- or down-regulated or showed no-significant relative change in expression level, 27 different categories could be derived for both AR and TC. Table 5 describes the distribution of genes across 26 of these categories (excluding the category for genes which were not called as significantly differentially expressed for any of the stage comparisons—*ns_ns_ns* in table nomenclature) as well as the number of DEGs within each category that contributed to enriched GO terms for the shoot transcriptomes. Firstly, while there were variations in outcomes for the different analysis programmes individually and in J1 and J3 combinations, overall the trends were similar. For the AR comparisons, it was mostly the expression categories which contained down-regulated but not up-regulated DEGs that could be associated with enriched GO terms. For example, category *ns_ns_down* for J3 contained 1153 DEGs, 88% of which could be assigned to 49 significantly enriched GO terms; category *ns_ns_up* for J3 contained 1016 DEGs, only 4% of these could be assigned to a single enriched GO term. The *up_ns_ns* category is the exception to this; for J3, 145 DEGs were present in this category, 77% of which contributed to 41 enriched GO terms. The TC comparisons were similar in terms of the overall trend for expression categories containing exclusively down-regulated DEGs being more likely to be associated with enriched GO terms.

Table 6 contains the equivalent information for the root transcriptomes. The data is also presented in terms of just the possible 8 combinations (excluding the *ns_ns* category) of the Early and Middle comparisons for which there is replication. For these Early and Middle comparisons, substantial numbers of DEGs were only detected by DESeq2, but a high proportions of these

Table 6. The total number of root differentially expressed genes at each expression category for each analysis method and the percentage of these DEGs associated with the total number enriched gene ontology (GO) for that expression category.

AR/TC	Expression category ¹	J1			J3			DESeq2			edgeR			limma-voom		
		DEGs		GO terms	DEGs		GO terms	DEGs		GO terms	DEGs		GO terms	DEGs		GO terms
		Total	%		Total	%		Total	%		Total	%		Total	%	
AR	ns_up	314	70	9	0	-	-	314	70	9	0	-	-	0	-	-
	ns_down	528	88	26	1	-	-	528	88	26	0	-	-	1	-	-
	up_ns	3	-	-	0	-	-	3	-	-	0	-	-	0	-	-
	up_up	2	-	-	0	-	-	2	-	-	0	-	-	0	-	-
	ns_down_down	330	89	35	0	-	-	329	89	34	0	-	-	1	-	0
	ns_down_ns	195	70	16	0	-	-	197	70	16	0	-	-	0	-	-
	ns_ns_down	1761	89	79	1376	85	69	1624	90	84	1478	84	68	1493	84	65
	ns_ns_up	1805	45	11	1527	68	15	1556	66	14	1614	50	13	1861	45	11
	ns_up_up	156	53	8	0	-	-	154	53	9	0	-	-	0	-	-
	ns_up_down	8	-	0	0	-	-	8	-	0	0	-	-	0	-	-
	ns_down_up	3	-	0	0	-	-	2	-	0	0	-	-	0	-	-
	ns_up_ns	150	67	5	0	-	-	152	68	5	0	-	-	0	-	-
	up_ns_ns	3	-	0	0	-	-	3	-	0	0	-	-	0	-	-
	up_up_ns	2	-	0	0	-	-	2	-	0	0	-	-	0	-	-
TC	ns_up	43	-	-	0	-	-	43	-	-	2	-	-	0	-	-
	ns_down	144	83	24	1	-	-	144	83	24	3	-	-	1	-	-
	up_ns	5	-	-	0	-	-	5	-	-	0	-	-	0	-	-
	ns_down_down	21	29	1	0	-	-	21	29	1	0	-	-	0	-	-
	ns_down_ns	120	86	30	0	-	-	120	86	30	3	-	0	0	-	-
	ns_ns_down	1625	90	67	942	85	50	1514	90	65	1275	89	62	1048	84	49
	ns_ns_up	1274	48	12	845	51	12	1071	54	15	920	50	13	1141	44	9
	ns_up_up	9	-	0	0	-	-	8	-	0	0	-	-	0	-	-
	ns_up_down	6	-	0	0	-	-	6	-	0	1	-	0	0	-	-
	ns_down_up	2	-	0	0	-	-	2	-	0	0	-	-	0	-	-
	ns_up_ns	28	-	0	0	-	-	29	-	0	0	-	-	0	-	-
up_ns_ns	5	-	0	0	-	-	5	-	0	0	-	-	0	-	-	

¹Expression categories which contained 0 DEGs have been omitted for clarity.

<https://doi.org/10.1371/journal.pone.0220518.t006>

could be associated with GO terms for the AR-*ns_up*, AR-*ns_down* and TC-*ns_down* expression categories. The *ns_down* categories were associated with c. 3 times as many GO terms as the *ns_up* categories. When the Early, Middle and Late comparisons were included, by far the majority of DEGs were contained within the AR/TC-*ns_ns_down* and *ns_ns_up* expression categories and for both of these categories across all the analysis methods a high proportion of the associated DEGs contributed to enriched GO terms though, noticeably, the *ns_ns_down* comparisons were associated with 6–7 times more GO terms than the *ns_ns_up* comparisons. Comparing AR and TC results for the Early and Middle root transcriptome comparisons, the numbers of DEGs called in the TC comparisons were only 23% of those called in the AR comparisons (DESeq2). For the Early, Middle and Late comparisons, the equivalent figure was about 60%.

Enriched GO terms and related gene models in relation to expression categories

Leaf transcriptome. Overall, when combining all of the 5 approaches (J1, J3, DESeq2, edgeR and limma-voom) 123 different GO terms were significantly enriched when the data

was analysed according to the 26 expression categories described in Table 5. Of these, 91 GO terms were enriched in the 2 expression categories *AR-up_ns_ns* and *AR-ns_ns_down*, and the orders of GO terms in Tables 7 and S3 have been adjusted to reflect this. The relationship between different leaf MF GO terms is illustrated in Fig 2 and for all MF, BP and CC GO terms illustrated in S4 Fig. Results for enriched GO terms and associated gene models in Groups 1–4 (Table 7) are described below. Details of GO terms and associated gene models in Groups 5–8 (S3 Table) can be found in S1 Results.

Considering GO terms enriched in *AR-up_ns_ns* but not in *AR-ns_ns_down* (Table 7, Groups 1 and 2), 2 related groups of GO terms were present: a) 7 MF GO terms associated with nucleoside/nucleotide binding, which were detected just by J1 and, b) 29 BP and MF GO terms associated with positive regulation of cellular processes and DNA binding/transcription factor activity, the majority of which were detected by all the analysis methods. The most-specific child GO term (referred to as the child GO term from this point onwards) for Group 1 was *ATP binding* (GO:0005524) which is clearly fundamental to many plant metabolic processes. Two of the 5 associated gene models (i.e., the 5 gene models which constituted the significant enrichment of the GO categories in Group 1 using BLAST2GO) have been annotated as ‘disease resistance’ proteins (S4A Table). The enrichment for these GO terms was only supported using one of the analysis approaches (J1), suggesting caution in interpretation of the significance. Group 2 contained 37 different gene models. The child GO terms for Group 2 were *positive regulation of transcription by RNA polymerase II* (BP, GO:0045944), *core promoter sequence-specific DNA binding* (MF, GO:0001046) and *DNA-binding transcription activator activity, RNA polymerase II-specific* (MF, GO:0001228). Not surprisingly, many of the associated gene models have been annotated as being potential transcription factors (S4A Table), though, this group also contained gene models associated with histones H1 and CENH3 as well as DNA and RNA polymerases. The majority of the Group 1 and 2 GO terms were largely absent from the other expression categories, indicating that the up-regulation of the associated gene models in the Early stage may have been quite specific to the onset of water-stress (at least for Group 2). However, three of the same GO terms were also enriched in expression category *TC-up_down_up*. Of the associated 17 gene models which constituted the enrichment in this case, 4 were in common with *AR-up_ns_ns* Group 2, including the Histone1 and CENH3, but the remaining 13 represent a different set of (largely) transcription factors, 6 of which were annotated as containing a WRKY domain and 2 heat stress/shock factors. WRKY domain and heat stress/shock containing transcription factors play multiple roles in plant stress responses and the high proportion of these in this group, in combination with the heat stress/shock factors, indicates a potential relevance to drought response.

Group 3 was enriched for BP, CC and MF GO terms in expression category *AR-up_ns_ns* and in many expression categories which showed down-, but not subsequent up-regulation in Middle or Late comparison stages (Tables 7 and S3 and Figs 2 and S4). The child term for the BP, CC and MF GOs were *cellular protein modification process* (GO:0006464), *plasma membrane* (GO:0005886) and *kinase activity* (GO:0016301), respectively, indicating an up-regulation of gene models involved in processes at the plasma membrane during the onset of water-stress—followed by a down-regulation of similar processes as the water-stress becomes more severe—(though not down-regulation of the same gene models significantly below the AR or TC reference significance levels). Group 3 *AR-up_ns_ns* contained 168 gene models within the enriched GO terms, compared to 1117 gene models for expression categories which showed down-, but not subsequent up-regulation in Middle or Late comparison stages. The annotations of the gene models indicated a representation of potential stress-response, transporter, transcription factor and kinase activities associated with both up- and down-regulated expression categories. While, both up- and down-regulated expression categories contained some

Table 7. Enriched gene ontology (GO) terms associated with selected expression categories of shoot differentially expressed genes.

Group ¹	Enriched GO term				Analysis method ²	
	i.d.	Category ³	Position ⁴	Description	AR-unn ⁵	AR-nnd ⁵
1	GO:0097367	MF	5	carbohydrate derivative binding	J1	-
	GO:0008144	MF	6	drug binding	J1	-
	GO:0017076	MF	6	purine nucleotide binding	J1	-
	GO:0032553	MF	6	ribonucleotide binding	J1	-
	GO:0030554	MF	7	adenyl nucleotide binding	J1	-
	GO:0032555	MF	7	purine ribonucleotide binding	J1	-
	GO:0035639	MF	7	purine ribonucleoside triphosphate binding	J1	-
	GO:0032559	MF	8	adenyl ribonucleotide binding	J1	-
	GO:0005524	MF	9	ATP binding	J1	-
2	GO:0009893	BP	5	positive regulation of metabolic process	J3,E	-
	GO:0009891	BP	6	positive regulation of biosynthetic process	J1,J3,D,E,L	-
	GO:0010604	BP	6	positive regulation of macromolecule metabolic process	J1,J3,D,E,L	-
	GO:0051173	BP	6	positive regulation of nitrogen compound metabolic process	J1,J3,D,E,L	-
	GO:0031325	BP	6	positive regulation of cellular metabolic process	J3,E	-
	GO:0010557	BP	8	positive regulation of macromolecule biosynthetic process	J1,J3,D,E,L	-
	GO:0031328	BP	8	positive regulation of cellular biosynthetic process	J1,J3,D,E,L	-
	GO:0045935	BP	8	positive regulation of nucleobase-containing compound metabolic process	J1,J3,D,E,L	-
	GO:0010628	BP	9	positive regulation of gene expression	J1,J3,D,E,L	-
	GO:0051254	BP	9	positive regulation of RNA metabolic process	J1,J3,D,E,L	-
	GO:1902680	BP	10	positive regulation of RNA biosynthetic process	J1,J3,D,E,L	-
	GO:1903508	BP	11	positive regulation of nucleic acid-templated transcription	J1,J3,D,E,L	-
	GO:0006366	BP	11	transcription by RNA polymerase II	J3	-
	GO:0045893	BP	12	positive regulation of transcription, DNA-templated	J1,J3,D,E,L	-
	GO:0006357	BP	12	regulation of transcription by RNA polymerase II	J3	-
	GO:0045944	BP	13	positive regulation of transcription by RNA polymerase II	J1,J3,D,E,L	-
	GO:0008289	MF	3	lipid binding	J1,D,L	-
	GO:0001067	MF	5	regulatory region nucleic acid binding	J1,J3,D,E,L	-
	GO:0003677	MF	5	DNA binding	J1,D,E,L	-
	GO:0003690	MF	6	double-stranded DNA binding	J1,J3,D,E,L	-
	GO:0044212	MF	6	transcription regulatory region DNA binding	J1,J3,D,E,L	-
	GO:0140110	MF	6	transcription regulator activity	J1,J3,D,E,L	-
	GO:1990837	MF	7	sequence-specific double-stranded DNA binding	J1,J3,D,E,L	-
	GO:0001047	MF	8	core promoter binding	J1,J3,D,E,L	-
	GO:0001046	MF	9	core promoter sequence-specific DNA binding	J1,J3,D,E,L	-
	GO:0001228	MF	9	DNA-binding transcription activator activity, RNA polymerase II-specific	J1,J3,D,E,L	-
	GO:0003700	MF	11	DNA-binding transcription factor activity	J1,J3,D,E,L	-
	GO:0000976	MF	14	transcription regulatory region sequence-specific DNA binding	J1,J3,D,E,L	-
	GO:0000981	MF	15	DNA-binding transcription factor activity, RNA polymerase II-specific	J1,J3,D,E,L	-
	3	GO:0009987	BP	2	cellular process	J1,J3,D,E,L
GO:0006807		BP	3	nitrogen compound metabolic process	J1,D,E	-
GO:0071704		BP	3	organic substance metabolic process	J1,D,E	J1,J3,D,E,L
GO:0044238		BP	4	primary metabolic process	J1,J3,D,E	J1,J3,D,E,L
GO:0044237		BP	4	cellular metabolic process	J1,J3,D,E	J3,L
GO:0044260		BP	5	cellular macromolecule metabolic process	J1,D,E	-
GO:0043412		BP	5	macromolecule modification	J1,J3,D,E	J3,L
GO:0044267		BP	6	cellular protein metabolic process	J1,D,E	J3,L
GO:0036211		BP	6	protein modification process	J1,J3,D,E	J3,L

(Continued)

Table 7. (Continued)

Group ¹	Enriched GO term				Analysis method ²	
	i.d.	Category ³	Position ⁴	Description	AR-unn ⁵	AR-nnd ⁵
	GO:0006464	BP	7	cellular protein modification process	J1,J3,D,E	J3,L
	GO:0016020	CC	3	membrane	J1,J3,L	J1,J3,D,E,L
	GO:0071944	CC	4	cell periphery	J1,J3,D,E,L	J1,J3,D,E,L
	GO:0005886	CC	5	plasma membrane	J1,J3,D,E,L	J1,J3,D,E,L
	GO:0005488	MF	2	binding	J1,J3,D,E,L	J1,J3,D,E,L
	GO:0016740	MF	3	transferase activity	J1,E,L	J1,J3,D,E,L
	GO:0097159	MF	3	organic cyclic compound binding	J1,J3,D,E,L	J1,J3,D,E,L
	GO:1901363	MF	3	heterocyclic compound binding	J1,J3,D,E,L	J1,J3,D,E,L
	GO:0016772	MF	4	transferase activity, transferring phosphorus-containing groups	J1,J3,D,L	J1,J3,D,E,L
	GO:0016301	MF	5	kinase activity	J1,J3,D,L	J1,J3,D,E,L
4	GO:0008152	BP	2	metabolic process	-	J1,J3,D,E,L
	GO:0050896	BP	2	response to stimulus	-	J1,J3,D,E,L
	GO:0009056	BP	3	catabolic process	-	J1,J3,D,E,L
	GO:0009058	BP	3	biosynthetic process	-	J1,J3,D,E,L
	GO:0009719	BP	3	response to endogenous stimulus	-	J1,J3,D,E,L
	GO:0019748	BP	3	secondary metabolic process	-	J1,J3,D,E,L
	GO:0005975	BP	4	carbohydrate metabolic process	-	J1,J3,D,E,L
	GO:0006091	BP	4	generation of precursor metabolites and energy	-	J1,J3,D,E,L
	GO:0006629	BP	4	lipid metabolic process	-	J1,J3,D,E,L
	GO:0015979	BP	4	photosynthesis	-	J1,J3,D,E,L
	GO:1901564	BP	4	organonitrogen compound metabolic process	-	J3
	GO:0019538	BP	5	protein metabolic process	-	J3
	GO:0005576	CC	2	extracellular region	-	J1,J3,D,E,L
	GO:0005623	CC	2	cell	-	J1,J3,D,E,L
	GO:0043226	CC	2	organelle	-	J3,D
	GO:0044464	CC	3	cell part	-	J1,J3,D,E,L
	GO:0030312	CC	5	external encapsulating structure	-	J1,J3,D,E,L
	GO:0043227	CC	5	membrane-bounded organelle	-	J3,D
	GO:0005737	CC	6	cytoplasm	-	D
	GO:0005618	CC	6	cell wall	-	J1,J3,D,E,L
	GO:0009579	CC	6	thylakoid	-	J1,J3,D,E,L
	GO:0043229	CC	6	intracellular organelle	-	J3,D
	GO:0012505	CC	6	endomembrane system	-	J3,D,E
	GO:0043231	CC	7	intracellular membrane-bounded organelle	-	J3,D
	GO:0044444	CC	7	cytoplasmic part	-	J3,D
	GO:0005794	CC	8	Golgi apparatus	-	J1,D,E
	GO:0009536	CC	8	plastid	-	J1,J3,D,E,L
	GO:0003824	MF	2	catalytic activity	-	J1,J3,D,E,L
	GO:0005215	MF	2	transporter activity	-	J1,J3,D,E,L
	GO:0016787	MF	3	hydrolase activity	-	J1,J3,D,E,L
	GO:0030246	MF	3	carbohydrate binding	-	J1,J3,D,E,L
	GO:0036094	MF	4	small molecule binding	-	J1,J3,D,E,L
GO:1901265	MF	4	nucleoside phosphate binding	-	J1,J3,D,E,L	
GO:0000166	MF	5	nucleotide binding	-	J1,J3,D,E,L	

¹Only Groups 1–4 are shown. Full details on all Groups 5–8 are given in S3 Table

²Analysis methods identifying significant DEGs associated with the indicated enriched GO term and expression category. D = DESeq2, E = edgeR and L = limma-voom.

³MF = Molecular Function; BP = Biological Process; CC = Cellular component.

⁴Ascending numbers indicate a more specific category in the GO terms hierarchies.

⁵Abbreviated expression categories. u = up; d = down; n = ns.

annotations relating to plant hormones, the down-regulated categories included a number of auxin and ethylene associated transcription factors as well as gibberellin and cytokinin metabolic enzymes. The down-regulated gene expression category also contained 51 gene models annotated as 'chloroplastic' compared to a single gene model in the up-regulated expression category.

Group 4 consisted of GO terms which were not enriched in the AR-up_ns_ns expression category but were enriched across most of the analysis methods in the AR-ns_ns_down expression category as well as, to a lesser extent, other down regulated expression categories in the Middle or Late comparison stages. The child terms for these GO categories were, BP—protein metabolic process (GO:0019538), CC—Golgi apparatus (GO:0005794) and plastid (GO:0009536), and MF—nucleotide binding (GO:0000166) (Tables 7 and S3 and Figs 2 and S4). Together, these GO terms suggest that, particularly at the Late comparison stage, there is a down regulation of genes associated with metabolism more generically, i.e. affecting secondary metabolic, carbohydrate, lipid, photosynthetic and protein metabolic processes (BP terms) along with a reduction in nucleotide binding activity (i.e., ATP-binding). The most specific CC terms enriched were *Golgi apparatus* and *plastid*, which is compatible with reductions in

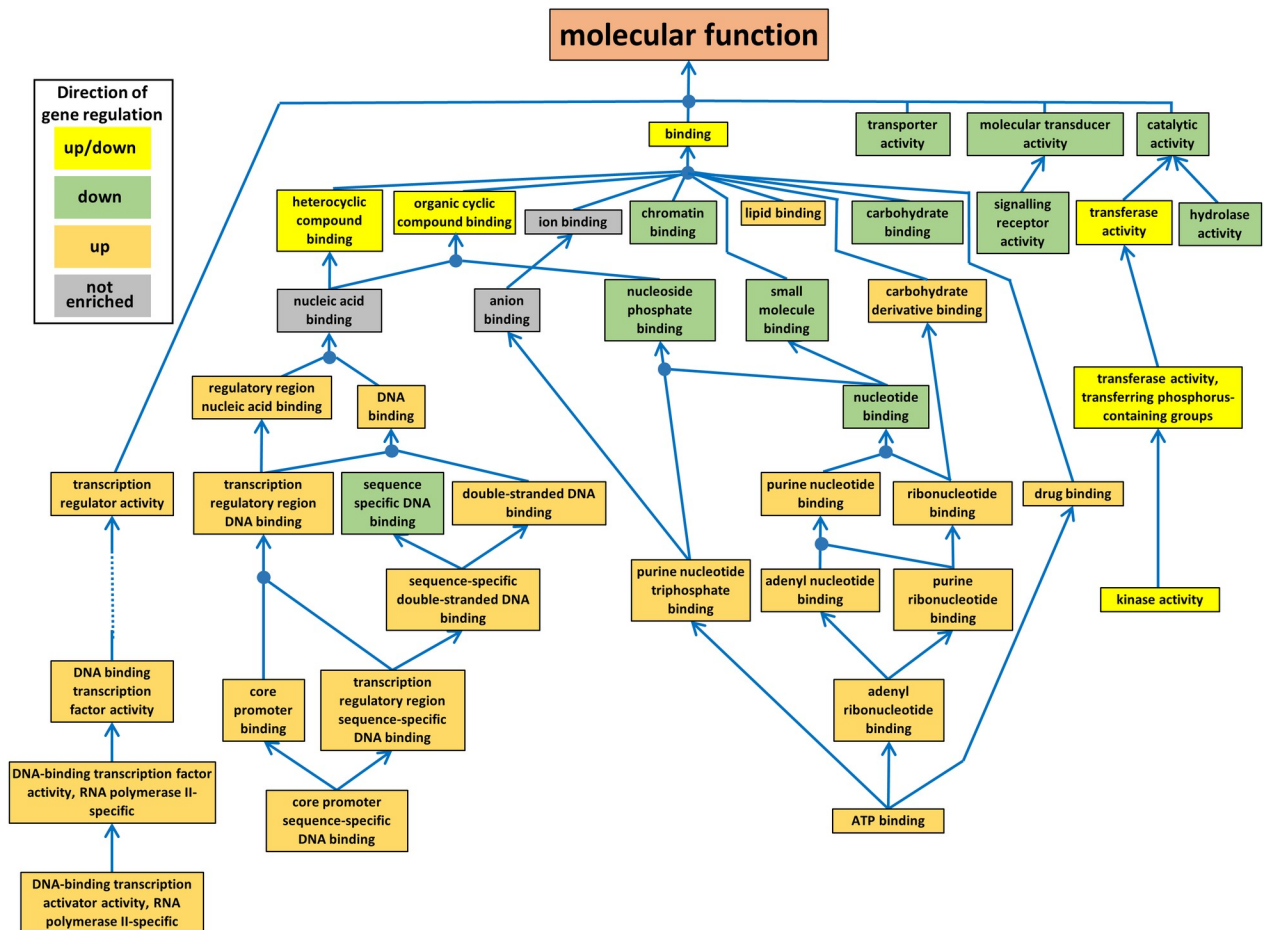


Fig 2. Enriched gene ontology (GO) terms associated with the Molecular Function GO category for shoot differentially expressed genes. The colour of the box indicates whether the enriched GO terms was associated with just up-regulated, just down-regulated or both up- and down-regulated DEGs across the 4 sampling points (35%, 15%, 5% and 1% EWC). Blue arrows indicate the direction of the GO hierarchy from child to parent and blue circles indicate where two or more child GO terms are associated with a single parent GO term.

<https://doi.org/10.1371/journal.pone.0220518.g002>

post-translational modification of proteins, protein trafficking and photosynthetic activity likely to accompany a reduction in overall cell metabolic activity. Group 4 enriched GOs contained 2124 gene models, 193 of which contained the annotation 'kinase'. Additional relatively highly represented terms included 'chloroplastic', 'synthase', 'transporter', 'binding', and 'resistance', all of which occurred in more than 50 of the gene models (S4A Table).

Root transcriptomes. Analysis of the root transcriptome identified 31 different enriched GO terms across the 8 Early and Middle expression categories and 110 enriched GO terms across the 26 Early, Middle and Late expression categories (described in Tables 8 and 9 and Figs 3 and S5 Expression Categories Roots A and B). For the Early and Middle comparisons (Table 8 and Fig 3), 3 Groups of GO terms were identified, relating to up-regulation (Group 1), up and down-regulation, (Group 2) and down-regulation at the Middle stage (Group 3). No GO terms were associated with early stage up- or down-regulation. As described earlier, all GO terms were associated with DEGs called only by DESeq2. Only 2 GO terms were contained within Group1, *plasma membrane* (CC, GO:0005886) and *transporter activity* (MF, GO:0005215). Of the associated 58 gene models, 39 had annotations containing 'transporter', 'transport', 'antiporter', 'permease' or 'channel' and a further 8 were described as membrane transport proteins in Uniprot, including 3 aquaporins. Annotations also included 4 auxin transporters and one auxin induced transmembrane protein (S4B Table).

Group 2 contained 7 enriched GO terms, with child GO terms of *carbohydrate metabolic process* (BP, GO:0005975), *cell wall* (CC, GO:0005618) and *catalytic activity* (MF, GO:0003824) (Table 8 and Fig 3). Group 2 AR-ns_up enriched GO terms contained 219 gene models. Focussing on the *carbohydrate metabolic process* and *cell wall* GO categories, the most obvious difference between the up- and down-regulated gene models was the number of annotations associated with 'germin' or 'peroxidase'. Of the 43 up-regulated gene models in these categories, none were annotated as 'germin' and only 2 as 'peroxidase'; in the down-regulated gene models, 46 (40%) were annotated as 'germin' and 22 (19%) as 'peroxidase'. This indicates substantially different metabolic activities for the up- and down-regulated gene-models within Group 2 acting on the cell wall (S4B Table).

Group 3a/ab (Table 8 and Fig 3) consisted of 439 down-regulated gene models with child GO terms of *photosynthesis* (BP, GO:0015979), *plastid* (CC, GO:0009536) and *carbohydrate binding* (MF, GO:0030246). Of the 439 gene models, 83 (19%) contained the annotation 'chloroplastic' or 'chlorophyll'. The average count data for these 83 gene models at 35% EWC from the root data was c. 90x less than that for the same gene models at the same stage from the shoot data. Thus, the enrichment for these GO terms may not be particularly significant in a biological context. When the chloroplast-associated gene models were omitted from the AR- and TC-ns_down expression categories, the enriched GO terms that remained (Group 3b/ab) were *metabolic process* (BP, GO:0008152), *cellular process* (BP, GO:0009987), *response to stimulus* (BP, GO:0050896), *catabolic process* (BP, GO:0009056), *response to stress* (BP, GO:0006950), *generation of precursor metabolites and energy* (BP, GO:0006091), *binding* (MF, GO:0005488) and *carbohydrate binding* (MF, GO:0030246)—and a further 2 were identified: *transferase activity, transferring phosphorus-containing groups* (MF, GO:0016772) and *kinase activity* (MF, GO:0016301). While most of these Group 3b/ab GO terms are quite general, these latter two groups were more defined and contained 53 gene models, the majority of which were annotated as receptor-kinases (S4 Table). This indicates a likely transmembrane, cell-surface location and down-regulation of cell-signalling functions.

Details of GO terms relating to different expression categories and analysis methods when the unreplicated 1% EWC stage for the root data is included can be found in S2 Results.

Table 8. Enriched gene ontology (GO) terms associated with selected expression categories of root differentially expressed genes.

Group ¹	Enriched GO term				Analysis method ²		
	i.d.	Category ³	Position ⁴	Description	AR-nu ⁵	AR-nd ⁵	TC-nd ⁵
1	GO:0005886	CC	5	plasma membrane	J1,D	-	-
1	GO:0005215	MF	2	transporter activity	J1,D	-	-
2	GO:0005975	BP	4	carbohydrate metabolic process	J1,D	J1,D	J1,D
2	GO:0005576	CC	2	extracellular region	J1,D	J1,D	J1,D
2	GO:0016020	CC	3	membrane	J1,D	J1,D	-
2	GO:0071944	CC	4	cell periphery	J1,D	J1,D	-
2	GO:0030312	CC	5	external encapsulating structure	J1,D	J1,D	J1,D
2	GO:0005618	CC	6	cell wall	J1,D	J1,D	J1,D
2	GO:0003824	MF	2	catalytic activity	J1,D	J1,D	J1,D
3ab	GO:0008152	BP	2	metabolic process	-	J1,D	J1,D
3ab	GO:0009987	BP	2	cellular process	-	J1,D	J1,D
3ab	GO:0050896	BP	2	response to stimulus	-	J1,D	J1,D
3a	GO:0019748	BP	3	secondary metabolic process	-	J1,D	-
3a	GO:0009628	BP	3	response to abiotic stimulus	-	J1,D	J1,D
3ab	GO:0009056	BP	3	catabolic process	-	J1,D	J1,D
3a	GO:0044237	BP	3	cellular metabolic process	-	J1,D	J1,D
3ab	GO:0006950	BP	3	response to stress	-	J1,D	-
3ab	GO:0006091	BP	4	generation of precursor metabolites and energy	-	J1,D	J1,D
3a	GO:0015979	BP	4	photosynthesis	-	J1,D	J1,D
3a	GO:0005623	CC	2	cell	-	J1,D	J1,D
3a	GO:0044464	CC	3	cell part	-	J1,D	J1,D
3a	GO:0005622	CC	4	intracellular	-	J1,D	J1,D
3a	GO:0044424	CC	5	intracellular part	-	-	J1,D
3a	GO:0043227	CC	5	membrane-bounded organelle	-	-	J1,D
3a	GO:0009579	CC	6	thylakoid	-	J1,D	J1,D
3a	GO:0005737	CC	6	cytoplasm	-	J1,D	J1,D
3a	GO:0044444	CC	7	cytoplasmic part	-	J1,D	J1,D
3a	GO:0043231	CC	7	intracellular membrane-bounded organelle	-	-	J1,D
3a	GO:0009536	CC	8	plastid	-	J1,D	J1,D
3ab	GO:0005488	MF	2	binding	-	J1,D	J1,D
3ab	GO:0030246	MF	3	carbohydrate binding	-	J1,D	-
3b	GO:0016772	MF	5	transferase activity, transferring phosphorus-containing groups	-	J1,D	-
3b	GO:0016301	MF	7	kinase activity	-	J1,D	-

¹Group3a = GO terms enriched only if DEGs with chloroplast-related annotations are included; Group3b = GO terms enriched only if DEGs containing chloroplast-related annotations are omitted; Group 3ab = GO terms which are enriched irrespective of the inclusion or omission of DEGs with chloroplast-related annotations.

²Analysis methods identifying significant DEGs associated with the indicated enriched GO term and expression category. D = DESeq2, E = edgeR and L = limma-voom.

³MF = Molecular Function; BP = Biological Process; CC = Cellular component.

⁴Ascending numbers indicate a more specific category in the GO terms hierarchies.

⁵Abbreviated expression categories. u = up; d = down; n = ns.

<https://doi.org/10.1371/journal.pone.0220518.t008>

KEGG pathways

Significantly up- and down-regulated gene models from ‘mirror-image’ expression categories (i.e., AR-up_ns_ns and AR-down_ns_ns; TC-ns_ns_up and TC-ns_ns_down; AR-up_down_up and AR-down_up_down, etc.) were compared in terms of the number of up- and down-regulated gene models which could be associated with particular enzyme codes from the KEGG

Table 9. KEGG enzyme activity codes for which significantly different numbers of associated differentially expressed genes were present in comparisons of ‘mirror image’ expression categories.

Tissue	Expression category comparison ¹	Number of DEGs in each expression category		Number of DEGs associated with KEGG enzyme activity codes in the contrasted expression categories			
				KEGG enzyme activity code	DEGs		P ²
					1	2	
	AR/TC-1_2	1	2				
Shoot	AR-uuu_ddd	105	166	EC:3.1.3.16—phosphatase	5	1	3.39E-02
	AR-nun_ndn	319	404	EC:3.6.1.15—phosphatase	3	26	<1.00E-04
				EC:3.6.1.3—adenylpyrophosphatase	1	15	1.50E-03
	AR-nmu_nnd	1139	1276	EC:1.11.1.7—lactoperoxidase	4	21	1.89E-03
				EC:3.6.1.15—phosphatase	17	50	2.70E-04
				EC:3.6.1.3—adenylpyrophosphatase	12	35	2.87E-03
	TC-nud_ndu	178	228	EC:3.6.1.15—phosphatase	2	12	2.73E-02
EC:3.6.1.3—adenylpyrophosphatase				0	10	3.08E-03	
TC-nun_ndn	423	518	EC:3.6.1.15—phosphatase	4	18	1.51E-02	
TC-nnu_nnd	866	872	EC:1.11.1.7—lactoperoxidase	3	14	9.34E-04	
Root	AR-nu_nd	314	528	EC:1.11.1.7—lactoperoxidase	3	21	9.68E-03
				EC:4.1.1.2—decarboxylase	0	20	6.73E-04
	AR-nuu_ndd	156	330	EC:1.11.1.7—lactoperoxidase	1	12	7.05E-02
				EC:4.1.1.2—decarboxylase	0	18	1.25E-03
	AR-nmu_nnd	1805	1761	EC:1.11.1.7—lactoperoxidase	9	44	<1.00E-04
				EC:3.6.1.15—phosphatase	39	86	<1.00E-04
				EC:3.6.1.3—adenylpyrophosphatase	33	63	1.28E-03
	TC-nnu_nnd	1274	1625	EC:3.1.3.16—phosphatase	18	9	1.93E-02
EC:1.11.1.7—lactoperoxidase				6	45	<1.00E-04	
EC:3.6.1.15—phosphatase				28	87	<1.00E-04	
EC:3.6.1.3—adenylpyrophosphatase				21	65	1.60E-04	

¹Abbreviated expression categories. *u* = up; *d* = down; *n* = ns.

²P-values were calculated using Fisher’s Exact Test comparing the total number of DEGs in each of expression categories 1 and 2 with the corresponding values associated with the KEGG enzyme activity codes.

<https://doi.org/10.1371/journal.pone.0220518.t009>

pathways and significant differences identified (Fisher’s Exact Test; $p < 0.05$). As illustrated in Table 9, five enzyme codes were identified across root and shoot samples, with EC3.6.1.15 (phosphatase), EC3.6.1.3 (adenylpyrophosphatase) and EC1.11.1.7 (lactoperoxidase) the most frequently represented. For all but 2 of the 17 enzyme codes occurrences (excluding AR-nu_nd and TC-nud_ndu in which up- and down-regulation are present on both sides of the comparison) represented by significantly different numbers of up- and down-regulated DEGs, there were a greater number of down-regulated than up-regulated DEGs.

Discussion

Water-stress, like many environmental abiotic stresses, is unlikely to have a sudden onset. The stress will increase gradually over time as available water diminishes in the soil. Consequently, plants exposed to water-stress will adapt gradually over time. Thus, when looking at the effects of increasing water-stress in terms of changes in the transcriptome, it is important to remember that we are sampling the plant transcriptome at discrete points along what is likely to be a continuum of changing patterns of gene expression. In addition, while there may be genes that have major effects which can act independently of genetic background or, at least, the effects of which can be observed across multiple genetic backgrounds, in seeking to improve the resilience of

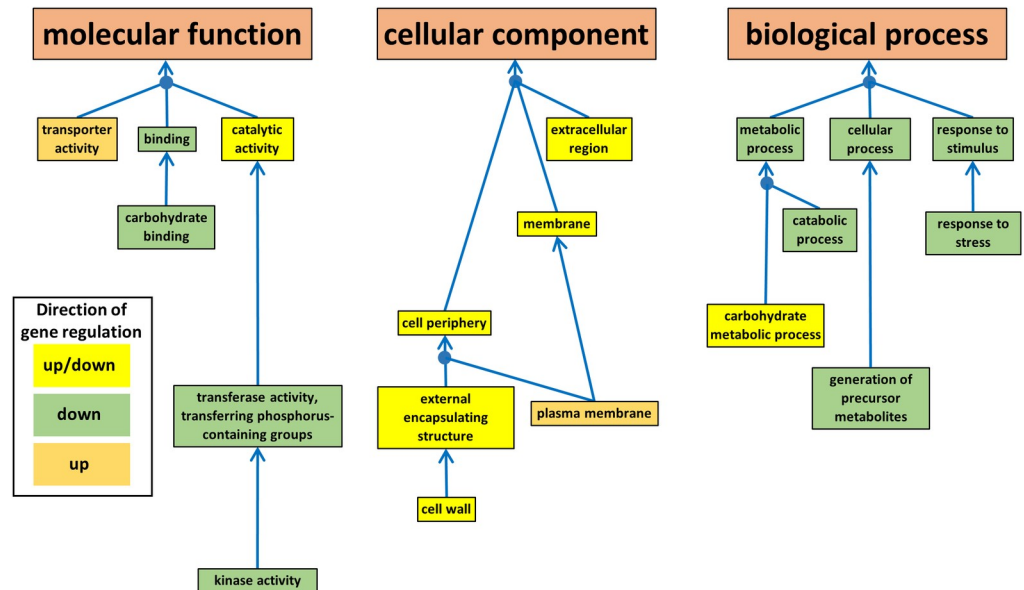


Fig 3. Enriched gene ontology (GO) terms associated with Molecular Function, biological process and cellular compartment GO categories for the root differentially expressed genes. The colour of the box indicates whether the enriched GO terms was associated with just up-regulated DEGs, just down-regulated DEGs or both up- and down-regulated DEGs across the 3 sampling points (35%, 15% and 5% EWC). Blue arrows indicate the direction of the GO hierarchy from child to parent and blue circles indicate where two or more child GO terms are associated with a single parent GO term.

<https://doi.org/10.1371/journal.pone.0220518.g003>

plants in the face of water limitation, we are trying to manipulate interacting biological processes and gene family effects [63–66]. For highly heterogeneous and outbreeding perennial populations, such as perennial ryegrass, which are widely distributed across multiple environments this is likely to be particularly true [2]. Thus, in the present study, the emphasis has not been on trying to identify candidate genes as such, but more to describe and compare, through the lens of GO terms, more generalised biological processes occurring in the leaves and roots of perennial ryegrass with increasing water stress. The genotype under study is, in some ways, atypical in that it is an inbred and, as a consequence, largely homozygous genotype. This genotype was chosen as it was used for the published genome and physical map assemblies for perennial ryegrass [41, 67] and so brings direct advantages to the analysis in terms of the quality of the available gene annotations. Additionally, as this study directly references the published perennial ryegrass genome assembly, it is hoped that this will also contribute to the development of a genomic platform for the further genetic dissection of the drought response in a wider selection of perennial ryegrass germplasm reflecting both agricultural and landscape adaptations.

The identification of GO terms associated with the progression of water-stress depends on the range of DEGs identified. This, itself, is dependent on the type of analysis undertaken. In this study, one aspect that we were interested in was whether differences in DEGs identified by different analysis software tools affected the range of associated GO terms. Additionally, we wished to see how using a ‘jury’ system (J1 and J3) affected the range of GO terms. The choice of the 3 programmes was partly that they incorporate different assumptions about data distributions, normalisation and significance testing but also because they are widely used and have performed well in larger-scale comparisons, including studies where replicate number has been limited [30–32]. Additionally, Costa-Silva et al. [29] suggested that combining the results of analysis programmes can provide more reliable results (with reliability being equated with qRT-PCR replication in that study).

Comparisons of DESeq2, edgeR, limma-voom, J1 and J3

Shoot transcriptomes. Table 10 describes the numbers and % of GO terms detected by each combination of methods. Focussing on the AR comparisons (a discussion of AR versus TC comparisons can be found in S1 Discussion), the largest proportion of GO terms were detected by all the methods. When the methods are considered individually, J1, DESeq2 and edgeR (in that order) detected the most enriched GO terms and limma-voom and J3 the least. It is, perhaps, slightly surprising that most enriched GO terms are generated by the DEGs in the J1 set and the least by the J3 set. J1 contains more DEGs than J3 by design, but the J1 set could be considered to be less stringently selected than the J3 set (called by any of the tools as opposed to all of the tools) so, possibly, more likely to accumulate false positives and less likely to be within the FDR criteria for the Fisher’s Exact Test. But this appears not to have been the case.

Looking more closely at the distribution of enriched AR expression category GO terms according to method (Table 7) Group 1 was exceptional in that only J1 DEGs were significantly enriched for the GOs in this section, the child term for which was *ATP binding* (GO:0005524). These connected GO terms (Fig 2) were only associated with 5 gene models and the probability values were close to the FDR threshold. Additionally, only 2 of these gene models were called as DEGs by all 3 programmes. So, this may suggest less confidence in these Group 1 GO terms as reflecting underlying biological processes specifically associated with the Early comparison stage. For the expression stages for which most DEGs were identified across

Table 10. The number and proportion of enriched gene ontology (GO) terms detected by the different analysis methods for the shoot data.

Analysis method combination	Enriched GO terms for shoot					
	AR+TC		AR		TC	
	n	%	n	%	n	%
J1,J3,D,E,L	110	32	80	36	30	23
J1,D	40	11	18	8	22	17
J1,D,E	20	6	15	7	5	4
D,E	20	6	8	4	12	9
D	19	5	2	1	17	13
J1	16	5	14	6	2	2
J1,D,E,L	16	5	9	4	7	5
E	16	5	12	5	4	3
J3	11	3	7	3	4	3
J3,D,E	11	3	6	3	5	4
J1,D,L	9	3	8	4	1	1
L	9	3	6	3	3	2
J3,E	7	2	3	1	4	3
J3,L	7	2	5	2	2	2
J3,D	7	2	5	2	2	2
J1,L	7	2	6	3	1	1
J1,J3,D,E	6	2	6	3	0	0
J1,E,L	4	1	4	2	0	0
J1,J3,D,L	3	1	2	1	1	1
J3,D,E,L	3	1	0	0	3	2
J1,E	3	1	3	1	0	0
J1,J3,D	3	1	0	0	3	2
J1,J3,L	1	<1	1	<1	0	0

<https://doi.org/10.1371/journal.pone.0220518.t010>

all methods (AR *up_ns_ns* and *ns_ns_down*) for Group 2, 23 out of the 29 GO terms were enriched across all the methods; for Group 4, 24 out of the 34 GO terms were enriched across all methods. For Group 3, which contains GO terms that were enriched in both expression stages, the proportions were slightly less with 6 and 12 out of 19 of the GO terms being enriched across all methods for AR *up_ns_ns* and *ns_ns_down*, respectively. It is also worth noting that using just a single programme would have suggested either up-regulation or down-regulation of some GO terms (particularly within the BP category) within Group 3 AR *up_ns_ns* and *ns_ns_down*, but not both up- and down-regulation. More generally, it is also clear that the overall pattern of biological processes, at least when described through enriched GO terms, is certainly nuanced by the choice of analysis method.

Root transcriptomes. For the root transcriptomes, for the Early and Middle stage comparisons (Table 8), the choice of programme makes a major difference to the number of enriched GO terms detected. DESeq2 identified 844 DEGs and 35 enriched GO terms at the Middle comparison stage, compared to only 1 DEG detected by either of edgeR or limma-voom (Table 3). If we compare these GO terms with those identified when the Late comparison is included, it is interesting that all of the non-chloroplast-associated GO terms just detected by DESeq2 at the Middle stage are also detected by all of the programmes, up- or down-regulated, in the AR/TC-*ns_ns_down* expression categories—which tends to support their detection. Conversely, the chloroplast-associated GO terms (i.e., Table 8, Group 3) were not detected in the AR/TC-*ns_ns_down* expression categories by any of the programmes and only in the AR/TC-*ns_down_ns* and AR-*ns_down_down* expression categories by DESeq2 (S5 Table). Thus, DESeq2 may have been able to identify more marginal changes in differential expression patterns at an earlier stage, but this may have also been at the risk of generating GO terms that are less well supported.

Patterns of enriched GO terms

During this discussion we will be considering GO terms which were enriched by any of the analysis methods, as described in Tables 7–9 and S3 and S5.

Probably the most striking aspect of this study of changes in gene expression in response to drought through enriched GO terms was the distinction between up- and down-regulated categories of gene models for the shoot transcriptome. With the exception of expression stage AR-*up_ns_ns*, very few GO terms were enriched at any of the other expression stages which indicated any up-regulation (18 expression categories, 33 enriched GO terms in total). This is a particularly stark comparison when we compare AR-*ns_ns_down* with AR *ns_ns_up*. Averaging across all analysis methods (Table 5) within AR-*ns_ns_down* there was an average of 1219 down-regulated DEGs, 87% of which could be associated with 42 enriched GO terms. The equivalent figures for AR *ns_ns_up* were 1080 up-regulated DEGs, 11% of which could be associated with 1 enriched GO term. A similar, but reverse, pattern can be seen comparing AR-*up_ns_ns* with AR *down_ns_ns* where the equivalent figures are 171 up-regulated DEGs, 74% of which are associated with 43 enriched GO terms and 89 down-regulated DEGs associated with 0 enriched GO terms. So, while there is an approximate balance in the overall numbers of up- and down-regulated DEGs across all expression categories (Table 4), there is a distinct imbalance in the degree to which they contribute to enriched GO terms and so indicate co-ordinated biological processes. Hong et al. [68], using mammalian cancer tumour datasets, showed that gene pairs with functional links in KEGG pathways tended to have positively correlated expression levels and proposed that analysing up- and down-regulated genes separately was more powerful than analysing all of the DEGs genes together—and this would also seem to be true in the present case. Due to different experimental designs and lack of

reporting of separate analysis for up- and down-regulated genes in some cases, it is difficult to make direct comparisons with previously published studies in plant systems. However, Wang et al [69] and Kokáš et al [11] comparing drought-stressed and control *Festuca mairei* and barley, respectively, reported that up- and down-regulated genes fell into distinct GO categories—though this is not quite the same as the results we present here. The imbalance detected for the shoot transcriptomes, in terms of the relationship between direction of DEG regulation and the number of enriched GO terms, was not so apparent for the root transcriptomes, though we are dealing with a much smaller set of DEGs. AR-*ns_up* and *ns_down* were more or less equivalent (particularly if the GO terms associated with the chloroplast-related DEGs are omitted). However, a comparison of AR/TC-*ns_ns_up* with *ns_ns_down* did indicate a higher percentage of DEGs contributing to more enriched GO terms in the latter expression category compared to the former. It will be interesting to see if this can be confirmed in future work.

It is an observation of this study that while the overall balance in terms of the numbers of up- and down-regulated DEGs at the different stages is apparent, the association of these two groups of DEGs with enriched GO terms is markedly unbalanced and this raises the question as to why so many DEGs in some expression categories were associated with so few or no enriched GO terms. Many studies of differential gene expression show this pattern of overall balance in terms of numbers of up- and down-regulated DEGs at time points and it is interesting to speculate as to whether this balance serves a function in terms of maintaining the molecular equilibrium of the cell, with the impact of the balancing overall level of gene expression ‘uncoupled’ from a more coordinated effect on cell phenotype, thus generating fewer enriched GO terms.

KEGG pathway activities

The comparison of ‘mirror-image’ expression categories (Table 9) identified 5 KEGG enzyme activity codes represented across the various expression categories associated with significantly different numbers of up- and down-regulated DEGs. As described earlier, the number of down-regulated DEGs associated with enzyme codes exceeded the number of up-regulated DEGs associated with expression codes the majority of the time. The most commonly occurring enzyme codes were EC3.6.1.15 (phosphatase) and EC3.6.1.3 (adenylypyrophosphatase) occurring across 6 and 5 expression categories, respectively. Both enzyme activities are nucleoside triphosphatases, EC3.6.1.3 being specifically an ATP phosphohydrolase and EC3.6.1.15 also having phosphohydrolase activity for nucleoside tri/diphosphates, thiamine diphosphate and flavin adenine dinucleotide. Across all expression categories described in Table 9, EC3.6.1.15 was associated with 268 down-regulated DEGs and 97 up-regulated DEGs with a further 14 DEGs showing evidence of both up- and down-regulation. The equivalent figures for EC3.6.1.3 were 178, 67 and 10 for down-, up and both down- and up regulated DEGs, respectively. Thus, the DEGS associated with both these key enzyme activities involved in driving fundamental cellular metabolic and cytoskeletal processes, showed overall down-regulation, which may well indicate the general slowing of cellular metabolic processes as the drought proceeds.

The other frequently represented enzyme activity code, occurring across 5 different expression categories (excluding AR-*nu_nd*) was EC1.11.1.7 (lactoperoxidase). This is an enzyme activity involved in the phenylpropanoid biosynthesis pathway and associated with the final stages of lignin biosynthesis in the development of secondary plant cell walls [70]. DEGs associated with this enzyme activity were also more often down- (157) rather than up-regulated (26) across shoot and root (Table 9). A number of previous studies have looked at the activities and/or expression profiles of enzymes from the phenylpropanoid pathway under drought stress and opposite or mixed trends, in terms of their regulation, have been reported [71–74]–

though there are considerable variations in the species under study, experimental methodologies and time courses across studies. However, the frequent occurrences of genes involved in the phenylpropanoid pathway in drought studies does indicate the significance of cell-wall metabolism in the response to water-stress.

The other two enzyme classes were EC3.1.3.16 (phosphatase– 2 occurrences) and EC4.1.1.2 (decarboxylase– 2 occurrence). EC3.1.3.16, includes enzymes which show protein-serine/threonine phosphatase activity against a wide range of proteins and was the only enzyme class which showed predominantly up-regulated DEGs in the shoot (AR-*uuu/ddd*) and the root (TC-*nnu/nnd*). The second enzyme class, EC4.1.1.2, was identified just in the root during the AR-*nu_nd* AR-*ns_down_down/ns_up_up* comparison stage. All 20 of the DEGs associated with this enzyme class were down-regulated and annotated as ‘germin’. EC4.1.1.2 is specifically an oxalate de-carboxylase and accumulation of malic (leaves) and fumaric (roots) acids in perennial ryegrass in response to PEG-induced water-stress has been reported [17]. Similar observations for malic, fumaric and oxalic acids have been reported as a drought response in wheat leaves[75]. The observed down-regulation of DEGs associated with oxalate decarboxylase activity in the present study could be part of the process of maintaining intracellular ionic balance under water stress as mediated by organic acid concentrations.

Conclusion

In comparing the performance of three RNAseq analysis tools, both individually and in combination, we have identified differences in terms of the range of DEGs identified and, in some cases, the inferences one might make in terms of the associated biological processes indicated by GO terms. In future work, it is likely that we will continue to use all 3 programmes, J1 and J3 and compare the performance profiles across a more diverse selection of germplasm to see if the most consistent and complete descriptions of transcriptome responses to diminishing water availability are generated by a single, or combinations of analysis methods. While this is not a conservative approach, relative levels of differential gene expression are not, in themselves, an experimental end but more a means with which sets of genes, post-translational products and broader biological processes can be identified for further experimentation using other approaches such as QTL analysis, proteomics, metabolomics and genetic modification.

Major differences were found in terms of the number of enriched GO terms at different stages depending on whether the gene model set was taken from up- or down-regulated DEGs, particularly for the leaf transcriptomes. For the leaf, up-regulated DEGs were associated with more enriched GO terms during the Early comparison stage and gene annotations indicated that many of these may have had transcription factor and membrane transporter activity. For all other stages, down-regulated DEGs generated more enriched GO terms with more general metabolic associations. For the root, few DEGs were identified at Early or Middle comparison stages (though more by DESeq2 than the other programmes) and the Late comparison stage was associated with a far larger number of DEGs. Gene annotations indicated that heavy-metal associated membrane transport and down-regulation of cytoskeleton-associated ATPase activities might be significant processes. However, because of lack of effective repetition at the Late comparison stage, these results have to be treated as preliminary. KEGG analysis also indicated that ATPase and lignin biosynthesis-associated peroxidase activities were affected by increasing water-stress.

Supporting information

S1 Table. *L. perenne* genome annotation table (BLAST2GO PRO table export.).
(XLSX)

S2 Table. Read and alignment statistics for the RNAseq replicates at each sampling point.
(XLSX)

S3 Table. GO terms associated with different expression categories for shoot DEGs.
(XLSX)

S4 Table. Gene models and annotations associated with enrichment of GO terms for indicated expression categories from leaf and root.
(XLSX)

S5 Table. GO terms associated with different expression categories for root DEGs using Early, Middle and Late expression comparisons.
(XLSX)

S6 Table. The number of shoot DEGs and their contribution to enriched GO terms for each analysis method arranged according to 'mirror-image' pairs.
(XLSX)

S1 Fig. Images of the four replicates of p226/135/16 sampled at each estimated water content (EWC) for RNAseq.
(PPTX)

S2 Fig. Count distributions derived from the leaf transcriptomes. DESeq2 rlog transformations plotted against density for each of the four replicates at each sampling point.
(PPTX)

S3 Fig. Count distributions derived from the root transcriptomes. DESeq2 rlog transformations plotted against density for each of the four replicates at each sampling point. Note the variation among root-derived samples, particularly at 1% EWC, compared to the leaf-derived samples.
(PPTX)

S4 Fig. GO-terms associated with leaf gene expression category comparisons.
(PPTX)

S5 Fig. GO-terms associated with root gene expression category comparisons.
(PPTX)

S1 Methods. R script codes implementing the pre-processing and generation of count matrices.
(PDF)

S2 Methods. R script code for the identification of DEGs using DESeq2.
(PDF)

S3 Methods. R script code for the identification of DEGs using edgeR.
(PDF)

S4 Methods. R script code for the identification of DEGs using limma-voom.
(PDF)

S1 Results. Description of enriched GO terms and associated gene models in Groups 5–8 of the leaf transcriptome.
(DOCX)

S2 Results. Enriched GO terms identified when including the Late comparisons for the root transcriptome.

(DOCX)

S1 Discussion. AR and TC comparisons.

(DOCX)

Acknowledgments

We would like to thank Colin Sauze and Max Friedersdorff, Aberystwyth University, for their assistance with the high performance computing infrastructure.

Author Contributions

Conceptualization: Ian Armstead, Narcis Fernandez-Fuentes.

Data curation: Albert Fradera-Sola, Ann Thomas, Narcis Fernandez-Fuentes.

Formal analysis: Albert Fradera-Sola, Ann Thomas, Ian Armstead, Narcis Fernandez-Fuentes.

Investigation: Albert Fradera-Sola, Narcis Fernandez-Fuentes.

Methodology: Ann Thomas, Matthew Hegarty, Ian Armstead, Narcis Fernandez-Fuentes.

Project administration: Narcis Fernandez-Fuentes.

Resources: Ann Thomas, Dagmara Gasior, John Harper, Matthew Hegarty.

Software: Albert Fradera-Sola.

Supervision: Narcis Fernandez-Fuentes.

Validation: Narcis Fernandez-Fuentes.

Writing – original draft: Ian Armstead, Narcis Fernandez-Fuentes.

Writing – review & editing: Albert Fradera-Sola, Ann Thomas, Dagmara Gasior, John Harper, Matthew Hegarty, Ian Armstead, Narcis Fernandez-Fuentes.

References

1. Aper J. GA, Cougnon M., Baert J. Drought effect on yield of perennial ryegrass (*Lolium perenne* L.). In: Sokolović D. HC, Radović J., editors. Quantitative Traits Breeding for Multifunctional Grasslands and Turf. Springer, Dordrecht; 2014. pp. 367–371.
2. Cyriac D, Hofmann RW, Stewart A, Sathish P, Winefield CS, Moot DJ. Intraspecific differences in long-term drought tolerance in perennial ryegrass. *PLoS One*. 2018; 13(4):e0194977. <https://doi.org/10.1371/journal.pone.0194977> PMID: 29617413
3. Bedada G, Westerbergh A, Muller T, Galkin E, Bdolach E, Moshelion M, et al. Transcriptome sequencing of two wild barley (*Hordeum spontaneum* L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. *Bmc Genomics*. 2014; 15. <https://doi.org/10.1186/1471-2164-15-995> PMID: 25408241
4. Dalal M, Sahu S, Tiwari S, Rao AR, Gaikwad K. Transcriptome analysis reveals interplay between hormones, ROS metabolism and cell wall biosynthesis for drought-induced root growth in wheat. *Plant Physiol Biochem*. 2018; 130:482–92. <https://doi.org/10.1016/j.plaphy.2018.07.035> PMID: 30081325
5. Ding XP, Li XK, Xiong LZ. Insight into Differential Responses of Upland and Paddy Rice to Drought Stress by Comparative Expression Profiling Analysis. *Int J Mol Sci*. 2013; 14(3):5214–38. <https://doi.org/10.3390/ijms14035214> PMID: 23459234
6. Dudhate A, Shinde H, Tsugama D, Liu SK, Takano T. Transcriptomic analysis reveals the differentially expressed genes and pathways involved in drought tolerance in pearl millet *Pennisetum glaucum* (L.)

- R. Br. PLoS One. 2018; 13(4):e0195908. <https://doi.org/10.1371/journal.pone.0195908> PMID: 29652907
7. Hittalmani S, Mahesh HB, Shirke MD, Biradar H, Uday G, Aruna YR, et al. Genome and Transcriptome sequence of Finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *Bmc Genomics*. 2017; 18. <https://doi.org/10.1186/s12864-017-3850-z> PMID: 28619070
 8. Hoang XLT, Nhi DNH, Thu NBA, Thao NP, Tran LSP. Transcription Factors and Their Roles in Signal Transduction in Plants under Abiotic Stresses. *Curr Genomics*. 2017; 18(6):483–97. <https://doi.org/10.2174/1389202918666170227150057> PMID: 29204078
 9. Hu L, Xie Y, Fan SJ, Wang ZS, Wang FH, Zhang B, et al. Comparative analysis of root transcriptome profiles between drought-tolerant and susceptible wheat genotypes in response to water stress. *Plant Sci*. 2018; 272:276–93. <https://doi.org/10.1016/j.plantsci.2018.03.036> PMID: 29807601
 10. Jonaviciene K, Studer B, Asp T, Jensen LB, Paplauskiene V, Lazauskas S, et al. Identification of genes involved in a water stress response in timothy and mapping of orthologous loci in perennial ryegrass. *Biol Plant*. 2012; 56(3):473–83. <https://doi.org/10.1007/s10535-012-0110-6>
 11. Kokas F, Vojta P, Galuszka P. Dataset for transcriptional response of barley (*Hordeum vulgare*) exposed to drought and subsequent re-watering. *Data Brief*. 2016; 8:334–41. <https://doi.org/10.1016/j.dib.2016.05.051> PMID: 27331111
 12. Kulkarni M, Soolanayakanahally R, Ogawa S, Uga Y, Selvaraj MG, Kagale S. Drought Response in Wheat: Key Genes and Regulatory Mechanisms Controlling Root System Architecture and Transpiration Efficiency. *Frontiers in chemistry*. 2017; 5:106-. <https://doi.org/10.3389/fchem.2017.00106> PMID: 29259968
 13. Kwasniewski M, Daszkowska-Golec A, Janiak A, Chwialkowska K, Nowakowska U, Sablok G, et al. Transcriptome analysis reveals the role of the root hairs as environmental sensors to maintain plant functions under water-deficiency conditions. *Journal of Experimental Botany*. 2016; 67(4):1079–94. <https://doi.org/10.1093/jxb/erv498> PMID: 26585228
 14. Liu S, Jiang Y. Identification of differentially expressed genes under drought stress in perennial ryegrass. *Physiologia Plantarum*. 2010; 139(4):375–87. <https://doi.org/10.1111/j.1399-3054.2010.01374.x> PMID: 20444191
 15. Pan L, Zhang XQ, Wang JP, Ma X, Zhou ML, Huang LK, et al. Transcriptional Profiles of Drought-Related Genes in Modulating Metabolic Processes and Antioxidant Defenses in *Lolium multiflorum*. *Frontiers in Plant Science*. 2016; 7:519. <https://doi.org/10.3389/fpls.2016.00519> PMID: 27200005
 16. Talukder SK, Azhaguvel P, Mukherjee S, Young CA, Tang Y, Krom N, et al. De Novo Assembly and Characterization of Tall Fescue Transcriptome under Water Stress. *Plant Genome*. 2015; 8(2). <https://doi.org/10.3835/plantgenome2014.09.0050>
 17. Foito A, Byrne SL, Shepherd T, Stewart D, Barth S. Transcriptional and metabolic profiles of *Lolium perenne* L. genotypes in response to a PEG-induced water stress. *Plant Biotechnology Journal*. 2009; 7(8):719–32. <https://doi.org/10.1111/j.1467-7652.2009.00437.x> PMID: 19702648
 18. Csiszar J, Galle A, Horvath E, Dancso P, Gombos M, Vary Z, et al. Different peroxidase activities and expression of abiotic stress-related peroxidases in apical root segments of wheat genotypes with different drought stress tolerance under osmotic stress. *Plant Physiol Biochem*. 2012; 52:119–29. <https://doi.org/10.1016/j.plaphy.2011.12.006> PMID: 22305075
 19. Habte E, Muller LM, Shtaya M, Davis SJ, von Korff M. Osmotic stress at the barley root affects expression of circadian clock genes in the shoot. *Plant Cell Environ*. 2014; 37(6):1321–37. <https://doi.org/10.1111/pce.12242> PMID: 24895755
 20. He XY, Zeng JB, Cao FB, Ahmed IM, Zhang GP, Vincze E, et al. HvEXPB7, a novel beta-expansin gene revealed by the root hair transcriptome of Tibetan wild barley, improves root hair growth under drought stress. *Journal of Experimental Botany*. 2015; 66(22):7405–19. <https://doi.org/10.1093/jxb/erv436> PMID: 26417018
 21. Li Y, Chen X, Chen Z, Cai R, Zhang H, Xiang Y. Identification and Expression Analysis of BURP Domain-Containing Genes in *Medicago truncatula*. *Frontiers in Plant Science*. 2016; 7:485-. <https://doi.org/10.3389/fpls.2016.00485> PMID: 27148311
 22. Challa GS, Li WL. De novo assembly of wheat root transcriptomes and transcriptional signature of longitudinal differentiation. *PLoS One*. 2018; 13:e0205582. <https://doi.org/10.1371/journal.pone.0205582> PMID: 30395610
 23. Moumeni A, Satoh K, Kondoh H, Asano T, Hosaka A, Venuprasad R, et al. Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC Plant Biol*. 2011; 11:174. <https://doi.org/10.1186/1471-2229-11-174> PMID: 22136218

24. Liu ZG, Gao K, Shan S, Gu RC, Wang ZK, Craft EJ, et al. Comparative Analysis of Root Traits and the Associated QTLs for Maize Seedlings Grown in Paper Roll, Hydroponics and Vermiculite Culture System. *Frontiers in Plant Science*. 2017; 8:436. <https://doi.org/10.3389/fpls.2017.00436> PMID: 28424719
25. Sanchez-Urdaneta AB, Pena-Valdivia CB, Trejo C, Aguirre JR, Cardenas E. Root growth and proline content in drought sensitive and tolerant maize (*Zea mays* L.) seedlings under different water potentials. *Cereal Res Commun*. 2005; 33(4):697–704. <https://doi.org/10.1556/crc.33.2005.2-3.137>
26. Seeve CM, Cho IJ, Hearne LB, Srivastava GP, Joshi T, Smith DO, et al. Water-deficit-induced changes in transcription factor expression in maize seedlings. *Plant Cell Environ*. 2017; 40(5):686–701. <https://doi.org/10.1111/pce.12891> PMID: 28039925
27. Silva EN, Ferreira-Silva SL, Viegas RA, Silveira JAG. The role of organic and inorganic solutes in the osmotic adjustment of drought-stressed *Jatropha curcas* plants. *Environ Exp Bot*. 2010; 69(3):279–85. <https://doi.org/10.1016/j.envexpbot.2010.05.001>
28. Takahashi H, Noda M, Sakurai K, Watanabe A, Akagi H, Sato K, et al. QTLs in barley controlling seedling elongation of deep-sown seeds. *Euphytica*. 2008; 164(3):761–8. <https://doi.org/10.1007/s10681-008-9720-7>
29. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One*. 2017; 12(12):e0190152. Epub 2017/12/22. <https://doi.org/10.1371/journal.pone.0190152> PMID: 29267363
30. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna*. 2016; 22(6):839–51. <https://doi.org/10.1261/rna.053959.115> PMID: 27022035
31. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform*. 2015; 16(1):59–70. <https://doi.org/10.1093/bib/bbt086> PMID: 24300110
32. Williams CR, Baccarella A, Parrish JZ, Kim CC. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinformatics*. 2017; 18:38. <https://doi.org/10.1186/s12859-016-1457-z> PMID: 28095772
33. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11(10):R106. <https://doi.org/10.1186/gb-2010-11-10-r106> PMID: 20979621
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
35. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, et al. EBSseq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*. 2013; 29(8):1035–43. <https://doi.org/10.1093/bioinformatics/btt087> PMID: 23428641
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
37. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014; 15(2):R29. <https://doi.org/10.1186/gb-2014-15-2-r29> PMID: 24485249
38. Conesa A, Gotz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008; 2008:619832. <https://doi.org/10.1155/2008/619832> PMID: 18483572
39. The Gene Ontology C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019; 47(D1):D330–D8. <https://doi.org/10.1093/nar/gky1055> PMID: 30395331
40. QuickGO. Available from: <https://www.ebi.ac.uk/QuickGO/>.
41. Byrne SL, Nagy I, Pfeifer M, Armstead I, Swain S, Studer B, et al. A synteny-based draft genome sequence of the forage grass *Lolium perenne*. *Plant J*. 2015; 84(4):816–26. <https://doi.org/10.1111/tpj.13037> PMID: 26408275
42. Hoagland DR AD. The water culture method for growing plants without soil. *California Agricultural Experimental Station Circular*. 1950; 347:1–39.
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
44. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data 2014. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
45. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015; 12(4):357–60. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142

46. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci U S A*. 2010; 107(21):9546–51. <https://doi.org/10.1073/pnas.0914005107> PMID: 20460310
47. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016; 17:13. <https://doi.org/10.1186/s13059-016-0881-8> PMID: 26813401
48. Van Belle G, Fisher L, Heagerty P, Lumley T. *Principal Component Analysis and Factor Analysis. Biostatistics: A Methodology for the Health Sciences. Wiley Series in Probability and Statistics*. Second ed: John Wiley & Sons; 2004.
49. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015; 33:290. <https://doi.org/10.1038/nbt.3122> PMID: 25690850
50. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
52. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010; 11(3):R25. <https://doi.org/10.1186/gb-2010-11-3-r25> PMID: 20196867
53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
54. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*. 2014; 30(9):1236–40. Epub 2014/01/29. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
55. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research*. 2012; 41(D1):D387–D95. <https://doi.org/10.1093/nar/gks1234> PMID: 23197656
56. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*. 2015; 44(D1):D279–D85. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716
57. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res*. 2006; 34(Database issue):D257–60. Epub 2005/12/31. doi: 34/suppl_1/D257 [pii] <https://doi.org/10.1093/nar/gkj079> PMID: 16381859
58. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013; 41(Database issue):D377–86. <https://doi.org/10.1093/nar/gks1118> PMID: 23193289
59. Lees J, Yeats C, Redfern O, Clegg A, Orengo C. Gene3D: merging structure and function for a Thousand genomes. *Nucleic Acids Res*. 2010; 38(Database issue):D296–300. Epub 2009/11/13. doi: gkp987 [pii] <https://doi.org/10.1093/nar/gkp987> PMID: 19906693
60. Wu CH, Nikolskaya A, Huang H, Yeh L-SL, Natale DA, Vinayaka CR, et al. PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research*. 2004; 32(Database issue):D112–D4. <https://doi.org/10.1093/nar/gkh097> PMID: 14681371
61. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000; 28(1):27–30. <https://doi.org/10.1093/nar/28.1.27> PMID: 10592173
62. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015; 43(Database issue):D204–12. <https://doi.org/10.1093/nar/gku989> PMID: 25348405
63. Bhargava S, Sawant K. Drought stress adaptation: metabolic adjustment and regulation of gene expression. *Plant Breeding*. 2013; 132(1):21–32. <https://doi.org/10.1111/pbr.12004>
64. Hu H, Xiong L. Genetic Engineering and Breeding of Drought-Resistant Crops. *Annual Review of Plant Biology*. 2014; 65(1):715–41. <https://doi.org/10.1146/annurev-arplant-050213-040000> PMID: 24313844
65. Monroe JG, Powell T, Price N, Mullen JL, Howard A, Evans K, et al. Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. *eLife*. 2018; 7:e41038. <https://doi.org/10.7554/eLife.41038> PMID: 30520727
66. Tardieu F. Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. *Journal of Experimental Botany*. 2011; 63(1):25–31. <https://doi.org/10.1093/jxb/err269> PMID: 21963615

67. Harper J, De Vega J, Swain S, Heavens D, Gasior D, Thomas A, et al. Integrating a newly developed BAC-based physical mapping resource for *Lolium perenne* with a genome-wide association study across a *L. perenne* European ecotype collection identifies genomic contexts associated with agriculturally important traits. *Annals of Botany*. 2019. <https://doi.org/10.1093/aob/mcy230> PMID: 30715119
68. Hong G, Zhang W, Li H, Shen X, Guo Z. Separate enrichment analysis of pathways for up- and downregulated genes. *Journal of the Royal Society, Interface*. 2014; 11(92):20130950. <https://doi.org/10.1098/rsif.2013.0950> PMID: 24352673
69. Wang JPP, Bughrara SS. Monitoring of gene expression profiles and identification of candidate genes involved in drought responses in *Festuca mairei*. *Mol Genet Genomics*. 2007; 277(5):571–87. <https://doi.org/10.1007/s00438-007-0208-2> PMID: 17323082
70. Meents MJ, Watanabe Y, Samuels AL. The cell biology of secondary cell wall biosynthesis. *Annals of Botany*. 2018; 121(6):1107–25. <https://doi.org/10.1093/aob/mcy005> PMID: 29415210
71. dos Santos AB, Bottcher A, Kiyota E, Mayer JLS, Vicentini R, Brito MD, et al. Water Stress Alters Lignin Content and Related Gene Expression in Two Sugarcane Genotypes. *J Agric Food Chem*. 2015; 63(19):4708–20. <https://doi.org/10.1021/jf5061858> PMID: 25936563
72. Ksouri N, Jiménez S, Wells CE, Contreras-Moreira B, Gogorcena Y. Transcriptional Responses in Root and Leaf of *Prunus persica* under Drought Stress Using RNA Sequencing. *Frontiers in Plant Science*. 2016; 7(1715). <https://doi.org/10.3389/fpls.2016.01715> PMID: 27933070
73. Li JQ, Fan FF, Wang LH, Zhan QW, Wu PJ, Du JL, et al. Cloning and expression analysis of cinnamoyl-CoA reductase (CCR) genes in sorghum. *PeerJ*. 2016; 4:e2005. <https://doi.org/10.7717/peerj.2005> PMID: 27231650
74. Liu W, Jin Y, Li M, Dong L, Guo D, Lu C, et al. Analysis of CmCADs and three lignifying enzymes in oriental melon ('CaiHong7') seedlings in response to three abiotic stresses. *Scientia Horticulturae*. 2018; 237:257–68. <https://doi.org/10.1016/j.scienta.2018.04.024>.
75. Guo R, Shi L, Jiao Y, Li M, Zhong X, Gu F, et al. Metabolic responses to drought stress in the tissues of drought-tolerant and drought-sensitive wheat genotype seedlings. *AoB Plants*. 2018; 10(2):ply016. <https://doi.org/10.1093/aobpla/ply016> PMID: 29623182