Original Article

# High-resolution temporal transcriptome sequencing unravels ERF and WRKY as the master players in the regulatory networks underlying sesame responses to waterlogging and recovery

Linhai Wang [a,1], Komivi Dossa [a,b,1], Jun You [a,1], Yanxin Zhang [a], Donghua Li [a], Rong Zhou [a], Jingyin Yu [a], Xin Wei [a,c], Xiaodong Zhu [a], Shiyang Jiang [a], Yuan Gao [a], Marie Ali Mmadi [a], Xiurong Zhang [a,*]

[a] Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, No.2 Xudong 2nd Road, Wuhan 430062, China
[b] Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, 01 BP 526 Cotonou, Benin
[c] College of Life Sciences, Shanghai Normal University, Shanghai 200234, China

ABSTRACT

Major crops are generally sensitive to waterlogging, but our limited understanding of the waterlogging gene regulatory network hinders the efforts to develop waterlogging-tolerant cultivars. We generated high-resolution temporal transcriptome data from root of two contrasting sesame genotypes over a 48 h period waterlogging and drainage treatments. Three distinct chronological transcriptional phases were identified, including the early-waterlogging, late-waterlogging and drainage responses. We identified 47 genes representing the core waterlogging-responsive genes. Waterlogging/drainage-induced transcriptional changes were mainly driven by ERF and WRKY transcription factors (TF). The major difference between the two genotypes resides in the early transcriptional phase. A chronological transcriptional network model predicting putative causal regulations between TFs and downstream waterlogging-responsive genes was constructed and some interactions were validated through yeast one-hybrid assay. Overall, this study unveils the architecture and dynamic regulation of the waterlogging/drainage response in a non-model crop and helps formulate new hypotheses on stress sensing, signaling and sophisticated adaptive responses.

## 1. Introduction

During their growth and development, plants are prone to various stressful conditions that demand a dynamic metabolic reprogramming. Waterlogging, which arises from excess soil water, decreases oxygen availability to plants and slows down aerobic respiration and photosynthesis [75]. Under the current climate change, floods are surging in many regions of the world, limiting agricultural production and threatening food security [26]. It has been estimated that waterlogging decreases crop productivity by approximately 20% [31]. Many wetland species have evolved a suite of traits to thrive under waterlogging conditions [9]. These traits sequentially appear in three stages. Initially, the plant rapidly induces a set of signal transduction components, which activate a metabolic adaptation involving fermentation pathways. And the third stage involves morphological changes such as the formation of aerenchyma and/or adventitious root [17,30,54]. Extensive studies on the genetic and molecular regulation of these flood-adaptive strategies have been performed, especially on wetland species [11,19,24, 45,52,64]. For example, the key genes regulating flood-adaptive quiescence (*SUB1A*) and escape (*SNORKEL1/2*) were unraveled in rice [24,72]. In addition, investigations on the model plant species *Arabidopsis thaliana* have led to breakthroughs in our understanding of low oxygen sensing and regulation. In particular, results from studies of Gibbs et al. [21] and Licausi et al. [41] demonstrated that oxygen sensing in plants operates via posttranslational regulation of the group VII family of ethylene response factors (ERF) by the N-end rule pathway. The prompt recovery from the post waterlogging period is also a key determinant of plant tolerance. Recently, Yeung et al. [73] discovered

---

three Arabidopsis genes namely *RESPIRATORY BURST OXIDASE HO-MOLOG D*, *SENESCENCE-ASSOCIATED GENE113* and *ORESARA1*, which function in a signaling network that regulates recovery processes following flooding.

Unfortunately, with the exception of rice, major and important crops are terrestrial species generally sensitive to waterlogging, and limited progress has been made so far to develop tolerant germplasm [5]. Sesame (*Sesamum indicum* L.) is a typical example of crop naturally adapted to dry conditions and highly susceptible to waterlogging [68,71]. Sesame is a source of high-quality oil with a unique abundance of antioxidants such as sesamin and sesamolin. It also has a component balance in oleic acid and linoleic acid concentrations [3]. Recently, the annual demand of sesame seeds has increased [13]. However, with the increasing frequency and intensity of floods in tropical areas, catastrophic yield losses are recorded in sesame fields, expanding the gap between seed production and demand [1,50]. It is estimated that sesame can suffer yield losses of more than 30% (in severe cases, 50–90%) when cultivated in waterlogged soils [23,50]. Waterlogging-tolerant sesame resources are rare and our insight into waterlogging responses at the morphological, physiological and molecular levels is still limited. In field conditions, high temperatures amplify the stress induced by waterlogging and most of sesame genotypes barely survive over 36 h of stress [57]. The accession ZZM2541 was initially identified by Feng et al. [18] and later confirmed by [58,59]), Wei et al. [71], as a tolerant genotype able to survive several days of waterlogging. The intra-species variation in waterlogging tolerance of sesame is of great importance because it could potentially point to the key molecular components that modulate tolerance to this stress. A detailed understanding of the molecular regulation of waterlogging responses in sesame is cardinal in the search for commanding genes that can be exploited to improve the existing germplasm. Previous transcriptome profiling studies revealed that a huge number of genes were altered upon waterlogging exposure [67,68]. The latter even investigated transcriptome changes during the recovery from waterlogging and identified a set of 3016 genes functioning in energy supply and cell repair or formation [68]. However, these studies analyzed this response at limited temporal resolution, therefore, the architecture and dynamics of the waterlogging/recovery gene regulatory networks are still unclear.

In the present study, we generated a high-quality and high-resolution time series RNA sequencing data from root of two contrasting sesame cultivars (ZZM2541 and Ezhi-2) during waterlogging and recovery stages. Our data help determine the sequence of transcriptional phases that occur during waterlogging and recovery. We also identified the major regulators operating during each transcriptional phase and constructed a dynamic model of the waterlogging/recovery gene regulatory networks. Comparison of the transcriptional responses in waterlogging tolerant and sensitive genotypes further shed light on the tolerance mechanism. Overall, we provide here, a novel insight into the structure of waterlogging/recovery regulatory network in a non-model crop.

## 2. Material and methods

### 2.1. Plant materials and stress treatment

Two genotypes of sesame (*Sesamum indicum* L.) were obtained from the China National Genebank, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences and used in this experiment. The genotype ZZM2541 (R2G) displays a strong tolerance to waterlogging stress while Ezhi No.2 (EG) is highly susceptible as demonstrated by Wei et al. [71]. The experiment was conducted in a greenhouse as described by Dossa et al. [16]. The average temperature and relative humidity during the experiment is presented in Supplementary file 1. Plants were grown in pots (25 cm diameter and 30 cm depth) containing 7 Kg of loam soil mixed with 10% compound fertilizer. The plants were irrigated every 3 days and the soil volumetric water content (vwc) was maintained at ~35%. A pot tray was placed under each pot to avoid water loss. A completely

randomized blocking design with four replicates was employed. 15 days after the initiation of flowering, half of the pots were waterlogged by standing in a plastic bucket filled with tap water up to 3 cm above the soil surface. Each pot contains three seedlings, which were maintained waterlogged for 36 h and afterwards, pots were drained to allow plants to recover for 12 h (vwc = ~35%). In parallel, half of the pots were kept under normal growth conditions (vwc = ~35%) during the whole experiment. Root samples (whole root) were collected from three individual plants (biological replicates) from different pots in the stress and control treatments at 11 time points following the flowchart presented in Fig.1. In total, 126 root samples were collected, snap frozen in liquid nitrogen and used for follow-up analyses (Supplementary file 2).

### 2.2. Transcriptome sequencing and data analysis

#### 2.2.1. RNA extraction, library preparation and RNA sequencing

Total RNA was extracted from 126 root samples using the TRIzol reagent (Invitrogen), and treated with DNase I and Oligo (dT) to isolate mRNAs. The concentration and quality were determined using an ultraviolet spectrophotometer and 2% denaturing agarose gels. The cDNA was synthesized using the mRNA fragments as templates. The short fragments (~300 bp) were ligated with adapters and the suitable fragments were selected for PCR amplification. The libraries were paired-end sequenced using the Illumina platform Hiseq 2000 [14].

#### 2.2.2. RNA-seq data analysis

The program FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was employed to determine the base quality of the raw reads (in FASTQ format) and we removed the paired-end reads containing more than 5% ambiguous residues (Ns) and those containing >10% bases with a Phred quality score of 10. Then, the raw reads were trimmed using Trimmomatic, version 0.32 [6]. After cleaning and quality reassessment with FastQC, approximately 31.6–56.8 million high-quality reads of 90-bp length remained in each sample. The high-quality reads were mapped to the sesame reference genome v1.0 (http://ocri-genomics.org/Sinbase/login.htm) [69] using the STAR software [12], allowing no more than one mis-match in the alignment. Approximately, 88.9–97.4% of the clean reads were uniquely mapped to the reference genome, with 94.3–98.6% of them uniquely mapped to genic regions. Using the featureCount package [40], the gene expression levels were calculated based on the number of unique matched reads to the sesame genome v1.0 [69] and were normalized to Transcripts Per Million (TPM). Further analyses were implemented based on genes with TPM > 1 following the diagram presented in Fig. 2a.

### 2.3. Differential gene expression analysis

The identification of the differentially expressed genes (DEGs) was performed with the R software (http://www.r-project.org) according to the pipeline described by Hickman et al. [25] with slight modifications. For each genotype, the DEGs after waterlogging and drainage treatments compared to the control treatment were identified using a generalized linear model with a log link function and a poisson distribution. The time after treatment and the treatment itself were considered as factors in a saturated model: fit.full = glm(counts ~ treatment + time + treatment:time, data = data, family = poisson(link = "log")). Next, we assessed the time effect on the total count for each gene using a reduced model: fit.reduced = glm(counts~times, data = EG_data, family = poisson(link = "log")). To evaluate the treatment effect on the total read count for each gene, the saturated model was compared with the reduced model using ANOVA with a $\chi^2$ test: anova(fit.reduced, fit.full, test = "Chisq"). Only genes, which displayed a minimum of a 2-fold difference in expression on at least one time point, were retained for further analysis after a Bonferroni correction of the $P$ value ($P \leq 0.05$) obtained from the $\chi^2$ test. All the DEGs were assigned to the time point of their first differential expression.
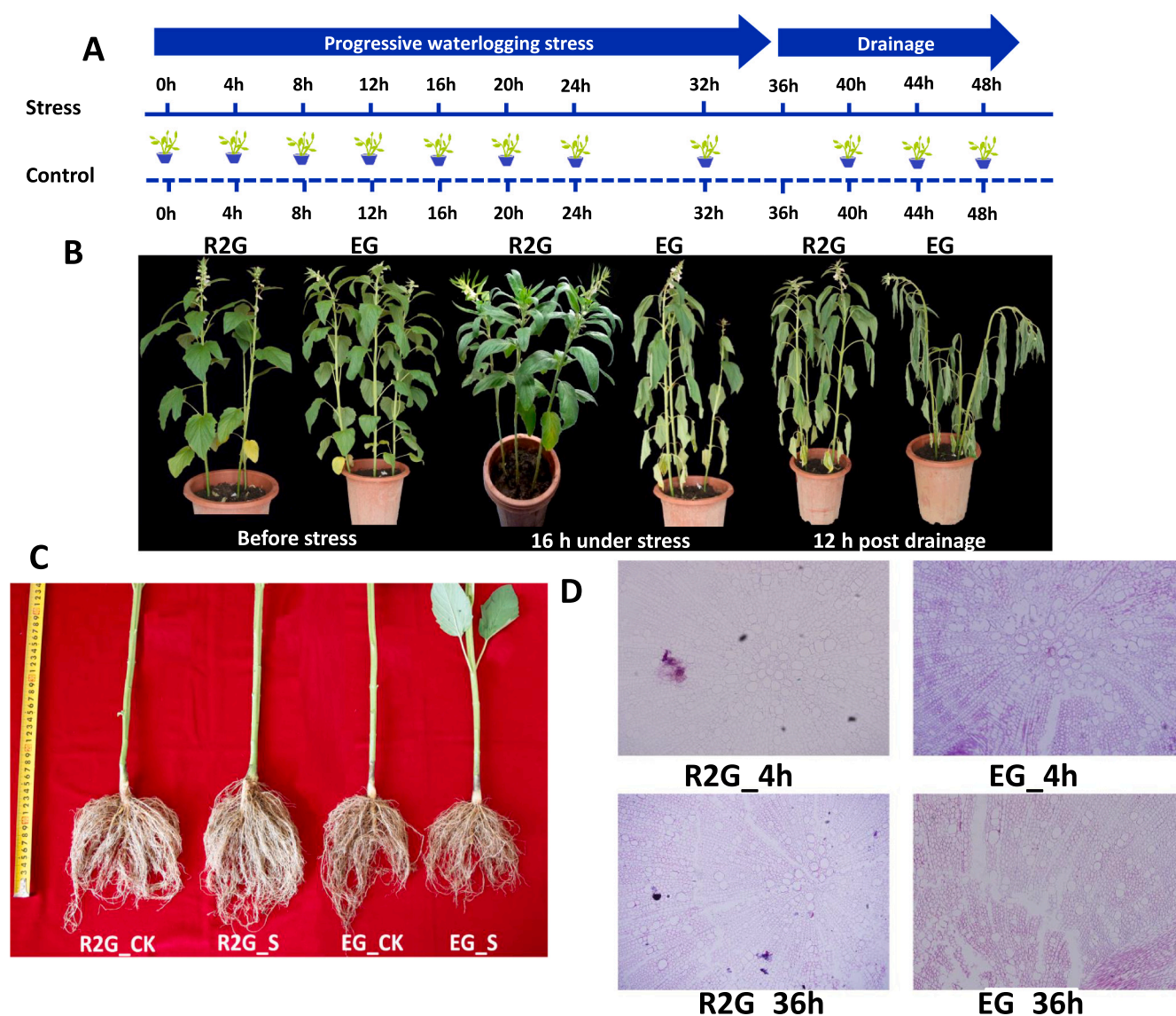
**Fig. 1.** Schematic representation of the experimental setup and analysis of the root samples. (**A**) Two sesame genotypes were grown in pots until the flowering period. Half of the plants remained in control conditions while the other half was submitted to 36 h of waterlogging followed by 12 h of drainage. Root samples were collected in triplicate at different time points in parallel in the control (CK) and stress (S) treatments. (**B**) Phenotypes of the two genotypes (R2G-tolerant and EG-sensitive) before, 16 h under waterlogging and 12 h after waterlogging stress. (**C**) Roots of single representative replicates of R2G and EG in CK and at 36 h post waterlogging stress. (**D**) Transverse root sections of R2G and EG after 4 h and 36 h waterlogging stress.

## 2.4. Time-serie clustering of gene expression profiles

Because of the high number of collecting time points, a lot of information related to the gene expression changes could be masked in a traditional clustering method such as principal component analysis. Therefore, we implemented the *tsne* reduction scatter plot, which is a non-linear dimensionality reduction method for embedding high dimensional data in a low-dimensional space [63]. The analysis was performed with the R software (http://www.r-project.org) using the '*tsne*' package v.0.1–3 (https://cran.r-project.org/web/packages/tsne/tsne.pdf).

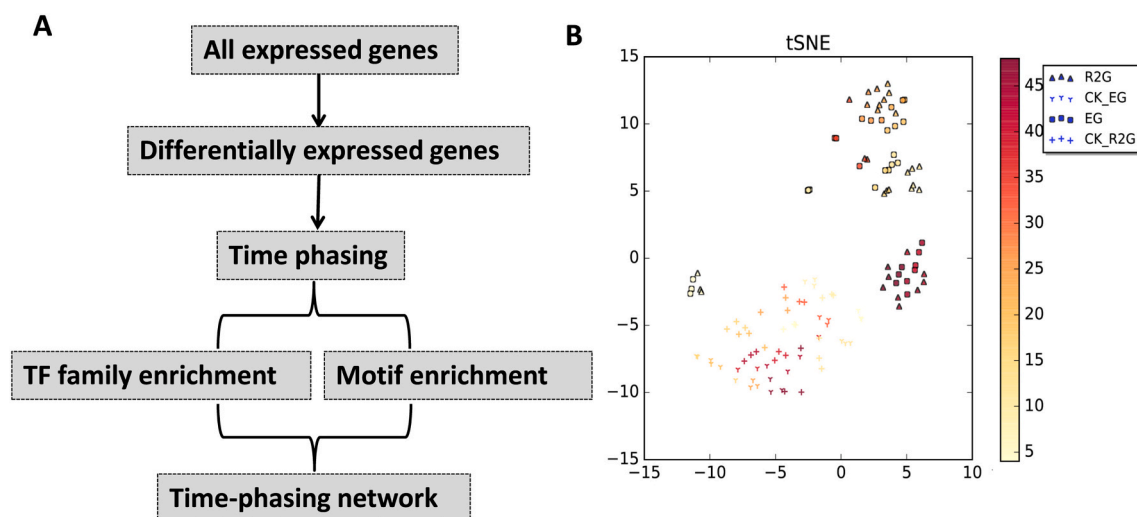## 2.5. Identifying the chronological transcriptional phases of waterlogging-drainage responses

To identify the different transcriptional phases under waterlogging/drainage treatments, the expression values of all DEGs at the different

collecting time points were employed for a hierarchical heatmap clustering in EG and R2G. Clustering of the chronological time points in the heatmaps helped group them into transcriptional phases. Accordingly, all DEGs were assigned to the transcriptional phases when they become first differentially expressed.

## 2.6. Enrichment analysis of TF family and cis-regulatory motifs

The sesame transcription factor (TF) database composed of 58 TF families and 2326 members was downloaded from the PlantTFDB v4.0 (http://planttfdb.cbi.pku.edu.cn/, [32]). Overrepresentation analysis of TF families among the DEGs in response to waterlogging/drainage treatments was performed using the cumulative hypergeometric distribution ($P \leq 0.05$) computed with the '*phyper*' function (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Hypergeometric.html) in the R software (http://www.r-project.org) with a Bonferroni correction of the *P* values [7].

**Fig. 2.** Overview of the analysis scheme of the transcriptome data. (**A**) Flowchart showing the steps in the RNA-seq data analysis. (**B**) Tsne scatter plot depicting the clustering patterns of the samples according to the genotypes and stress treatments based on the gene expression profiles. In the diagram, the transverse ordinates represent the first and second principal components; the symbols in the graph represent the samples, and the color depth represents the duration of the stress application (in hour) from light yellow to crimson to indicate a process from 0 to 48 h. CK-R2G, CK-EG represent root samples collected at different time points under control conditions from the genotypes R2G and EG, respectively. R2G and EG represent root samples collected at different time points under stress treatments from the two sesame genotypes R2G and EG, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The next step consisted in detecting the enriched cis-regulatory motifs within the promoters of the regulated genes by these enriched TF families. First, all the sesame TF binding motifs were downloaded from the JASPAR website (http://jaspar.genereg.net/, [33]). Then, sequences of 500 bp upstream from the transcription start site of all DEGs were retrieved from the sesame genome v1.0 [69]. The promoter regions were scanned for presence/absence of the motifs using the FIMO tool v5.0.3 [22] with a threshold ($P < 0.01$). All promoters were grouped into transcriptional phases based on the appertaining transcriptional phase of the corresponding gene. Then, we performed an enrichment analysis of each motif within the promoters of the DEGs belonging to a particular transcriptional phase using the cumulative hypergeometric distribution ($P \leq 0.05$) as described above.

### 2.7. Gene ontology

Gene ontology (GO) enrichment analysis was conducted in the software AgriGO v2.0 (http://bioinfo.cau.edu.cn/agriGO/, [62]). A hypergeometric test with a significance threshold of 0.05 after a Bonferroni correction was used to identify overrepresented GO categories. Similarly, the kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses were conducted using the Blast2GO software [10].

### 2.8. Regulatory network construction

To predict the regulatory network connecting the TFs to the transcriptional phases, we selected TFs within the DEGs that have annotated DNA binding motifs obtained from the JASPAR database as described above. Thereafter, the regulated genes belonging to each transcriptional phase were gathered together to test for overrepresentation of each DNA binding motif using the hypergeometric distribution ($P \leq 0.05$). According to the transcriptional phases in which a particular TF binding motif was enriched, a directional edge was drawn from the corresponding TF to that phase. The Cytoscape software v.3.6.1 [56] was employed to display the resulting networks.

### 2.9. Analysis of CAMTA5 homologs in plants

The genomic, coding and protein sequences of the gene *LOC105178907* (*CAMTA5*) were obtained from Sinbase (http://sesame-bioifo.org/Sinbase/index.html, [70]). The homologs in Arabidopsis, tomato and rice were searched using BLASTp [2]. The gene exon-intron structures were constructed using the website GSDS v2.0 (http://gsds.cbi.pku.edu.cn/, [29]) and the conserved motifs prediction was performed in the SMART tool (http://smart.embl-heidelberg.de/, [38]).

### 2.10. Yeast-one-hybrid assay

The 500 bp sequences upstream from the transcription start site of 19 predicted target genes for seven TFs were isolated (Supplementary file 3) and cloned into the bait vector pHIS2 (Clontech, USA) between EcRI/*Sma*I and *Eco*RI/*Sac*I sites. The full cDNA sequences of the TFs were PCR amplified (Supplementary file 3) and cloned into the prey vector pGADT7-Rec2 (Clontech, USA) using EcoRI/*Xho*I sites. The constructs were confirmed by sequencing at Tsingke company (Wuhan, China). The co-transformed yeasts strain Y187 containing the bait and prey were cultivated on the synthetic dextrose minimal medium (SD) lacking leucine, tryptophan and histidine (SD/−Leu/−Trp/-His) supplemented with 0 or 3 mM 3-amino-1,2,4-triazole (3-AT). The pictures were taken seven days later. Yeasts co-transformed with pGAD-53 and pGBD53 served as positive controls. The negative control was pGADT7-Rec2 and p53HIS2 co-transformation. The interaction between prey and bait was observed according to the growth of the yeast transformants in a series of 10-fold dilution.

### 2.11. Root anatomy experiment

Root samples were collected along the taproot (top 2 cm) and stored in formalin–alcohol–glacial acetic acid (90:5:5, *v*/v/v) for 24 h. Dehydration, staining and transverse sections were processed following descriptions of Dossa et al. [13,14]. Leica scanner Aperio CS2 (Leica Biosystems, Germany) was used for imaging of the root sections and image processing was performed with the Aperio Imagescope software (Leica Biosystems, Germany).

### 2.12. Biochemical analysis

We took 1 g fresh root samples harvested from plants under stress and non stress conditions (in triplicate), crushed and extracted with 1 ml of 50 mM sodium phosphate (pH 7.4) buffer and kept on ice for 20 min. The mixture was centrifuged at 12000 *g* for 15 min and quantified using a spectrophotometer. Measurements of ascorbate peroxidase (APX), catalase (CAT), peroxidase (POD), superoxide dismutase (SOD), and malondialdehyde (MDA) were performed using the kits provided by Nanjing Jiancheng bioengineering Institute, China, following manufacturer's instructions [76].

### 2.13. qRT-PCR validation

qRT-PCR was performed with three independent biological replicates and three technical replicates of each biological replicate using a Light Cycler 480 II (Roche, Switzerland). Experimental details are reported in work of Dossa et al. [13,14]. The relative expression level of the selected genes was normalized to the expression level of the endogenous gene *SiH3* based on the $2^{-\Delta\Delta Ct}$ method [43]. The primer list is presented in Supplementary file 4.

### 2.14. Statistical analysis

Data were analyzed with the R (www.r-project.org). One-way analysis of variance was performed by comparing the two genotypes, followed by Tukey HSD test for mean comparison.

## 3. Results

### 3.1. High-resolution gene expression pattern under time-course waterlogging/drainage

We sequenced a high-resolution time series transcriptome data from root of two contrasting sesame genotypes (EG and R2G) at 11 consecutive time points within a 48 h period waterlogging/drainage treatments (Fig. 1a). The tolerant genotype R2G exhibited low wilting signs throughout the experiment and rapidly recovered after drainage. On the contrary, the sensitive genotype EG was strongly wilted (Fig. 1b). The root biomass of the stressed R2G plants was increased as compared to the non stressed plants while the opposite was observed in EG (Fig. 1c). We further analyzed the anatomy of the taproot samples collected from plants after 4 h and 36 h waterlogging treatment. We observed that R2G presented a well-organized taproot structure after 4 h stress with very few intercellular spaces resulting from the collapse of the parenchyma cells after 36 h waterlogging (Fig. 1d). In contrast, after 4 h waterlogging stress, EG showed a high loss of cortical cells and later, a disintegration of the internal root structure was observed. These results confirm that R2G had higher capacity to sustain waterlogging stress as compared to EG. Waterlogging induces oxidative stress and a strong activity of antioxidant enzymes, including ascorbate peroxidase (APX), catalase (CAT), peroxidase (POD), superoxide dismutase (SOD), is essential for buffering excessive reactive oxygen species (ROS). The malondialdehyde (MDA) content is associated with lipid peroxidation via an increased generation of ROS, therefore, MDA level is a good indicator for stress tolerance [76]. In this study, we observed globally a significantly higher induction of APX, CAT, POD in R2G than EG throughout the waterlogging stress duration. SOD induction was not significantly different between the two genotypes. However, the MDA level was significantly higher in EG than R2G, indicating that EG suffered more from oxidative stress induced by waterlogging than the tolerant genotype R2G (Supplementary file 5).

Overall, the RNA-sequencing of the 126 root samples (Supplementary file 2) yielded over 633 Gb clean data with 97% of bases scoring Q30 and above (Supplementary file 6). Gene expression levels were normalized to Transcripts Per Million (TPM) and further analyses were implemented based on genes with TPM > 1 following the flowchart in Fig. 2a. A generalized linear model was employed to identify genes whose transcript levels differed significantly under waterlogging/drainage treatments and under control conditions [25]. In total, we obtained a set of 6201 unique differentially expressed genes (DEGs) (Supplementary file 7), including 4912 genes in R2G and 6032 genes in EG. The greater preponderance of affected genes in EG underscores its sensitivity to waterlogging [68]. We randomly selected 14 DEGs and estimated their expression fold change under temporal waterlogging stress based on qRT-PCR. Globally, the qRT-PCR result was similar to the RNA-seq report with a Pearson correlation coefficient of $R^2 = 0.78$ (Supplementary file 8), validating the RNA-seq based gene expression profiling and DEG detection in this study.Based on the gene expression profile, the *tsne* reduction analysis was implemented and as shown in Fig. 2b, the samples could be roughly divided into four groups. The first group gathered together CK_R2G and CK_EG showing that the two genotypes displayed similar transcriptional reprogramming over time under control conditions. The second and third groups included samples collected during the early and late waterlogging stresses, respectively. Finally, the last group encompassed samples from the post drainage stage in R2G and EG. Altogether, the observed clustering pattern denotes that sesame differentially readjusts the transcriptional activity upon waterlogging and drainage treatments as compared to the control conditions.

### 3.2. Key transcriptional phases in sesame response to waterlogging/drainage

We took advantage of our dense time series transcriptome data to unravel the major phases in the transcriptional readjustment under waterlogging/drainage treatments. The hierarchical heatmap clustering was performed based on the DEGs. Although a slight difference could be observed in the transcriptional time-phasing results of R2G and EG, globally, three key chronological transcriptional phases were highlighted: phase 1 = 4, 8 and 12 h; phase 2 = 16, 20, 24 and 32 h and phase 3 = 40, 44, 48 h (Fig. 3a,b). Based on the chronology of the time series in the different phases, we deduce that phase 1 and phase 2 represent the early and late transcriptional responses to waterlogging, respectively, while phase 3 constitutes the post-drainage transcriptional readjustment. Next, we assigned the DEGs to the different transcriptional phases according to the time point they become first differentially expressed (Supplementary file 7). Consistently in the two genotypes, phase 1 involved more than half of the DEGs (58% and 56% in EG and R2G, respectively), suggesting that waterlogging stress induces a burst of transcriptional activity at the early stage. A lesser proportion of the DEGs were assigned to phase 2 (39% in EG and R2G) and only 3 and 5% of the DEGs in EG and R2G, respectively, were assigned to phase 3. These patterns of transcriptome reprogramming under waterlogging/drainage treatments were quite similar between the two genotypes, therefore, a deep excavation of the enriched pathways and the gene models within each transcriptional phase would help illuminate the features underlying the sensitivity/tolerance of these two genotypes. For example, only 46, 32 and 10% of the DEGs were shared by both genotypes during the transcriptional phases 1, 2 and 3, respectively, indicating that each genotype has a specific response to waterlogging/drainage during the time-course (Fig. 3c). Nonetheless, we identified 47 conserved genes constantly differentially expressed throughout all the seven time points under waterlogging stress in the two genotypes, which represent the core waterlogging-responsive (CWR) genes despite the tolerance levels in sesame (Supplementary file 9). Only ten of the CWR genes were up-regulated by waterlogging. Overall, the sesame CWR genes encode various enzymes (ADH, trehalose 6-phosphate phosphatase, flavonol synthase, cytochrome P450, naphthoate synthase, chlorophyll synthase, peroxiredoxin, etc.) important in anaerobic metabolism and cellular homeostasis.

### 3.3. Enriched pathways in the different waterlogging/drainage transcriptional phases

To get insight into the biological functions of the DEGs involved in each transcriptional phase, the kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis was performed (Fig. 4). Upon exposure to waterlogging stress (phase 1), genes belonging to various biological classes, mainly related to the terms of metabolic pathways and biosynthesis of secondary metabolites were differentially expressed in the tolerant genotype R2G. These major KEGG terms along with several minor KEGG terms such as nitrogen metabolism, fatty acid metabolism, ABC transport, glutathione metabolism, etc., were absent in the sensitive genotype. This observation denotes a rapid reaction to waterlogging stress at the early stages in R2G. The second transcriptional phase involved similar KEGG terms like the first one, but novel terms such as fatty acid biosynthesis, biotin metabolism, ascorbate and aldarate metabolism, etc., were also enriched in R2G. We noticed a conspicuous similarity between the second transcriptional phase-enriched KEGG terms in the sensitive genotype EG and those from the first transcriptional phase in R2G, which implies that the adaptive responses to waterlogging in the sensitive genotype were delayed until 16 h after the stress application. Finally, the drainage stage (phase 3) involved several genes related to energy production, a mechanism which is quite conserved in both genotypes, though the amplitude of implicated genes was slightly higher in R2G. Overall, our results indicate a major difference in the transcriptional responses at the onset of the waterlogging treatment between the two genotypes, which may underpin their contrasting tolerance levels.

### 3.4. Overview of the major transcription factors families active during the different waterlogging /drainage transcriptional phases

Transcription factors (TF) are the major operators of the transcriptional reprogramming under abiotic stress [15,39]. The DEGs assigned to each transcriptional phase were screened in order to identify the enriched TF families active in response to waterlogging/drainage. Overall, 23 TF families, including 14 in R2G and 21 in EG were found principally regulating gene expression during sesame responses to waterlogging/drainage (Supplementary file 10). The WRKY and ERF members were the most dominant TFs detected throughout waterlogging/drainage responses in the two genotypes (Fig. 5). Besides, the genes belonging to other TF families such as HSF, GRAS, MYB and ARF were also overrepresented. Since the genotype R2G is highly tolerant to waterlogging stress, we regarded its phase-specific enriched TF families
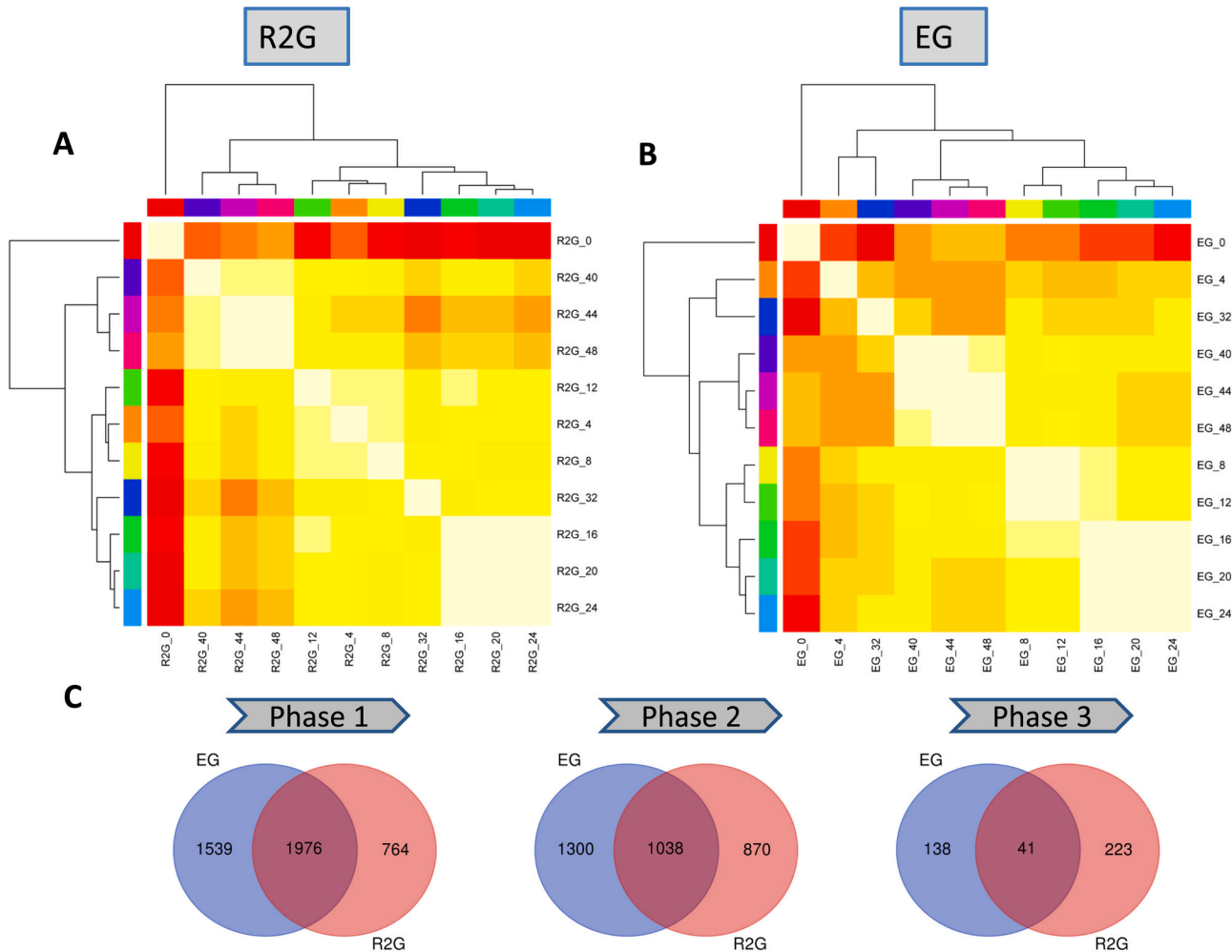


**Fig. 3.** The waterlogging/drainage response consists of three transcriptional phases. (**A,B**) Heatmap hierarchical clustering of the differentially expressed genes (DEGs) from the different time points in two sesame genotypes R2G and EG. Color ranges from cream to red show high to low correlation coefficients. Based on the chronology of the events, three transcriptional phases were highlighted: phase 1 = 4, 8 and 12 h; phase 2 = 16, 20, 24 and 32 h and phase 3 = 40, 44, 48 h. The DEGs were assigned to the different transcriptional phases according to the time point they become first differentially expressed (**C**). Venn diagram depicting the shared and unique DEGs at the three transcriptional phases in two sesame genotypes R2G and EG. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
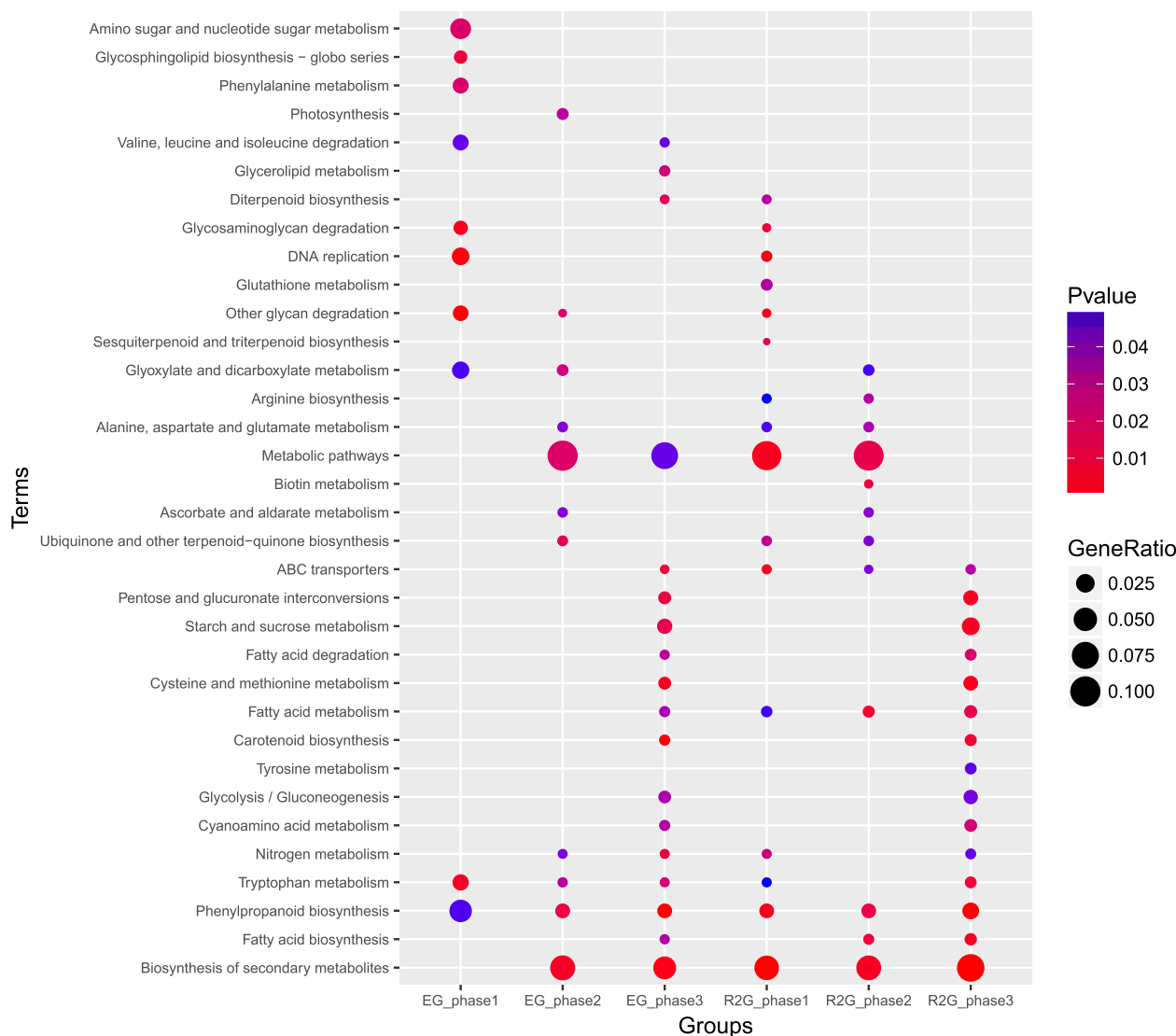
**Fig. 4.** The kyoto encyclopedia of genes and genomes enrichment analysis of the differentially expressed genes at the three transcriptional phases in two sesame genotypes R2G and EG.

as the model transcriptional regulation to survive waterlogging stress. Compared to the model in R2G, the sensitive genotype EG showed a delay in triggering the efficient transcriptional regulation. For example, several ERF TFs were prompted to modulate gene expression at the onset of waterlogging stress (phase 1) in R2G, but this mechanism was observed only much later (phase 2) in EG. Another example concerns the STAT gene (*LOC105165486*) which was specifically activated during phase 2 and phase 3 in R2G and EG, respectively. This result relates that a timely and accurate gene expression regulation during the transcriptional phases is essential for survival to waterlogging/drainage in sesame.

*3.5. Insight into the functional ERF groups active under waterlogging/ drainage*

Previous reports on Arabidopsis and rice highlighted the importance of ERF TFs, particularly group VII-ERFs, in sensing the low-oxygen induced by flooding and the induction of adaptive responses [21,24,41,72]. In this study, the abundance of ERF TFs differentially expressed at the early stage of waterlogging stress in R2G suggests that this genotype may rapidly sense the stress and quickly trigger downstream signaling networks involved in the acclimation response. This

mechanism was not observed in the sensitive genotype and we deduced that a rapid sensing of waterlogging stress via the ERF TFs may be crucial for tolerating the stress. We further investigated the ERF groups active in R2G during the transcriptional phase 1 based on the classification of the sesame AP2/ERF superfamily provided by Dossa et al. [15]. Surprisingly, we did not find any member of group VII differentially expressed, which was further confirmed through qRT-PCR analysis (Supplementary file 11). Instead, members of the groups III and IX were overrepresented (Fig. 6). Within these early waterlogging-responsive regulators, four genes including *LOC105175593*, *LOC105163408*, *LOC105164449* and *LOC105168482* were strongly induced just after 4 h waterlogging, denoting their early roles upstream of signaling pathways. Similarly, when we analyzed the ERF TFs regulated later (phase 2) in the sensitive genotype, we observed that members of groups III and IX were also abundant along with only one gene (*LOC105157874* (*SiRAP2.2*)) from group VII (Fig. 6a). Interestingly, the two genotypes shared 11 ERF TFs highly active under waterlogging stress. These results indicate that groups III and IX ERF genes are likely to be the main drivers of the activation of waterlogging-responsive genes in sesame. ERF TFs were also implicated in the transcriptional responses during the recovery from waterlogging in the two genotypes, with the majority being downregulated (Fig. 6b).

### 3.6. Major functional WRKF TFs active under waterlogging/drainage

A total of 60% of WRKY TFs present in the sesame genome were strongly implicated in waterlogging and drainage responses in our study (Supplementary file 7). WRKY TFs displayed a more positive regulation than ERF TFs under waterlogging/drainage in sesame. Available transcriptome data in other plant species indicate that several genes encoding WRKY Ts are induced to high levels upon flooding [35,37,44]. Based on the classification of the sesame WRKY family [39], we searched for overrepresented WRKY TFs active during the different transcriptional phases. Members from all of the seven sesame WRKY TFs were represented, showing the importance of this gene family in waterlogging/drainage responses in sesame (Fig. 6c,d). In particular, the members of group II-b were abundant within DEGs at the onset of the waterlogging stress, implying their early roles in flood adaptive responses. Members of group I were mainly involved in the transcriptional phase 2. The two genotypes shared 13 and 8 WRKY TFs during the first and second transcriptional phases (Fig. 6c,d), respectively, representing the core WRKY members necessary for sequential transcriptional regulation under waterlogging in sesame.

### 3.7. Predicting the target regulated genes through cis-regulatory motifs enrichment analysis

The identification of overrepresented cis-regulatory elements is a key step in shedding light on the regulatory mechanisms active in a particular environmental condition [25]. In order to elucidate the waterlogging/drainage regulatory network in sesame, we screened for overrepresented regulatory motifs in the 500 bp promoter regions of the target genes assigned to each of the three transcriptional phases under

waterlogging and drainage. As expected, the main enriched TF binding motifs were WRKY and ERF during the transcriptional phases 1 and 2, correlating with the activity of these major regulators (Fig. 7). However, we also observed a high number of regulated genes harboring the DOF motifs in their promoters, although the DOF TFs were not significantly enriched within the main regulators during the phases 1 and 2. In addition, after drainage, the cis-regulatory motifs related to CAMTA, TCP and EIL were predominant in the promoters of the regulated genes, which did not correspond to the regulators identified during this particular transcriptional phase. These results indicate a complex sesame waterlogging/drainage regulatory network in which, a particular TF may regulate gene expression in its appertaining/different and single/multiple transcriptional phases. Moreover, some members of less-represented TF families (for example, the DOF genes *LOC105163550* and *LOC105174708*) may also be as crucial as regulators belonging to the enriched TF families for their roles in gene expression regulation.

### 3.8. Regulatory networks underlying waterlogging/drainage response in sesame

Next, we constructed the gene regulatory networks predicting directional interactions between waterlogging/drainage responsive regulators and all regulated genes associated with the three transcriptional phases using the TF DNA binding motif information. Fig. 8 presents the generated regulatory networks underlying waterlogging/drainage responses in EG and R2G, in which the circular nodes represent the key regulators connected by an edge to a transcriptional phase (represented by a rectangle in the network) when the corresponding DNA binding motif was overrepresented in that phase. The size of the nodes is proportional to the number of the inferred regulated genes
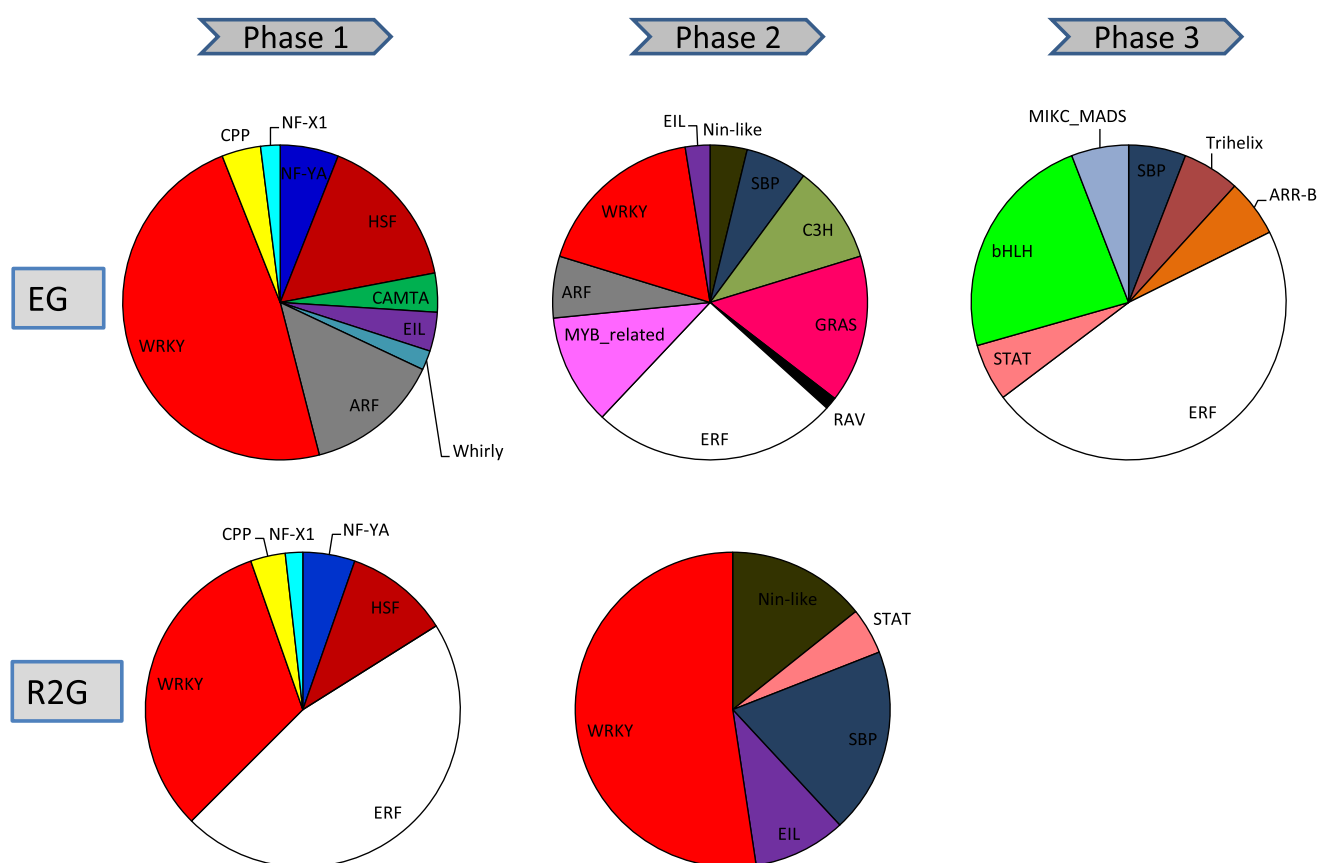


**Fig. 5.** Several TF families are involved in waterlogging/drainage responses in sesame. The pie charts depict the proportion of the major TF families enriched (hypergeometric distribution; $P \leq 0.05$) within the differentially expressed genes activated during the three transcriptional phases in two sesame genotypes R2G and EG.
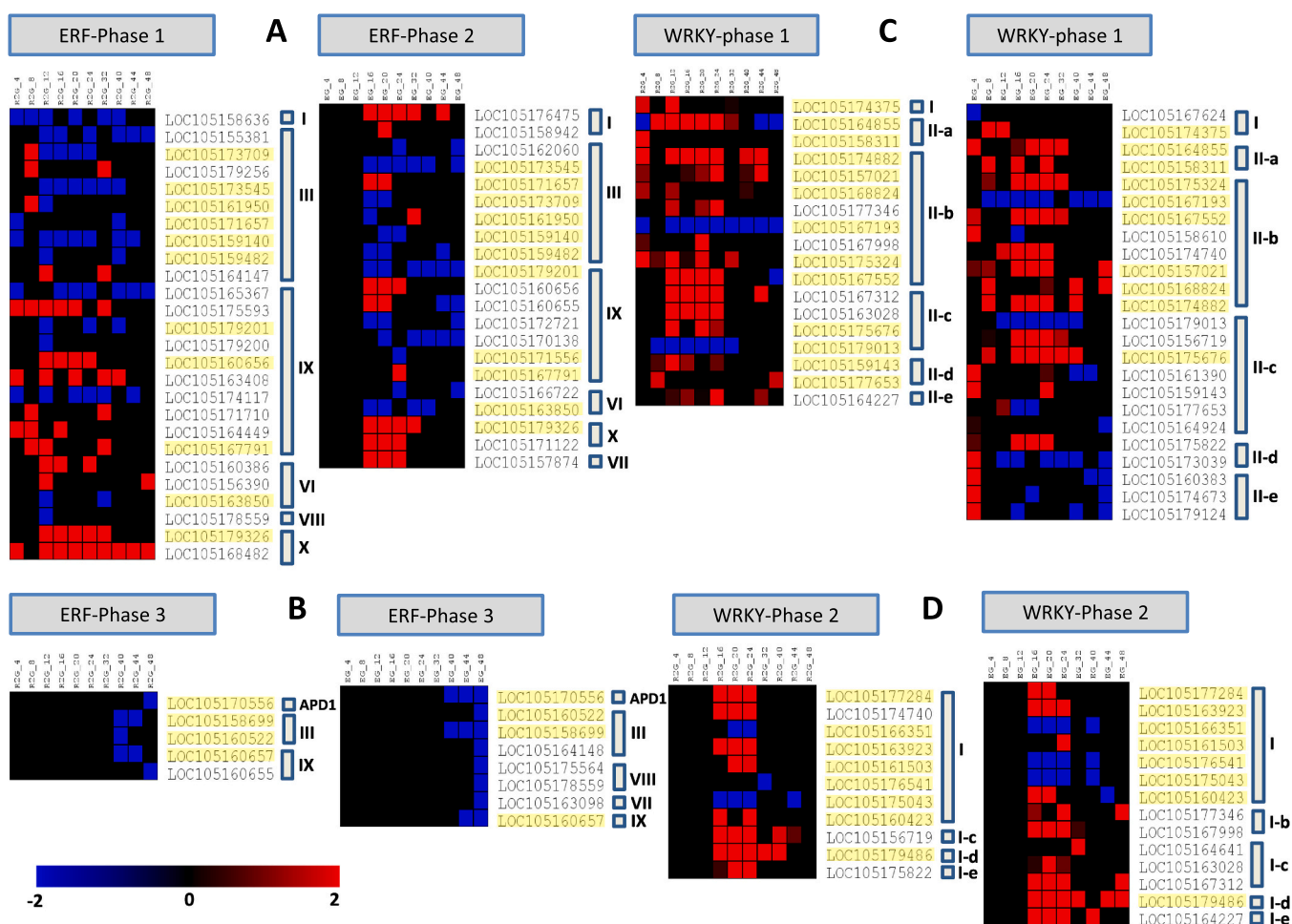
**Fig. 6.** Various ERF and WRKY TFs groups are implicated in the waterlogging-drainage responses in sesame. (A,B) Heatmap displaying the log2 fold change of the differentially expressed genes from the ERF family in two sesame genotypes R2G and EG during phases 1–3. (C,D) Heatmap displaying the log2 fold change of the differentially expressed genes from the WRKY family in two sesame genotypes R2G and EG during phases 1–2. At each time point, up-regulation is shown in red, down-regulation in blue and no-regulation is displayed in black color. The genes highlighted in yellow are those conserved in the two genotypes. The Roman numerals represent the different ERF and WRKY groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

harboring the corresponding TF binding motifs in the promoter region and nodes are colored according to the transcriptional phase where they first became differentially expressed. As suspected, our predicted network models showed that TFs may regulate genes from several transcriptional phases. For instance, the genes *LOC105178907* (CAMTA), *LOC105178559* (ERF), *LOC105175742* (TCP), *LOC105159163* (bZIP), etc., were predicted to modulate gene expression in two or three different phases.

Consistently in the two genotypes, the ERF and WRKY genes represent the largest TF families, ubiquitously active in the transcriptional regulation of waterlogging/drainage responses in sesame. The two genotypes shared similar regulators (38) but the regulation patterns greatly differed, which may partly explain their contrasting responses to waterlogging/drainage. The predicted regulatory networks highlighted the main TFs acting upstream of waterlogging signaling cascade including two key ERF TFs (*LOC105159140* (DEAR), and *LOC105178559* (ERF9)) in R2G (Supplementary file 12). Several TFs belonging to phases 1 and 2 regulated genes from following phases. In addition, a particular transcriptional regulation mechanism was also predicted, in which regulators may target genes from previous transcriptional phases. Exceptionally, TFs belonging to phase 3 (drainage) mainly regulated genes from phases 1 and 2, indicating a mechanism towards the restoration/recovery of a functional metabolism which was

slowed down under waterlogging stress [68]. The predicted regulatory networks also highlighted some master players (*LOC105174708* (Dof), *LOC105163550* (Dof), *LOC105178280* (HD-ZIP), *LOC105171783* (MYB), *LOC105179594* (Trihelix), *LOC105173131* (MYB-related), *LOC105167312* (WRKY), *LOC105179013* (WRKY) and *LOC105161503* (WRKY)), which are predicted to regulate significant numbers of target genes. Other constitutive regulators (*LOC105175742* (TCP), *LOC105171768* (TCP) and *LOC105178907* (CAMTA)) participating in all the transcriptional phases of waterlogging/drainage responses were also predicted. Distinctly, the constitutive switch *LOC105178907* (CAMTA) was conserved in the two genotypes (Fig. 8). This gene displayed a positive response to waterlogging stress as its expression level was highly increased upon exposure to the stress. *LOC105178907* is annotated as a calmodulin-binding transcription activator 5 (*CAMTA5*) which is present in various plant species, including mocots and dicots (Arabidopsis (*AT4G16150*), rice (*LOC_Os07g30774*), tomato (*Solyc12g99340*)). Although the gene structure of *CAMTA5* has considerably changed among these species, the functional protein motifs are well conserved (Supplementary file 13).

In order to provide a solid experimental support to the predicted interactions in the regulatory networks presented in Fig. 8, we performed a yeast-one-hybrid assay using the promoter regions of 19 target genes of seven selected TFs (*LOC105163550, LOC105167312,*
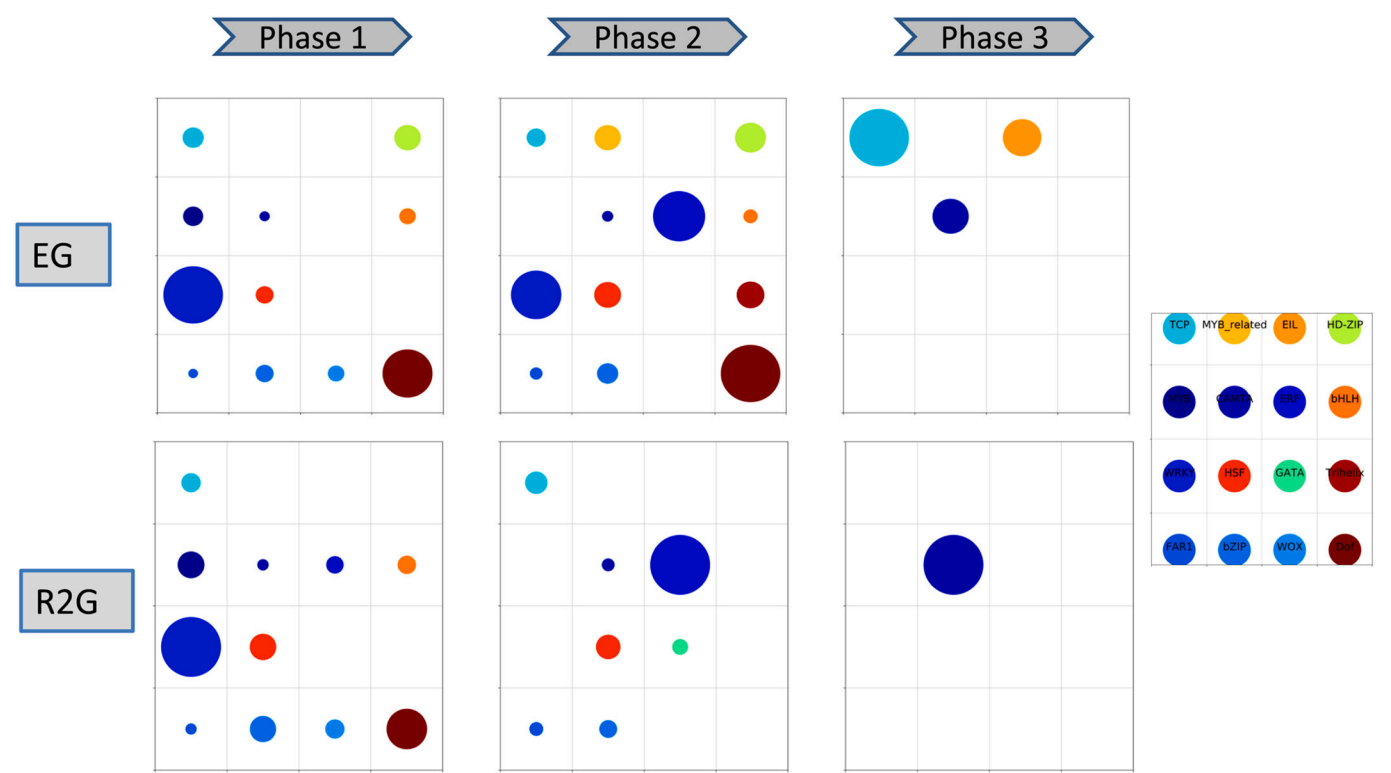
**Fig. 7.** Identification of the overrepresented transcription factors binding motifs. The bubble charts indicate known TF DNA binding motifs overrepresented in gene promoters (hypergeometric distribution; P ≤ 0.05) belonging to the three transcriptional phases in R2G and EG.
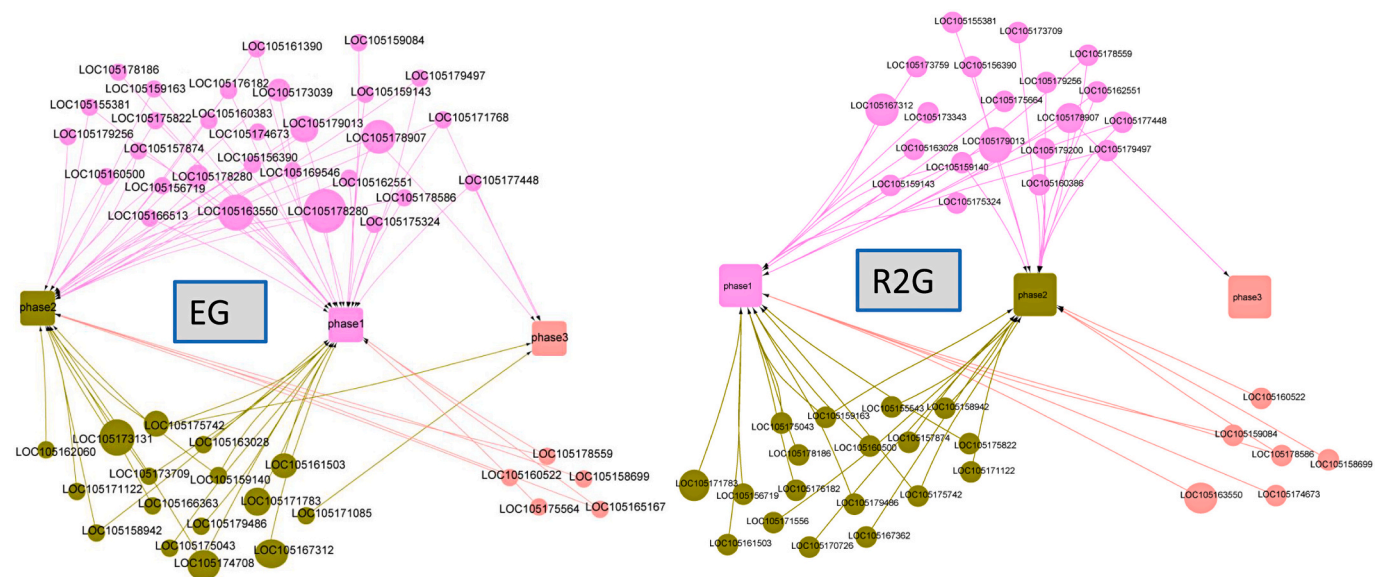


**Fig. 8.** Predicted directional interactions in the waterlogging-drainage gene regulatory network. Network plot of inferred connections between waterlogging-drainage-induced TF and genes in the three transcriptional phases. The promoter sequences of genes associated with a transcriptional phase were tested for over-representation of DNA motifs shown to be bound to TFs that are differentially transcribed following waterlogging-drainage treatments in two sesame genotypes R2G and EG. Each TF with a known motif is represented by a colored circle and is plotted at the time point that its corresponding gene is first differentially expressed. Each transcriptional phase is represented by a rectangle. An edge between a TF and a phase indicates significant enrichment of the corresponding binding motif in that phase. The size of each TF node is proportional to the number of phases in which its binding site is overrepresented and the number of predicted regulated downstream genes. To facilitate the interpretation of the network, nodes are grouped and colored according to the transcriptional phase where they first become differentially expressed.

LOC105179013, LOC105179594, LOC105173131, LOC105171768 and LOC105178907). Except for two targets (LOC105161887 and LOC105171509) which did not show successful interactions with the TF LOC105163550, our results demonstrated that all the selected TFs could bind to the 17 other corresponding target genes (Fig. 9).

### 3.9. The gene pools underpinning waterlogging/drainage differential responses in R2G and EG

To uncover the key genes underpinning the differential responses to waterlogging/drainage in R2G and EG, we compared their transcriptomes at the different time points grouped into transcriptional phases. In total, 474 genes were differentially expressed between the two genotypes including 335, 88, 51 genes during phases 1, 2 and 3, respectively (Supplementary file 14). This result further supports the premise that the R2G waterlogging tolerance mechanism stems from the early stage of exposure to the stress. The DEGs were enriched in the KEGG terms related to biosynthesis of secondary metabolites, protein processing in endoplasmic reticulum, phenylpropanoid biosynthesis, etc. (Supplementary file 15). GO enrichment analysis of these DEGs demonstrated their involvement in several biological processes related

to metabolic activity, response to abiotic stress and biotic defense (Supplementary file 16). It is worth noting that R2G regulated 173 genes early (4 h) and constantly throughout the waterlogging duration but these genes were regulated only after 8 h in the sensitive genotype (Supplementary file 17). These genes encode important genes involved in the respiration process, antioxidants, aquaporins, etc., which may represent crucial components of the early responses and tolerance to waterlogging in R2G.

## 4. Discussion

Sesame is highly sensitive to waterlogging stress and unlike drought, a few hours of waterlogging (lasting over 36 h) are detrimental to the crop's survival, growth and yield, particularly, during the crop establishment and flowering stages [1,50,57]. In this short time, the plant undergoes enormous transcriptional changes, as reported by Wang et al. [68], who found that 72% of the global transcriptome is altered. Hence, elucidating waterlogging stress transcriptional responses in sesame requires a high-resolution temporal transcriptome data as generated for the first time in the present work. According to findings by Tamang et al. [60], the post-drainage stage is also a crucial process in crop survival
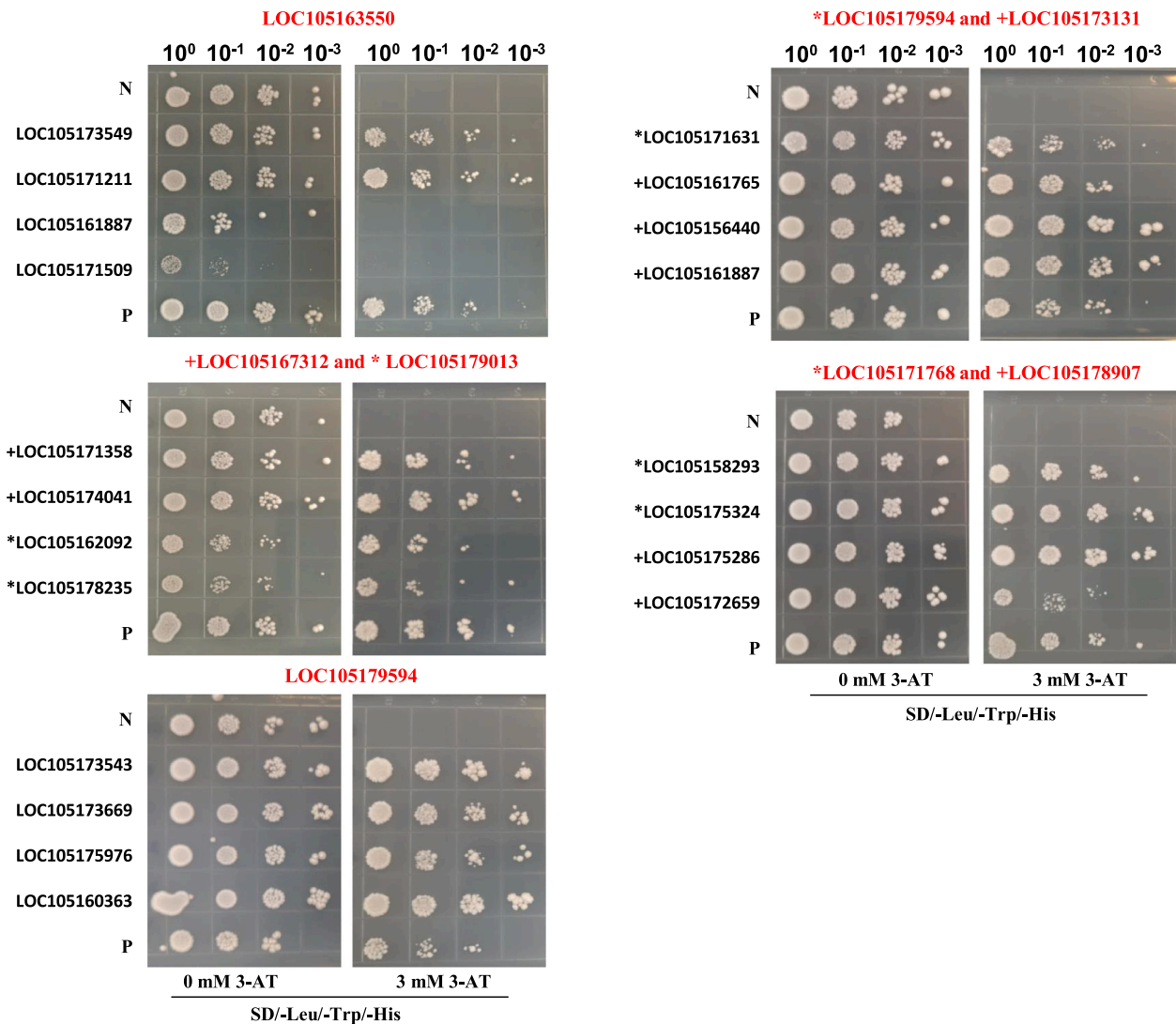


**Fig. 9.** Yeast one hybrid assay for the interactions between seven selected transcription factors (LOC105163550, LOC105167312, LOC105179013, LOC105179594, LOC105173131, LOC105171768 and LOC105178907) and their predicted target genes. The co-transformed yeasts strain Y187 containing the bait and prey were cultivated on the synthetic dextrose minimal medium (SD) lacking leucine, tryptophan and histidine (SD/−Leu/−Trp/-His) supplemented with 0 or 3 mM 3-amino-1,2,4-triazole (3-AT). N = negative control and P = positive control. The symbols + and * show correspondence between TFs and their targets.

after waterlogging. Transcriptomic readjustment during the post-waterlogging stage has been largely neglected [73]. In this study, we extended our time-course transcriptome data over 3 time points after water withdrawal to provide a comprehensive view of waterlogging/drainage transcriptional responses in a non-model crop. Furthermore, the comparison of the transcriptome reprogramming in genotypes displaying contrasting responses to waterlogging stress offers a novel insight into the tolerance mechanism to this stress. Transcriptome is very dynamic and even under normal growth conditions, in the lapse of a few hours, the expression levels of hundreds of genes are altered. For example, by comparing the transcriptomes from the different time points to 0 h in the control conditions (Fig. 1), we observed that over 11% of the global transcriptome was significantly changed in a period of 48 h. To preclude some background noises during the DEG detection and related biases in the result's interpretation, in our study, the gene expression under waterlogging/drainage treatment was compared to the controls from the same sampling time point [55]. Although this approach greatly increased the sampling size and costs, it guaranteed high quality DEG records [25].

### 4.1. The early transcriptional phase is crucial for tolerating waterlogging stress in sesame

Taking advantage of our dense time-series data, we were able to discern three distinct transcriptional phases, involving different biological processes during the waterlogging/drainage responses. Phase 1, as the early transcriptional response, engages genes acting as the stress sensors and functioning upstream of flooding signaling cascades. It is a fundamental step in plant acclimation to flooding as the initiation of adaptive responses requires an accurate and a timely perception of water inundation [53]. We observed a transcriptional burst driven by several TFs mainly, ERF and WRKY TFs during this phase leading to intense changes in metabolic pathways. The key finding in this study is the difference between the tolerant (R2G) and sensitive (EG) genotypes in their early transcriptional responses to waterlogging stress. Several TF families, including CPP, NF-X1, NF-YA, WRKY and HSF, were commonly present during the transcriptional phase 1 in the two genotypes, however, the ERF genes, which were highly regulated upon exposure to waterlogging in R2G were only activated lately in EG. During the last decade, extensive studies have demonstrated the roles of ERF TFs in the stress signal perception and regulation of flood-responsive genes in plants ( [11,19,21,24,41,45,52,64]. From our results, we suspect that the tolerant genotype perceives early waterlogging stress through the ERF TFs and rapidly triggers adaptive transcriptional regulation. The sensitive genotype EG behaves like the majority of sesame cultivars under waterlogging stress. Therefore, much attention needs to be focused on the understanding of the molecular mechanisms leading to an early stress sensing and a timely activation of the waterlogging core responsive genes detected in this study.

### 4.2. Group ERF-VII genes in sesame and Arabidopsis are differentially regulated

Flooding severely restricts gas diffusion in plant organs and causes low oxygen availability. Members of group ERF-VII, including *HRE1*, *HRE2*, *RAP2.3*, *RAP2.2* and *RAP2.12* have emerged as important regulators of the responses to waterlogging-induced low-oxygen via an evolutionary conserved N-end rule pathway in Arabidopsis [51]. Further evidences in rice with the genes *SUB1A* and *SNORKEL1/2* have propelled the ERF-VII group as a cornerstone in plant hypoxia responses [24,72]. The sesame genome contains three ERF-VII genes (*SiHRE2*, *SiRAP2.2* and *SiRAP2.12*) characterized by the conserved Met-Cys initiating motif (Supplementary file 18), therefore, could be substrates of the N-end rule pathway [51]. In the present study, the gene *RAP2.12* was not affected by waterlogging stress, which is consistent with the reports in Arabidopsis [41,44]. But, *RAP2.12* as the main sensor of the

oxygen level was constitutively expressed in Arabidopsis root ( [21,41], an observation that we did not note for its homolog in the sesame root. Also, the other key gene (*RAP2.2*) involved in sensing the oxygen level in Arabidopsis [20] was not affected by hypoxia [44]. But, in our study, *SiRAP2.2* was induced (2 fold) after 16 h of waterlogging in both genotypes (Supplementary file 7). Moreover, the genes *HRE1* and *HRE2* are known to be strongly induced upon exposure to hypoxia in Arabidopsis [42]. Surprisingly, the counterpart of *HRE2* in sesame, was not induced by waterlogging. More importantly, *HRE2* transcript level is low in the Arabidopsis root [42], but a constitutive expression of this gene was noted in the sesame root. Our results indicate that the regulations of the ERF-VII genes under flooding in sesame and Arabidopsis are quite different. Van Veen et al. [65] also revealed how ERF-VIIs have evolved to behave in varying ways across species. A primary sensor of the ambient oxygen level must be present before the oxygen level decreases [51], hence, we suspect that *SiHRE2* might act as the primary oxygen sensor in sesame. The inspection of our previous transcriptome data from root of waterlogged distinct genotypes [67,68] showed that *SiHRE2* was unaltered and constitutively expressed, providing additional support to our hypothesis. *SiHRE2* is subtly more expressed in the tolerant genotype than the sensitive one and it is tempting to speculate that this may underpin the quick and stronger hypoxic response in R2G [65]. Nevertheless, the roles of these three ERF-VII genes in sensing the oxygen levels and their potential interactions with other ERF groups implicated in the early transcriptional phase are pending further in-depth investigations.

### 4.3. WRKY TFs, the overlooked working force active under waterlogging stress

In the present study, we found that many WRKY TFs act as positive regulators of the waterlogging responses in sesame, especially at the onset of waterlogging. Similar observations have been reported in several plant species, including Arabidopsis and rice ( [8,27,44,66]. Heterologous expression of the sunflower transcription factor (TF) *HaWRKY76* in Arabidopsis was shown to confer submergence tolerance [48]. Unfortunately, the roles of the majority of WRKY family members in plant waterlogging responses are still poorly understood. We observed that 13 and 8 differentially expressed WRKY TFs were shared by the two assayed genotypes during phases 1 and 2, respectively. The significant number of common WRKY-DEGs and importantly, the conserved timing of their differential regulation in R2G and EG indicate that these core WRKY TFs play fundamental roles in the sophisticated waterlogging acclimation in sesame. Hsu et al. [27] have linked the activity of WRKY genes, particularly the gene *WRKY22*, to the induction of the innate immunity in Arabidopsis upon exposure to waterlogging. However, whether this proposed model is widely conserved in plants has not been validated. We checked the homolog of *WRKY22* in sesame (*LOC105173039*) and found that it was not differentially expressed in the tolerant genotype under waterlogging/drainage treatments. It may be possible that another WRKY TF in sesame plays the same role of *WRKY22*, therefore, a particular attention should be paid to the core WRKY-DEGs.

WRKY and ERF represent the two major flooding-activated TF families in plants [28,44,68]. However, whether these two gene families interact in the same signaling pathways under flooding is unknown. The results from our study hint at no important interaction between WRKY and ERF TFs in flooding responses. In fact, if the regulation of WRKY TFs was related to the ERF TFs, which are the primary regulators of flooding stress, we would have noticed a marked difference between R2G and EG in the transcriptional phases of WRKY TFs activation, similar to ERF genes. Hence, we deduce that WRKY and ERF gene expression regulations are triggered by distinct flooding-associated signals. Recently, Arbona et al. [4] have demonstrated the importance of ABA as a central hormonal signal in roots of flooded Carrizo citrange. The induction of WRKY TFs has been described in some cases as ABA-dependent [49],

therefore, the potential link between WRKY TFs and ABA in flooding adaptive responses will be worth thorough experimentations.

### 4.4. Construction of transcriptional networks helps discover major players in the waterlogging/drainage responses

Although extensive molecular studies have been performed in waterlogging responses in plants, a comprehensive insight into the waterlogging transcriptional regulatory network is still lacking. Integrating TF DNA binding motif enrichment data with our chronological transcriptional network model predicted putative causal regulations between TFs and downstream waterlogging-responsive genes. In total, we successfully identified 59 candidate regulatory factors, many of which are well conserved in the two sesame genotypes, a clue of the biological relevance of our predicted model. A similar computational analysis performed by Hickman et al. [25] highlighted the potential of combining time series expression data with motif analysis to infer key regulators and their targets in gene regulatory networks. They identified several yet unknown important regulators of pathogen and insect resistance in Arabidopsis. In our study, we identified CAMTA TFs, particularly *LOC105178907* (CAMTA5)*,* as important regulators of waterlogging/drainage responses in sesame. So far, functional characterizations of CAMTA family members provided evidences of their involvement in plant abiotic stress responses such as drought, salinity, cold [34,47,74]. Interestingly, Kidokoro et al. [34] has recently demonstrated that *AtCAMTA5* could target DREB1B/C genes and modulate cold responses in *Arabidopsis thaliana*. We also found several DREB genes predicted as targets for *LOC105178907*, particularly at the onset of waterlogging stress. There is no clear evidence of the implication of *CAMTA5* in waterlogging stress responses, therefore, future detailed investigations on *LOC105178907* and its homologs may provide a novel insight into waterlogging/drainage transcriptional regulation in plants. Besides, the timing of the differential expression of the waterlogging regulators in sesame is still unclear but we suspect that it is related to the specific functions of each TF according to the stress intensity. The result of the yeast-one-hybrid assays evidencing the interactions between TFs and targets further confirm that the predicted regulatory network will serve as a solid foundation for understanding the complex regulation of waterlogging responses in plants.

### 4.5. Future works

As a complement to our transcriptional regulatory network analysis, we generated dense time series leaf metabolome data. We expect that the integration of these two omics datasets will further inform on the biological processes underpinning each transcriptional phase in waterlogging/drainage responses in sesame. These datasets will also be of great importance for our ongoing genome-wide association study, which aims at unraveling the potential allelic variations and genes associated with variations in waterlogging tolerance using a large panel of sesame germplasm. Our predicted regulatory network highlighted several switches with unknown functions in plant waterlogging responses. Therefore, future experimentations aiming at elucidating the mechanisms involving these genes would significantly advance our understanding on plant responses to waterlogging. Genetic manipulation of the regulators detected in this study could be a proper approach [61]. Sesame resilience to genetic manipulation is still significantly enough to justify the use of a heterologous system such as *Arabidopsis thaliana* [77]. Arabidopsis genetic transformation could be a proxy to study the overexpression of the up-regulated regulators from our predicted network, but it will not be possible for the down-regulated regulators. Klok et al. [35] combined Arabidopsis hairy root culture and low-O$_2$ stress to identify genes involved in hypoxic responses. We could follow similar procedures in sesame since the hairy root culture approach has been demonstrated to be very amenable in several recalcitrant plants [46]. Combining sesame hairy root culture and target gene editing

technologies such as the CRISPR-Cas9 technology in order to modulate the expression levels of the regulators [36] will help screen and pinpoint the most important master players in the predicted regulatory networks by assessing the morphological, molecular and metabolic responses to either hypoxic stress or waterlogging/recovery treatments.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.11.022.

### Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. The raw RNA-seq data were submitted at NCBI SRA under the accession number: PRJNA516309 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA516309/).

### Author contributions

X.Z. and L.W. designed the project. L.W., K.D., J.Y., Y.Z., D.L., R.Z., J.Y., X.W., X.Z., S.J., Y.G., M.A.M., X.Z. executed the experiment. L.W., K.D., J.Y. analyzed the results. K.D., J.Y., L.W. wrote the manuscript. All authors contributed to reading and editing the manuscript.

### Declaration of Competing Interest

The authors declare no conflict of interest.

### Acknowledgments

### References

[1] M.H. Ali, Effect of different durations of water-logging at different growth stages on seed yield of sesame, Int. J. Appl. Sci. 1 (2018) 68–76.

[2] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[3] K.R. Anilakumar, A. Pal, F. Khanum, A.S. Bawas, Nutritional, medicinal and industrial uses of sesame (Sesamum indicum L.) seeds, Agric. Conspec. Sci. 75 (2010) 159–168.

[4] V. Arbona, S.I. Zandalinas, M. Manzi, M. González-Guzmán, P.L. Rodriguez, A. Gómez-Cadenas, Depletion of abscisic acid levels in roots of flooded Carrizo citrange (*Poncirus trifoliata* L. Raf. × *Citrus sinensis* L. Osb.) plants is a stress-specific response associated to the differential expression of PYR/PYL/RCAR receptors, Plant Mol. Biol. 93 (2017) 623.

[5] J. Bailey-Serres, T.D. Colmer, Plant tolerance of flooding stress – recent advances, Plant Cell Environ. 37 (2014) 2211–2215.

[6] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.

[7] E. Breeze, E. Harrison, S. McHattie, L. Hughes, R. Hickman, C. Hill, S. Kiddle, Y. Kim, C.A. Penfold, D. Jenkins, C. Zhang, K. Morris, C. Jenner, S. Jackson, B. Thomas, A. Tabrett, Legaie Roxane, J.D. Moore, D.L. Wild, S. Ott, D. Rand, J. Beynon, K. Denby, A. Mead, V. Buchanan-Wollaston, High-resolution temporal profiling of transcripts during Arabidopsis leaf senescence reveals a distinct chronology of processes and regulation, Plant Cell 23 (2011) 873–894.

[8] M.T. Campbell, C.A. Proctor, Y. Dou, A.J. Schmitz, P. Phansak, G.R. Kruger, et al., Genetic and molecular characterization of submergence response identifies Subtol6 as a major submergence tolerance locus in maize, PLoS One 10 (3) (2015), e0120385, https://doi.org/10.1371/journal.pone.0120385.

[9] T.D. Colmer, L.A.C.J. Voesenek, Flooding tolerance: suites of plant traits in variable environments, Funct. Plant Biol. 36 (2009) 665–681.

[10] B.A. Conesa, S. Götz, J. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.

[11] T. Dawood, X. Yang, E.J. Visser, Beek TAte, P.R. Kensche, S.M. Cristescu, S. Lee, K. Flokova, D. Nguyen, C. Mariani, et al., A co-opted hormonal cascade activates dormant adventitious root primordia upon flooding in Solanum dulcamara, Plant Physiol. 170 (2016) 2351–2364.

[12] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR : ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.

[13] K. Dossa, D. Diouf, L. Wang, X. Wei, Y. Zhang, M. Niang, D. Fonceka, J. Yu, M. A. Mmadi, L.W. Yehouessi, B. Liao, X. Zhang, N. Cisse, The emerging oilseed crop *Sesamum indicum* enters the "Omics" era, Front. Plant Sci. 8 (2017) 1154.

[14] K. Dossa, D. Li, L. Wang, X. Zheng, J. Yu, X. Wei, R. Zhou, D. Fonceka, D. Diouf, B. Liao, N. Cissé, X. Zhang, Transcriptomic biochemical and physio-anatomical investigations shed more light on responses to drought stress in two contrasting sesame genotypes, Sci. Rep. 7 (2017) 8755.

[15] K. Dossa, X. Wei, D. Li, Y. Zhang, L. Wang, D. Fonceka, et al., Insight into the AP2/ ERF transcription factor superfamily in sesame (Sesamum indicum) and expression profiling of the DREB subfamily under drought stress, BMC Plant Biol. 16 (2016) 171.

[16] K. Dossa, J. You, L. Wang, Y. Zhang, D. Li, et al., A new arsenal for deciphering waterlogging/drainage responses: 195 dense time-series RNA-seq dataset from sesame, Sci Data 6 (2019) 204.

[17] D.E. Evans, Aerenchyma formation, New Phytol. 161 (2003) 35–49.

[18] X.Y. Feng, X.R. Zhang, T.H. Xiao, Identification and evaluation of sesame germplasm resources for tolerance to waterlogging in China, Chin Oil Crops 3 (1991) 12–15 (in Chinese).

[19] T. Fukao, K. Xu, P.C. Ronald, J. Bailey-Serres, A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice, Plant Cell 18 (2006) 2021–2034.

[20] P. Gasch, M. Fundinger, J.T. Müller, T. Lee, J. Bailey-Serres, Angelika Mustroph, Redundant ERF-VII transcription factors bind to an evolutionarily conserved Cis -motif to regulate hypoxia-responsive gene expression in Arabidopsis, Plant Cell 28 (2016), 160–18.

[21] D.J. Gibbs, S.C. Lee, N.M. Isa, S. Gramuglia, T. Fukao, G.W. Bassel, C.S. Correia, F. Corbineau, F.L. Theodoulou, J. Bailey-Serres, M.J. Holdsworth, Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants, Nature 479 (2011) 415–418.

[22] C.E. Grant, T.L. Bailey, W.S. Noble, FIMO: scanning for occurrences of a given motif, Bioinformatics 27 (2011) 1017–1018.

[23] A.A. Hassan, N.N. Karim, A.A. Sarkar, M.H. Ali, Effect of waterlogging on the growth and yield of summer grown sesame (Sesamum indicum), Indian J. Agric. Sci. 71 (4) (2001) 271–272.

[24] Y. Hattori, K. Nagai, S. Furukawa, X.J. Song, R. Kawano, H. Sakakibara, J. Wu, T. Matsumoto, A. Yoshimura, H. Kitano, et al., The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water, Nature 460 (2009) 1026–1030.

[25] R. Hickman, C. Marcel, Van Verk, Anja J.H. Van Dijken, Marciel Pereira Mendes, Irene A. Vroegop-Vos, Lotte Caarls, Merel Steenbergen, Ivo Van der Nagel, Gert Jan Wesselink, Aleksey Jironkin, Adam Talbot, Johanna Rhodes, Michel De Vries, Robert C. Schuurink, Katherine Denby, Corné M.J. Pieterse, Saskia C.M. Van Wees, Architecture and dynamics of the jasmonic acid gene regulatory network, Plant Cell 29 (2017) 2086–2105.

[26] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, Watanabe Satoshi, H. Kim, S. Kanae, Global flood risk under climate change, Nat. Clim. Chang. 3 (2013) 816–821.

[27] F.-C. Hsu, M.Y. Chou, S.J. Chou, Y.R. Li, H.P. Peng, M.C. Shih, Submergence confers immunity mediated by the WRKY22 transcription factor in Arabidopsis, Plant Cell 25 (2013) 2699–2713.

[28] F.C. Hsu, M.Y. Chou, H.P. Peng, S.J. Chou, M.C. Shih, Insights into hypoxic systemic responses based on analyses of transcriptional regulation in *Arabidopsis*, PLoS One 6 (2011), e28888.

[29] B. Hu, J. Jin, A.-Y. Guo, H. Zhang, J. Luo, G. Gao, GSDS 2.0: an upgraded gene feature visualization server, Bioinformatics 31 (8) (2015) 1296–1297.

[30] M.B. Jackson, T.D. Colmer, Response and adaptation by plants to flooding stress, Ann. Bot. 96 (2005) 501–505.

[31] M.B. Jackson, The impact of flooding stress on plants and crops, 2013. Available at, http://www.plantstress.com/Articles/waterlogging_i/waterlog_i.htm (accesses January 2019).

[32] J.P. Jin, F. Tian, D.C. Yang, Y.Q. Meng, L. Kong, J.C. Luo, G. Gao, PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants, Nucleic Acids Res. 45 (D1) (2017) D1040–D1045.

[33] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J.A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S.R. Kulkarni, G. Tan, D. Baranasic, D.J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W.W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework, Nucleic Acids Res. 46 (2018) D260–D266.

[34] S. Kidokoro, K. Yoneda, H. Takasaki, F. Takahashi, K. Shinozaki, K. Yamaguchi-Shinozaki, Different cold-signaling pathways function in the responses to rapid and gradual decreases in temperature, Plant Cell 29 (4) (2017) 760–774.

[35] E.J. Klok, I.W. Wilson, Dale Wilson, Scott C. Chapman, Rob M. Ewing, Shauna C. Somerville, W. James Peacock, Rudy Dolferus, S. Elizabeth, Dennis expression profile analysis of the low-oxygen response in Arabidopsis root cultures, Plant Cell 14 (2002) 2481–2494.

[36] S. Konermann, M.D. Brigham, A.E. Trevino, J. Joung, O.O. Abudayyeh, C. Barcena, P.D. Hsu, N. Habib, J.S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Nature. 517 (2015) 583–588.

[37] J. Kreuzwieser, J. Hauberg, K.A. Howell, A. Carroll, H. Rennenberg, A.H. Millar, J. Whelan, Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia, Plant Physiol. 149 (2009) 461–473.

[38] I. Letunic, P. Bork, 20 years of the SMART protein domain annotation resource, Nucleic Acids Res. 46 (2017) D493–D496.

[39] D. Li, P. Liu, J. Yu, L. Wang, K. Dossa, Y. Zhang, R. Zhou, X. Wei, X. Zhang, Genome-wide analysis of WRKY gene family in the sesame genome and identification of the WRKY genes involved in responses to abiotic stresses, BMC Plant Biol. 17 (2017) 152.

[40] Y. Liao, G.K. Smyth, W. Shi, featurecounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014) 923–930.

[41] F. Licausi, M. Kosmacz, D.A. Weits, B. Giuntoli, F.M. Giorgi, L.A.C.J. Voesenek, P. Perata, J.T. van Dongen, Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization, Nature 479 (2011) 419–422.

[42] F. Licausi, J.T. van Dongen, B. Giuntoli, G. Novi, A. Santaniello, P. Geigenberger, P. Perata, HRE1 and HRE2, two hypoxia-inducible ethylene response factors, affect anaerobic responses in Arabidopsis thaliana, Plant J. 62 (2010) 302–315.

[43] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method, Methods 25 (2001) 402–408.

[44] A. Mustroph, M.E. Zanetti, C.J.H. Jang, H.E. Holtan, P.P. Repetti, D.W. Galbraith, T. Girke, J. Bailey-Serres, Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 18843–18848.

[45] R. Narsai, J.M. Edwards, T.H. Roberts, J. Whelan, G.H. Joss, B.J. Atwell, Mechanisms of growth and patterns of gene expression in oxygen deprived rice coleoptiles, Plant J. 82 (2015) 25–40.

[46] N.N. Ono, L. Tian, The multiplicity of hairy root cultures: prolific possibilities, Plant Sci. 180 (2011) 439–446.

[47] N. Pandey, A. Ranjan, P. Pant, R.K. Tripathi, F. Ateek, H.P. Pandey, U.V. Patre, S. V. Sawant, CAMTA 1 regulates drought responses in *Arabidopsis thaliana*, BMC Genomics 14 (2013) 216.

[48] J. Raineri, K.F. Ribichich, R.L. Chan, The sunflower transcription factor HaWRKY76 confers drought and flood tolerance to *Arabidopsis thaliana* plants without yield penalty, Plant Cell Rep. 34 (12) (2015) 2065–2080.

[49] P.J. Rushton, I.E. Somssich, P. Ringler, Q.J. Shen, WRKY transcription factors, Trends Plant Sci. 15 (2010) 247–258.

[50] P.K. Sarkar, A. Khatun, A. Singha, Effect of duration of water-logging on crop stand and yield of sesame, Int. J. Innov. Appl. Stud. 14 (1) (2016) 1–6.

[51] R. Sasidharan, A. Mustroph, Plant oxygen sensing is mediated by the N-end rule pathway: a milestone in plant anaerobiosis, Plant Cell 23 (2011) 4173–4183.

[52] R. Sasidharan, A. Mustroph, A. Boonman, M. Akman, A.M.H. Ammerlaan, T. Breit, M.E. Schranz, L.A.C.J. Voesenek, P.H. van Tienderen, Root transcript profiling of two Rorippa species reveals gene clusters associated with extreme submergence tolerance, Plant Physiol. 163 (2013) 1277–1292.

[53] R. Sasidharan, L.A.C.J. Voesenek, Ethylene-mediated acclimations to flooding stress, Plant Physiol. 169 (2015) 3–12.

[54] M. Sauter, Root responses to flooding, Curr. Opin. Plant Biol. 16 (2013) 282–286.

[55] D. Spies, C. Ciaudo, Dynamics in transcriptomics: advancements in RNA-seq time course and downstream analysis, Comput. Struct. Biotechnol. J. 13 (2015) 469–477.

[56] G. Su, J.H. Morris, B. Demchak, G.D. Bader, Biological network exploration with Cytoscape 3, Curr. Protoc. Bioinformatics 47 (2014) 1–24.

[57] J. Sun, X. Zhang, Y. Zhang, L. Wang, B. Huang, Effects of waterlogging on leaf protective enzyme activities and seed yield of sesame at different growth stages, Chin J Appl Environ Biol. 15 (6) (2009) 790–795.

[58] J. Sun, X.R. Zhang, Y.X. Zhang, B. Huang, Z. Che, Comprehensive evaluation of waterlogging tolerance of different sesame varieties, Chin. J. Oil Crop Sci. 30 (2008) 518–521.

[59] J. Sun, X.R. Zhang, Y.X. Zhang, L.H. Wang, D.H. Li, Evaluation of yield characteristics and waterlogging tolerance of sesame germplasm with different plant types after waterlogging, J. Plant Genet. Resour. 11 (2010) 139–146.

[60] B.G. Tamang, J.O. Magliozzi, M.A.S. Maroof, T. Fukao, Physiological and transcriptomic characterization of submergence and reoxygenation responses in soybean seedlings, Plant Cell Environ. 37 (2014) 2350–2365.

[61] M. Taylor-Teeples, L. Lin, M. de Lucas, G. Turco, T.W. Toal, A. Gaudinier, N. F. Young, G.M. Trabucco, M.T. Veling, R. Lamothe, P.P. Handakumbura, G. Xiong, C. Wang, J. Corwin, A. Tsoukalas, L. Zhang, D. Ware, M. Pauly, D.J. Kliebenstein, K. Dehesh, I. Tagkopoulos, G. Breton, J.L. Pruneda-Paz, S.E. Ahnert, S.A. Kay, S. P. Hazen, S.M. Brady, An Arabidopsis gene regulatory network for secondary cell wall synthesis, Nature 517 (2015) 571–575.

[62] T. Tian, Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, W. Xu, Z. Su, agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update, Nucleic Acids Res. 45 (Web Server issue) (2017) W122–W129.

[63] L.J.P. Van der Maaten, G.E. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[64] H. van Veen, A. Mustroph, G.A. Barding, M. Vergeer-van Eijk, R.A.M. Welschen-Evertman, O. Pedersen, E.J.W. Visser, C.K. Larive, R. Pierik, J. Bailey-Serres, et al., Two rumex species from contrasting hydrological niches regulate flooding tolerance through distinct mechanisms, Plant Cell 25 (2013) 4691–4707.

[65] H. van Veen, M. Akman, D.C. Jamar, D. Vreugdenhil, M. Kooiker, P. van Tienderen, L.A. Voesenek, M.E. Schranz, R. Sasidharan, *Group* VII Ethylene Response Factor diversification and regulation in four species from flood-prone environments, Plant Cell Environ. 37 (2014) 2421–2432.

[66] V.E. Vana, N. Marini, C. Busanello, C. Pegoraro, J.A. Fernando, L.C. Da Maia, A. Costa de Oliveira, Regulation of rice responses to submergence by WRKY transcription factors, Biol. Plant. 62 (2018) 551–560.

[67] L. Wang, D. Li, Y. Zhang, Y. Gao, J. Yu, X. Wei, X. Zhang, Global gene expression responses to waterlogging in roots of sesame (*Sesamum indicum* L.), Acta Physiol. Plant. 34 (2012) 2241.

[68] L. Wang, D. Li, Y. Zhang, Y. Gao, J. Yu, X. Wei, X. Zhang, Tolerant and susceptible sesame genotypes reveal waterlogging stress response patterns, PLoS One 11 (2016) e0149912.

[69] L. Wang, S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, et al., Genome sequencing of the high oil crop sesame, Genome Biol. 15 (2014) R39.

[70] L. Wang, J. Yu, D. Li, X. Zhang, Sinbase: an integrated database to study genomics, genetics and comparative genomics in *Sesamum indicum*, Plant Cell Physiol. 56 (2014), e2.

[71] W. Wei, L. Wang, X. Ding, Y. Zhang, Y. Gao, & Zhang, X Morpho-anatomical and physiological responses to waterlogging of sesame (*Sesamum indicum* L.), Plant Sci. 208 (2013) 102–111.

[72] K. Xu, X. Xu, T. Fukao, P. Canlas, R. Maghirang-Rodriguez, S. Heuer, A.M. Ismail, J. Bailey-Serres, P.C. Ronald, D.J. Mackill, Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice, Nature 442 (2006) 705–708.

[73] E. Yeung, H. van Veen, D. Vashisht, A.L.S. Paiva, M. Hummel, Tom Rankenberg, Bianka Steffens, Anja Steffen-Heins, Margret Sauter, Michel de Vries, Robert C. Schuurinkg, Jérémie Bazin, Julia Bailey-Serres, Laurentius A.C.J. Voesenek, Rashmi Sasidharan, A stress recovery signaling network for enhanced flooding tolerance in *Arabidopsis thaliana*, PNAS 115 (2018) E6085–E6094.

[74] R. Yue, C. Lu, T. Sun, T. Peng, T. Han, X. Han, J. Qi, S. Yan, S. Tie, Identification and expression profiling analysis of calmodulin-binding transcription activator genes in maize (*Zea mays* L.) under abiotic and biotic stresses. Front, Plant Sci. 6 (2015) 576.

[75] A. Zabalza, J.T. van Dongen, A. Froehlich, et al., Regulation of respiration and fermentation to control the plant internal oxygen concentration, Plant Physiol. 149 (2009) 1087–1098.

[76] Y. Zhang, D. Li, R. Zhou, et al., Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress, BMC Plant Biol. 19 (2019) 66, https://doi.org/10.1186/s12870-019-1665-6.

[77] R. Zhou, P. Liu, D. Li, X. Zhang, X. Wei, Photoperiod response-related gene SiCOL1 contributes to flowering in sesame, BMC Plant Biol. 18 (2018) 343.