

# SCIENTIFIC REPORTS



OPEN

## Elucidation of the molecular responses to waterlogging in *Sesbania cannabina* roots by transcriptome profiling

Cheng-Gang Ren<sup>1</sup>, Cun-Cui Kong<sup>1</sup>, Kun Yan<sup>2</sup>, Hua Zhang<sup>2</sup>, Yong-Ming Luo<sup>2</sup> & Zhi-Hong Xie<sup>1</sup>

*Sesbania cannabina*, a multipurpose leguminous crop, is highly resistant to waterlogging stress. However, the scant genomic resources in the genus *Sesbania* have greatly hindered further exploration of the mechanisms underlying its waterlogging tolerance. Here, the genetic basis of flooding tolerance in *S. cannabina* was examined by transcriptome-wide gene expression changes using RNA-Seq in seedlings exposed to short-term (3 h) and long-term (27 h) waterlogging. After de-novo assembly, 213990 unigenes were identified, of which 145162 (79.6%) were annotated. Gene Ontology and pathway enrichment analyses revealed that the glycolysis and fermentation pathways were stimulated to produce ATP under hypoxic stress conditions. Energy-consuming biosynthetic processes were dramatically repressed by short and long term waterlogging, while amino acid metabolism was greatly induced to maintain ATP levels. The expression pattern of 10 unigenes involved in phenylpropanoid biosynthesis, glycolysis, and amino acid metabolism revealed by qRT-PCR confirmed the RNA-Seq data. The present study is a large-scale assessment of genomic resources of *Sesbania* and provides guidelines for probing the molecular mechanisms underlying *S. cannabina* waterlogging tolerance.

Flooding and waterlogging are common phenomena with the potential to become an increasingly serious threat to the environment, which is progressively deteriorated by humans and their anthropogenic activities, resulting in global climate change. Adverse abiotic stresses can heavily damage crop production worldwide during the heavy rain season in lowland fields. Because of the limited diffusion of gas under water, waterlogging creates a low oxygen (hypoxia) environment in the root areas of crops<sup>1</sup>.

Higher plants are aerobic organisms. Since the diffusion of molecular oxygen is reduced in water, soil waterlogging is a serious obstacle to plant growth and development and may make plants hypoxic. The response of plants to external hypoxia has been intensively studied for decades. Alternative to oxidative respiration, plants employ glycolysis to generate ATP and ethanolic fermentation to produce the NAD<sup>+</sup> required for sustaining the EMP pathway<sup>2</sup>. Not only metabolic adaptations which confer tolerance to anoxia, prolong tissue survival, there are also morphological adaptations to waterlogging by plants, such as formation of aerenchyma<sup>3</sup>. Understanding the genetic basis of tolerance variation and the underlying genes and processes will be the key to discovering novel tolerance mechanisms and ultimately translating these to crops<sup>4</sup>.

Proteomics research has identified many anaerobically induced polypeptides (ANPs)<sup>5</sup>. ANPs are essential for the tolerance to hypoxic conditions in many plants<sup>6,7</sup>. Further studies showed that the majority of ANPs are involved in the glycolysis and fermentation pathways<sup>8</sup>. A set of anaerobic peptides including aldolase, enolase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, sucrose synthase, and alcohol dehydrogenase have been identified as being selectively induced under hypoxia in maize<sup>9</sup>. Microarray studies were performed to analyze the response to hypoxia in *Arabidopsis thaliana*<sup>10</sup>, cotton<sup>6</sup>, maize<sup>11</sup>, and other plants<sup>12</sup>. The expression levels of a zinc finger-like protein, SKP1/ASK1-like protein, and 20S proteasome subunit  $\alpha$ -3 increased markedly after 2 h of minimal oxygen supply in maize<sup>13</sup>. A low oxygen-sensing N-end rule proteolytic pathway<sup>14</sup> and a gene, *Sicyp51*, believed to confer tolerance to hypoxia were identified recently<sup>15</sup>. SNORKEL1

<sup>1</sup>Key Laboratory of Biology and Utilization of Biological Resources of Coastal Zone, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, 264003, China. <sup>2</sup>Key Laboratory of Coastal Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, China. Correspondence and requests for materials should be addressed to Z.-H.X. (email: zhxie@yic.ac.cn)

Sample	Raw Reads	Clean Reads	Clean	Error	Q20	Q30	GC
			Bases	(%)	(%)	(%)	(%)
R1	46891272	43971630	6.6G	0.02	96.1	90.69	43.71
R2	48598584	45774804	6.87G	0.02	96.43	91.34	43.98
WLR3H1	51310282	49550926	7.43G	0.01	97.02	92.65	44.37
WLR3H2	64341606	61288682	9.19G	0.02	96.2	90.84	43.89
WLR27H1	55278170	51429496	7.71G	0.03	94.78	87.72	45.29
WLR27H2	54607108	50723740	7.61G	0.03	94.66	87.4	44.91
Total	321027022	302739278	45.41G				

**Table 1.** Summary of sequences analysis. R, Untreated root; WLR, Waterlogged root; H, Hour; Q20, The percentage of bases with a Phred value > 20; Q30, The percentage of bases with a Phred value > 30.

	Number of	Percentage
	Unigenes	(%)
Annotated in NR	155725	72.77
Annotated in NT	143553	67.08
Annotated in KO	61329	28.65
Annotated in SwissProt	3772	1.76
Annotated in PFAM	105129	49.12
Annotated in GO	106734	49.87
Annotated in KOG	58137	27.16
Annotated in all Databases	1054	0.49
Annotated in at least one Database	170345	79.6
Total Unigenes	213990	100

**Table 2.** BLAST analysis of non-redundant unigenes against public databases.

and 2 in rice and MYB, and AP2/ERF transcription factors (TFs) in *Arabidopsis*<sup>16–18</sup> were found to regulate gene expression under low oxygen. In *Glycine max*, waterlogging stress specific genes were found to be enriched for lipoxygenase 2, ethylene forming enzyme, matrixin family proteins, and 12-oxophytodienoate reductase 2. The Ethylene response factor (ERF) represented the highest number of significantly expressed TFs under hypoxic conditions followed by bHLH, MYB, NAC, and WRKY<sup>19</sup>. Rapid changes in a large number of transcripts involving not only well-known ANPs, but also those not previously known to be involved in waterlogging tolerance, indicate that plants activate complex mechanisms in response to waterlogging<sup>6,20</sup>.

The RNA-Seq approach has a higher sensitivity than microarray analysis and includes both low- and high-level gene expression<sup>21</sup>. These advantages resulted in the increased application of RNA-Seq to elucidate the response of plants to various environmental stresses, such as cold<sup>22</sup>, salt<sup>23,24</sup>, and drought<sup>23,25</sup>. RNA-Seq has also been successfully used to analyze the responses to waterlogging stress in maize<sup>26</sup>, sesame<sup>27</sup>, and rape<sup>28</sup>. However, transcriptional studies of manure crops are rare, despite their potential in environment-friendly agricultural practices.

The genus *Sesbania* has been used increasingly as a green manure crop in rice cropping systems because of its ability to fix atmospheric nitrogen through symbiosis with rhizobia and its adaptability to flooded soils<sup>29–31</sup>. Of approximately 20 species distributed in these regions, *S. cannabina* is one of the most common species, and its ability to promote the drainage of ill-drained fields and paddy-upland rotated fields has been demonstrated<sup>32</sup>. To better understand the molecular mechanisms underlying the response of *S. cannabina* to soil waterlogging, the transcriptional profiles of waterlogged *S. cannabina* roots were analyzed by RNA-Seq. Genes expressed in the control and treatment groups were compared to elucidate the species-specific responses of *S. cannabina* and to identify genes or strategies associated with waterlogging resistance. Our results will improve our understanding of the response of flood-tolerant manure crops to soil waterlogging stress.

## Results

**Transcriptome sequencing and assembly.** To obtain a comprehensive view of waterlogging tolerance in *S. cannabina*, 30-day old seedlings of *S. cannabina* were waterlogged. High-throughput RNA sequencing (RNA-Seq) was performed on waterlogged roots (3 h, and 27 h), non- waterlogged control roots (CK) using the Illumina platform. Illumina sequencing data were deposited in the NCBI GEO database under accession number GSE92670. In total, 321,027,022 Illumina PE raw reads were generated (Table 1). After removing adaptor sequences, ambiguous nucleotides, and low-quality sequences, 302,739,278 clean reads were recorded. Assembly of clean reads resulted in 213,990 unigenes in the range of 201–16,929 bp with a N50 length of 1055 bp (Table S1).

**Sequence annotation.** Next, the unigenes were annotated by alignment with seven public databases (Table 2). Analyses showed that 155,725 unigenes (72.77%) had significant matches in the Nr database, 143,553

Species	Number of annotations	Plant_nr
<i>Glycine max</i>	31715	63716
<i>Glycine soja</i>	30692	50066
<i>Cicer arietinum</i>	28198	29056
<i>Medicago truncatula</i>	21235	90439
<i>Phaseolus vulgaris</i>	16644	32585
<i>Lotus japonicus</i>	6093	8640
<i>Vitis vinifera</i>	1773	74877
<i>Gossypium raimondii</i>	1098	86601
<i>Citrus sinensis</i>	1054	66922
<i>Pyrus x</i>	784	38836

**Table 3.** The top-10 Blast hits for the assembled unigenes based on NCBI non-redundant (nr) protein database search.

(67.08%) in the Nt database, and 105,129 (49.12%) in the Pfam database. In total, 170,345 unigenes (79.6%) were successfully annotated in at least one of the Nr, Nt, Swiss-Prot, KEGG, GO, KOG, and Pfam databases, with 1054 unigenes (0.49%) in all seven databases. Most of the unigenes could be mapped to legumes sequences, in which 31,715, 30,692, 28,198 and 21,235 showed high sequence similarity with that of *Glycine max*, *Glycine soja*, *Cicer arietinum* and *Medicago truncatula*, respectively (Table 3).

For GO analysis, there were 106,734 unigenes divided into three ontologies (Fig. 1A). For the biological process (BP) category, genes involved in ‘cellular process’ (61,631), ‘metabolic process’ (58,709), and ‘single-organism process’ (46,928) were highly represented. The cellular component (CC) category mainly comprised proteins involved in ‘cell’ (32,127), ‘cell part’ (32,126), and ‘organelle’ (21,218). Within the molecular function (MF) category, ‘binding’ (61,098), ‘catalytic activity’ (50,782), and ‘transporter activity’ (7502) were highly represented.

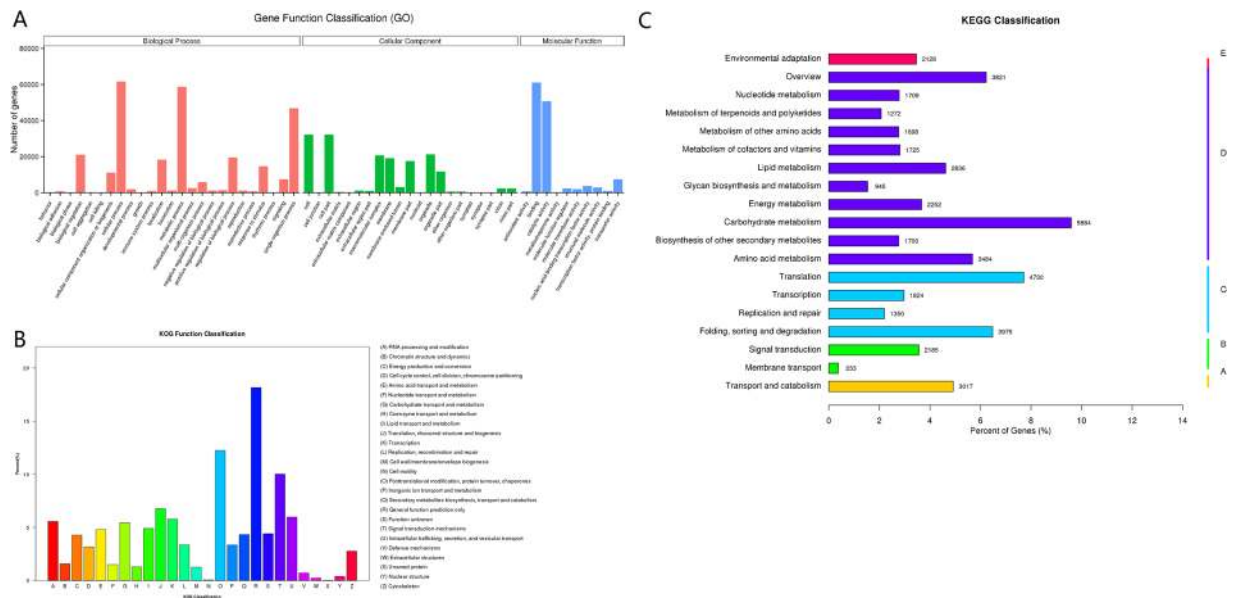
In addition, all unigenes were subjected to a search against the KOG database for functional prediction and classification. In total, 31,137 unigenes were assigned to the KOG classification and divided into 26 specific categories (Fig. 1B). The ‘general functional prediction only’ (10554) was the largest group, followed by ‘post-translational modification, protein turnover, chaperone’ (7119), ‘signal transduction mechanisms’ (5822), ‘translation, ribosomal structure and biogenesis’ (3935), and ‘intracellular trafficking, secretion, and vesicular transport’ (3471). Only a few unigenes were assigned to ‘cell motility’ (42) and ‘unnamed protein’ (15).

Unigene metabolic pathway analysis was also performed using the KEGG annotation system. This process predicted a total of 130 pathways representing 61,329 unigenes (Fig. 1C). The pathways involving the highest number of unique transcripts were ‘carbohydrate metabolism’ (5884), followed by ‘translation’ (4730) and ‘folding, sorting and degradation’ (3975).

**Differential expression analysis of assembled transcripts in response to waterlogging treatments in the early and late stages.** Differential expression analysis was firstly performed based on pairwise comparison between different waterlogging treatment times (CK, 3 h and 27 h). DEGs were defined as genes that were significantly enriched or depleted by the treatments. A total of 2719 transcripts, which accounted for approximately 1.27% of the total transcripts were differentially expressed in *S. cannabina* in response to the early stages (3 h) of waterlogging (Table S2). Meanwhile, a total of 9738 transcripts, which accounted for approximately 4.5% of the total transcripts were differentially expressed in *S. cannabina* in response to the latestages (27 h) of waterlogging (Table S2). Commonly up- and down-regulated transcripts were identified between the early and late stages of waterlogging to determine the degree of overlap (Fig. 2A,B). A total of 970 transcripts were upregulated and 1,749 transcripts were downregulated between the CK and 3 h waterlogged groups. The maximum number of commonly up- (4,171) and down- (5,567) regulated transcripts was observed between the CK and 27 h waterlogged groups.

To explore patterns of co-regulation of the DEGs in *S. cannabina* in response to the early and late stages of waterlogging, the expression profiles of these genes were clustered using the hierarchical clustering algorithm and visualized in a heat map (Fig. S1). The results showed distinct patterns of expression of DEGs in response to different stages of waterlogging stress.

**Functional classification of DEGs.** To gain a better understanding of the adaptive mechanism shift in *S. cannabina* in response to waterlogging, biological function analysis of DEGs was performed by GO enrichment using the whole transcriptome as the background (Fig. 3). GO analysis was performed for upregulated DEGs in CK compared with those in the 3 h waterlogging group (Table S3). In the MF category, the top two enriched terms were oxidoreductase activity and transporter activity, followed by enzyme inhibitor activity. In the CC category, ‘integral component of membrane’ and ‘intrinsic component of membrane’ were the two dominant enriched terms. In BP, ‘oxidoreduction coenzyme metabolic process’ was the most enriched, and various terms of metabolic processes were also enriched, including ‘pyridine-containing compound metabolic process’, ‘pyruvate metabolic process’ and ‘glucose metabolic process’ (corrected p-value < 0.005). For the downregulated DEGs (Table S4), the most enriched terms were oxidoreductase activity in the MF category. Consistent with this, ‘oxidation-reduction process’ was the top BP enriched term in downregulated DEGs, which taken together with ‘oxidoreductase activity’ as the most enriched in upregulated DEGs implied that different sets of oxidoreductase



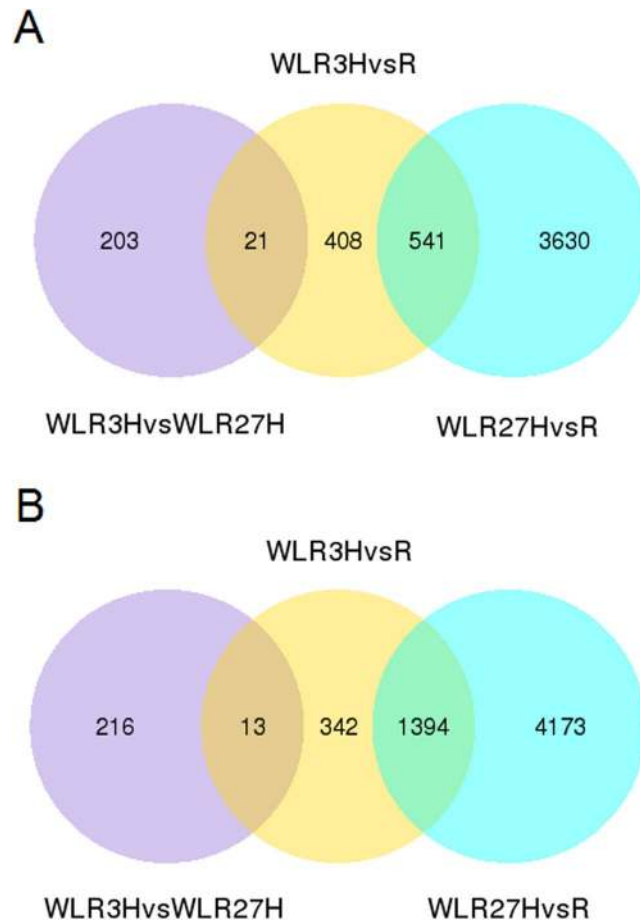
**Figure 1.** Functional annotation and classification of *S. cannabina* transcriptome. (A) GO categorization of non-redundant unigenes. Each annotated sequence was assigned at least one GO term. (B) COG annotation of putative proteins. (C) KEGG annotation of putative proteins. The capital letters against the colored bars indicate five main categories, (A) cellular processes; (B) environmental information processing; (C) genetic information processing; (D) metabolism; and (E) organism systems.

genes participate in either maintaining redox equilibrium in a normal environment or in alleviating the oxidative stress caused by waterlogging. Other enriched terms included carbohydrate metabolic process and hydrolase activity, acting on glycosyl bonds.

Comparison of DEGs in 27 h waterlogging with those in CK identified ‘metabolic process’ and ‘catalytic activity’ as the top two GO enrichment terms in the downregulated DEGs (Table S5). These two terms included >2000 DEGs, which were all repressed. Additionally, 100% of DEGs involved in ‘hydrolase activity’, ‘cellulose synthase (UDP-forming) activity’ and ‘cellulose biosynthetic process’ were also downregulated. Changes in transcript levels suggested that the energy-demanding processes of cellulose and cell wall biosynthesis were greatly inhibited in the *S. cannabina* response to the late stage of waterlogging. For the GO enrichment analysis of the upregulated DEGs, ‘transcription factor complex’, ‘glucose metabolic process’ and ‘enzyme regulator activity’ were the dominant terms (Table S6).

KEGG pathway enrichment analysis for DEGs also revealed common patterns and significant differences in response/adaptation between the early and late stages of waterlogging in *S. cannabina*. The top five enriched pathways by DEGs in the 3 h waterlogging treatment (Table S7) ( $FDR \leq 0.05$ ), were phenylalanine metabolism, cysteine and methionine metabolism, phenylpropanoid biosynthesis, tyrosine metabolism and sulfur metabolism. DEGs in the 27 h waterlogging treatment was also analyzed (Table S8) ( $FDR \leq 0.05$ ). The top two enriched pathways by DEGs in the 27 h waterlogging treatment were biosynthesis of secondary metabolites and phenylpropanoid biosynthesis. Phenylpropanoid biosynthesis was the third and second enriched pathway by DEGs in 3 h and 27 h waterlogged plantlets, respectively. There were 58 and 186 unigenes annotated as involved in this pathway in 3 h and 27 h waterlogged plantlets, respectively, with 45 shared unigenes showing changed expression both in the early and late stage of waterlogging stress (Fig. S2). Ten upregulated DEGs were annotated as encoding enzymes involved in phenylpropanoid biosynthesis. Among them, five DEGs (Cluster-8728.18814, Cluster-8728.82713, Cluster-8728.82715, Cluster-8728.84811, Cluster-8728.109527) were involved in the early stage of waterlogging and six DEGs (Cluster-8728.90349, Cluster-8728.158848, Cluster-8728.18814, Cluster-8728.94869, Cluster-8728.117414, Cluster-8728.94871) were involved in the late stage of waterlogging, with one shared DEG annotated as 4-coumarate-CoA ligase-Cluster-8728.18814 showing 4.75- and 6.54-fold increased expression, respectively (Table S9). Phenylpropanoid biosynthesis is regulated by biotic and abiotic stimuli, and phenylpropanoid-based polymers such as lignin, suberin, and tannin contribute substantially to the stability and robustness of plants in the face of mechanical or environmental damage<sup>33–35</sup>. On the other hand, their downregulation suggested the harmful effects of waterlogging. There were 58 downregulated DEGs in the early stage of waterlogging and 176 downregulated DEGs in the late stage of waterlogging (Fig. S2).

Most of the ANPs identified were found to be enzymes of glycolysis or sugar-phosphate metabolism<sup>8</sup>. There were 32 unigenes annotated as encoding enzymes involved in glycolysis/gluconeogenesis pathways in the early stage of waterlogging and no unigenes were significantly enriched in the late stage of waterlogging (Fig. 4, Tables S7 and S8). There were five DEGs annotated as encoding glucose-6-phosphate isomerase (GPI): Cluster-8728.113805, Cluster-8728.113801, Cluster-8728.113802, Cluster-8728.120033, and Cluster-8728.112865 had more than 2-fold increased expression. The activities of the corresponding enzymes 6-phosphofructokinase (6-PFK), fructose-bisphosphate aldolase class I (FBA), glyceraldehyde-3-phosphate

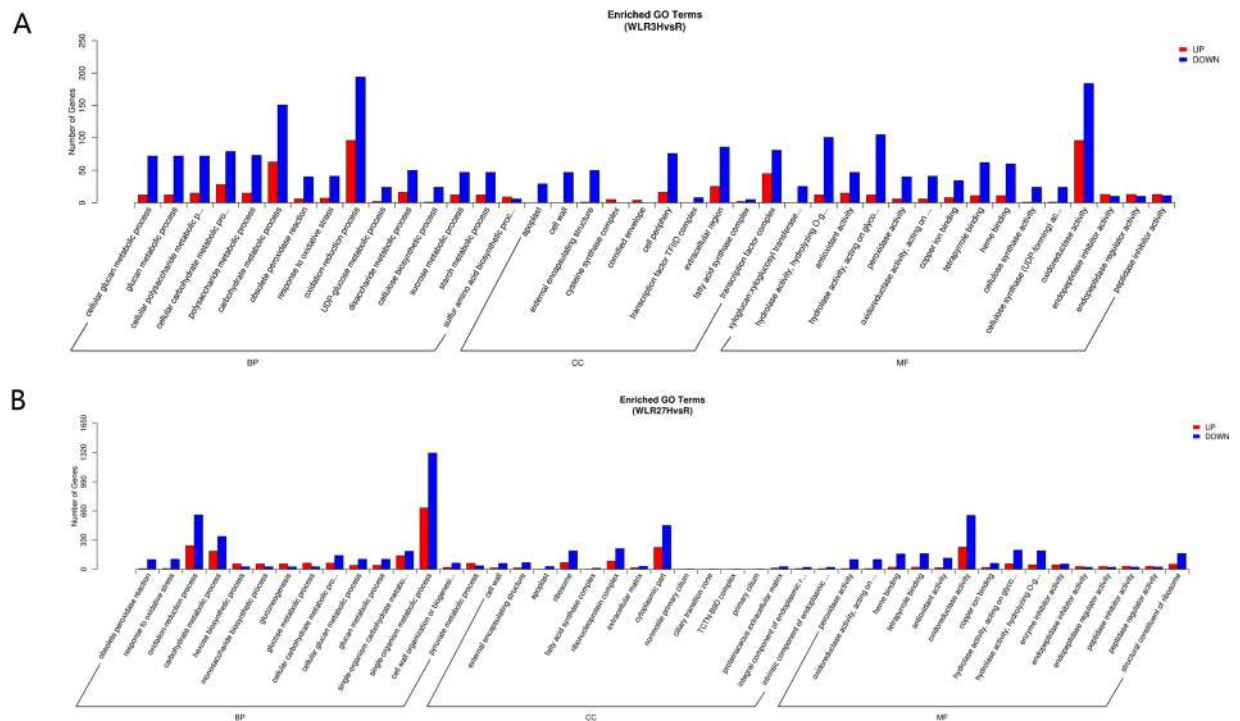


**Figure 2.** Venn diagrams of differentially expressed transcripts under waterlogging treatment in *S. cannabina* plantlets. **(A)** Numbers of DEGs exclusively upregulated in each treatment; **(B)** numbers of DEGs exclusively downregulated in each treatment. The numbers of DEGs with common or opposite expression change tendencies between treatments are shown in the overlapping regions. The total numbers of up- or down-regulated genes in each treatment are the sum of the numbers in each circle.

dehydrogenase (GAPDH), phosphoglycerate mutase (PGAM), and pyruvate kinase (PDC) were in turn upregulated. Supporting our results, lactic acid fermentation is activated during the initial stages of root hypoxia in many plants. In gray poplar, lactate dehydrogenase transcripts increase as an initial reaction to O<sub>2</sub> deprivation, whereas they decrease after approximately 5 h because of the decrease in cytosolic pH caused by lactic acid<sup>36</sup>. In plants, lactic acid fermentation is usually followed by alcoholic fermentation, which is the major fermentation process for plants. In our results, five upregulated DEGs were annotated as encoding alcohol dehydrogenase (ADH). The activation of alcoholic fermentation resulted in energy production in 3 h waterlogged *Sesbania cannabina* plantlets. However, none of these DEGs showed changed expression in 27 h waterlogged plantlets.

Since low oxygen induced TFs could mediate the expression of anaerobic responsive genes, we analyzed for TFs that were differentially regulated by both early and late stages of waterlogging. Total 176 waterlogging-regulated TFs were classified based on their assigned protein families in early stage, counted for 6.1% of the DEGs. Of these, 60 TFs were up-regulated and 116 TFs were downregulated (Fig. 5A; Table S10). Total 593 waterlogging-regulated TFs were classified in late stage, counted for 6.1% of the DEGs. Among them, 286 TFs were up-regulated and 307 TFs were down-regulated (Fig. 5B; Table S10). Genes belonging to the AP2-EREBP, WRKY, MYB and bHLH family represent most of the differentially expressed TFs. The bHLH family represented the highest number of significantly expressed TF in early stage of waterlogging, while ethylene response element binding protein (EREBP) represented the highest number of significantly expressed genes in late stage of waterlogging (Fig. 5).

**Verification of RNA-Seq data by real-time quantitative RT-PCR.** The differential expression of selected transcripts from RNA-Seq data was validated by real-time PCR analysis. The expression analysis was performed for selected genes belonging to the phenylpropanoid biosynthesis pathway, glycolysis pathway, and amino acid metabolism. In addition, WRKY20 and NAD(P)H dehydrogenase, which showed differential expression, were also selected randomly from the transcriptome dataset and validated by real-time PCR analysis. A comparative analysis of all the selected genes showed that the expression patterns in real-time PCR analysis were similar to those observed in the RNA-Seq data (Fig. S3). The Pearson correlation coefficient was calculated by SPSS to assess the correlation between different platforms (Fig. S4). We observed a good concordance in the expression patterns



**Figure 3.** GO functional classification of the control and waterlogging treated plantlets. (A) comparison between 3 h treatment and CK; (B) comparison between 27 h treatment and CK.

of DEGs obtained by the RNA-Seq and qRT-PCR methods, as indicated by the correlation coefficient (0.942 and 0.933), confirming that our experimental results were valid.

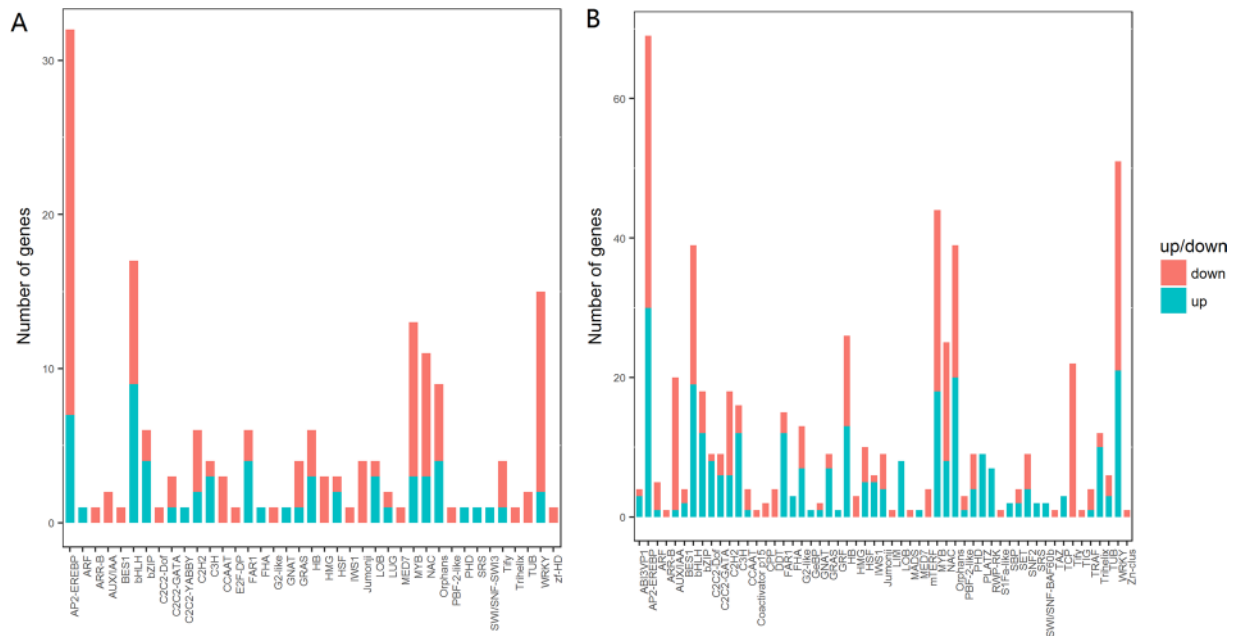
## Discussion

**Effects on energy-consuming biosynthetic processes.** Low oxygen conditions in plants shift metabolism from oxidative phosphorylation to anaerobic fermentation to maintain ATP production<sup>2</sup>. Our results supported the longstanding notion that waterlogging promotes anaerobic respiration, as observed by the upregulation of DEGs encoding enzymes involved in glycolysis and fermentation (Fig. 4). On the other hand, plants can activate additional responses to low oxygen conditions, including the downregulation of energy consuming processes. In the present study, genes involved in the biosynthesis of carotenoids, flavonoids, phenylpropanoids, and fatty acid elongation were downregulated in the roots of waterlogged *S. cannabina* (Table S9). These results, together with the finding that waterlogging resulted in a decrease in total root carbohydrate and activities of starch synthesis-related enzymes, ADP-glucose pyrophosphorylase (AGPase) and starch synthase (Fig. S5), led to the conclusion that *S. cannabina* roots responded to waterlogging by regulating energy consumption and production.

**Effects on carbon metabolism and amino acid metabolism.** When plants lack circulatory system to transport oxygen produced by photosynthesis to heterotrophic roots, oxygen-dependent mitochondrial respiration in the root is greatly limited under waterlogging conditions. As verified by qRT-PCR, many genes including well-known hypoxic genes associated with glycolysis and fermentation (ADH and PDC) were induced in the early stage of waterlogging, which indicated that the glycolysis and fermentation pathway was activated to maintain ATP production under stress conditions. As a result, the demand for carbohydrates increased and carbon metabolism significantly increased in waterlogged *S. cannabina* roots. Under hypoxic conditions, the acceleration of carbon metabolism is critical for plant survival<sup>37,38</sup>. The switch from mitochondrial respiration to fermentation is likely to strongly affect energy and carbon metabolism. The up-regulated expression of genes coding for the glycolytic pathway enzymes under waterlogging treatment can be considered an adaptive response of *S. cannabina* to flooding stress. Consistent with our results, glycolytic enzymes are key factors in the early response to flooding in soybean<sup>39,40</sup>.

In a comparison of transcriptome responses to low-oxygen in *Arabidopsis*, cotton, and poplar, amino acid metabolism changes were common among the three dicotyledonous species, although there was almost no overlap between specific processes<sup>41</sup>. Waterlogging also led to considerable changes in the levels of amino acids in *S. cannabina* roots (Table S9). In response to the 3 h waterlogging treatment, genes involved in cysteine and methionine metabolism, tyrosine metabolism, beta-alanine metabolism, arginine and proline metabolism, and glycine, serine, and threonine metabolism were upregulated. With the exception of tyrosine, genes involved in the metabolism of these amino acids were also upregulated in the 27 h waterlogging group. Kreuzwieser *et al.* proposed that hypoxia inhibits the TCA cycle and activates the glycolysis and fermentation pathways, resulting in the accumulation of amino acids closely derived from glycolysis intermediates<sup>34</sup>. As seen also in *Arabidopsis*<sup>42</sup>,





**Figure 5.** Class-wise counts of transcription factors identified in DEGs under waterlogging. **(A)** TFs (up- and down-regulated) were found in 3 h waterlogging. **(B)** TFs (up- and down-regulated) were found in 27 h waterlogging.

of the total transcripts were predicted to be related to transcription, indicating that waterlogging stress induced high transcriptional activity. Further illustrating the active responses of the transcription factors is under our experimental sets.

## Conclusion

In the present study, high-throughput RNA-sequencing technology was used to analyze differential gene expression profiles of *S. cannabina* roots subjected to short-(3 h) and long- (27 h) term waterlogging stress. The expression of 2719 and 9738 DEGs was reprogrammed by short and long exposure to waterlogging, respectively (Fig. S6). Transcript comparison in *S. cannabina* under short and long term waterlogging conditions using NGS technology shed light on the molecular basis of the responses of this waterlogging-tolerant legume. Under hypoxic stress conditions, the glycolysis and fermentation pathway was activated to produce ATP. A series of energy-consuming biosynthetic processes were dramatically repressed in response to short and long term waterlogging, whereas amino acid metabolism was induced to help maintain ATP. The assembled transcript sequences and comparison analysis with public transcriptome data revealed the key genes and metabolic pathways playing important roles in waterlogging tolerance. The present study provides a foundation for future studies on *S. cannabina* at the cellular and genomic levels.

## Methods

**Plant material and treatments.** Seeds of *S. cannabina* (Retz.) Pers. (Shandong Academy of Agricultural Sciences, China) from a single population were sterilized in 10% hydrogen peroxide for 10 min and rinsed several times with distilled water before use, then germinated at 28 °C for 48 h and sown (thinned to three uniform seedlings after germination) in a single 2-Liter pot containing autoclaved zonalite, arranged randomly in a greenhouse with day/night temperatures of 30/22 °C, 60% ± 2% relative humidity, and a photoperiod of 14/10 h light/dark. The photon flux density was approximately 800  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Hoagland solution was supplied regularly and the pots were weighed every week to adjust the water content. The same conditions were used for the treatments. The pots were placed in watertight containers and filled with room temperature water maintaining a water level above the soil surface as the treatment. Untreated plantlets were placed in the same container as the control. Thirty-day-old plantlets were used for waterlogging treatment. The 27 h sample was waterlogged first, after 24 hours, the 3 h sample was treated, when the treatment was all done, all samples from three treatments (0 h, 3 h, and 27 h) for RNA isolation were harvested at same time point, snap frozen in liquid nitrogen and stored at -80 °C.

**RNA isolation, cDNA library construction, and sequencing.** Root samples from 3 seedlings were pooled for a single biological replicate; total RNA from 2 biological replicates was isolated using TRIzol (Invitrogen, USA) following the manufacturer's instructions. RNA was checked on 1% agarose gels to avoid possible degradation and contamination, and was then examined by a Nano Photometer spectrophotometer (IMPLEN, CA, USA) for RNA purity. Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA) was used to measure RNA concentration, and RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used to evaluate RNA integrity. Only RNA samples that passed the quality tests were chosen for RNA-Seq analyses.



RNA-Seq library construction was performed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, USA) following manufacturer's instructions. Four index codes were added to assign sequences for each sample. In brief, mRNA was purified from 3 µg total RNA using poly-T oligo-attached magnetic beads (Life technologies, CA, USA), and then fragmented using divalent cations under elevated temperature in the NEB proprietary fragmentation buffer. Double-stranded cDNAs were synthesized using random hexamers and M-MuLV Reverse Transcriptase (RNase H<sup>-</sup>), followed by DNA Polymerase I and RNase H. After adenylation of the 3' ends of DNA fragments, Illumina PE adapter oligonucleotides were ligated to prepare for hybridization. To select cDNA fragments of the preferred 200 bp in length, the library fragments were purified using an AMPure XP system (Beckman Coulter, Beverly, CA, USA). cDNA fragments with ligated adaptor molecules on both ends were selectively enriched using Illumina PCR Primer Cocktail in a 12-cycle PCR reaction. The products were purified (AMPure XP system) and quantified using the Agilent high-sensitivity DNA assay on the Agilent Bioanalyzer 2100 system (Agilent, Santa Clara, CA, USA). The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. RNA-seq libraries were sequenced on an Illumina HiSeq. 4000 platform to generate 150 bp paired-ended (PE) reads. This procedure was carried out by Novogene Bioinformatics Technology Co. Ltd (Tianjing, China).

**De novo transcriptome assembly and annotation.** Raw reads were pre-processed to remove low quality regions and adapter sequences using NGS QC Toolkit (v2.3)<sup>53</sup>. At the same time, Q20, Q30, GC-content, and the sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality. Transcriptome assembly was achieved using Trinity<sup>54</sup> with `min_kmer_cov` set to 2 by default and all other parameters set to default. Gene function was annotated based on the following seven databases, Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences) and Swiss-Prot (a manually annotated and reviewed protein sequence database) using NCBI blast 2.2.28 + software with a cutoff E-value of 10<sup>-5</sup>; KOG/COG (Clusters of Orthologous Groups of proteins) using NCBI blast 2.2.28 + software with a cutoff E-value of 10<sup>-3</sup>; KO (KEGG Ortholog database) using KAAS, KEGG Automatic Annotation Server with a cutoff E-value of 10<sup>-10</sup>; Pfam (Protein family) using HMMER 3.0 package with a cutoff E-value of 0.01; and GO (Gene Ontology) annotation was performed by using Blast2GO software with a cutoff E-value of 10<sup>-6</sup>.

**Differential expression analysis.** Gene expression levels were estimated by RSEM<sup>55</sup> for each sample as follows: firstly, clean data were mapped back onto the assembled transcriptome by bowtie2 using default settings (mismatch 0). Then, the read count for each gene was obtained from the mapping results and normalized to fragments per kilo base of transcript sequence per millions base pairs sequenced is recommended for paired-end reads (FPKM). Differential expression analysis of two treatments was performed using the DESeq<sup>56</sup> R package (1.10.1). The resulting P values were adjusted using the Benjamini's approach<sup>57</sup> for controlling the false discovery rate. FDR adjusted P < 0.05 and absolute value of log<sub>2</sub> ratio > 1 were set as the threshold for significance of gene expression differences between treatments. GO enrichment analysis of DEGs was implemented by the Goseq R package based on Wallenius non-central hyper-geometric distribution<sup>58</sup>, which can adjust for gene length bias in DEGs. KOBAS<sup>59</sup> software was used to test the statistical enrichment of DEGs in KEGG pathways.

**Quantitative real-time PCR.** The expression patterns of 12 genes involved in phenylpropanoid biosynthesis, glycolysis pathway, and amino acid metabolism were analyzed using qRT-PCR. Three independent biological replicates of root tissue were used for RNA extraction for qRT-PCR assays. The reverse transcription reactions were carried out by GoScript<sup>™</sup> Reverse Transcription System (Promega, USA). Gene-specific primers were designed according to the reference unigenes sequences using Primer Premier 5.0 (Table S11). All reactions were performed on an iCycler iQ real-time PCR detection system (BIO-RAD) with three technical replicates. Each reaction was performed in a total volume of 20 µl including 0.2 µM primer pairs, 2 µl diluted cDNA, and 10 µl 2 × SYBR Green PCR Master Mix (TaKaRa Bio Inc., Dalian, China). The amplification reactions were incubated at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s, 58 °C for 15 s, and 72 °C for 20 s. *S. cannabina* actin and tubulin gene were used as normalizer (Table S11), and relative gene expression levels were calculated using the 2<sup>-ΔΔC<sub>t</sub></sup> method.

**Carbohydrate content analysis.** Frozen root tissue (100 mg) was used to quantify the total carbohydrate content using a method described by Juntawong *et al.*<sup>1</sup>, with three replicates for each sample. Non-structural carbohydrates were extracted and hydrolyzed by adding 3 ml of 3 M HCl and incubated in a boiling water bath for 3 h. The extract was neutralized by Na<sub>2</sub>CO<sub>3</sub>. The anthrone method was used to determine total carbohydrate content relative to a standard series of glucose. In brief, the extract (200 µl) and distilled water (300 µl) were mixed with 2 ml of 0.14% (w/v) anthrone solution in 95% H<sub>2</sub>SO<sub>4</sub>, incubated in a boiling water bath for 8 min, and rapidly cooled on ice. The absorbance was quantified at 630 nm.

**Assay of activities of the enzymes.** The crude extracts of AGPase or starch synthase were prepared according to Nakamura *et al.*<sup>60</sup> with modifications. In brief, 0.3 g of root tissues was homogenised on ice in a 1 ml pre-cooled solution composed of 100 mM Hepes-NaOH at pH 7.4, 5 mM MgCl<sub>2</sub>, 3 mM EDTA, 12.5% (v/v) glycerol, 5% (v/v) polyvinylpyrrolidone, and 50 mM β-mercaptoethanol. The homogenate was centrifuged at 4 °C at 12,000 × g for 5 min. The supernatant was collected for AGPase or starch synthase activities analysis. For AGPase activity, a 30 µl aliquot of the crude enzyme extract was mixed well with 110 µl reaction solution comprising 100 mM Hepes-NaOH at pH 7.4, 5 mM MgCl<sub>2</sub>, 3 mM pyrophosphoric acid, 4 mM dithiothreitol (DTT) and 1.2 mM adenosine diphosphate glucose (ADPG), incubate in a 30 °C water bath for 20 min, then terminated by treatment for 30 s in a boiling water. The reaction solution was cooled rapidly in an ice-water, and then

centrifuged for 10 min at  $10,000 \times g$  at  $4^\circ\text{C}$ . A  $100\ \mu\text{l}$  aliquot of the supernatant was added to  $5.2\ \mu\text{l}$  solution composed of  $5.76\ \text{mM}$  nicotinamide adenine dinucleotide phosphate (NADP),  $0.07\ \text{U}$  glucose-6-P-dehydrogenase and  $0.08\ \text{U}$   $\beta$ -glucuronidase, incubate in a  $30^\circ\text{C}$  water bath for 10 min, terminated by treatment for 30 s in a boiling water. A control was conducted in parallel in the reaction system composed of a  $20\ \mu\text{l}$  aliquot of the crude enzyme extract and the sterile deionised water to correct the background glucose release. The  $\text{OD}_{340\ \text{nm}}$  in the reaction solution were measured and normalized by  $\text{OD}_{340\ \text{nm}}$  resulting from the control. The enzyme activity was then estimated with normalized  $\text{OD}_{340\ \text{nm}}$  in the reaction solution against the standard concentration curve established at  $340\ \text{nm}$  with a series of  $100\ \text{mM}$  Hepes-NaOH solutions with gradient concentrations of NADPH. Each reaction was replicated 3 times.

To examine starch synthase activity, a  $20\ \mu\text{l}$  aliquot of the enzyme extract was mixed with  $40\ \mu\text{l}$  solution composed of  $50\ \text{mM}$  Hepes-NaOH at  $\text{pH}\ 7.4$ ,  $15\ \text{mM}$  DTT,  $0.7\ \text{mg}$  amylopectin and  $1.6\ \text{mM}$  ADPG, incubate 20 min, then in a  $100^\circ\text{C}$  water bath, and cooled rapidly in an ice-water bath. It was further mixed with a  $20\ \mu\text{l}$  solution containing  $50\ \text{mM}$  Hepes-NaOH at  $\text{pH}\ 7.4$ ,  $1.2\ \text{U}$  pyruvate kinase,  $4\ \text{mM}$  phosphoenolpyruvate,  $10\ \text{mM}$   $\text{MgCl}_2$  and  $200\ \text{mM}$  KCl, incubate 20 min, terminated by  $100^\circ\text{C}$  water bath. It was immediately cooled in an ice-water bath, and then centrifuged at  $10,000 \times g$  at  $4^\circ\text{C}$  for 10 min. A  $60\ \mu\text{l}$  aliquot of the supernatant was mixed with  $40\ \mu\text{l}$  solution containing  $50\ \text{mM}$  Hepes-NaOH at  $\text{pH}\ 7.4$ ,  $2\ \text{mM}$  NADP,  $10\ \text{mM}$  glucose,  $0.35\ \text{U}$  glucose-6-P dehydrogenase,  $1.4\ \text{U}$  hexokinase and  $20\ \text{mM}$   $\text{MgCl}_2$ , incubated for 10 min. The starch synthase activity was then determined with the  $\text{OD}_{340\ \text{nm}}$  in the reaction solution against a standard concentration curve established with gradient concentrations of NADPH at  $340\ \text{nm}$ . Each reaction was replicated 3 times.

## References

- Juntawong, P. *et al.* Elucidation of the molecular responses to waterlogging in *Jatropha* roots by transcriptome profiling. *Front Plant Sci.* **5**, 658, doi:10.3389/fpls.2014.00658 (2014).
- Bailey-Serres, J. *et al.* Making sense of low oxygen sensing. *Trends Plant Sci.* **17**, 129–138 (2012).
- Elena, L., Hans, V. & Pierdomenico, P. Plant responses to flooding stress. *Current Opinion in Plant Biology* **33**, 64–71 (2016).
- Pucciariello, C., Voesenek, L. A., Perata, P. & Sasidharan, R. Plant responses to flooding. *Front Plant Sci.* **5**, 226, doi:10.3389/fpls.2014.00226 (2014).
- Cesar, V. M., Mary, E. R., Jih-Jing, L. & Robert, A. K. Constitutive and inducible aerobic and anaerobic stress proteins in the *echinocloa* complex and rice. *Plant Physiol.* **101**, 217–226 (1993).
- Christianson, J. A., Llewellyn, D. J., Dennis, E. S. & Wilson, I. W. Global gene expression responses to waterlogging in roots and leaves of cotton (*Gossypium hirsutum* L.). *Plant Cell Physiol.* **51**, 21–37 (2010).
- Ellis, M. H., Dennis, E. S. & Peacock, W. J. Arabidopsis roots and shoots have different mechanisms for hypoxic stress tolerance. *Plant Physiol.* **119**, 57–64 (1999).
- Zuxin, Z., Danfeng, Z. & Yonglian, Z. Transcriptional and post-transcriptional regulation of gene expression in submerged root cells of maize. *Plant Signal Behav.* **4**, 132–135 (2009).
- Sachs, M. M., Freeling, M. & Okimoto, R. The anaerobic proteins of maize. *Cell* **20**, 761–767 (1980).
- Klok, E. J. *et al.* Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. *Plant Cell* **14**, 2481–2494 (2002).
- Chang, W. W. *et al.* Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment and identification of proteins by mass spectrometry. *Plant Physiol.* **122**, 295–318 (2000).
- Sasidharan, R. *et al.* Root transcript profiling of two *Rorippa* species reveals gene clusters associated with extreme submergence tolerance. *Plant Physiol.* **163**, 1277–1292 (2013).
- Zhang, Z. X., Tang, W. H., Tao, Y. S. & Zheng, Y. L. cDNA microarray analysis of early response to submerging stress in *Zea mays* roots. *Russ J Plant Physiol.* 5243–5249 (2005).
- Gibbs, D. J. *et al.* Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature* **479**, 415–418 (2011).
- Tang, W., Zhang, Z., Zou, X. & Zheng, Y. Functional genomics of maize submergence tolerance and cloning of the related gene *Sicyp51*. *Sci China C Life Sci.* **48**, 337–345 (2005).
- Hoeren, F. U., Dolferus, R., Wu, Y., Peacock, W. J. & Dennis, E. S. Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (ADH1) by low oxygen. *Genetics* **149**, 479–490 (1998).
- Hattori, Y. *et al.* The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* **460**, 1026–1030 (2009).
- Yang, C. Y., Hsu, F. C., Li, J. P., Wang, N. N. & Shih, M. C. The AP2/ERF transcription factor ATERF73/HRE1 modulates ethylene responses during hypoxia in Arabidopsis. *Plant Physiol.* **156**, 202–212 (2011).
- Chen, W. *et al.* Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. *Front Plant Sci.* **7**, 1044, doi:10.3389/fpls.2016.01044 (2016).
- Lenssen, J. P. M., Menting, F. B. J. & Van der Putten, W. H. Plant responses to simultaneous stress of waterlogging and shade: amplified or hierarchical effects? *New Phytologist* **157**, 281–290 (2003).
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
- Pang, T., Ye, C., Xia, X. & Yin, W. De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. *BMC Genomics* **14**, 488, doi:10.1186/1471-2164-14-488 (2013).
- Fan, X. *et al.* Gene expression profiling of soybean leaves and roots under salt, saline-alkali and drought stress by high-throughput Illumina sequencing. *Gene* **512**, 392–402 (2013).
- Postnikova, O. A., Shao, J. & Nemchinov, L. G. Analysis of the alfalfa root transcriptome in response to salinity stress. *Plant Cell Physiol.* **54**, 1041–1055 (2013).
- Dong, C. *et al.* Gene expression profiling of *Sinapis alba* leaves under drought stress and rewatering growth conditions with Illumina deep sequencing. *Mol. Biol. Rep.* **39**, 5851–5857 (2012).
- Zhai, L. *et al.* Genome-wide identification and analysis of microRNA responding to long-term waterlogging in crown roots of maize seedlings. *Physiol. Plant* **147**, 181–193 (2013).
- Wang, L. *et al.* Tolerant and susceptible sesame genotypes reveal waterlogging Stress Response Patterns. *PLoS ONE* **11**, e0149912, doi:10.1371/journal.pone.0149912 (2016).
- Lee, Y. *et al.* Global gene expression responses to waterlogging in leaves of rape seedlings. *Plant Cell Rep.* **33**, 289–299 (2014).
- Becker, M., Ladha, J. K. & Ottow, J. C. G. Growth and  $\text{N}_2$  fixation of two stem-nodulating legumes and their effect as green manure on lowland rice. *Soil Biol. Biochem.* **22**, 1109–1119 (1990).
- Becker, M. & George, T. Nitrogen fixing response of stemand root-nodulating *Sesbania* species to flooding and mineral nitrogen. *Plant Soil* **175**, 189–196 (1995).
- Ladha, J. K., Pareek, R. P. & Becker, M. Stem nodulating legume-*Rhizobium* symbiosis and its agronomic use in lowland rice. *Adv Soil Sci.* **20**, 147–192 (1992).

32. Itoh, J., Shioya, T., Nouguchi, R. & Kogano, K. Improvement of soil properties by cultivation of tropical leguminous crops in heavy clayey paddy fields. *Bull Hokuriku Natl Agric Exp Stn.* **34**, 27–41 (1992).
33. La Camera, S. *et al.* Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. *Immunol Rev.* **198**, 267–284 (2004).
34. Wang, L. H. *et al.* Global gene expression responses to waterlogging in roots of sesame (*Sesamum indicum* L.). *Acta Physiol Plant* **34**, 2241–9 (2012).
35. Fini, A. *et al.* Drought stress has contrasting effects on antioxidant enzymes activity and phenylpropanoid biosynthesis in *Fraxinus ornus* leaves: An excess light stress affair? *J Plant Physiol.* **169**(10), 929–39 (2012).
36. Kreuzwieser, J. U. R. *et al.* Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia. *Plant Physiol.* **149**, 461–473 (2009).
37. Lin, K. R., Weng, C., Lo, H. & Chen, J. Study of the root antioxidative system of tomatoes and eggplants under waterlogged conditions. *Plant Sci.* **167**, 355–365 (2004).
38. Narsai, R., Rocha, M., Geigenberger, P., Whelan, J. & Van Dongen, J. T. Comparative analysis between plant species of transcriptional and metabolic responses to hypoxia. *New Phytol.* **190**, 472–487 (2011).
39. Nanjo, Y. *et al.* Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings. *Plant Mol Biol.* **77**, 129–144 (2011).
40. Oh, M. & Komatsu, S. Characterization of proteins in soybean roots under flooding and drought stresses. *J Proteomics* **114**, 161–181 (2015).
41. Christianson, J. A., Llewellyn, D. J., Dennis, E. S. & Wilson, I. W. Comparisons of early transcriptome responses to low-oxygen environments in three dicotyledonous plant species. *Plant Signal Behav.* **5**, 1006–1009 (2010).
42. Branco-Price, C., Kaiser, K. A., Jang, C. J. H., Larive, C. K. & Bailey-Serres, J. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* **56**, 743–755 (2008).
43. Yang, O. *et al.* The *Arabidopsis* basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene* **436**, 45–55 (2009).
44. Tran, L.-S. P. *et al.* Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* **16**, 2481–2498 (2004).
45. Jiang, Y. & Deyholos, M. K. Functional characterization of *Arabidopsis* NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol. Biol.* **69**, 91–105 (2009).
46. Cui, M. H. *et al.* An *Arabidopsis* R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. *FEBS Lett.* **587**, 1773–1778 (2013).
47. Jiang, Y., Yang, B. & Deyholos, M. K. Functional characterization of the *Arabidopsis* bHLH92 transcription factor in abiotic stress. *Mol. Gen. Genom.* **282**, 503–516 (2009).
48. Valliyodan, B. *et al.* Expression of Root-Related Transcription Factors Associated with Flooding Tolerance of Soybean (*Glycine max*). *International Journal of Molecular Sciences* **15**, 17622–17643 (2014).
49. Yang, A., Dai, X. & Zhang, W.-H. A R2R3-type MYB gene, OsMYB2, is involved in salt, cold, and dehydration tolerance in rice. *J Exp. Botany* **63**, 2541–2556 (2012).
50. Mahajan, S. & Tuteja, N. Cold, salinity and drought stresses: An overview. *Arch. Biochem. Biophys.* **444**, 139–158 (2005).
51. Rebecca, H.-N. *et al.* High-throughput *in vivo* analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biol.* **5**, e237, doi:10.1371/journal.pbio.0050237 (2007).
52. Yan, K. *et al.* Physiological adaptive mechanisms of plants grown in saline soil and implications for sustainable saline agriculture in coastal zone. *Acta Physiol. Plant* **35**, 2867–2878 (2013).
53. Patel, R. K. & Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One* **7**, e30619, doi:10.1371/journal.pone.0030619 (2012).
54. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652 (2011).
55. Li, B. & Dewey, C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
56. Wang, L. *et al.* DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).
57. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat.* **29**, 1165–1188 (2001).
58. Young, M. D. *et al.* Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, R14, doi:10.1186/gb-2010-11-2-r14 (2010).
59. Mao, X. *et al.* Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793 (2005).
60. Nakamura, Y., Yuki, K., Park, S. Y. & Ohya, T. Carbohydrate metabolism in the developing endosperm of rice grains. *Plant Cell Physiol.* **30**, 833–839 (1989).
61. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res.* **40**, D109–D114 (2012).

## Acknowledgements

This work is financed by One Hundred-Talent Plan of Chinese Academy of Sciences (Y529041011), the National Natural Science Foundation of China (31601238, 31370108 and 31570063), the High-tech Industrialization Cooperation Funds of Jilin province and the Chinese Academy of Science, Shandong Key Research and Development Program (2016CYJS05A01-1), Agricultural Scientific and Technological Innovation Project of Shandong Academy of Agricultural Sciences (+CXGC2016B10), Yantai Key Project of Research and Development Plan (2016ZH074).

## Author Contributions

C.-G.R. conceived and designed the experiments. C.-G.R., Z.-H.X., Y.-M.L. and H.Z. supervised the experiments. C.-G.R., C.-C.K. and K.Y. performed the experiments. C.-G.R. analyzed the data and wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-07740-5

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017