Genes Genom (2016) 38:949–960
DOI 10.1007/s13258-016-0439-x

Online ISSN 2092-9293
Print ISSN 1976-9571

RESEARCH ARTICLE

Novel drought-responsive regulatory coding and non-coding transcripts from *Oryza Sativa* L.

Seon-Ju Shin[1] · Hongryul Ahn[2] · Inuk Jung[3] · Sungmin Rhee[2] · Sun Kim[2,3,4] · Hawk-Bin Kwon[1]

Received: 26 April 2016 / Accepted: 30 May 2016 / Published online: 9 June 2016

Abstract Drought-stress can cause major economic loss and is a serious issue to address in agriculture. Defining the molecular pathways in how a plant responds to drought-stress may prove valuable in developing new drought-resistant plants. In this study, we identified several novel drought-responsive regulatory coding and noncoding transcripts in rice, *Oryza sativa* L., using the next generation sequencing (NGS) technique and bioinformatics analyses. We produced comprehensive NGS RNA sequencing data for mRNA, small RNA, and long noncoding RNA at the experimental conditions of aeration without watering for 1 and 6 h in compared with the 0 h control. We performed bioinformatics analysis to identify novel drought-responsive transcription factors (TFs), novel drought-responsive micro RNA (miRNA), and novel drought-responsive long noncoding RNAs (lncRNAs). These transcripts were validated by quantitative real-time PCR (qRT-PCR). We identified 18 TFs, ten lncRNAs and one miRNA as being novel drought-responsive regulatory transcripts. Computational analysis using a gene regulatory network showed that these transcripts were related to regulation of response to drought-stress.

✉ Hawk-Bin Kwon
hbkwon@sunmoon.ac.kr

[1] Department of Biomedical Sciences, Sunmoon University, Asan 336-708, Korea

[2] Department of Computer Science and Engineering, Seoul National University, Gwanak-Gu, Seoul 151-744, Republic of Korea

[3] Interdisciplinary Program in Bioinformatics, Seoul National University, Gwanak-Gu, Seoul 151-747, Republic of Korea

[4] Bioinformatics Institute, Seoul National University, Gwanak-Gu, Seoul 151-747, Republic of Korea

Keywords Rice · Drought · Coding and noncoding transcripts · NGS · Bioinformatics

Introduction

Plants have developed sophisticated protective mechanisms to overcome various abiotic environmental stresses that include extreme temperatures, drought and high salinity (Sadhukhan et al. 2014). These abiotic changes can have significant detrimental effects on plant growth and crop yields and are a major reason for crop losses (Atkinson and Urwin 2012).

Plant protective mechanisms against these stresses rely on a range of mechanisms, including adjustment of physiological and cellular processes via regulation of gene expression and transcription, post-transcriptional regulation, translation and post-translational modifications for various target genes and proteins (Ingram and Bartels 1996; Covarrubias and Reyes 2010; Cramer et al. 2011; Bailey-Serres et al. 2012; Lyzenga and Stone 2012; Howell 2013). The gene regulatory elements involved in such protective mechanisms include transcription factors (TFs) and non-coding RNAs, such as (miRNAs), siRNAs and long non-coding RNAs (lncRNAs).

To date, many drought-responsive TFs have been identified such MYB, AP2/EREBP, bHLH, HSF, NAC and WRKY families (Wang et al. 2011; Jeong et al. 2010; Silveira et al. 2015). Other than transcription factors, various proteins have been discovered that mediate drought-stress protection and these include molecular chaperons, osmotic adjustment proteins, ion channels, membrane transporters, and anti-oxidation/detoxification proteins (Hu et al. 2006).

Besides proteins, there is research into networks of non-coding RNAs (ncRNAs) in characterizing the drought

protection pathways. The ncRNAs are a group of functional RNA molecules that are not translated into protein, and they are divided into two categories. The first includes ncRNAs that have general housekeeping functions and these include ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear/small nucleolar RNA (sn/snoRNA). The second category includes microRNA (miRNA), small interfering RNA (siRNA) and long non-coding RNA (lncRNA) that have regulatory roles in plants for diverse cellular processes such as nutrient homeostasis (Sunkar et al. 2007), growth, development and differentiation (Ben Amor et al. 2009) and abiotic stress response (Sunkar et al. 2007; Matsui et al. 2013). A number of microRNAs are known to be related to drought stress. Among these miR168, miR172 and miR396 from rice are included (Zhao et al. 2007; Liu et al. 2008; Zhou et al. 2010).

miRNAs are small non-coding RNAs found in both animals and plants. Mature miRNAs, many of which are conserved across plant families, are usually 20–25 nucleotides (nt) in length and are derived from stem-loop regions of approximately 70-ntRNA precursors. miRNAs are produced by the activities of Dicer-like RNase III enzymes (Bartel 2004; Murchison and Hannon 2004; Kim et al. 2005; Jones-Rhoades et al. 2006; Kohli et al. 2014). miRNAs can bind to target sequences on mRNAs by an exact or near-exact complementary base pairing, which enables them to direct the cleavage and destruction of the mRNA (Rhoades et al. 2002; Chen 2005) or repress translation of the target mRNA (Jones-Rhoades et al. 2006; Kohli et al. 2014). As such, miRNAs mostly function as negative post-transcriptional regulators. The majority of miRNAs target genes that encode transcriptional factors or key enzymes that play important roles in cell development and in plants in response to various biotic and abiotic stresses (Jones-Rhoades et al. 2006; Schmmer et al. 2012).

In contrast to miRNAs, lncRNAs are mRNA-like transcripts of over 200 nucleotides long that are not translated into proteins. There is evidence that lncRNAs have important functions in gene regulation .(Zhang et al. 2013; Jin et al. 2014). The biogenesis of lncRNAs is very similar to that of protein-coding mRNAs in that most of the lncRNAs are transcribed by RNA polymerase II in eukaryotes, although there are several reports that many lncRNAs are transcribed by RNA polymerase III (Zhang et al. 2013). lncRNAs share many of the features of mRNAs, such as 5′-capping, splicing, and 3′-polyadenylation (Carninci et al. 2005; Liu et al. 2012). The origin of lncRNA transcription includes intronic, exonic, intergenic and intragenic areas of the genome. These also include promoter and enhancer regions and other 3′-and 5′ untranslated regions (UTRs) of genes (Nie et al. 2012; Liu et al. 2012). lncRNAs can be transcribed in either the sense or antisense direction (Nie et al. 2012; Matsui et al. 2013). Although the functions of various lncRNA remain largely unknown, there is evidence that a number of lncRNA may be involved in regulating transcription (Ponting et al. 2009). These include transcriptional interference (Zhang et al. 2013), an enhancement of target genes accessibility to RNA polymerase (Hirota et al. 2008), blocking of the pre-initiation complex from the target gene promoter (by formation of a RNA-double strand DNA triplex) (Martianov et al. 2007), and control of transcription factor localization in addition to inhibition of RNA polymerase activities (Mariner et al. 2008; Nguyen et al. 2001; Willingham et al. 2005). There are also reports of lncRNAs also affecting post-transcriptional events. These include alternative splicing, RNA transport, RNA translation, and RNA degradation (Zhang et al. 2013; Jouannet and Crespi 2011). There are also reports that lncRNAs may also have roles in drought stress response (Zhang et al. 2014) and epigenetic control of gene expression (Heo and Sung 2011; Rinn et al. 2007; Tsai et al. 2010; Zhao et al. 2008).

The goal of this study was to identify drought-responsive coding and non-coding regulatory transcripts at the whole genome level. In this study, we report 18 transcription factors, 10 lncRNAs and one miRNA that were responsive to drought-stress and their message level changes were verified by qRT-PCR. All of these transcripts have not been previously reported as drought-responsive transcripts in the literature. Since these transcripts were drought-responsive, our findings provide useful information in investigating how rice plant responds to drought-stress. A typical study on the biological function of potential regulatory elements is to modulate expression by knocking out or amplifying the element. Validation of the biological function of these transcripts remains to another study. To produce a testable hypothesis on the biological function of these drought-responsive transcripts, we performed network-based analysis of these regulatory transcripts by performing clustering and gene ontology analysis for genes that were potentially targeted by these regulatory transcripts.

Materials and methods

Plant materials and stress treatments

Rice plants, *O.sativa* L. Nipponbare, were acquired from the Rural Development Administration of Korea (RDA). The rice plants were cultured in Yoshida solution at pH 5.8 (Yoshida et al. 1976) and maintained in a temperature-controlled culture room at 29 °C under 16 h/8 h light/dark conditions. Rice plants at the three-leaf stage were subjected to drought-stress for 1 and 6 h by removal of the

culture solution. Untreated plants were used as control. After treatment, entire plants were immediately transferred into liquid nitrogen.

Next generation sequencing (NGS) for coding and noncoding transcripts

Total RNA was extracted from control and drought-stressed rice plants using Tri-Reagent (MRC, Cincinnati OH, USA) according to the manufacturer's instructions. This was followed by prepping the mRNAs using oligo dT magnetic beads on the total RNA. First and second cDNAs were then synthesized. After the cDNA ends were repaired, the 3′ ends were adenylated and the adaptors were ligated. After PCR amplification of cDNAs, NGS was conducted using a TruSeq^{TM} RNA sample preparation Guide (Illumina, San Diego CA, USA). To identify miRNA from rice plants upon drought-stress, total small RNAs (mRNA of <30 bp in length) were purified from control rice plants and drought-treated rice plants. We sequentially attached 5′ and 3′ adaptors to small RNA fragments. Single stranded cDNA was synthesized using a room temperature reaction. cDNA was PCR-amplified using common primers. The small RNA library was gel purified. After library validation, NGS was performed using the MiSeq^{TM} Sequencing System (Illumina). To identify lncRNA's of more than 200 bp in length, rRNAs were removed from total RNA using Ribo-Zero^{TM} Magnetic Kits (Epicentre, Madison WI, USA). Fragments of 250–500 bp including adaptor size were selected and a template cDNA library was made. Deep sequencing was done using the MiSeq^{TM} Sequencing System (Illumina).

Functionally unknown drought-responsive transcription factor analysis

To measure mRNA expression, the sequenced short reads were aligned to the rice plant genome (Os-Nipponbare-Reference-IRGSP-1.0) using TopHat aligner (Trapnell et al. 2009). Gene expressions were calculated with fragments per kilo bases of genes for per million reads (FPKM) normalization using Cufflinks (Trapnell et al. 2010). After statistical test for differentially expressed gene at 0–1 and 0–6 h using Cuffdiff, a sub-package of Cufflinks, genes with p-values of less than 0.05, a minimum expression value of 1 FPKM and with at least twofold change expression difference between time points were selected as differentially expressed genes (DEGs). Among DEGs, genes annotated as TF from PlantTFDB (Jin et al. 2014) were selected as putative drought-responsive TFs. After literature searches, TFs not previously reported in drought experiment studies were selected as candidate, functionally unknown drought-responsive TFs. We performed a Gene Ontology (GO) enrichment analysis of gene sets using Rice Oligonucleotide Array Database (Cao et al. 2012) and selected GO biological process terms with p value <0.05 and proportion >2 % as enrichment biological functions.

Novel drought-responsive miRNA analysis

After removing the adaptor sequences of the miRNA, NGS data were analyzed for identifying novel miRNA according to primary criterion (Meyers et al. 2008). At first, short reads matching to known miRNAs annotated in the miRBase database (Kozomara and Griffiths-Jones 2014) were removed, and the remaining read sequences were mapped onto the reference genome using Bowtie (Langmead et al. 2009). The remaining reads were further filtered using Rfam database (Griffiths-Jones et al. 2003) to remove non-coding RNAs. Next, we selected putative miRNAs that possess known miRNA characteristics, such as a hairpin structure, from the mapping results. For this process, read pairs were selected that mapped to ± sequences separated by 200–300 bp to identify miRNA existence. The next step was calculation of RNA secondary structure with sequences onto which the selected pairs were mapped using RNAfold, a sub-package of ViennaRNA (Lorenz et al. 2011). The miRNAs whose hairpin structures were confirmed by using the MIRCheck package (Jones-Rhoades and Bartel 2004) were selected for putative novel miRNAs. After selection of putative novel miRNAs, reads that mapped to repeat sequence of rice genome were further removed. Under the assumption that the expression level of a putative novel miRNA should be higher than a certain level, only reads with high read counts from more than two rice plant samples were selected. After that, expression levels of novel miRNA found in the previous step and known RNAs annotated in the miRBase database were computed according to protocol of a previous study (Jiang et al. 2015). After statistical test for differentially expressed miRNA at 0–6 h using DEGseq (Wang et al. 2010), a well-known sequence based DE statistical analysis tool, miRNAs with p-values of less than 0.05, a minimum raw read count of 10, and with at least twofold change expression difference between time points were selected as differentially expressed miRNAs (DEmiRNAs). Finally DEmiRNAs that were not reported in previous drought-related studies were listed as novel drought-responsive miRNAs.

Novel drought-responsive long non-coding RNA analysis

Assembly of lncRNAs was done using the ribosome RNA depleted RNA-seq data. First, the sequences from RNA-seq data were aligned to the rice genome (IRGSP 1.0) using TopHat (Trapnell et al. 2009). The annotated genes were

masked to quantify the intergenic non-coding genes using Cufflinks (Trapnell et al. 2010). From the assembled candidate lncRNAs, we removed the assembled transcripts of <200 bp in length. The remaining candidates were tested for coding potential using Coding Potential Calculator (CPC) (Kong et al. 2007). The sequences of the candidate lncRNAs were aligned to known coding genes using BLAST (Altschul et al. 1990). From the result, lncRNAs with significant alignment were further removed from the candidate list.

Quantitative real-time PCR (qRT-PCR)

The expression level of coding transcripts, lncRNA and miRNA were quantified by qRT-PCR. We used two genes as qRT-PCR references. The drought stress inducible *Dip1* (Dehydration-stress inducible protein 1, AY587109, Yi et al. 2010) was used as reference for drought stress (Fig. 2d) and elongation factor 1α (EF-1α) was used as internal RT-PCR reference. The primer sequences of these two genes are as follows: Dip1-S (5′-GGCTACAGAGGA AGTGAGCAGCCC-3′), Dip1-AS (5′-TTAAGCGCTGCT CTTGTGCTCGCC-3′), EF-1α-S (5′-AGCGTGTGATT-GAGAGGTTC-3′), EF-1α-AS (5′-ACTTCCACAGGGCA ATATCG-3′). qRT-PCR of coding transcripts and lncRNA was performed using a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules CA, USA). Total RNAs were extracted by Tri-Reagent (MRC). cDNAs were synthesized from 2 µg total RNA with TOPscript™ RT MIX (Enzynomics, Seoul, Korea). qRT-PCR reactions were carried out using TOPreal™ qPCR 2X SYBR Pre-MIX (Enzynomics). The condition for qRT-PCR reaction was as follows: initial denaturation at 95 °C for 15 min, 45 amplification cycles including denaturation at 95 °C for 15 s, annealing at 60 °C for 15 s and elongation at 72 °C for 20 s. To confirm specificity, a melting curve analysis was performed at the end of the PCR amplifications. All experimental samples were run in duplicate. Expression value of coding transcripts and lncRNAs was presented as Ct (threshold cycles) values. The abundance of the coding transcripts and lncRNAs was normalized against the amount that of elongation factor 1α (EF-1α) and the relative abundance was determined by the comparative threshold cycle method, $2^{-\Delta\Delta Ct}$, using CFX Manager™ Software (Bio-Rad). MicroRNA qRT-PCR was performed using CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules CA, USA) according to TaqMan® miRNA assays (Applied Biosystems, USA). cDNAs were synthesized from 1 µg of total RNA with TaqMan® MicroRNA Reverse Transcription Kit components (Applied Biosystems, USA) and qRT-PCR reactions were carried out using TaqMan® Universal Master MixII (Applied Biosystems, USA). The abundance of the miRNA was normalized against the amount that of bdi-miR168-5p (TCGCTTGGTGCAGATCGGGAC) using CFX Manager™ Software (Bio-Rad).

Transcription factor regulatory network construction

To characterize biological functions of novel drought-responsive TFs, we constructed a TF regulatory network. 1893 microarray datasets normalized by the Robust Multi-array Average (Irizarry et al. 2003) method were downloaded from the OryzaExpress Gene Expression Network (http://bioinf.mind.meiji.ac.jp/Rice_network_public/script). TF-target gene regulatory relationships were inferred based on co-expression relationships, calculating Pearson's correlation coefficients (PCCs) between the novel TF genes and putative target genes. Highly positive/negative correlated gene pairs ($|PCC| > 0.5$) were selected for regulatory relationships (i.e., edges of the TF regulatory network). The TF regulatory network was displayed using Prefuse Force-Directed Layout in Cytoscape (Saito et al. 2012), and the nodes in the network were colored according to logarithm of expression fold-change at 6 h compared with 0 h.

Identification of putative miRNA targets

To characterize the biological functions of novel drought-responsive miRNAs, we first constructed a miRNA regulatory network. Using psRNAtarget (Dai and Zhao 2011), a tool for finding plant small RNA target based on sequence characteristics, we obtained preliminary candidates of target genes for miRNAs. From those candidates, negatively correlated genes with miRNAs (i.e. when miRNAs were up/down regulated with more than twofold-change at 0-to-6 h and target genes changed in the opposite manner with more than twofold-change at 0-to-6 h) were selected as final miRNA targets.

Results

Identification of drought-responsive genes and their biological functions

To identify drought-responsive novel or function-unknown coding and noncoding transcripts from rice plant, *O. sativa* L., we used 4.8 gigabytes of sequencing reads with NGS. Among the total 37,896 genes, genes with $p < 0.05$, at least 1 FPKM expression at any time point and a minimum of twofold change were selected as differentially expressed genes (DEGs). A total of 1080 and 4017 DEGs were found at 0–1 h and 0–6 h data points, respectively. Histograms and volcano plots of expression fold changes are shown in Fig. 1. The DEG analysis showed that the number of DEG increased as drought-stress continued. This observation

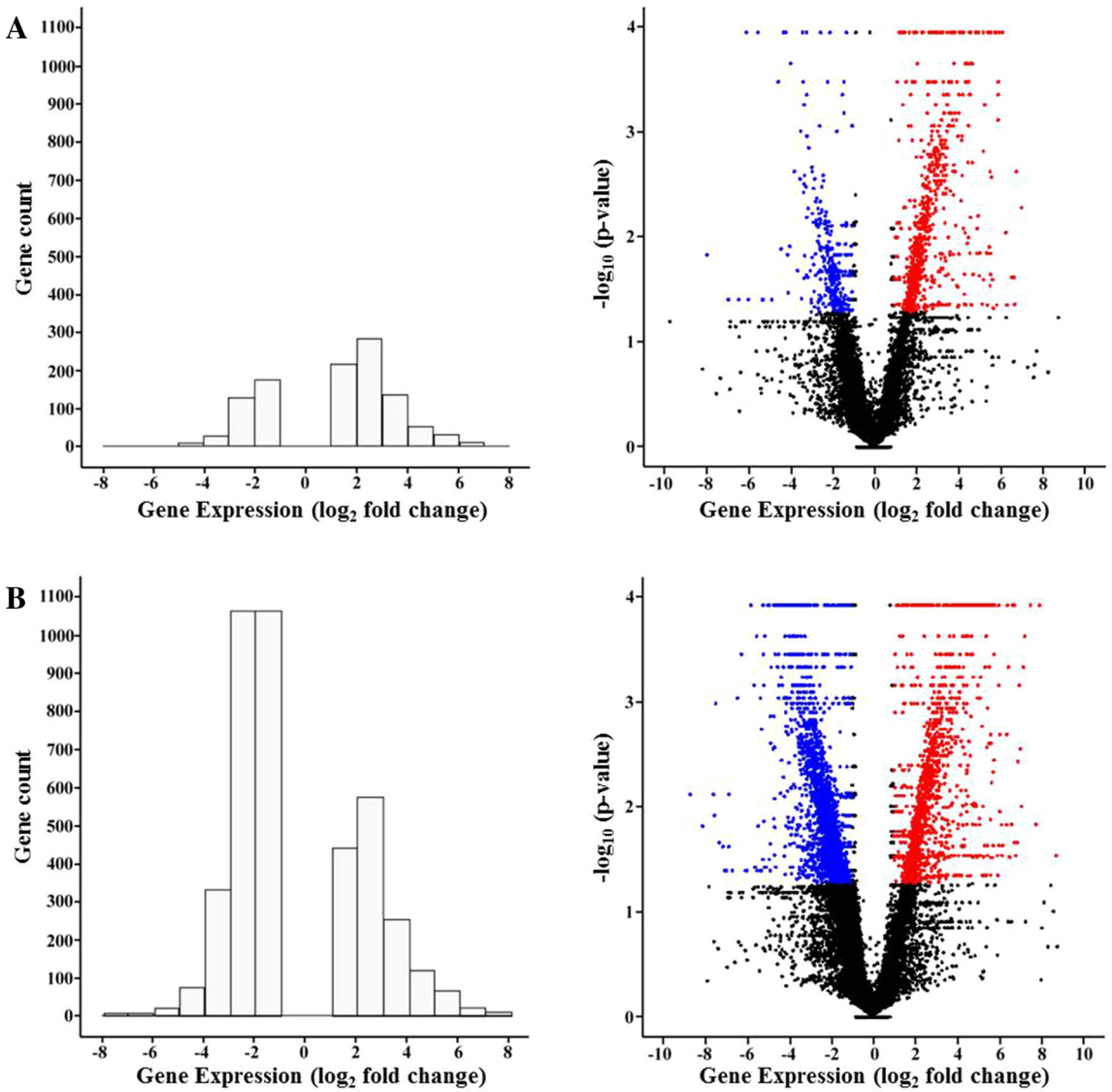

Fig. 1 **a** For 0-to-1 h time duration, the number of DEGs vs. gene expression $\log_2$ fold change (*left* figure) and volcano plot (*right* figure). Each dot in the volcano plot represents a gene by its fold change and its significance level by $-\log_{10}$(*p*-value) **b** For 0–1 h time duration, the number of DEGs vs. Gene expression $\log_2$ fold change (*left* figure) and volcano plot (*right* figure)

showed that when the plant was exposed to a longer period of drought-stress, transcriptional changes were bigger for our experiment.

Among the 4017 DEGs from the 0-to-6 h data set, 1472 genes were activated and 2545 genes were suppressed. We investigated the known functions of the activated/suppressed differentially expressed genes under drought-stress by Gene Ontology (GO) enrichment analysis. The result is shown in Table 1.

Identification of functionally unknown drought-responsive transcription factors (TFs)

The plant transcription factor specific database PlantTFDB (Jin et al. 2014) reports 2048 TFs in rice plant. Among these, 236 genes were differentially expressed as part of the 0–6 h group. After filtering out TFs previously reported in drought-stress studies, we selected 68 TFs as candidates with unknown functions for further study. We performed qRT-PCR on the 68 TFs and validated 18 of them (Table 2). The expression level of all 18 coding TFs was increased after 1 and 6 h drought treatment compared to 0 h control.

Expression of 11 out of 18 candidates (Os02g0649300, Os02g0638650, Os12g0123700, Os05g0421600, Os05g0583000, Os06g0166400, Os10g0478300, Os02g0579000, Os09g0558800, Os06g0670300, Os01g0674000) gradually and continuously increased after 1 h and 6 h drought-stress treatment. The expression of the remaining seven candidates (Os03g0820300, Os02g0764700, Os03g0180900, Os01g0863300, Os04g0429

Table 1 Enriched functions of activated and suppressed DEGs under drought-stress

Type	GO ID	GO description	Gene number	Pro-portion	P-value
Activated DEGs	GO:0006350	Transcription	73	3	0
	GO:0006950	Response to stress	50	2	0
	GO:0008152	Metabolic process	155	7	0
	GO:0045449	Regulation of transcription	114	5	0
	GO:0055114	Oxidation reduction	86	4	0
	GO:0006355	Regulation of transcription, DNA-dependent	71	3	0.0007
	GO:0006810	Transport	53	2	0.0108
	GO:0042309	Homoiothermy	81	4	0.017
	GO:0050826	Response to freezing	81	4	0.017
	GO:0006468	Protein amino acid phosphorylation	73	3	0.0371
Suppressed DEGs	GO:0005975	Carbohydrate metabolic process	75	3	0
	GO:0006468	Protein amino acid phosphorylation	155	5	0
	GO:0006508	Proteolysis	62	2	0
	GO:0008152	Metabolic process	190	7	0
	GO:0055114	Oxidation reduction	170	6	0
	GO:0006810	Transport	92	3	0.0002
	GO:0006952	Defense response	58	2	0.0025
	GO:0042309	Homoiothermy	117	4	0.0302
	GO:0050826	Response to freezing	117	4	0.0302
	GO:0006412	Translation	65	2	0.0312
	GO:0045449	Regulation of transcription	92	3	0.0436
	GO:0006355	Regulation of transcription, DNA-dependent	79	3	0.0471

Gene ontology biological process terms with p-value <0.05 and proportion >2 % were selected

050, Os01g0859100, Os12g0123800) increased after 1 h drought-stress treatment and then slightly decreased after 6 h treatment (Fig. 2a).

Identification of novel drought-responsive lncRNA transcripts

lncRNAs are defined as transcripts whose length is greater than 200 bp and have none or very weak coding potentials. From ribosome depleted control and drought-treated RNA-seq samples, a total of 104 lncRNA transcripts were identified. The average expression level of these lncRNAs at 0, 1 and 6 h was 51.5, 41.39 and 26.72 FPKM, respectively, demonstrating a majority having decreased expression over the course of the drought treatment. Differential expression analysis using a fold-change of four identified 12 down-regulated lncRNAs between 0 and 1 h, and 24 lncRNAs between 1 and 6 h. From 0 to 6 h, there were 38 down-regulated lncRNAs. 19 lncRNAs were up-regulated between 0 and 1 h, and 35 lncRNAs between 1 and 6 h. 63 lncRNAs were up-regulated between 0 and 6 h. The expression level of 12 lncRNAs were upregulated by ten-fold at 6 h compared to 0 h. The expression level of 7 lncRNAs were up-regulated by tenfold at 1 h compared to 0 h. As a result, we selected 10 novel lncRNAs for validation via qRT-PCR (Table 4). The workflow for identifying drought-responsive regulatory elements for transcription factors, miRNAs and lncRNAs and their qRT-PCR validated results is depicted in Figs. 2b and 3, respectively.

Identification of novel drought-responsive miRNAs

We identified 19 drought-responsive miRNA at 0–6 h from NGS sequencing data. We further investigated whether these 19 DEmiRNAs were novel drought-responsive miRNA that were not yet identified or miRNAs that were in miRBase already but were not known as being drought-responsive.

After filtering out short reads that were mapped to known miRNAs annotated in miRBase or another type of RNA such as an rRNA or tRNA, we identify novel miRNAs using the remaining reads. We found 11, 9, and 9 potential miRNAs from each of 0, 1, and 6 h samples. From those miRNAs, only one miRNAs was expressed in more than ten read counts for at least two samples. Osa-miR9898 in Table 3 shows the sequence, genomic location, read counts at three samples, and Fig. 4 shows predicted hairpin structure of the

Table 2 Novel/function-unknown drought-stress responsive coding transcripts

No.	Gene ID	Description	Fold change	*p*-value	No of edges
1	Os02g0649300	Similar to Short highly repeated interspersed DNA	250.6	0.0001	42
2	Os03g0820300	Similar to ZPT2-14	57.1	0.0001	226
3	Os02g0764700	Similar to ethylene-responsive transcription factor 4	43.9	0.0001	252
4	Os03g0180900	Tify domain containing protein	22.8	0.0003	311
5	Os01g0863300	Similar to MCB2 protein	19.0	0.0001	318
6	Os02g0638650	Pathogenesis-related transcriptional factor and ERF domain containing protein	18.6	0.0002	5
7	Os12g0123700	No apical meristem	14.9	0.0001	0
8	Os05g0421600	No apical meristem	13.3	0.0016	226
9	Os05g0583000	Similar to WRKY8	7.3	0.0072	2
10	Os06g0166400	Similar to TINY-like protein	6.6	0.0022	261
11	Os04g0429050	Similar to OSIGBa0093L02.1 protein	6.5	0.0001	0
12	Os10g0478300	Similar to Y19 protein	5.8	0.004	124
13	Os01g0859100	Similar to WIP1 protein	5.3	0.0033	72
14	Os12g0123800	No apical meristem	5.2	0.0055	107
15	Os02g0579000	No apical meristem	4.3	0.0052	31
16	Os09g0558800	Non-protein coding transcript	4.2	0.0284	0
17	Os06g0670300	Homeodomain-like containing protein	3.8	0.0167	224
18	Os01g0674000	Homeodomain-like containing protein	3.6	0.0141	158

Fold change is the ratio of expression level at 0–6 h. *p*-value indicates significance of differential expression of the gene

[a] No *of edges* number of target genes

novel miRNA. The hairpin sequence we found is an extended version of the osa-miR169 g hairpin sequence that is originally predicted as a paralogue of an Arabidopsis hairpin sequence by a computational method (Jones-Rhoades and Bartel 2004). The extension length is 13 nucleotides to the 5′ side and 12 nucleotides to the 3′ side. As shown in Fig. 4, the novel miRNA and an existing miRNA, osa-miR169 g, are arranged in phase in a hairpin sequence. This result is consistent with the previous research in *Arabidopsis* where multiple small RNAs originated and are arranged in phase within a hairpin sequence (Zhang et al. 2010). The expression level of os-miR9898 was down-regulated by 5.5-fold at 6 h compared to 0 h (Fig. 2c). As a result, we selected one novel miRNA for validation via qRT-PCR (Fig. 4).

Computational validation of gene regulatory function using a gene regulatory network

The number of novel coding and non-coding regulatory transcripts were too small (18 TFs, one miRNA, and 10 lncRNAs) for a GO enrichment analysis module (Table 5). To overcome this problem, we performed a computational validation of functions based on gene regulatory network analysis. First, a TF-target gene regulatory network of the selected 18 TFs was constructed. Visualization of the network using Cytoscape revealed clusters that were up-regulated or down-regulated (Fig. 5). T-test on how significantly target gene clusters were regulated showed that the 6 core TFs—Os01g0859100, Os01g0863300, Os02g0649300, Os02g0764700, Os03g0180900, and Os03g0820300—formed up-regulated cluster groups. The enriched GO biological terms of these up-regulated target-gene clusters were related to regulation of transcription, phosphorylation, and response-to-stress among others. Four core TFs—Os05g0421600, Os05g0583000, Os06g0166400, and Os06g0670300 - formed down-regulated cluster groups. Their enriched GO terms were proteolysis, metabolic process, oxidation–reduction, and regulation of transcription.

Discussion

In this study, NGS identified novel or function-unknown coding and noncoding transcripts from response of rice plant, *O. sativa* L., to drought-stress. We identified 4457 differentially expressed genes that responded to drought-stress. Among them, 236 genes were annotated to be TFs and 68 TFs had unknown functions related to drought-stress. From these 68 transcript factors, we found various domains with known DNA-sequence targets (ethylene-, dehydration-responsive element, pathogenesis-related transcriptional, hypothetical, zinc finger and WRKY), but some had no known DNA-binding domains.

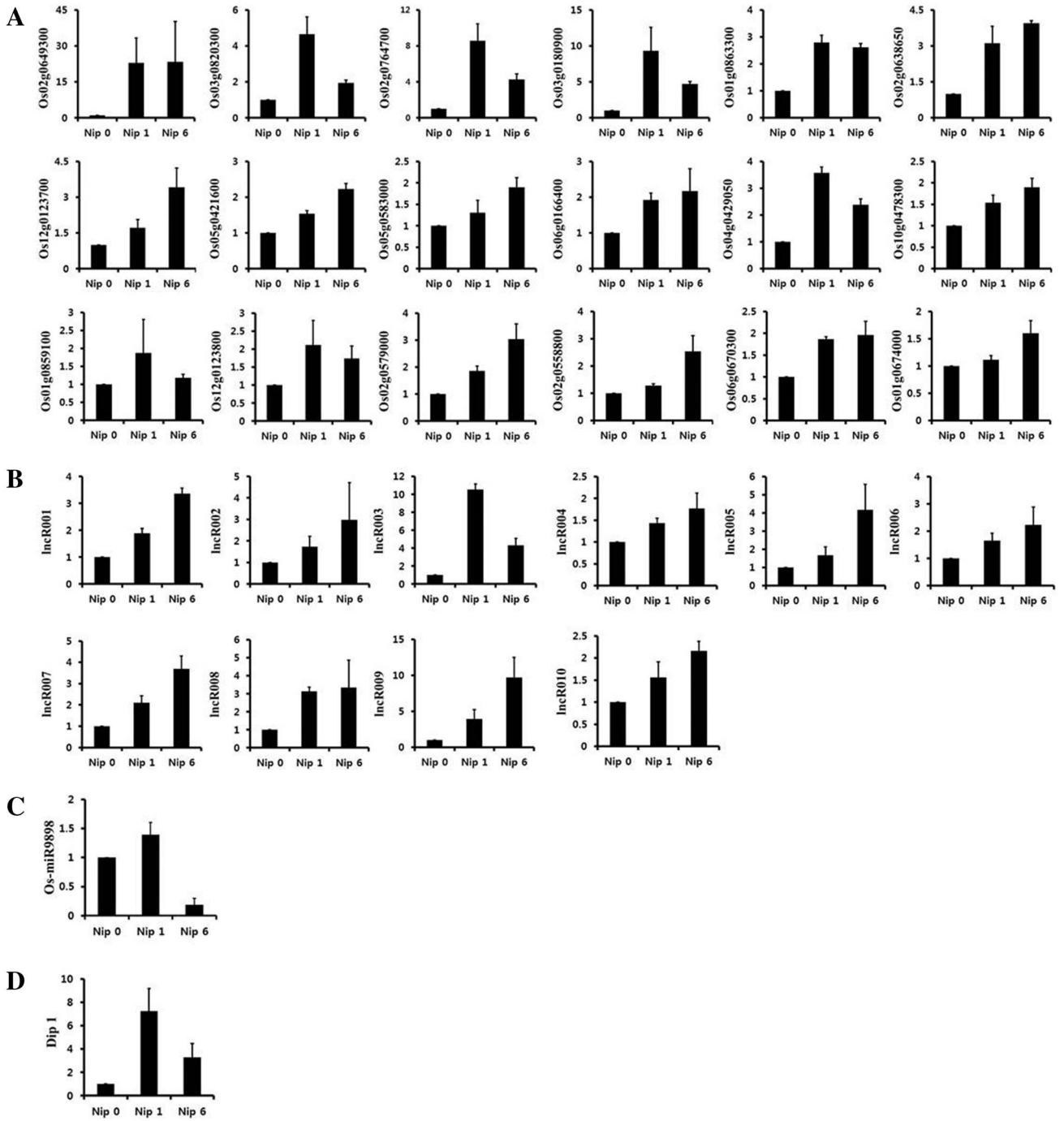

Fig. 2 Time series transcripts expression profiling. **a** Novel coding transcripts qRT-PCR. **b** Novel long noncoding RNA transcripts qRT-PCR. **c** Novel miRNA transcript qRT-PCR. **d** Drought stress inducible *Dip1* qRT-PCR

To confirm the expression profile of the novel or function-unknown transcripts identified from mRNA deep sequencing and sequence data analysis, we performed qRT-PCR with 68 coding transcripts, mostly transcription factors, and selected 18 novel or function-unknown TFs for further analysis. Based on RNA-seq, we also identified 9 miRNAs and 104 novel lncRNAs that were novel and differentially expressed in response to drought-stress. Of these selected transcripts, 10 lncRNAs were validated by qRT-PCR and selected as novel drought-responsive regulatory transcripts. The identity of these experimentally validated novel regulatory transcripts is the major finding of our paper. Characterizing the biological function of these regulatory elements should be completed in a study

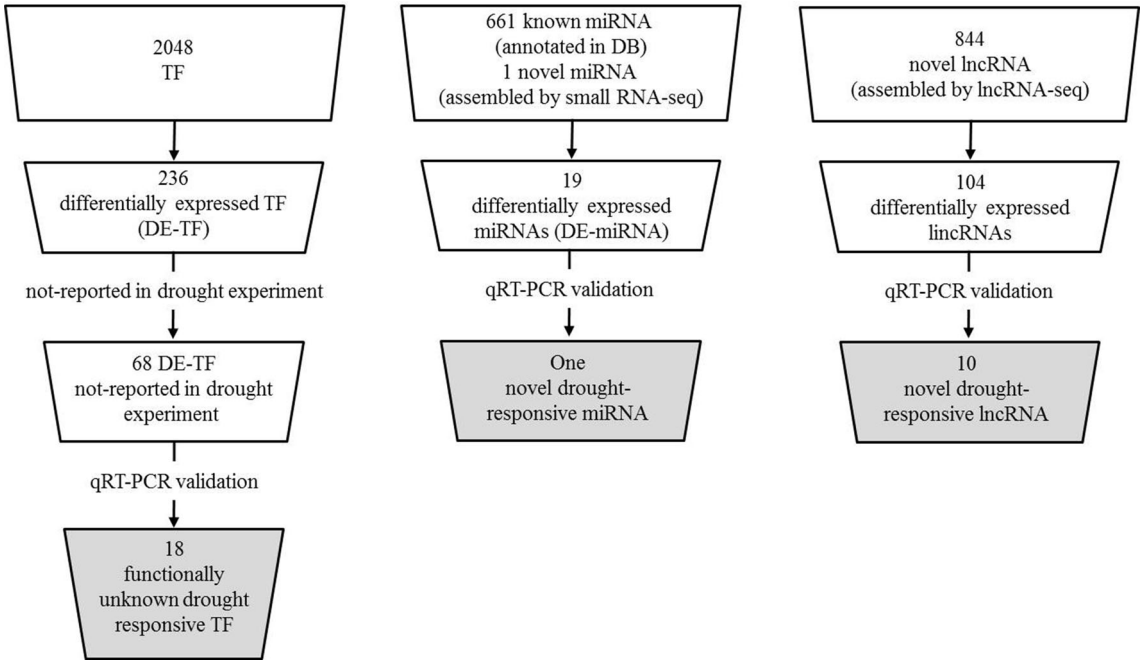

Fig. 3 Identification protocol of functionally unknown and novel drought-responsive coding/non-coding transcripts. *left*, *middle*, and *right* figures are procedures for TF, miRNA, and lncRNA, respectively

Table 3 Novel drought-responsive miRNA and their target genes. Fold change is the ratio of expression level at 0–6 h

Name (sequence)	Fold change	Target gene	Target gene description
osa-miR9898	0.07	Os01g0105800	Similar to Iron sulfur assembly protein 1
(UGAGGGCUAGAGCCUGCCUCUGG)		Os03g0324300	Similar to cDNA clone:J033123C05, full insert sequence
Locus: chr04:19420063-19420213		Os09g0555500	Similar to Chloroplast phytoene synthase 3

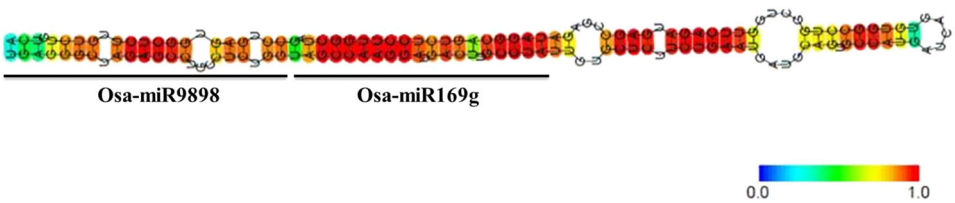

Fig. 4 Hairpin structure of novel un-annotated drought-responsive miRNA hairpin structure. *Color* represents the folding intensity between match pairs

of knocking out or amplifying the candidate transcripts in the rice plant and given the large number of our candidates, it is beyond the scope of our present study.

To help establish a working hypothesis of these novel regulatory elements in silico, we looked at the family functions of these novel regulatory transcripts by the way of computation and database searches. Since our number of candidate regulatory elements was too small for a GO enrichment-based analysis, we expanded our analysis to also include regulatory elements and their target genes that were inferred computationally using large gene expression data sets and coupled with prior knowledge of known regulatory mechanisms. From this analysis, our 18 novel TFs were involved in pathways for protein amino acid phosphorylation, response to stress, regulation of transcription, response to freezing, suppressing proteolysis, controlling metabolic processes, oxidation–reduction, and regulation of transcription. Our one miRNAs was also candidates for regulating genes involved in metabolic process and photosynthesis.

Table 4 Drought-responsive novel long noncoding RNAs. Fold change is the ratio of expression level at 0–6 h

No.	Name	Coordinate	Length (nt)	Fold change
1	OslncR001	chr01:6258403-6258719	316	34.57
2	OslncR002	chr04:517678-520999	3321	27.49
3	OslncR003	chr09:12748787-12749549	762	19.94
4	OslncR004	chr06:2397329-2397623	294	19.44
5	OslncR005	chr03:29153589-29153884	295	14.06
6	OslncR006	chr05:23013356-23013588	232	13.50
7	OslncR007	chr03:18087324-18087531	207	12.06
8	OslncR008	chr01:35205268-35205654	386	11.49
9	OslncR009	chr11:3926165-3926472	307	9.33
10	OslncR010	chr04:4649935-4650140	205	4.82
11	OslncR011	chr03:4249729-4250038	309	5.69

Table 5 Enriched functions of target gene module of activating/suppressing TF

Type	GO ID	GO description	Gene number	Proportion (%)	*p*-value
Target gene module of activating TF	GO:0006350	Transcription	39	4	0
	GO:0006355	Regulation of transcription, DNA-dependent	37	4	0
	GO:0006468	Protein amino acid phosphorylation	47	5	0
	GO:0006950	Response to stress	22	2	0
	GO:0045449	Regulation of transcription	59	6	0
	GO:0042309	Homoiothermy	33	3	0.012
	GO:0050826	Response to freezing	33	3	0.012
Target gene module of suppressing TF	GO:0006508	Proteolysis	11	3	0.0001
	GO:0008152	Metabolic process	30	9	0.0006
	GO:0055114	Oxidation reduction	17	5	0.012
	GO:0045449	Regulation of transcription	18	5	0.0199

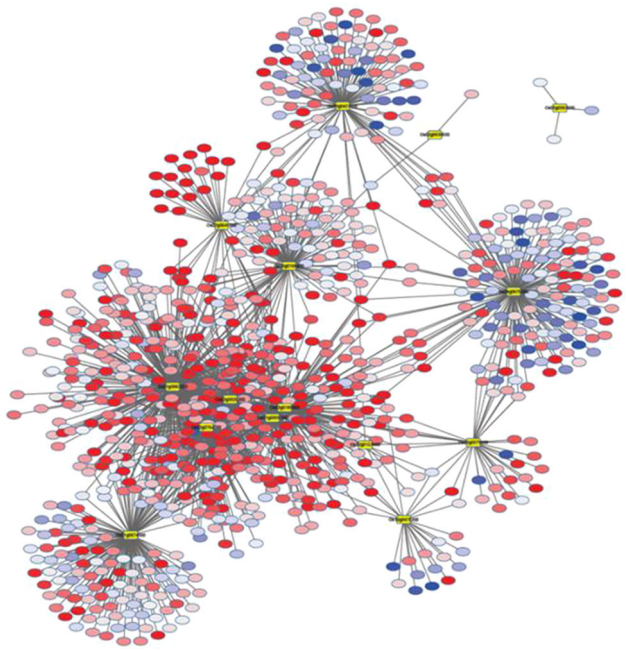

Fig. 5 Gene regulatory networks of novel drought-responsive TF and miRNAs. TF-target gene regulatory network of 18 functionally unknown drought-responsive TFs

From whole genome deep sequencing, it is known that lncRNAs are from 5′, 3′UTR, introns and intergenic regions (Nie et al. 2012). The candidates lncRNAs from our drought-response screen were all larger that 200 bp and coded for only long intergenic lncRNAs (lincRNAs). These intergenic noncoding RNA were between 200 and 500 bp long, as we also included very long intergenic lncRNAs (vlincRNA).

In summary, our report provides a model for study of adaption pathways activated with environmental stress and also identifying candidate genes or transcripts that may be important in developing abiotic and other biotic stress resistant crop plants by genetic modification (GM crops).

Acknowledgments This work was supported by grants from the Next Generation BioGreen 21 Program (No. PJ01121101) funded by the Rural Development Administration, Korea.

Funding This study was funded by Grants from the Next Generation BioGreen 21 Program (No. PJ01121101) funded by the Rural Development Administration, Korea.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63:3523–3543

Bailey-Serres J, Fukao T, Gibbs DJ, Holdsworth MJ, Lee SC, Licausi F, Perata P, Voesenek LA, Dongen JT (2012) Making sense of low oxygen sensing. Trends Plant Sci 17:129–138

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M (2009) Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Res 19:57–69

Cao P, Jung KH, Choi D, Hwang D, Zhu J, Ronald PC (2012) The rice oligonucleotide array database: an atlas of rice gene expression. Rice 5:17

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R et al (2005) The Transcriptional landscape of the mammalian genome. Science 309:1559–1563

Chen X (2005) microRNA biogenesis and function in plants. FEBS Lett 579:5923–5931

Covarrubias AA, Reyes JL (2010) Post-trascriptional gene regulation of salinity and drought responses by plant microRNAs. Plant Cell Environ 33:481–489

Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. BMC Plant Biol 11:163

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39(suppl 2):W155–W159

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. Nucleic Acids Res 31:439–441

Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 313:76–79

Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 456:130–134

Howell SH (2013) Endoplasmic reticulum stress responses in plants. Annu Rev Plant Biol 64:477–499

Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, Xiing L (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proc Natl Acad Sci USA 103:12987–12992

Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol Plant Mol Biol 47:377–403

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264

Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y, Kim M, Reuzeau C, Kim JK (2010) Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. Plant Physiol 153:185–197

Jiang P, Zhou N, Chen X, Zhao X, Li D, Wang F, Bi L, Zhang D (2015) Integrative analysis of differentially expressed micro-RNAs of pulmonary alveolar macrophages from piglets during H1N1 swine influenza A virus infection. Sci rep 5:8167

Jin J, Liu J, Wang H, Wong L, Chua NH (2013) PLncDB: plant long non-coding RNA database. Bioinformatics 29:1068–1071

Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res 42((Database issue)2):D1182–D1187

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol 57:19–53

Jouannet V, Crespi M (2011) Long nonprotein-coding RNAs in plants. Prog Mol Subcell Biol 51:179–200

Kim J, Jung JH, Reyes JL, Kim YS, Kim SY, Chung KS, Kim JA, Lee M, Lee Y, Narry Kim V, Chua NH, Park CM (2005) microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. Plant J 42:84–94

Kohli D, Joshi G, Deokar AA, Bhardwaj AR, Agarwal M, Katiyar-Agarwal S, Srinivasan R, Jain PK (2014) Identification and characterization of Wilt and salt stress-responsive microRNAs in chickpea through high-throughput sequencing. PLoS One 9:e108851

Kong L, Zhang L, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35:W345–W349

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73

Langmead B, TrapnellnC Pop M, Salzberq SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Liu HH, Tian X, Li YJ, Wu CA, Zheng CC (2008) Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. RNA 14:836–843

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. Plant Cell 24:4333–4345

Lorenz R, Bernhart SH, Höner Zu , Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6(1):26

Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. J Exp Bot 63:599–616

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 29:499–509

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666–670

Matsui A, Nguyen AH, Nakaminami K, Seki M (2013) *Arabidopsis* non-coding RNA regulation in abiotic stress responses. Int J Mol Sci 14:22642–22654

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poething RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu J (2008) Criteria for annotation of plant MicroRNAs. Plant Cell 20:3186–3190

Murchison EP, Hannon GJ (2004) miRNAs on the move: miRNA biogenesis and the RNAi machinery. Curr Opin Cell Biol 16:223–229

Nguyen VT, Kiss T, Michels AA, Bensaude O (2001) 7SK Small nuclear RNA binds to and inhibits the activity of CDK9/Cyclin T complexes. Nature 414:322–325

Nie L, Wu HJ, Hsu JM, Chang SS, Labaff AM, Li CW, Wang Y, Hsu JL, Hung MC (2012) Long non-coding RNAs: versatile master

regulators of gene expression and crucial players in cancer. Am J Transl Res 4:127–150

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629–641

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell 110:513–520

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323

Sadhukhan A, Panda SK, Sahoo L (2014) The cowpea RING ubiquitin ligase VuDRIP interacts with transcription factor VuDREB2A for regulating abiotic stress responses. Plant Physiol Biochem 83:51–56

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to cytoscape plug-ins. Nat Methods 9:1069–1076

Silveira RD, Abreu FR, Mamidi S, McClean PE, Vianello RP, Lanna AC, Carneiro NP, Brondani C (2015) Expression of drought tolerance genes in tropical upland rice cultivars (*Oryza sativa*). Genet Mol Res 14:8181–8200

Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25:1105–1111

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329:689–693

Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26:136–138

Wang D, Pan Y, Zhao X, Zhu L, Fu B, Li Z (2011) Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. BMC Genom 12:149

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 309:1570–1573

Yi N, Kim YS, Jeong MH, Oh SJ, Jeong JS, Park SH, Jung H, Choi YD, Kim JK (2010) Functional analysis of six drought-inducible promoters in transgenic rice plants throughout all stages of plant growth. Planta 232:743–754

Yoshida S, Forno DA, Cock JH, Gomez KA (1976) Laboratory manual for physiological studies of Rice. Manila, IRRI

Zhang W, Gao S, Zhou X, Xia J, Chellappan P, Zhou X, Zhang X, Jin H (2010) Multiple distinct small RNAs originate from the same microRNA precursors. Genome Biol 11:R81

Zhang J, Mujahid H, Hou Y, Nallamilli BR, Peng Z (2013) Plant long ncRNAs: a new frontier for gene regulatory control. Am J Plant Sci 4:1038–1045

Zhang W, Han Z, Guo O, Liu Y, Zheng Y, Wu F, Jin W (2014) Identification of maize long non-coding RNAs responsive to drought stress. PLOS One 9:e98958