INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
19–0634/2019/22–6–1524–1532
DOI: 10.17957/IJAB/15.1230
http://www.fspublishers.org

Full Length Article

Identification of Drought-Responsive Regulatory Genes by Hierarchical Selection of Expressed Sequence Tags and their Expression under Drought Stress in Rice

Rizky Dwi Satrio[1], Miftahul Huda Fendiyanto[1], Ence Darmo Jaya Supena[2], Suharsono[2] and Miftahudin[2*]
[1]*Plant Biology Graduate Program, Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University (IPB University), Kampus IPB Darmaga, Bogor 16680, Indonesia*
[2]*Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University (IPB University), Kampus IPB Darmaga, Bogor 16680, Indonesia*
[*]For correspondence: miftahudin@apps.ipb.ac.id; miftahudinm@gmail.com

Abstract

The key components of drought-stress signaling that are responsible for drought tolerance in rice have not been identified yet. Although many expressed sequence tags (ESTs) of rice grown under drought stress have been identified and deposited in the GenBank, they have not been optimally used for functional analyzes. This research was thus established to analyze an EST dataset in order to understand the global gene expression profile induced by drought stress in rice, to discover novel drought-tolerance genes using *in silico* study, and to confirm them through qRT-PCR analysis. A dataset of drought-response-related ESTs in rice was downloaded from the NCBI, preprocessed and grouped into clusters for functional analysis to obtain stress-responsive contigs. These contigs were then hierarchically selected to identify those related to transcription factors. The selected contigs were then incorporated into the drought-tolerant QTL on rice physical maps. The gene expression of each selected locus was evaluated in rice cv. Hawara Bunar and IR64 under drought stress. Gene Ontology analysis showed that the contigs consisted of genes encoding proteins involved in metabolic and cellular processes, related to catalytic and binding function, and located in the nucleus, cell membrane, and endomembrane organelles. Ten candidate genes potentially involved in regulating tolerance of drought stress were identified. They were located on rice chromosomes 1, 2, 3, 4, 6 and 9, and colocalized with the QTL for the drought-tolerance trait. Based on *in silico* and qRT-PCR analyzes, *Os03g0160100*, *Os03g0815100*, *Os06g0601000* and *Os09g0553100* were considered to be candidates for uncharacterized drought-responsive regulatory genes that potentially being used for manipulating drought tolerance in rice.

Keywords: Transcriptome; *In silico*; Expressed sequence tag (EST); qRT-PCR; Rice; Drought

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple food crops globally and consumed by almost half of the world's population. The growth and productivity of rice are strongly influenced by the availability of water, which is a highly limited resource for rice production under drought conditions. Water deficit leads to the loss of rice production worldwide. Given the background of a decreasing supply of water for rice cultivation throughout the world, there is an increasingly urgent need to improve the adaptability of rice to drought stress (Pandey and Shukla, 2015).

The complexity of the mechanism of drought tolerance in rice has made it difficult to characterize drought-tolerance traits that need to be understood to enhance the performance of the drought stress response. This has impeded the improvement of drought-tolerant varieties of rice. Several phenotypic traits, which are the combination of various morphological, biochemical, and genetic characters as well as molecular factors, make quantitative contributions to drought tolerance. As the most basic function, the molecular mechanism of drought tolerance in rice is a challenging project to elucidate (Lenka *et al.*, 2010).

Drought stress can induce the expression of a considerable number of genes, initiating various responses that activate adaptive programs or genes related to the downstream response (Farooq *et al.*, 2009a, b). Drought-responsive genes encode proteins that contribute to a variety of biological processes, molecular functions, or cellular components (Zhang *et al.*, 2016). These can be grouped into two classes: those that play a direct role in protecting the plant against drought stress and those that control the expression of genes or activate signal transduction in response to drought stress (Shanker *et al.*, 2014; Silveira *et al.*, 2015). Several drought-induced genes encode transcription factors (TFs) that are essential for

To cite this paper: Satrio, R.D., M.H. Fendiyanto, E.D.J. Supena, Suharsono and Miftahudin, 2019. Identification of drought-responsive regulatory genes by hierarchical selection of expressed sequence tags and their expression under drought stress in rice. *Intl. J. Agric. Biol.* 22: 1524–1532

regulating drought tolerance. Despite the fact that many genes that respond to drought stress have been identified, no studies have yet identified the major genes or key drought stress signaling factors that play essential roles for drought tolerance in rice (Farooq *et al.,* 2009b; Silveira *et al*., 2015; Shankar *et al*., 2016; Borah *et al*., 2017).

The global analysis of gene expression, also known as transcriptome analysis, has been broadly applied using various technologies, such as cDNA libraries, microarrays and RNA-seq, which are based on Sanger sequencing, hybridization and high-throughput sequencing, respectively (Wang *et al*., 2009). Transcriptome studies in rice have significantly improved our ability to discover and explore novel genes that play key roles in drought tolerance and increased our ability to focus on previously identified genes that might be related to specific molecular responses induced by drought stress (Silveira *et al*., 2015).

Single-transcript studies were first performed decades ago, before omics technologies had been established. In the 1980s, Sanger sequencing, which produces low-throughput data, began to be used to sequence random transcripts from cDNA libraries termed expressed sequence tags (ESTs). Sequences of ESTs represent the products of gene expression at the transcriptional level, which are the objects of interest in transcriptomic study. During the 1990s, it became popular to use ESTs as an efficient technique to study responses at the molecular level without sequencing the whole genome of an organism. Exponential growth in the ESTs data deposited in public databases then occurred in the 2000s, enabling the discovery of genes or alternative transcripts, aiding gene structure or functional analysis, complementing genome annotations, and facilitating proteome study (Lowe *et al.*, 2017).

ESTs are short sequences generated by single-pass sequencing of cDNA libraries picked at random. They should be processed and annotated to provide biologically relevant information. Raw ESTs contain low-quality sequences and vector regions that have to be removed to obtain a qualifiable sequence for further analyzes. Considering the random selection of cDNA clones from the library, a clustering step is required to obtain the unique consensus sequences or unigenes that are not redundant. Finally, annotation of the structural or functional unigenes is required to add biologically relevant information to the sequences (Huang *et al*., 2017). ESTs can now potentially be used to obtain a more comprehensive understanding of global gene expression because of the increase in the number of data sequences deposited in GenBank and development of more accurate and efficient software for transcriptomic data processing. They can also enable information to be obtained about novel genes that play key roles in the molecular mechanism of drought tolerance in plants.

Here, we processed an EST dataset from the database to study rice responses to drought stress at the transcriptomic level. There is an abundance of EST sequences of rice grown under drought stress deposited in the NCBI, which enables studies on various genotypes, treatment levels, specific tissues, and developmental stages. The objectives of the present study were to understand the global gene expression profile induced by drought stress, to discover novel drought-response-related genes through *in silico* study and to confirm the validity of the selected gene candidates using expression analysis in rice.

Materials and Methods

EST Data Resources and Preprocessing

The drought stress-related ESTs in rice were downloaded from NCBI using the keywords "(rice [Organism]) AND drought." They were preprocessed to remove parts of the sequence that might cause incorrect matching in the assembly step using SeqClean software. A plasmid dataset from UniVec (ftp://ftp.ncbi.nih.gov/pub/UniVec/) NCBI and the *Escherichia coli* genome (NC_000913.2) were used to screen out contaminants from the ESTs. The preprocessed ESTs were then clustered based on their sequence similarity. Subsequently, the software programs PHRAP and CAP3 were used, and their assembly performances were compared to obtain the best contigs. Only contigs produced from the better software were used for further analyzes.

Locus Identification

The contigs were mapped into the Nipponbare IRGSP 1.0 rice genome (Sakai *et al.*, 2013) to determine their positions along all 12 chromosomes using ReadsMap software. They were also aligned into rice full-length cDNA (RAP-DB) to identify the physical locations of loci using local BLASTN software with maximum E-value of less than 1e−5. The loci were determined based on the Top-Hit BLASTN results using rice cDNA database were used as gene loci.

Functional Analysis of EST Contigs

Functional analysis of the contigs produced from the clustering step was performed using Blast2GO v5.021 (Conesa and Gotz, 2007). Contigs containing homologous sequences with annotated proteins in the NCBI nr database were used for functional characterization based on a maximum E-value of less than 1e−3 using BLASTX. Sequences were also imported into InterProScan. CDD, HAMAP, HMMPanther, HMMPfam, HMMPir, FPrintScan, BlastProDom, Gene3D, SFLD, SuperFamily, Coils, MobiDBLite, Phobius, SignalPHMM and TMHMM were used as the databases. The contig sequences were translated into six reading frames and then aligned to the dataset of selected databases. According to the mapping of the corresponding Top-Hit of BLASTX and InterProScan results to Gene Ontology (GO), the GO accession numbers were assigned to the contigs for the classification of functional groups.

Determination of Drought-responsive Regulatory Gene Candidates

Three selection steps were performed to determine the candidates for drought-responsive genes in rice. First, GO analysis was narrowed on the subhierarchy of the biological process term. The predicted contigs involved in the stress response were selected. Nucleotides of the selected contigs in fasta format were extracted for further analyzes (Ramalingam *et al.*, 2006). Second, stress-responsive contigs were aligned to the sequences of rice genes involved in the transcriptional control in PlantTFDB4.0 (Jin *et al*., 2017), PlnTFDB (Pérez-Rodríguez *et al.*, 2010) and GRASSIUS (Mukundi *et al.*, 2017) using local BLASTX software. Third, the drought-responsive contigs that putatively encode TFs proteins were incorporated into the drought stress-tolerant QTL in a rice database. The QTARO (Yonemaru *et al.*, 2010) database was explored to collect data on the physical positions of QTL controlling traits related to drought stress tolerance in rice. The physical positions of the selected contigs and the collected QTL were mapped along the 12 rice chromosomes using the ggplot2 package in R software.

Plant Growth and Stress Treatments

Seeds of IR64 and Hawara Bunar (HB), drought-sensitive and -tolerant rice cultivars, respectively, were germinated and grown in Yoshida culture medium (Yoshida *et al.*, 1976) for 14 days. To determine the relative expression levels of the genes of interest under water-limiting stress, plants at the four-leaf stage were removed from the medium and air-dried for 3 h, or were grown under low-osmotic conditions using 20% polyethylene glycol (PEG) 6000 for 24 h. The shoots and roots were collected separately after drought stress treatment for the isolation of RNA. Each stress treatment was repeated as biological triplicates, giving a total of 36 experimental units in this study.

RNA Isolation and cDNA Synthesis

Total RNA isolation was performed in biological triplicates from the shoot or root tissues from 14-day-old rice seedlings using TRIzol reagent (Invitrogen, California, USA). The quantity of RNA was measured using GenQuant (Bio-Rad, California, USA) and the quality was visualized by agarose gel electrophoresis. The genomic DNA was removed by treatment with DNAse I (Thermo Scientific, Massachusetts, USA). Each sample was reverse-transcribed into cDNA using RevertAid First-strand cDNA Synthesis Kit (Thermo Scientific, Massachusetts, USA).

Primer Design and qRT-PCR Analysis

The ten selected drought-responsive contigs derived from previous EST analysis were used in the primer design and qRT-PCR analysis. The primers for qRT-PCR analysis were designed using the Primer3 program with the following parameters: melting temperature (Tm), 57–63°C; GC content, 40%–60% and amplicon size, 100–250 bp (Table 1). The actin transcript was used as an internal control for expression analysis. Each reaction was performed in technical triplicates in the StepOnePlus Real-Time PCR System (Applied Biosystems, California, USA) using DyNAmo Flash SYBR Green qPCR Kit (Thermo Scientific, Massachusetts, USA). The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression levels of the selected genes (Livak and Schmittgen, 2001).

Results

Expressed Sequence Tag Processing

A total of 12,129 EST sequences related to drought and rice were curated from the NCBI database and preprocessed in order to generate 11,847 (97.67%) clean ESTs with an average length of 514 bp. Using PHRAP software, the high-quality ESTs were clustered into 6,504 unigenes, including 1,680 (25.83%) contigs and 4,824 (74.26%) singletons. Moreover, using CAP3 software, the high-quality ESTs were also clustered into 7,734 unigenes, including 1,298 (16.76%) contigs and 6,436 (83.12) singletons. Although both of these software programs produced more than a thousand contigs, those produced by PHRAP were better than those produced by CAP3. Specifically, the contigs from PHRAP were more numerous and had longer average lengths, longer maximum lengths and fewer singletons than the CAP3 contigs (Fig. 1). Based on the assembly and clustering results, only PHRAP contigs were used for subsequent analyzes. Singletons produced from the assembler, which are potentially products of low-quality sequences, were not used for further analyzes in this study.

Global Expression Profile of Drought-Responsive Genes in Rice

Locus identification and mapping of contigs derived from ESTs into the rice chromosomes. A BLASTN search using rice Nippponbare cDNA sequences revealed that the contigs generated from the ESTs were mapped onto all 12 of the rice chromosomes at both plus and minus strands. For the EST sequences in chromosomes 2, 3, 4, 9 and 11, the distribution was predominantly in the plus strand, while those in chromosomes 1, 5, 6, 7, 8, 10 and 12 were predominantly in the minus strand. The number of drought-responsive sequences in the chromosomes was not always positively correlated with the chromosome length. In terms of the abundance of sequences, the greatest number were located in chromosome 3, despite this being shorter than chromosomes 1 and 2. Most of the contigs were located on these three chromosomes, with a total of 207 (13.15%), 195 (12.38%) and 210 (13.34%) sequences on chromosomes 1–3, respectively (Fig. 2).

Table 1: List of primers used for qRT-PCR analysis

Locus ID	Primer Forward	Primer Reverse	Fragment Length (bp)
Os03g0128700	GAGTGGAGTGCCACCATTTT	CCTTTGCACTGTCAGATATC	190
Os03g0160100	TCCTTCACCGTCCTAATTGC	CAGGTGCCATCCACTCCGGA	237
Os03g0815100	AGAAGGCGCTCGTGTTCTAC	TTCACCTCCTTCCCCTGCTG	115
Os04g0584600	AAACTGGAGCGAGAGGAACA	CCGACTCCCATATTGCCCTT	248
Os06g0601000	TGAGGATGATGACGATGGAA	GCCAGTTTTTCAGCTCTGGC	201
Os06g0622700	TCTAGAGGCCGAGTGTCGTC	TGGGTAGACCGGGCACCGGG	220
Os09g0323700	CGTGTACAAGACGTGGGTGT	AAGAGGTCGGAGTTGGCGAG	181
Os09g0553100	AGAGGCACAGGAAGGTGCT	TGACGGTCTTGCGGCGGGC	208
Os01g0867300	GAGATGACGCTGGAGGACTT	ATCGCGGCGCGGTCCATCGG	206
Os02g0218550	CAGCACCACCGTCTCCAC	CCAACCTCGTCCCCGCTGGA	197
OsAct (actin)	GAAGGATGCCTATGTTGGTGA	GAAAGTCTCAAACATGATCT	235

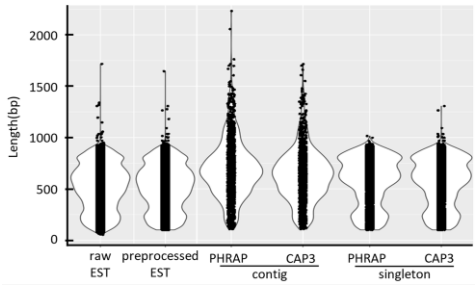

Fig. 1: Distribution of expressed sequence tags, contigs, and singleton lengths. The EST raw data related to drought stress in rice were downloaded from NCBI, consisting of 12,129 sequences. Preprocessing was performed using SeqClean software to generate 11,847 clean preprocessed ESTs. Clustering and assembly were performed using PHRAP and CAP3 software to generate contigs from preprocessed ESTs and unassembled ESTs were categorized as singletons

Functional Analysis of Global Transcripts

A BLASTX search revealed that there were 1,438 (85.59%) contigs with high similarity to proteins in the NCBI nr database. Of these, 1,347 (93.67%) showed high similarity to proteins with a recognized function, while only 90 (6.25%) showed similarity to deduced proteins of unknown function. Species distribution analysis showed that the best BLASTX hits of 967 (67.24%), 522 (36.30%) and 60 (4.17%) of the EST sequences were with the *O. sativa* japonica group, *O. sativa* indica group and *O. sativa* database, respectively, while 264 (18.35%) of the annotated sequences showed similarity to EST sequences from other plant species.

An InterProScan search revealed that there were 981 (58.39%) and 465 (27.67%) contigs distributed in 586 families and 310 domains, respectively. The "P-loop containing nucleoside triphosphate hydrolase", "protein kinase-like domain superfamily," "glycosidase hydrolase superfamily", "NAD(P)-binding domain superfamily," and "leucine-rich repeat domain superfamily" were the most abundant families, containing 29, 24, 11, 11 and 11 contigs, respectively. In terms of the domains, the most abundant were "protein kinase domain", "EF-hand domain", "bifunctional inhibitor/plant lipid transfer protein/seed storage helical protein domain", "serine-threonine/tyrosine-protein kinase/catalytic domain", and "RNA recognition motif domain", containing 16, 8, 8, 8 and 7 contigs, respectively.

The EST accessions produced from BLASTX and InterProScan analyzes were used to redeem symbols or gene names utilizing mapping files provided by the NCBI and EBI. The identified gene names were then searched in the species-specific entries of the gene-product list of the GO database for GO Mapping. GO annotation was then performed by applying an annotation rule to the retrieved ontology terms. This rule searches for the most specific annotations based on a particular reliability threshold using BLAST2GO software (Conesa and Gotz, 2007).

A total of 1,298 (77.26%) annotated contigs belonged to one or more ontological categories. Of these, 939 (55.98%), 978 (58.21%) and 923 (54.94%) sequences were assigned to the GO categories of biological process (BP), molecular function (MF) and cellular component (CC), respectively. These three categories were grouped predominantly into four, two and eight level I subgroups, respectively (Fig. 3–5).

GO of the category BP is a valuable tool for the functional analysis of gene datasets. The BP categories were here grouped into four subgroups, namely, metabolic process, cellular process, regulation of gene expression and response to stress terms, consisting of 737 (78.49%), 744 (79.23%), 100 (10.65%) and 140 (14.91%) sequences, respectively. Twelve processes belonged to the metabolic process subgroup, whereas three processes belonged to the cellular process subgroup (Fig. 3).

The assignment of GO terms within the MF category revealed representatives within each of several categories, with the highest numbers of annotations occurring in two subgroups, namely, binding activity and catalytic activity, consisting of 625 (63.97%) and 600 (61.41%) sequences, respectively. Seven MF terms were found to belong to the binding activity subgroup, while only three MF terms belonged to the catalytic activity subgroup.

Within the CC category, cell (80.82%) was the most strongly represented subgroup, followed by intracellular (72.91%), cytoplasm (55.04%), intracellular membrane-bound organelle (51.25%), membrane (47.24%) and cytoplasmic part (44.53%) terms. Most of the sequences were localized in the cell, suggesting that they may contribute to cytosol, membrane, nucleus and plastid components or functions.

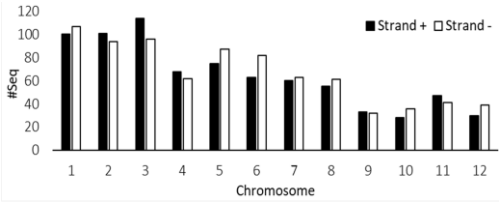

Fig. 2: Contigs generated from ESTs were mapped in the rice genome. Contigs were mapped in the Nipponbare IRGSP 1.0 genome using ReadsMaps software. The contigs were spread over all 12 of the rice chromosomes in both plus and minus strands

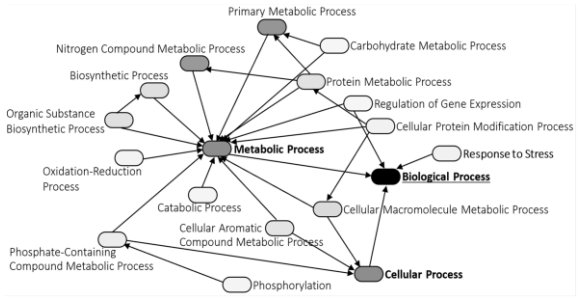

Fig. 3: Global transcript expression Gene Ontology of drought-responsive genes related to the biological process (BP) category. Metabolic and cellular processes dominated within the BP category related to drought response. The response to stress term was an outlier of both metabolic and cellular processes, which may act as a regulator of both processes. Darker colors indicate higher numbers of sequences

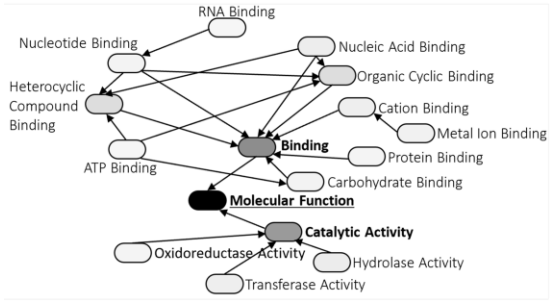

Fig. 4: Global transcript expression of drought-responsive genes related to the molecular function (MF) term in Gene Ontology. Catalytic activity and binding were the dominant terms in the category of molecular function related to drought induction response. Catalytic activities were related to enzymatic functions while binding activities were related to regulatory processes. Darker colors indicate higher numbers of sequences

Identification of Drought-Responsive Regulatory Candidate Genes

The analysis of drought-responsive regulatory candidate genes was performed by gradual selection, starting with the extraction of contigs in a functional analysis step, then selection of the contigs exhibiting similarity to TF sequences, followed by their incorporation into QTL maps

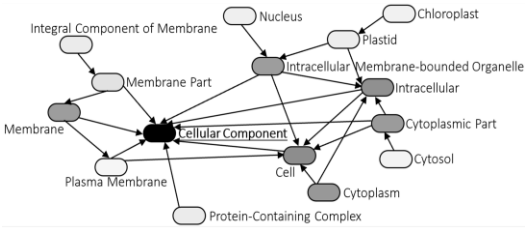

Fig. 5: Global transcript expression of drought-responsive genes related to the cellular component (CC) term in Gene Ontology. Darker colors indicate higher numbers of sequences

for drought-tolerant traits in rice. Response to stress term, which consisted of 140 stress-responsive contigs were then hierarchically selected from the BP term. A BLASTX search using the rice plant TF database was also performed to select drought-responsive transcription factors. Upon the identification of rice TFs using the TFDB 4.0, PlnTFDB, and GRASSIUS TF databases, 26 contigs with the best matches were predicted, with E-values of less than 1e−5. The TF families identified were bHLH, bZIP, WRKY, LSD, NAC and MYB.

The roles of the 26 selected genes were then determined by incorporating them into the drought-tolerant QTL on physical maps. Based on the QTARO database, the QTL for drought-tolerant traits in rice were distributed along all 12 of the rice chromosomes. However, only ten selected genes that were determined from the previous step were incorporated into the QTL for drought-tolerant traits in rice (Table 2). Fig. 6 shows the distribution of the selected contigs along the 12 rice chromosomes.

To confirm the validity of the drought-responsive regulatory candidate genes, ten primer pairs were generated to quantify the gene expression using qRT-PCR (Table 1). The results showed that the expression levels of all ten genes were increased or decreased in response to drought stress depending on the genetic background, specific tissue, or type of treatment (Fig. 7).

Two clusters of genes were obtained based on the expression patterns of the selected genes. The first cluster consisting of *Os09g0323700, Os02g0218550* and *Os03g0128700* was predominantly upregulated in the root or shoot tissue of the IR64 cultivar under both air-drying and PEG treatments. The second cluster consisting of *Os04g0584600, Os01g0867300, Os03g0815100, Os03g0160100, Os06g0622700, Os06g0601000* and *Os09g0553100* was predominantly upregulated in the root or shoot tissue of the HB cultivar under air-drying and PEG treatments. This second cluster showed a better pattern and was recommended to undergo further characterization, considering that the genes were relatively upregulated in the rice cv. HB, which is a drought-tolerant variety. However, the genes in the second cluster exhibited different patterns of expression between root and shoot tissues in the HB cultivar under PEG treatment. *Os06g0601000* and

Table 2: Drought-responsive regulatory candidate genes with co-localized QTLs in the corresponding rice chromosomes. The highlighted loci (*Os03g0160100*, *Os03g0815100*, *Os06g0601000*, *Os09g0553100*) were upregulated under drought stress conditions in the rice cv. Hawara Bunar, which is a drought-tolerant variety, as revealed by qRT-PCR

No.	Contig	Locus	Strand	Chr	TF description	Traits for QTL	References for QTL
1	Contig519	*Os01g0867300*	-	1	bZIP family	plant height no stress	Hemamalini *et al.* (2000)
2	Contig668	*Os02g0218550*	-	2	OsIBCD020180	spikelet fertility	Lafitte *et al.* (2004)
3	Contig1142	*Os03g0128700*	+	3	OsIBCD019262	days to heading, days to 50% flowering	Moncada *et al.* (2001); Bernier *et al.* (2007)
4	Contig829	***Os03g0160100***	+	3	OsIBCD020180	days to flowering, root volume and weight low moisture, days to 50% flowering, days to flowering	Venuprasad *et al.* (2002); Bernier *et al.* (2007); Lafitte *et al.* (2002)
5	Contig1400	***Os03g0815100***	+	3	NAC family	panicle length	Nguyen *et al.* (2004)
6	Contig1218	*Os04g0584600*	+	4	OsIBCD019262	root fresh weight	Qu *et al.* (2008)
7	Contig827	***Os06g0601000***	+	6	HB-other family	penetrated root thickness	Moncada *et al.* (2001)
8	Contig101	*Os06g0622700*	+	6	bZIP family	penetrated root thickness	Moncada *et al.* (2001)
9	Contig173	*Os09g0323700*	+	9	MYB_related family	relative water content	Price *et al.* (2002)
10	Contig1360	***Os09g0553100***	-	9	bHLH family	relative water content, days to heading	Babu *et al.* (2003)

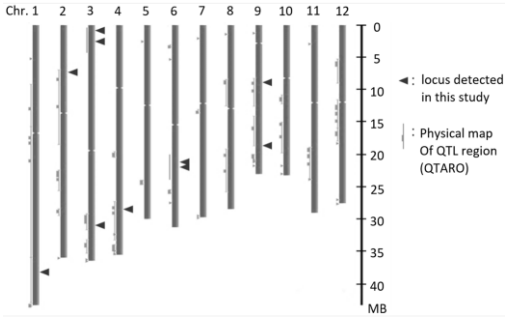

Fig. 6: Candidates for drought-responsive transcription factors were incorporated into the drought-tolerant QTL in rice. Ten candidate genes potentially involved in the regulation of drought stress tolerance were found in rice chromosomes 1, 2, 3, 4, 6 and 9

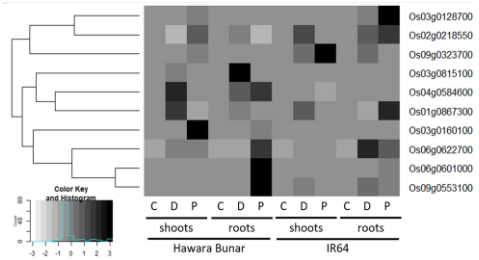

Fig. 7: Expression patterns of ten drought-responsive candidate genes in rice were measured using qRT-PCR. All of the genes were upregulated in at least one experiment under drought stress treatments. C: control, D: air-drying treatment for 3 h, P: 20% PEG6000 treatment for 24 h

Os09g0553100 were highly upregulated in the root tissue of HB, while *Os03g0160100* and *Os03g0815100* were highly upregulated in the shoot tissue of HB under PEG treatment and its root tissue under air-drying treatment, respectively.

Discussion

Drought-responsive ESTs in rice were retrieved from the NCBI, preprocessed to obtain high-quality ESTs and clustered using PHRAP or CAP3 software. We used these two software tools because of the different procedures that they employ. Specifically, PHRAP is an assembler that implements the comparison, alignment and assembly of a group of DNA sequences. The software compares a set of sequences by matching the regions for which reliable qualification is achieved. It then attempts to expand the alignment into the contigs if a match has been found. In contrast, CAP3 performs three different tasks: detection of fragment overlap, formation of the contig and generation of the consensus sequence. First, the overlaps between each pair of input fragments are determined. Second, the contigs are assembled by adding the best overlapping fragments one at a time. Third, the consensus sequences are generated from the contigs, merging each pairwise alignment into multiple alignments (Minetti *et al*., 2012). In the present study, better contig quality was achieved using the PHRAP software, in terms of the contig number and maximum and average lengths of the sequence. In subsequent analyzes, we only used contigs produced from the PHRAP software.

The contigs derived from ESTs were then subjected to GO-based functional analysis to determine their functional roles. The BP category included the subgroups of metabolic process, cellular process, regulation of gene expression and response to stress (Fig. 3). The metabolic and cellular processes included the largest numbers of EST sequences, indicating that a large number of metabolic processes occur in the drought-stressed tissue. Response to stress, which may reflect a regulatory process, was not a dominant subgroup within the BP category. On the other hand, the metabolic processes of primary and nitrogen compounds, which might be indirectly involved in the stress response, were abundant in the BP category. It was also indicated that only a small proportion of sequences directly involved in responses and regulatory processes are required as they can in turn affect many biological processes to establish the traits of stress tolerance.

The metabolic process terms of primary, carbohydrate, and nitrogen compound metabolism were strongly associated with carbohydrate partitioning, protein synthesis, and osmotic adjustment in rice under drought conditions. Sugar and proline are well-known as osmolytes that play roles in controlling cellular water potential in plants (Zhang *et al*., 2018). Oxidation-reduction processes were also

reported to be related to the mechanism of control of reactive oxygen species (ROS) detoxification in plant cells (Nakashima *et al.*, 2014; Arum *et al*., 2018).

In the MF category, only two subgroups were identified: binding activity and catalytic activity (Fig. 4). The sequences with binding function were strongly associated with regulatory processes. Carbohydrate, protein, cation, organic cyclic and heterocyclic binding functions can be possessed by receptor proteins or other proteins functioning in signaling cascades. Nucleotide binding or nucleic acid binding function found in this analysis is closely related to functioning as a TF or regulatory protein influencing gene expression. On the other hand, catalytic activities play a role as a functional protein or enzymatic activity. ROS scavenging is one example of a process involved in oxidoreductase activity as a mechanism of drought tolerance at the cellular level (Nakashima *et al*., 2014; Liu *et al*., 2019).

Within the CC category, the subgroups of cell, intracellular, cytoplasm, intracellular, membrane-bound organelle, membrane and cytoplasmic were identified (Fig. 5). The results revealed a greater diversity of proteins involved in membranes that act as receptors or signaling proteins rather than proteins functioning in the nucleus and potentially acting as transcriptional regulators of drought stress tolerance in rice.

In terms of the results previously obtained by GO analyses in other plant species treated under various stress conditions, some similarities and some differences were identified compared with the findings obtained here. In the case of chickpea under drought or salinity stress (Varshney *et al*., 2009; Deokar *et al*., 2011), *Arabidopsis pumila* under salt stress (Huang *et al*., 2017) and also in cotton (Lin *et al*., 2013), metabolic and cellular processes were also particularly identified within the BP category. In addition, in these previous studies, the response to stress term was only identified at a low rate. Moreover, in a library of ESTs constructed from a variety of tissues and organs under a range of conditions representing challenges typically encountered by common wheat, it was also found that the most abundant BP term was metabolic process, followed by transport, translation, redox, biosynthetic and other processes (Manickavelu *et al.*, 2012). In a different study in wheat, it was also found that cellular process was the most abundant term within the BP category, followed by the physiology and regulation of biological processes (Kawaura *et al.*, 2009). Furthermore, in a recent transcriptomic study in rice treated with aluminum stress, it was also found that cellular process was the most abundant term within the BP category (Zhang *et al*., 2019). As such, a general pattern has been identified showing that metabolic and cellular processes are almost always the most abundant BP terms identified in stress-treated tissue in plants.

The results for the MF category in this study are also identical to the results of previous studies. Specifically, binding and catalytic functions were the terms with the highest proportions in the MF category, which matches previous findings in chickpea, *A. pumila*, cotton and common wheat EST libraries generated from different tissues and stress conditions (Kawaura *et al*., 2009; Varshney *et al*., 2009; Deokar *et al*., 2011; Lin *et al*., 2013; Huang *et al*., 2017).

Similar proportions were also found in ontology terms of CC from the same studies. This finding of a similarity of these phenomena indicated that water stress induced the same pattern of transcriptome dynamics across various species. However, differences of particular terms and their proportions among the low-abundance terms were identified, which may have been caused by differences in transcriptome dynamics or just a difference in the GO version that was used. This suggests the need to repeat functional analyses periodically, in view of the constant updating of GO through the release of new versions.

To analyze the contribution of each genomic region containing genes expressed in response to drought stress and conferring tolerance to it, we mapped contigs generated from the previous step onto the 12 rice chromosomes. The results showed that the contigs were mapped in all of the 12 rice chromosomes at both plus and minus strands. Previous studies aimed at revealing genes responsible for drought tolerance through genetic mapping approaches also showed similar results. Loci of QTL for drought-tolerant traits in rice deposited in the QTARO database (Yonemaru *et al*., 2010) showed that drought-tolerant QTL were also located in all 12 of the rice chromosomes (Fig. 6). The distributions of contigs and QTL on chromosomes are important information for discovering novel genes. In the present study, we used this information to narrow down a set of contigs to identify novel drought-responsive regulatory candidate genes.

Drought-responsive candidate genes were identified by a gradual selection approach to narrow down the contigs generated from the previous step. The first step involved extracting the 140 stress-responsive contigs associated with the response to stress term within the BP category. Second step involved selecting 26 contigs with high similarity to rice TFs. Third step involved incorporating selected contigs into the QTL for drought-tolerant traits in rice.

In the last few years, many genes suspected of being responsible for drought tolerance or those involved in the response to drought stress in rice have been reported and some excellent reviews of these studies have been published (Shinozaki and Yamaguchi-Shinozaki, 2007; Nakashima *et al.*, 2014). The TFs act as the main regulators of gene expression at the transcription level, playing an important role in stress responses. The most well-known model of the mechanism of drought tolerance at the molecular level involves two pathways related to the presence of abscisic acid (ABA). In this model, the response to drought stress in rice occurs through the ABA-dependent pathway or ABA-independent pathway. The ABA-dependent pathway involves the interaction of a cis-acting element, called the

ABA-responsive element (ABRE) and TFs, called ABRE-binding protein (AREB)/ABRE-binding factors (ABFs). The AREB/ABFs regulate the transcription of downstream genes known as the AREB/ABF regulon. In contrast, the ABA-independent pathway involves dehydration-responsive element binding protein (DREB), NAM, ATAF, and CUC as major regulons. Minor regulons, namely, MYB/MYC, WRKY, and nuclear factor-Y (NF-Y) TFs, are also known to be involved in drought responses at the molecular level (Joshi *et al*., 2016). Considering the complexity of the molecular pathways involving TFs, many of which may not yet have been revealed, we decided to select the EST contigs with high similarity to TF proteins. We also selected EST contigs based on their location in the published QTL associated with drought-tolerant traits in rice. These steps of gradual selection led to only ten contigs being obtained, which were found on rice chromosomes 1, 2, 3, 4, 6 and 9 (Table 2). These might correspond to drought-responsive genes that encode regulatory proteins.

These ten candidate genes were identified based on *in silico* analysis, so their expression needed to be confirmed in a biological system. For this, we here used two rice accessions as genetic backgrounds with different levels of drought tolerance: IR64 as a drought-sensitive accession and HB as a drought-tolerant one. We also used two different tissues, namely, root and shoot, for spatial study, and three different types of stress, namely, control, 3 h of air-drying, and 24 h of exposure to 20% PEG6000, for different treatment study. Expression analysis using qRT-PCR showed that all ten genes were upregulated by the drought stress in at least one experiment (Fig. 7). This validated the results obtained from the *In silico* EST analyzes.

Two main clusters showed different dominant expression patterns in the different genetic backgrounds. Three genes were dominantly expressed in IR64, while seven were in HB. The expression analysis revealed interesting patterns regarding the genes that were highly expressed in the drought-tolerant cultivar HB in both root and shoot under drought treatment. Specifically, the results revealed that the loci *Os06g0601000* and *Os09g0553100* were highly expressed in HB root tissue under 24 h of treatment with 20% PEG6000, *Os03g0815100* was highly expressed in HB root tissue under 3 h of treatment with air-drying and that *Os03g0160100* was highly expressed in HB shoot tissue under 24 h of treatment with 20% PEG6000. These were thus suggested to be genes responsible for drought tolerance.

Os06g0601000, *Os09g0553100* and *Os03g0815100* are genes that have not been characterized, but were predicted to encode a homeodomain-like containing protein, a bHLH-domain-containing protein and an NAC-domain-containing protein, respectively. The GO terms corresponding to these proteins in the BP category were "response to water deprivation", "regulation of abscisic acid-activated signaling pathway", "stomatal movement", and "regulation of jasmonic acid mediated signaling pathway", whereas that in the MF category was "DNA binding" and that in the CC category was "nucleus." We thus suggested that these proteins might act as transcriptional regulators and be involved in the mechanism of drought tolerance in rice. The proteins were slightly different with the *Os03g0160100* that has been well characterized as mitogen-activated protein kinase kinase kinase, disease resistance, or defense/stress response (Wang *et al*., 2015). Although the BLASTX analysis indicated that *Os03g0160100* is similar to TFs, this protein appears to act as a protein kinase. However, it still exerts a function related to the response to stress.

Conclusion

The genes *Os06g0601000*, *Os09g0553100*, and *Os03g0815100*, which were highly expressed in rice root, and *Os03g0160100*, which was highly expressed in rice shoot under drought treatment, were deduced to be responsible for drought tolerance and potentially encoding regulatory proteins. To the best of our knowledge, the first three genes mentioned above are identified here for the first time as potential contributors to drought tolerance in rice. Further characterization of these genes is needed to clearly understand their roles and utilize them to improve rice and establish drought-tolerant varieties.

Acknowledgments

This work was financially supported by the "*Program Magister menuju Doktor untuk Sarjana Unggul*" research grant from the Ministry of Research, Technology and Higher Education, Indonesia, F.Y. 2016 to Dr. Miftahudin.

References

Arum, L.S., A. Supriyadi, E. Novianti, W.I.D. Fanata and T.A. Siswoyo, 2018. Drought stress induced the expression level of ascorbate peroxidase in the late seedlings of melinjo (*Gnetum gnemon*). *Intl. J. Agric. Biol*., 20: 1303–1308

Babu, R.C., B.D. Nguyen, V. Chamarerk, P. Shanmugasundaram, P. Chezhian, P. Jeyaprakash, S.K. Ganesh, A. Palchamy, S. Sadasivam, S. Sarkarung, L.J. Wade and H.T. Nguyen, 2003. Genetic analysis of drought resistance in rice by molecular markers: association between secondary traits and field performance. *Crop Sci*., 43: 1457–1469

Bernier, J., A. Kumar, V. Ramaiah, D. Spaner and G. Atlin, 2007. A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Sci*., 47: 507–516

Borah, P., E. Sharma, A. Kaur, G. Chandel, T. Mohapatra, S. Kapoor and J.P. Khurana, 2017. Analysis of drought-responsive signaling network in two contrasting rice cultivars using transcriptome-based approach. *Sci. Rep*., 7: 1–21

Conesa, A. and S. Gotz, 2007. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Intl. J. Plant Genomics*, 6: 1–12

Deokar, A.A., V. Kondawar, P.K. Jain, S.M. Karuppayil, N.L. Raju, V. Vadez, R.K. Varshney and R. Srinivasan, 2011. Comparative analysis of expressed sequence tags (ESTs) between drought-tolerant and -susceptible genotypes of chickpea under terminal drought stress. *BMC Plant Biol*., 11: 1–20

Farooq, M., A. Wahid, N. Kobayashi, D. Fujita and S.M.A. Basra, 2009a. Plant drought stress: effects, mechanisms and management. *Agron. Sustain. Dev*., 29: 185–212.

Farooq, M., A. Wahid, O. Ito, D.J. Lee and K.H.M. Siddique, 2009b. Advances in drought resistance of rice. *Crit. Rev. Plant Sci*., 28: 199–217.

Hemamalini, G.S., H.E. Shashidhar and S. Hittalmani, 2000. Molecular marker assisted tagging of morphological and physiological traits under two contrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). *Euphytica*, 112: 69–78

Huang, X., L. Yang, Y. Jin, J. Lin and F. Liu, 2017. Generation, annotation, and analysis of a large-scale expressed sequence tag library from *Arabidopsis pumila* to explore salt-responsive genes. *Front. Plant Sci.*, 8: 1–15

Jin, J.P., F. Tian, D.C. Yang, Y.Q. Meng, L. Kong, J. Luo and G. Gao, 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucl. Acids Res*., 45: 1040–1045

Joshi, R., S.H. Wani, B. Singh, A. Bohra, Z.A. Dar, A.A. Lone, A. Pareek and S.L. Singla-Pareek, 2016. Transcription factors and plants response to drought stress: current understanding and future directions. *Front. Plant Sci*., 7: 1–15

Kawaura, K., K. Mochida, A. Enju, Y. Totoki, A. Toyoda, Y. Sakaki, C. Kai, J. Kawai, Y. Hayashizaki, M. Seki, K. Shinozaki and Y. Ogihara, 2009. Assessment of adaptive evolution between wheat and rice as deduced from full-length common wheat cDNA sequence data and expression patterns. *BMC Genom.*, 10: 1–11

Lafitte, H.R., A.H. Price and B. Courtois, 2004. Yield response to water deficit in an upland rice mapping population: associations among traits and genetic markers. *Theor. Appl. Genet*., 109: 1237–1246

Lafitte, H.R., B. Courtois and M. Arraudeau, 2002. Genetic improvement of rice in aerobic systems: progress from yield to genes. *Field Crops Res*., 75: 171–190

Lenka, S.K., A. Katiyar, V. Chinnusamy and K.C. Bansal, 2010. Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant Biotechnol. J*., 9: 315–327

Lin, M., D. Lai, C. Pang, S. Fan, M. Song and S. Yu, 2013. Generation and analysis of a large-scale expressed sequence tag database from a full-length enriched cDNA library of developing leaves of *Gossypium hirsutum* L. *PLoS One*, 8: 1–15

Liu, J., M. Hasanuzzaman, H. Wen, J. Zhang, T. Peng, H. Sun and Q. Zhao, 2019. High temperature and drought stress cause abscisic acid and reactive oxygen species accumulation and suppress seed germination growth in rice. *Protoplasma*, 256: 1–11

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

Lowe, R., N. Shirley, M. Bleackley, S. Dolan and T. Shafee, 2017. Transcriptomics technologies. *PLoS Comput. Biol*., 13: 1–23

Manickavelu, A., K. Kawaura, K. Oishi, T. Shin, Y. Kohara, N. Yahiaoui, B. Keller, R. Abe, A. Suzuki, T. Nagayama, K. Yano and Y. Ogihara, 2012. Comprehensive functional analyses of expressed sequence tags in common wheat (*Triticum aestivum*). *DNA Res.*, 19: 165–177

Minetti, G., G. Leguizamon and E. Alba, 2012. SAX: a new and efficient assembler for solving DNA fragment assembly problem. *In: 13th Argentine Symposium on Artificial Intelligence*, Vol. 41, pp: 177–188

Moncada, P., C.P. Martinez, J. Borrero, M. Chatel, H. Gauch, E. Guimaraes, J. Tohme and S.R. McCouch, 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. *Theor. Appl. Genet*., 102: 41–52

Mukundi, E., F. Gomez-Cano, W.Z. Ouma and E. Grotewold, 2017. Design of knowledge bases for plant gene regulatory networks. *Meth. Mol. Biol.*, 1629: 207–223

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

Nguyen, T.T., N. Klueva, V. Chamareck, A. Aarti, G. Magpantay, A.C. Millena, M.S. Pathan and H.T. Nguyen, 2004. Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Mol. Genet. Genomics*, 272: 35–46

Pandey, V. and A. Shukla, 2015. Acclimation and tolerance strategies of rice under drought stress. *Rice Sci*., 22: 147–161

Pérez-Rodríguez, P., D.M. Riaño-Pachón, L.G. Corrêa, S.A. Rensing, B. Kersten and B. Mueller-Roeber, 2010. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucl. Acids Res*., 38: 827–827

Price, A.H., J. Townend, M.P. Jones, A. Audebert and B. Courtois, 2002. Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Mol. Biol.*, 48: 683–695

Qu, Y., P. Mu, H. Zhang, C.Y. Chen, Y. Gao, Y. Tian, F. Wen and Z. Li, 2008. Mapping QTLs of root morphological traits at different growth stages in rice. *Genetica*, 133: 187–200

Ramalingam, J., M.S. Pathan, O. Feril, Miftahudin, K. Ross, X.F. Ma, A.A. Mahmoud, J. Layton, M.A. Rodriguez-Milla, T. Chikmawati, B. Valliyodan, R. Skinner, D.E. Matthews, J.P. Gustafson and H.T. Nguyen, 2006. Structural and functional analyses of the wheat genomes based on expressed sequence tags (ESTs) related to abiotic stresses. *Genome*, 49: 1324–1340

Sakai, H., S.S. Lee, T. Tanaka, H. Numa, J. Kim, Y. Kawahara, H. Wakimoto, C.C. Yang, M. Iwamoto, T. Abe, Y. Yamada, A. Muto, H. Inokuchi, T. Ikemura, T. Matsumoto, T. Sasaki and T. Itoh, 2013. Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol*., 54: 1–6

Shankar, R., A. Bhattacharjee and M. Jain, 2016. Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep*., 6: 1–15

Shanker, A.K., M. Maheswari, S.K. Yadav, S. Desai, D. Bhanu and N.B. Attal, 2014. Drought stress responses in crops. *Funct. Integr. Genomics*, 1–12

Shinozaki, K. and K. Yamaguchi-Shinozaki, 2007. Gene networks involved in drought stress response and tolerance. *J. Exp. Bot*., 58: 221–227

Silveira, R.D., F.R. Abreu, S. Mamidi, P.E. McClean, R.P. Vianello, A.C. Lanna, N.P. Carneiro and C. Brondani, 2015. Expression of drought tolerance genes in tropical upland rice cultivars (*Oryza sativa*). *Genet. Mol. Res*., 14: 8181–8200

Varshney, R.K., P.J. Hiremath, P. Lekha, J. Kashiwagi, J. Balaji, A.A. Deokar, V. Vadez, Y. Xiao, R. Srinivasan, P.M. Gaur, K.H.M. Siddique, C.D. Town and D.A. Hoisington, 2009. A comprehensive resource of drought- and salinity responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics*, 10: 1–18

Venuprasad, R., H.E. Shashidhar, S. Hittalmani and G.S. Hemamalini, 2002. Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. *Euphytica*, 128: 293–300

Wang, S., J. Lim, S. Kim, S. Cho, S. Yoo, H. Koh, Y. Sakuraba and N. Paek, 2015. Mutation of SPOTTED LEAF3 (SPL3) impairs abscisic acid responsive signalling and delays leaf senescence in rice. *J. Exp. Bot*., 66: 7045–7059

Wang, Z., M. Gerstein and M. Snyder, 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10: 57–63

Yonemaru, J., T. Yamamoto, S. Fukuoka, Y. Uga, K. Hori and M. Yano, 2010. Q-TARO: QTL annotation rice online database. *Rice*, 3: 194–203

Yoshida, S., D.A. Forno and J.H. Cock, 1976. *Laboratory Manual for Physiological Studies of Rice*, International Rice Research Institute (IRRI), Los Banos, Laguna, PHL

Zhang, P., Z. Ding, Z. Zhong and H. Tong, 2019. Transcriptomic analysis for indica and japonica rice varieties under aluminum toxicity. *Intl. J. Mol. Sci*., 20: 1–16

Zhang, Y., Q. Chen, J. Lan, Y. Luo, X. Wang, Q. Chen, B. Sun, Y. Wang, R. Gong and H. Tang, 2018. Effects of drought stress and rehydration on physiological parameters and proline metabolism in kiwifruit seedling. *Intl. J. Agric. Biol*., 20: 2891–2896

Zhang, Z., Y. Li and B. Xiao, 2016. Comparative transcriptome analysis highlights the crucial roles of photosynthetic system in drought stress adaptation in upland rice. *Sci. Rep*., 6: 1–13

[Received 16 Apr 2019; Accepted 22 Jul 2019; Published (online) 22 Dec 2019]