Genome-wide identification of *DUF506* gene family in rice and expression profiling under abiotic stresses (#87907)

First submission

Guidance from your Editor

Please submit by **28 Jul 2023** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files
Download and review all files from the materials page.

10 Figure file(s)
12 Table file(s)
1 Raw data file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Genome-wide identification of *DUF506* gene family in rice and expression profiling under abiotic stresses

Wei Dong [Equal first author, 1], **Jian Tu** [Equal first author, 1], **Wei Deng** [1], **Jianhua Zhang** [1], **Yuran Xu** [1], **Anyu Gu** [1], **Hua An** [1], **Kui Fan** [2], **Rui Wang** [2], **Jianping Zhang** [2], **Limei Kui** [Corresp., 1], **Xiaolin Li** [Corresp. 1]

[1] Yunnan Academy of Agricultural Sciences, Food Crops Research Institute, Kunming, China

[2] Yunnan Grain Industry Group Co., Ltd, Kunming, China

Corresponding Authors: Limei Kui, Xiaolin Li
Email address: kuilimei@yaas.org.cn, xiaolinli@163.com

The Domains of unknown function 560 (DUF560), also called PDDEXK_6 family, are omnipresent plant proteins and confirmed to play critical roles in Arabidopsis root development as well as ABA and abiotic responses. However, genome-wide identification and expression pattern analysis in rice are still insufficient. Here, 10 OsDUF506 genes were identified and classified into four subfamilies based on the phylogenetic relationship. Segmental duplication was essential to expanding *OsDUF506s*, which experienced purifying selective pressure. *OsDUF506s*, except *OsDUF50609* and *OsDUF50610*, have colinear gene pairs with five monocot species indicating that they were conserved in evolution. In addition, the conserved domains, gene structures, SNPs distribution, and targeting miRNAs were systematically investigated. Massive cis-regulatory elements were found in promoter regions, indicating that *OsDUF506s* may be essential in hormone regulation and abiotic stress response. Therefore, we analyzed transcriptome data induced by plant hormones and performed qRT-PCR on 8 *OsDUF506s* under drought, cold, and phosphorus-deficient stresses. The results showed that most of the *OsDUF506s* ubiquitously respond to ABA and JA treatment, as well as drought and cold conditions. In conclusion, our findings provided insights into the evolution and function of *OsDUF506s*, which could benefit crop breeding in the future.

Genome-Wide Identification of DUF506 Gene Family in Rice and Expression Profiling Under Abiotic Stresses

Wei Dong[1,†], Jian Tu1[,†], Wei Deng[1], Jianhua Zhang[1], Yuran Xu[1], Anyu Gu[1], Hua An[1], Kui Fan[2], Rui Wang[2], Jianping Zhang[2], Limei Kui[1,*], Xiaolin Li[1,*]

[1] Food Crops Research Institute, Yunnan Academy of Agricultural Sciences (YAAS), Kun-ming, Yunnan Province, China
[2] Yunnan Grain Industry Group Co., Ltd, Kunming, Yunnan Province, China

Corresponding Author:
Xiaolin Li[1]
No.2238 Beijing Road, Kunming, Yunnan Province, 650051, China
Email address: xiaolinli@163.com
Limei Kui[2]
No.2238 Beijing Road, Kunming, Yunnan Province, 650051, China
Email address: klm@yaas.org.cn

†These two authors contributed equally to the work.

Abstract

The Domains of unknown function 560 (DUF560), also called PDDEXK_6 family, are omnipresent plant proteins and confirmed to play critical roles in Arabidopsis root development as well as ABA and abiotic responses. However, genome-wide identification and expression pattern analysis in rice are still insufficient. Here, 10 *OsDUF506* genes were identified and classified into four subfamilies based on the phylogenetic relationship. Segmental duplication was essential to expanding *OsDUF506s*, which experienced purifying selective pressure. *OsDUF506s*, except *OsDUF50609* and *OsDUF50610*, have colinear gene pairs with five monocot species indicating that they were conserved in evolution. In addition, the conserved domains, gene structures, SNPs distribution, and targeting miRNAs were systematically investigated. Massive cis-regulatory elements were found in promoter regions, indicating that *OsDUF506s* may be essential in hormone regulation and abiotic stress response. Therefore, we

analyzed transcriptome data induced by plant hormones and performed qRT-PCR on 8 *OsDUF506s* under drought, cold, and phosphorus-deficient stresses. The results showed that most of the *OsDUF506s* ubiquitously respond to ABA and JA treatment, as well as drought and cold conditions. In conclusion, our findings provided insights into the evolution and function of *OsDUF506s*, which could benefit crop breeding in the future.

Introduction

Domains of unknown functions(DUFs) are batches of gene families with conserved domains but unknown functions that widely exist in eukaryotes (Bateman et al. 2010). The number of DUF families recorded in the Pfam database reached 4716(https://www.ebi.ac.uk/interpro/entry/pfam/, accessed on 29 March 2023). Although most DUF families are still unknown, some DUF families have been investigated. In *Oryza sativa*,*OsDUF1618*(Wang et al. 2014),*OsDUF221*(Ganie et al. 2017),*OsDUF1110*(Harada et al. 2016),*OsDUF810*(Li et al. 2018),*OsDUF668*(Zhong et al. 2019a),*OsDUF231*(Zhong et al. 2019b),*OsDUS936*(Li et al. 2017) have been characterized. Previous studies showed that DUF genes involved different biological functions in rice. For instance, *SWOLLEN TAPETUM AND STERILITY 1* (*STS1*) containing DUF726 domain involved in sporopollenin biosynthesis by interacting with *Polyketide Synthase 2* (*OsPKS2*) and *Acyl-CoA Synthetase 12* (*OsACOS12*)(Yuan et al. 2022). Another DUF726 protein, encoded by *Leaked and Delayed Degraded Tapetum 1* (*OsLDDT1*), participated in fatty acid synthesis and anther epidermis formation(Sun et al. 2023). *ROLLED and ERECT LEAF 2* (REL2) containing DUF630 and DUF632 conserved domain involved in regulating leaf morphology, the functional loss of the protein led to rolling leaves(Yang et al. 2016).DUF genes were also demonstrated to be involved in different biotic and abiotic responses. For example, *Oryza sativa Stress Responsive DUF740 Protein*(*OsSRDP*) gene belonged to DUF740 family, its overexpressed transgenic plants, driven by promoter AtRd29A, revealed stronger resistance to drought, salinity, and cold stresses as well as rice blast fungus(Jayaraman et al. 2022).*DUF966-stress repressive gene 2*(*OsDSR2*) gene, belonging to DUF966 family, was involved in negative regulation of salt and drought stress responses(Luo et al. 2014).

DUF506 family, also called PDDEXK_6 family, is a group of plant proteins that are distant homologs of the PD-(D/E)XK nuclease superfamily. The nuclear structure is retained as α-β-β-α-β and includes the typical PDDEXK motifs II and III in modified forms as xDxxx motif located in the second core beta-strand, where x is any hydrophobic residue, and a D/E)X(D/N/S/C/G) pattern. Motif III's missing positively charged residue is possibly replaced by a conserved arginine in motif IV located in the proceeding alpha-helix(Knizewski et al. 2007). So far,

DUF506 proteins have not been systemically and functionally characterized. Previous research merely identified that the expressions of 13 *AtDUF506s* were ubiquitous in organs and related to abiotic stresses and ABA response(Ying 2021). The comparative microarray data showed that *AT2G20670* was inhibited by *B. cinereal*, heat, salinity, and osmotic stress(Sham et al. 2019). Recent studies revealed that *REPRESSOR OF EXCESSIVE ROOT HAIR ELONGATION 1* (*AtRXR1*) gene encoded AT3G25240 protein and was strongly induced by phosphorus limitation, which suppressed root hairs(RHs) extension by interacting with RabD2c GTPase. Moreover, its function under phosphorus limitation is conserved both in monocot and dicot, which is supported by the similar function of *Brachypodium distachyon DUF506*(Ying et al. 2022). Similar to *AtRXR1*, its homologous gene *AtRXR3*(*AT1G62420*) also inhibited RHs elongation but with a different mechanism. *AtRXR3* repressed RHs elongation via *ROOT HAIR DEFECTIVE6-LIKE4*(*RSL4*) and interacted with cytosolic CaMs(Ying & Scheible 2022). Current studies on DUF506s in Arabidopsis have suggested that DUF506 family members are significant in plant growth and abiotic resistance, but DUF506s functions in rice have rarely been investigated. The reports showed that *LOC_Os01g68650*, the closest homologous gene of *AtRXR1*, was upregulated under drought stress, and its expression in *OsDT11* overexpression line was higher than wild line, indicating that this gene may participate in drought response and enhanced by *OsDT11*(Zhao et al. 2020). *LOC_Os01g54340* was revealed as a nitrogen-sensitive gene and rapidly repressed by nitrogen starvation(Hsieh et al. 2018). Until now, the *OsDUF506* family has not been genome-wide identified and the functions are still unknown.

In this study, we identified the whole *OsDUF506* family members in *Oryza sativa* by bioinformatic methods. The phylogeny, conserved motifs, cis-acting regulatory elements, non-synonymous SNPs distribution, target miRNA, synteny, and tissue expression specificity were analyzed. Transcriptome and qRT-PCR were used to explore the expression pattern of *OsDUF506s* under plant hormones treatments as well as drought, cold, and phosphorus-deficient stresses. Our study provided a more comprehensive identification and classification of *OsDUF506s*, expanded our recognition of the functions under abiotic stresses, and served as the basis of molecular breeding in rice.

Materials & Methods

Identification of *DUF506* members in 10 Plant Species

All the genome databases were downloaded from the EnsemblPlants database (http://plants.ensembl.org, accessed on 18 October 2022). Thirteen Arabidopsis DUF506 protein sequences were obtained from UniProtKB/Swiss-Prot (SwissProt) database (https://www.uniprot.org, accessed on 18 October 2022) and used as queries to perform protein

blast search in 10 plant species by using TBtools (Chen et al. 2020). Meantime, the DUF506 typical domain (PF04720, PDDEXK_6) was downloaded from the PFAM database (http://pfam.xfam.org, accessed on 18 October 2022) and was used to search for DUF506 by using HMMER tool(https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan, accessed on 18 October 2022). All the candidates were integrated and edited to remove the redundant. Then the candidates were submitted to NCBI-CDD(https://www.ncbi.nlm.nih.gov/cdd/, accessed on 18 October 2022) to test for the existence of the complete DUF506 conserved domain. The molecular weight (Mw), isoelectric point(pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of DUF506 members were predicted with ExPASy (http://web.expasy.org/protparam, accessed on 18 October 2022). The subcellular localizations were predicted with WoLF PSORT (https://wolfpsort.hgc.jp, accessed on 18 October 2022).

Phylogenetic relationship, structure, and conserved motifs analysis of *OsDUF506* members

A total of 130 *DUF506s* was used for aligning with the method of MUSCLE(default parameters) and constructing an ML phylogenetic tree(default parameters) with MEGA 11 software by setting bootstrap to 1000 and JTT+G model(Tamura et al. 2021). The result was displayed by ChiPlot (https://www.chiplot.online,accessed on 18 October 2022). The conserved motifs were predicted with MEME (https://meme-suite.org/meme/doc/meme.html, accessed on 18 October 2022) and visualized by TBtools(Chen et al. 2020).

Prediction of CREs of *OsDUF506* members

The 2000bp upstream sequences of promoters were used for searching CREs with PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html, accessed on 18 October 2022) and visualized by TBtools(Chen et al. 2020).

Analysis of SNPs of *OsDUF506* members

The *OsDUF506* sequences were used to query for nsSNPs from the SNP-Seek database against Nipponbare reference (https://snp-seek.irri.org, accessed on 18 October 2022). the SNP-index was used to evaluate the subpopulation specificity of nsSNPs. The SNP-index is calculated as $SNP_{GJ}\text{-index}=N_{ref}/N_{all} \times 100\%$, $SNP_{H}\text{-index}=N_{h}/N_{all} \times 100\%$, $SNP_{XI}\text{-index}=1\text{-}SNP_{GJ}\text{-index}\text{-}SNP_{H}\text{-index}$, N_{ref} represented the number of varieties sharing the same allele with reference, N_h represented the number of varieties with heterozygous allele, N_{all} represented the total number of varieties with determined alleles at the SNP locus.

Syntenic analysis of *OsDUF506* members

The duplication events and syntenic relationship of *DUF506s* between rice and other plants were

obtained by MCScanX(Wang et al. 2012). The results were visualized by TBtools(Chen et al. 2020).The nonsynonymous (Ka) and synonymous (Ks) calculations were performed by the simple Ka/Ks calculator kit of TBtools(Chen et al. 2020).

Predict analysis of miRNAs interacting with *OsDUF506* members

The miRNAs targeting *OsDUF506s* were predicted by psRNATarget (https://www.zhaolab.org/psRNATarget/analysis?function=2, accessed on 18 October 2022) and visualized by ChiPlot(https://www.chiplot.online,accessed on 18 October 2022). The expressions of the predicted miRNAs were obtained from the PmiRExAt database(http://pmirexat.nabi.res.in/searchdb.html, accessed on 18 October 2022) and visualized by TBtools(Chen et al. 2020).

Expression analysis of *OsDUF506* members in different tissues and induced by plant hormones

The expression data in different tissues and induced by 50μM abscisic acid(ABA), 10μM gibberellin 3(GA_3), 10μM auxin(IAA), 100μM jasmonic acid(JA), 1μM brassinolide(BL), and 1μM trans-Zeatin(tZ) were obtained from RiceXPro database (https://ricexpro.dna.affrc.go.jp/quick-guide.html, accessed on 18 October 2022), the normalized signal intensity values(log_2) were used for constructing heatmap, and the scale method of normalized was used to intuitively reflect the expressing changes of particular genes at different treating time points by TBtools(Chen et al. 2020).

Plant growth conditions and abiotic stresses treatments

Japonica rice variety Yunkegeng 5 was used in this expression analysis. The plants were grown in a climate chamber for 14 days in Yoshida nutrient solution maintained at 28℃ with a photoperiod of 14h-light and 10h-dark. For drought stress, plants were shifted to a nutrient solution of 20% PE6000 for 3h. For cold stress, plants were shifted to the climate chamber under 4℃ treatment for 3h. For phosphorus deficiency stress, plants were transferred to phosphorus-deficient Yoshida nutrient solution for 7 days. The shoots of the treated and control plants were harvested and stored at -80℃ with three biological replicates from each.

Expression analysis of *OsDUF506* members by qRT-PC

Primer design and specificity check was performed by Primer-BLAST of NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi, accessed on 20 February 2023). The

total RNA was extracted by TaKaRa MiniBEST Plant RNA Extraction Kit, cDNA was synthesized using Vazyme HiSript III 1st Strand cDNA Synthesis Kit(+gDNA wiper). The qRT-PCR was accomplished by Vazyme ChamQ SYBR Color qPCR Master Mix (Without ROX) in LightCycler96 system under the PCR condition of 95℃ for 60s，45 cycles of 95℃ for 10s, 54 to 60℃ for 20s and 72℃ for 20s. The relative expressions data were calculated by $2^{-\triangle\triangle CT}$ method, and the reference used in this study was *OsActin*. The primer sequences were listed in Table S9. The significant differences level was analyzed by unpaired t-test with GraphPad Prism 8.0 software.

Results

Identification and phylogenetic relation of *DUF506* members in *Oryza sativa* and other nine plant species

Using the previous method, we identified 10 *DUF506*s in rice and named them *OsDUF50601* to *OsDUF50610*.Moreover,13,25,13,20,8,4,22,8,7 *DUF506s* were identified in *Arabidopsis thaliana*, *Glycine max*, *Solanum tuberosum*, *Gossypium raimondii*, *Zea mays*, *Hordeum vulgare*, *Triticum aestivum*, *Sorghum bicolor* and *Ananas comosus* respectively(Table S1).A total of 130 *DUF506s* was used to construct an ML phylogenetic tree and divided into 4 subfamilies according to the evolutionary distance referred to previous literature(Ying 2021). The IIIb subfamily held the most significant number of *DUF506s*, whereas the IIIa subfamily had the smallest. Three members belonged to subfamily I(*OsDUF50602*, *OsDUF50609,* and *OsDUF50610*), two belonged to subfamily II(*OsDUF50601* and *OsDUF50606*), only *OsDUF50603* belonged to subfamily IIIa, while the rest four members composed the largest subfamily IIIb(Figure 1).

Characterization and Conserved Motifs of Os*DUF506* members

The 10 *OsDUF506s* were located on chromosome1,3,5,7,10,11. The summary of characteristics was shown in Table 1. Their protein lengths varied from 269 to 507. The molecular weight, theoretical isoelectric points, and aliphatic index were predicted between 29861.74 to 54428.55 Da, 5.58 to 9.12, and 62.39 to 86.85, respectively. The instability index of proteins was over 40, and GRAVY values were negative, suggesting they were unstable hydrophilic proteins. Except for the subcellular localization of members from subfamily II (*OsDUF50601* and *OsDUF50606*) was predicted in the nucleus, the others were all localized in the chloroplast. *OsDUF506s* exhibited 1 to 3 exons, and genes in one subfamily shared a similar exon/intron structure (Figure 2B). The *OsDUF506s* from subfamily I exhibited only one CDS region, and subfamilies II and

IIIa exhibited three CDS regions. Due to the number of CDS regions, subfamily IIIb was divided into 2 branches. The result indicated that members of different subfamilies might function differently.

Fifteen conserved protein motifs of *OsUDF506s* were identified (Figure 2C). All the members of *DUF506* in rice and Arabidopsis shared motifs 1,2,3, and 5. Motifs 4 and 11 were most conserved in subfamily I. *OsDUF50602* and *OsDUF50610* ultimately shared the same motifs with *AT1G62420* and *AT3G25240*, indicating they might have the same biological function. Members in subfamily II shared motifs 8 and 11. The *OsDUF50602* from subfamily IIIa had the same motif construction as *AT2G39650*. All the members in subfamily IIIb shared motifs 4,6 and 7. Three-quarters possessed motif 9, and motif 13 only existed in Arabidopsis genes. The result showed that *DUF506s* in the same subfamily displayed similar motif characteristics, indicating a similar function.

Analysis of Cis-Acting Regulatory Elements (CREs) of *OsDUF506* members

To analyze the prospective function of *OsDUF506s*, we searched the 2000 bp sequence upstream of *OsDUF506s*, and 321 potential CREs were predicted (Figure 3). *OsDUF50604* possessed the largest number of CREs, while *OsDUF50603* and *OsDUF50609* possessed the least (Table S2). Nineteen types of CREs were identified and grouped into three functional categories: hormone response, abiotic stress response, and plant growth and metabolism. Hormone response-related CREs included MeJA-response elements (TGACG-motif and CGTCA-motif), abscisic acid response elements (ABRE), auxin response elements (TGA and AuxRR-core), salicylic acid response elements (TCA) and gibberellin response elements (GARE-motif and P-box). Each member contains at least 2 types of hormone response elements, MeJA-response elements, and abscisic acid response elements were abundant in *OsDUF506s*. In terms of abiotic stress response, light response elements, anaerobic induction elements, and anoxic specific inducible elements were the most abundant elements. Low-temperature response (LTR) elements existed in subfamily II and IIIb. Drought inducible elements (MSB) existed in *OsDUF50601*, *OsDUF50602*, *OsDUF50607,* and *OsDUF50609*, suggesting that they might be associated with drought stress. CREs related to plant growth and metabolism were relatively fewer. CAT-boxes relating to meristem expression were found in half of *OsDUF506s*, while Motif I in *OsDUF50601*, *OsDUF50605*, and *OsDUF50606* was involved in root specificity. Only a few *OsDUF506* members possessed CREs associated with circadian control, cell cycle regulation, zein metabolism regulation, endosperm expression, and flavonoid biosynthetic genes regulation. These results indicated that *OsDUF506s* might be essential in hormone regulation and abiotic stress response.

Analysis of SNPs in *OsDUF506* members

According to the Rice SNP-Seek Database,78 SNPs were found in *OsDUF506s*, 30 of which were non-synonymous SNP (nsSNP)(Table 2). *OsDUF50609* and *OsDUF50610* possessed 13 and 6 nsSNPs respectively, while the other members possessed 1 to 3 nsSNP respectively, indicating that they were relatively conserved. To explore the subpopulation-specific variants, 30 nsSNP genotyping data of 2644 rice varieties from nine subpopulations (five *Xian/indicia(XI)* subpopulations and four *Geng/japonica(GJ)* subpopulations) were analyzed. SNP_{GJ}-index represented the proportion of varieties with the same allele as the reference Nipponbare to the total number of varieties. Three nsSNPs (OsDUF50609.1, OsDUF50609.7, and OsDUF50610.6) from subfamily I and one nsSNP (OsDUF50607.1) from subfamily IIIb were suggested to be specific between Indica and Japonica varieties because the SNP_{GJ}-index values in Indica subpopulations were less than 5%. In contrast, the values in Japonica subpopulations were more than 85% (Figure 4 and Table S3).

Synteny Analysis of *OsDUF506* members

Gene duplications are significant for gene family evolution. The synteny analysis result showed three segmental duplications (*OsDUF50601-OsDUF50606*, *OsDUF50604-OsDUF50608,* and *OsDUF50605-OsDUF50607*) and one tandem duplication (*OsDUF50609-OsDUF50610*) in rice (Figure 5), which indicated segmental duplication was the core dynamic of expansion during *OsDUF506s* evolution. Ka/Ks ratios of all the duplications were below 0.5 (Table S4), demonstrating that *OsDUF506s* have gone through purifying selective pressure during evolution. In addition, these duplicated events were speculated to happen at least 20.48 million years ago (Table S4).

Besides, the duplicated events of *DUF506* members among rice, the dicot model plant (*Arabidopsis thaliana*), and other monocotyledonous species (*Zea mays*, *Hordeum vulgare*, *Triticum aestivum*, *Sorghum bicolor*, and *Ananas comosus*) were also identified. *OsDUF50604*, *OsDUF50605*, *OsDUF50607*, and *OsDUF50608* in rice and *AT3G22970* and *AT4G14620* in *Arabidopsis* from subfamily IIIb formed 6 homologous gene pairs and showed multiple collinearities (Figure 5). In five monocotyledonous species, except *OsDUF50609* and *OsDUF50610* on chromosome 11, the other *OsDUF506s* all had homologous genes. 10 colinear gene pairs were found between *Oryza sativa* and *Hordeum vulgare,* 11 pairs with *Zea mays* and *Sorghum bicolor* respectively, 12 pairs with *Ananas comosus*, and 30 pairs with *Triticum aestivum*(Figure 6 and Table S5). The results suggested that these *DUF506s* might be differentiated from the same ancestral type and function similarly. The *OsDUF506s* from subfamily IIIb were involved in gene pairs with *Arabidopsis thaliana and* those five monocots respectively, indicating that they might be critical to the evolution of the DUF506 family.

Predicted miRNAs Analysis

Sixty-nine predicted miRNAs were identified to target *OsDUF506s* and might involve in expression regulations (Figure 7A). All the *OsDUF506s* were targeted by multiple miRNAs, suggesting these genes were strictly regulated by the combination of multiple miRNAs. The *osa-miRNA2927*, *osa-miRNA5075*, *osa-miRNA1848*, *osa-miRNA2925*, and *osa-miRNA5809* targeted multiple *OsDUF506s* respectively, indicating that these miRNAs were critical to *OsDUF506s*. The lengths of matured miRNAs were between 19 and 24 nucleotides (Table S6). Most of the miRNAs inhibited *OsDUF506* expressions by cleavage, only *osa-miR2880*, *osa-miR5340*, *osa-miR2926*, *osa-miR156j-3p*, *osa-miR1875*, *osa-miR444b.1*, *osa-miR444c.1*, and *osa-miR5832* inhibited *OsDUF506* expressions by translation repression.

The expression analysis revealed that the majority of the miRNAs expressed quite low in rice tissues(Figure 7B). The osa-miR1874-3p specifically targeting *OsDUF50605* showed the highest expression level in embryo. The osa-miR444b.1 targeting *OsDUF50606* highly expressed in all tissues except anther, indicating the low expression of *OsDUF50606* in these tissues, which may inhibit the function of *OsDUF50606* in plant growth. The expressing levels of miRNAs under drought, salt, and cold stress did not showed significant changes, suggesting that they did not participate in abiotic stress responses.

Expression Analysis of *OsDUF506* members in different tissues and induced by plant hormones

To investigate the expression specificities of *OsDUF506s*, the expression values in leaf blade, leaf sheath, root, stem, inflorescence, anther, pistil, lemma, palea, ovary, embryo, and endosperm were analyzed (Figure 8; Table S7). *OsDUF50602* from subfamily I expressed in all tissues with the highest expressions in the embryo and endosperm 7 days after flowering, while *OsDUF506010* expresses extremely low in all tissues. The segment duplication gene pair *OsDUF50601-OsDUF50606* from subfamily II showed completely different expression modes, suggesting they might be involved in functional redundancy. Compared to the genes from subfamily IIIa, the four genes from subfamily IIIb showed higher expression levels in most tissues. The segment duplication gene pair *OsDUF50604-OsDUF50608* was highly expressed in anther and leaf sheath respectively, while another segment pair *OsDUF50605-OsDUF50607* was highly expressed in leaf blade and leaf sheath respectively, revealing that they might be involved in functional differentiation.

The CREs analysis of *OsDUF506s* suggested that they were widely involved in hormonal regulation, thus we further analyzed their expression profiles in root and shoot treated with 6 plant hormones abscisic acid (ABA), gibberellic acid (GA_3), indole-3-acetic acid (IAA),

jasmonic acid (JA), brassinolide (BL), and trans-zeatin (tZ). *OsDUF506s* showed different regulating modes (Figure 9 and Table S8). *OsDUF50602* was significantly upregulated induced by ABA and JA, while it showed opposite regulation in root and shoot induced by tZ. *OsDUF50601* and *OsDUF50606*, from subfamily II, showed the same regulation under ABA treatment. In root, except *OsDUF50605* was upregulated by ABA induction, and other members from subfamily IIIb were downregulated by ABA, IAA, and JA. Under BL treatment, *OsDUF50604* and *OsDUF50608* were upregulated, however, *OsDUF50605* and *OsDUF50607* were downregulated in root. Interestingly, *OsDUF50607* was significantly upregulated by ABA induction in shoot, while the opposite was observed in root. *OsDUF50605* also showed opposite regulations in shoot and root by ABA induction.

Expression Analysis of *OsDUF506* members under drought stress, cold stress, and phosphorus-deficient stress

To explore the responses of *OsDUF506s* under abiotic stresses, the relative expressions of 8 *OsDUF506s* under drought, cold, and phosphorus-deficient stresses were analyzed by qRT-PCR(Figure 10). Under drought condition, the expressions of *OsDUF50601*, *OsDUF50603*, *OsDUF50604*, *OsDUF50607*, and *OsDUF50608* were significantly downregulated, whereas only *OsDUF50602* was significantly upregulated with a more than 2-fold increase. By contrast, *OsDUF506s* were more sensitive to cold stress. Under cold stress, 6 *OsDUF506s* were upregulated except *OsDUF50606* and *OsDUF50607*, *OsDUF50601*, and *OsDUF50504* showed a 4-fold increase in gene expression. Under phosphorus-deficient condition within a week, *OsDUF50601*, *OsDUF50604,* and *OsDUF50607* were significantly downregulated, and only *OsDUF50605* was significantly upregulated but less than 2-fold. The result demonstrated the majority of *OsDUF506s* were induced by drought and cold stresses, suggesting these genes may be involved in these stress responses.

Discussion

DUFs are a specific type of gene with conserved domains but unknown functions. DUF506 family belongs to the PD-(D/E)XK nuclease superfamily, which contains numerous enzymes that participate in significant cellular processes(Knizewski et al. 2007). Most families are certificated to be different restriction endonucleases with the functions referring to repair of damaged DNA, resolve of holliday junctions, and excess cleaving in DNA recombination(Bujnicki 2003). The family persists in the characteristic motif II of PD-(D/E)XK but lacks a functional known domain(Knizewski et al. 2007). So far, the investigation of *DUF506* family is only performing in *Arabidopsis*(Ying 2021). In this study, we identified 10 *OsDUF506s* with intact DUF506 domain

in rice and divided them into four subfamilies based on the previous studies in *Arabidopsis* (Table 1). 120 *DUF506* genes from five other monocots and four dicots were also identified and analyzed phylogenetic relationship with *OsDUF506s*. The phylogenetic results showed they were different from *DUF1618*s which only existed in monocot divergence, *OsDUF506s* were present in both monocots and dicots（Figure 1）, indicating that *DUF506* was an ancient gene family which originated before the dicotyledon-monocotyledon differentiation (Wang et al. 2014). *DUF506* members from monocots or dicots preferred to gather in the same branch （Figure 1）, indicating a substantial divergence in *DUF506*s between monocots and dicots. Moreover, genome size did not correlate with the number of *DUF506* genes. For example, in dicots, *Arabidopsis thaliana* and *Solanum tuberosum* possessed the same quantity of *DUF506* genes, but their genome sizes differed significantly.

Gene duplication is widespread in plants and contributes to genome expansion and the evolution of new functions, the majority of types of gene duplicates are whole-genome duplications and tandem duplications(Panchy et al. 2016). However, in *OsDUF506* duplication events, segmental duplication accounted for 75%, while tandem duplication only accounted for 25% (Table S4), suggesting that segmental duplication was critical for expanding *OsDUF506* family members. The dynamic processes between gene duplications and gene losses contribute to genomes differences (Holland et al. 2017). Research on rice duplications revealed that 85% of duplicates experienced loss or subfuncionalization or neofunctionalization during 50-70 million years of evolution(Throude et al. 2009) *OsDUF50602* and *OsDUF50603* did not form duplicate gene pairs(Figure 5), indicating that they might undergo gene losses. The mechanisms of duplicate gene loss not only include deletion of duplicate sequence but also include pseudogenization which means to be silenced and genetic redundancy(Ho-Huu et al. 2012; Lynch & Conery 2000; Thibaud-Nissen et al. 2009).The duplicated gene with low expression may experience pseudogenization and pseudogenes often originate from tandem duplicates(Yang et al. 2011).The tandem duplicate pair *OsDUF50609-OsDUF50610* originated 58.69 million years ago supports this viewpoint. They barely expressed in any tissue according to the gene expression profiles of the RiceXPro database (Figure 8) and Rice Genome Annotation Project database (http://rice.uga.edu/expression.shtml, accessed on 18 October 2022). Besides, compared to other members, *OsDUF50609* and *OsDUF50610* contained the most nsSNP mutations(Table 2) and excessive nonsynonymous nucleotide mutants are one of the typical features of pseudogenes(Balakirev & Ayala 2003). DUF family seems to be rich in pseudogenes, for instance, DUF1311, DUF 1124, and DUF 3054(Thibaud-Nissen et al. 2009).On the other side, *OsDUF50609* and *OsDUF50610* exhibited a general loss bias in Gramineae species generally, such as *Zea mays*, *Hordeum vulgare*, *Triticum aestivum* and *Sorghum bicolor*(Figure 6). Except for the assumed pseudogenes, the other *OsDUF506*s all have colinear gene pairs with the other

five monocots (Figure 6), revealing they were conserved in evolution and expansion of *DUF506* family in plants and they might have originated from the same ancestor. Previous studies suggested that in both *Arabidopsis* and rice, more than one-half of duplicated genes have diverged in gene expression(Blanc & Wolfe 2004; Yim et al. 2009). The duplicated two gene pairs from subfamily IIIb (*OsDUF50604-OsDUF50608* and *OsDUF50605-OsDUF50607*) revealed different tissue specificity (Figure 8), and duplicates respond conversely under phosphorus-deficient conditions (Figure10), suggesting that they might be neofunctionalized.

MicroRNAs (miRNAs) are fundamental noncoding riboregulators for gene expression. In plants, miRNA silences genes by guiding RNA cleavage or translation inhibition(Song et al. 2019). They cooperate closely with target genes and transcription factors to regulate plant growth and resistance; it could be an effective strategy for precisely improving rice varieties by derepressing specific genes using CRISPR/Cas9 (Lin et al. 2021; Nadarajah & Kumar 2019). We predicted the miRNAs targeting *OsDUF506s* and analyzed their expression profiles in tissues and under drought, cold, and salt stresses (Figure 7). The osa-miR444b.1, which specifically targeting *OsDUF50606*, was the only one highly expressed in tissues except for anthers and under the four stresses. It was functionally unknown. One copy of the segmental duplicates from subfamily II (*OsDUF50606*) expressed extremely low in all tissues and showed no significant changes in the expression level under those abiotic stresses, while the other copy(*OsDUF50601*) was constitutively expressed and strongly induced by drought, cold, and phosphorus-deficient stresses(Figure 8 and Figure 10). Further verification is needed to determine whether the expression difference originates from the translation inhibition of osa-miR444b.1.

The results of expression induced by plant hormones showed that *OsDUF506s* were more sensitive to ABA and JA treatments than to IAA and GA_3 treatments, which was consistent with the presence of hormone responsive CREs observed in their promoters（Figure 3 and Figure 9）. Although the promoter region of some *OsDUF506s* also consisted of a few GA_3 and IAA CREs, they did not show a significant and specific response trend under corresponding hormone treatment.

The CREs analysis also showed that *OsDUF50601*, *OsDUF50602*, *OsDUF50607*, and *OsDUF50609* contained one MBS respectively, which was MYB transcription factor binding site that responded to drought. However, only *OsDUF50602* was upregulated under drought condition(Figure 3 and Figure 10）, indicating that their promotors might recruit different MYBs to active *OsDUF50602* expression and inhibit *OsDUF50601* and *OsDUF50607* expressions to regulate drought tolerance. Perhaps it may be due to the vital function of hormones under the drought condition. ABA is an important plant hormone regulating water status and stomatal movement. When plants suffer from drought environment, plants synthesize ABA. Increasing ABA could induce plants to close their stomatal and retain water(Lim et al.

2015). On the other hand, JA and its derivatives with a low level under normal condition accumulates at high levels and transmit over long distances under abiotic stress(Wang et al. 2021). Under ABA and JA treatment, *OsDUF50602* was highly upregulated, while *OsDUF50604* and *OsDUF50607* were downregulated in root (Figure 9). However, whether those two hormones induce the drought responses of these *OsDUF506* genes needs further verification. Moreover, a previous study revealed the expression of *OsDUF50602* was higher in *OsDT11* OE lines than in the wild line under drought treatment, indicating that the drought response of *OsDUF50602* might be enhanced in the particular genetic background (Zhao et al. 2020).

The *OsDUF506s* responded more to cold stress than to drought stress. The expression result under cold stress demonstrated that except for the slight downregulation of *OsDUF50606* and *OsDUF50607*, the other *OsDUF506s* were significantly upregulated, three genes with LTR elements, *OsDUF5060601*, *OsDUF5060604*, and *OsDUF50608*, had the highest level (Figure 10). *DUF506s* from other species also revealed an active response to cold stress. For instance, in *Arabidopsis* and *Brachypodium*, *At1g62420*, *At3g25240*, *Bradi2g58590*, and Bradi2g62310 were strongly induced by cold stress(Ying 2021). The above results showed that *DUF506* genes play a crucial role in cold response with different mechanisms.

Recent studies showed five *AtDUF506* genes belonging to subfamilies I and II (*At1g62420, At3g07350, At3g25240, At2g20670*, and *At4g32480*) were strongly upregulated by P-limitation(Ying et al. 2022; Ying & Scheible 2022). However, although the *OsDUF506s*, which belong to subfamilies I and II, shared the highly similar exon/intron structure and conserved motifs (Figure 3), only *OsDUF50601* extremely significantly downregulated by P-limitation, indicating that monocot and dicot species were different in phosphorus responding signal pathway.

Conclusions

In this study, 10 *OsDUF506* family members in *Oryza sativa* were identified and divided into four subfamilies. We analyzed the phylogenetic relationship, gene structures, conserved motif, CREs, SNP distribution, and targeting miRNA, which filled the gap of *DUF506* family in rice. The results of public expression profiles and RT-qPCR data in tissues and under plant hormones and abiotic stresses demonstrated that *OsDUF506s* were actively involved in ABA and JA response and had different expression patterns under drought and cold, which laid the foundation for further functional analysis of *OsDUF506* family.

References

458 Balakirev ES, and Ayala FJ. 2003. Pseudogenes: are they "junk" or functional DNA? *Annu Rev*
459 *Genet* 37:123-151. 10.1146/annurev.genet.37.040103.103949
460 Bateman A, Coggill P, and Finn RD. 2010. DUFs: families in search of function. *Acta*
461 *Crystallogr Sect F Struct Biol Cryst Commun* 66:1148-1152.
462 10.1107/S1744309110001685
463 Blanc G, and Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy
464 during Arabidopsis evolution. *Plant Cell* 16:1679-1691. 10.1105/tpc.021410
465 Bujnicki JM. 2003. Crystallographic and bioinformatic studies on restriction endonucleases:
466 inference of evolutionary relationships in the "midnight zone" of homology. *Curr Protein*
467 *Pept Sci* 4:327-337. 10.2174/1389203033487072
468 Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, and Xia R. 2020. TBtools: An
469 Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant*
470 13:1194-1202. 10.1016/j.molp.2020.06.009
471 Ganie SA, Pani DR, and Mondal TK. 2017. Genome-wide analysis of DUF221 domain-
472 containing gene family in Oryza species and identification of its salinity stress-responsive
473 members in rice. *PLoS One* 12:e0182469. 10.1371/journal.pone.0182469
474 Harada K, Yamashita E, Inoue K, Yamaguchi K, Fujiwara T, Nakagawa A, Kawasaki T, and
475 Kojima C. 2016. Plant-specific DUF1110 protein from Oryza sativa: expression,
476 purification and crystallization. *Acta Crystallogr F Struct Biol Commun* 72:480-484.
477 10.1107/S2053230X16007573
478 Ho-Huu J, Ronfort J, De Mita S, Bataillon T, Hochu I, Weber A, and Chantret N. 2012.
479 Contrasted patterns of selective pressure in three recent paralogous gene pairs in the
480 Medicago genus (L.). *BMC Evol Biol* 12:195. 10.1186/1471-2148-12-195
481 Holland PW, Marletaz F, Maeso I, Dunwell TL, and Paps J. 2017. New genes from old:
482 asymmetric divergence of gene duplicates and the evolution of development. *Philos*
483 *Trans R Soc Lond B Biol Sci* 372. 10.1098/rstb.2015.0480
484 Hsieh PH, Kan CC, Wu HY, Yang HC, and Hsieh MH. 2018. Early molecular events associated
485 with nitrogen deficiency in rice seedling roots. *Sci Rep* 8:12207. 10.1038/s41598-018-
486 30632-1
487 Jayaraman K, Sevanthi AM, Raman KV, Jiwani G, Solanke AU, Mandal PK, and Mohapatra T.
488 2022. Overexpression of a DUF740 family gene (LOC_Os04g59420) imparts enhanced
489 climate resilience through multiple stress tolerance in rice. *Front Plant Sci* 13:947312.
490 10.3389/fpls.2022.947312
491 Knizewski L, Kinch LN, Grishin NV, Rychlewski L, and Ginalski K. 2007. Realm of PD-
492 (D/E)XK nuclease superfamily revisited: detection of novel families with modified
493 transitive meta profile searches. *BMC Struct Biol* 7:40. 10.1186/1472-6807-7-40
494 Li L, Ye T, Guan Y, Lv M, Xie C, Xu J, Gao X, Zhu J, Cai L, and Xu Z. 2017. Genome-wide
495 identification and analyses of the rice OsDUF936 family. *Biotechnology &*
496 *Biotechnological Equipment* 32:309-315. 10.1080/13102818.2017.1413421
497 Li LH, Lv MM, Li X, Ye TZ, He X, Rong SH, Dong YL, Guan Y, Gao XL, Zhu JQ, and Xu ZJ.
498 2018. [The Rice OsDUF810 Family: OsDUF810.7 May be Involved in the Tolerance to

Salt and Drought]. *Mol Biol (Mosk)* 52:567-575. 10.1134/S0026898418040122

Lim CW, Baek W, Jung J, Kim JH, and Lee SC. 2015. Function of ABA in Stomatal Defense against Biotic and Drought Stresses. *Int J Mol Sci* 16:15251-15270. 10.3390/ijms160715251

Lin Y, Zhu Y, Cui Y, Chen R, Chen Z, Li G, Fan M, Chen J, Li Y, Guo X, Zheng X, Chen L, and Wang F. 2021. Derepression of specific miRNA-target genes in rice using CRISPR/Cas9. *J Exp Bot* 72:7067-7077. 10.1093/jxb/erab336

Luo C, Guo C, Wang W, Wang L, and Chen L. 2014. Overexpression of a new stress-repressive gene OsDSR2 encoding a protein with a DUF966 domain increases salt and simulated drought stress sensitivities and reduces ABA sensitivity in rice. *Plant Cell Rep* 33:323-336. 10.1007/s00299-013-1532-0

Lynch M, and Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151-1155. 10.1126/science.290.5494.1151

Nadarajah K, and Kumar IS. 2019. Drought Response in Rice: The miRNA Story. *Int J Mol Sci* 20. 10.3390/ijms20153766

Panchy N, Lehti-Shiu M, and Shiu SH. 2016. Evolution of Gene Duplication in Plants. *Plant Physiol* 171:2294-2316. 10.1104/pp.16.00523

Sham A, Al-Ashram H, Whitley K, Iratni R, El-Tarabily KA, and AbuQamar SF. 2019. Metatranscriptomic Analysis of Multiple Environmental Stresses Identifies RAP2.4 Gene Associated with Arabidopsis Immunity to Botrytis cinerea. *Sci Rep* 9:17010. 10.1038/s41598-019-53694-1

Song X, Li Y, Cao X, and Qi Y. 2019. MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. *Annu Rev Plant Biol* 70:489-525. 10.1146/annurev-arplant-050718-100334

Sun Z, Liu K, Chen C, Chen D, Peng Z, Zhou R, Liu L, He D, Duan W, Chen H, Huang C, Ruan Z, Zhang Y, Cao L, Zhan X, Cheng S, and Sun L. 2023. OsLDDT1, encoding a transmembrane structural DUF726 family protein, is essential for tapetum degradation and pollen formation in rice. *Plant Sci* 329:111596. 10.1016/j.plantsci.2023.111596

Tamura K, Stecher G, and Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38:3022-3027. 10.1093/molbev/msab120

Thibaud-Nissen F, Ouyang S, and Buell CR. 2009. Identification and characterization of pseudogenes in the rice gene complement. *BMC Genomics* 10:317. 10.1186/1471-2164-10-317

Throude M, Bolot S, Bosio M, Pont C, Sarda X, Quraishi UM, Bourgis F, Lessard P, Rogowsky P, Ghesquiere A, Murigneux A, Charmet G, Perez P, and Salse J. 2009. Structure and expression analysis of rice paleo duplications. *Nucleic Acids Res* 37:1248-1259. 10.1093/nar/gkn1048

Wang L, Shen R, Chen LT, and Liu YG. 2014. Characterization of a novel DUF1618 gene family in rice. *J Integr Plant Biol* 56:151-158. 10.1111/jipb.12130

Wang Y, Mostafa S, Zeng W, and Jin B. 2021. Function and Mechanism of Jasmonic Acid in Plant Responses to Abiotic and Biotic Stresses. *Int J Mol Sci* 22. 10.3390/ijms22168568

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, and Paterson AH. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40:e49. 10.1093/nar/gkr1293

Yang L, Takuno S, Waters ER, and Gaut BS. 2011. Lowly expressed genes in Arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation. *Mol Biol Evol* 28:1193-1203. 10.1093/molbev/msq298

Yang SQ, Li WQ, Miao H, Gan PF, Qiao L, Chang YL, Shi CH, and Chen KM. 2016. REL2, A Gene Encoding An Unknown Function Protein which Contains DUF630 and DUF632 Domains Controls Leaf Rolling in Rice. *Rice (N Y)* 9:37. 10.1186/s12284-016-0105-6

Yim WC, Lee BM, and Jang CS. 2009. Expression diversity and evolutionary dynamics of rice duplicate genes. *Mol Genet Genomics* 281:483-493. 10.1007/s00438-009-0425-y

Ying S. 2021. Genome-Wide Identification and Transcriptional Analysis of Arabidopsis DUF506 Gene Family. *Int J Mol Sci* 22. 10.3390/ijms222111442

Ying S, Blancaflor EB, Liao F, and Scheible WR. 2022. A phosphorus-limitation induced, functionally conserved DUF506 protein is a repressor of root hair elongation in plants. *New Phytol* 233:1153-1171. 10.1111/nph.17862

Ying S, and Scheible WR. 2022. A novel calmodulin-interacting Domain of Unknown Function 506 protein represses root hair elongation in Arabidopsis. *Plant Cell Environ* 45:1796-1812. 10.1111/pce.14316

Yuan G, Zou T, He Z, Xiao Q, Li G, Liu S, Xiong P, Chen H, Peng K, Zhang X, Luo T, Zhou D, Yang S, Zhou F, Zhang K, Zheng K, Han Y, Zhu J, Liang Y, Deng Q, Wang S, Sun C, Yu X, Liu H, Wang L, Li P, and Li S. 2022. SWOLLEN TAPETUM AND STERILITY 1 is required for tapetum degeneration and pollen wall formation in rice. *Plant Physiol* 190:352-370. 10.1093/plphys/kiac307

Zhao M, Ju Y, Zhao B, Li X, Dai L, Qu J, Chu Z, and Ding X. 2020. Transcriptional profiling analysis of OsDT11-mediated ABA-dependent signal pathway for drought tolerance in rice. *Plant Biotechnology Reports* 14:613-626. 10.1007/s11816-020-00637-2

Zhong H, Zhang H, Guo R, Wang Q, Huang X, Liao J, Li Y, Huang Y, and Wang Z. 2019a. Characterization and Functional Divergence of a Novel DUF668 Gene Family in Rice Based on Comprehensive Expression Patterns. *Genes (Basel)* 10. 10.3390/genes10120980

Zhong R, Cui D, and Ye ZH. 2019b. Evolutionary origin of O-acetyltransferases responsible for glucomannan acetylation in land plants. *New Phytol* 224:466-479. 10.1111/nph.15988

Table 1(on next page)

Table 1. Characteristics of *DUF506* genes in rice

Table 1. Characteristics of *DUF506* genes in rice

Gene ID	Subfamily	Locus ID	Size (aa)	Protein MW (Da)	pI	Instability index	Aliphatic index	GRAVY	Chromosome	Subcellular localization
OsDUF50601	II	LOC_Os01g54340	269	29861.74	7.61	47.02	65.76	-0.46	1	Nucleus
OsDUF50602	I	LOC_Os01g68650	293	30429.34	8.97	52.76	75.22	-0.134	1	Chloroplast
OsDUF50603	IIIa	LOC_Os01g74250	337	37616.96	7.25	60.92	86.85	-0.325	1	Chloroplast
OsDUF50604	IIIb	LOC_Os03g06680	307	33011.21	6.68	45.12	80.59	-0.416	3	Chloroplast
OsDUF50605	IIIb	LOC_Os03g58230	405	43737.47	9.12	50.54	68.52	-0.489	3	Chloroplast
OsDUF50606	II	LOC_Os05g44300	306	32169.30	6.76	56.42	62.39	-0.267	5	Nucleus
OsDUF50607	IIIb	LOC_Os07g08390	507	54428.55	9.07	59.08	66.07	-0.481	7	Chloroplast
OsDUF50608	IIIb	LOC_Os10g28210	301	32041.06	6.52	50.73	79.60	-0.347	10	Chloroplast
OsDUF50609	I	LOC_Os11g25020	286	30211.19	9.03	42.62	76.71	-0.108	11	Chloroplast
OsDUF50610	I	LOC_Os11g25040	306	32464.54	5.58	50.32	82.06	-0.176	11	Chloroplast

1

Table 2(on next page)

Table 2. Non-synonymous SNP distribution of OsDUF506s in 3024 rice varieties

Table 2. Non-synonymous SNP distribution of *OsDUF506s* in 3024 rice varieties

Subfamily	Gene ID	No. total SNP	nsSNP ID	Ref allele	Alt allele	Position
I	*OSDUF50602*	2	OSDUF50602.1	C	T	Chr1:39867256
	OsDUF50609	23	OsDUF50609.1	G	T	Chr11:14248387
			OsDUF50609.2	C	A	Chr11:14248411
			OsDUF50609.3	C	T	Chr11:14248432
			OsDUF50609.4	A	G	Chr11:14248446
			OsDUF50609.5	C	T	Chr11:14248498
			OsDUF50609.6	A	G	Chr11:14248539
			OsDUF50609.7	C	G	Chr11:14248616
			OsDUF50609.8	C	G	Chr11:14248636
			OsDUF50609.9	C	T	Chr11:14248665
			OsDUF50609.10	G	T	Chr11:14248737
			OsDUF50609.11	C	T	Chr11:14248861
			OsDUF50609.12	A	C	Chr11:14248932
			OsDUF50609.13	T	C	Chr11:14249448
	OsDUF50610	9	OsDUF50610.1	G	T	Chr11:14267840
			OsDUF50610.2	C	T	Chr11:14267960
			OsDUF50610.3	G	A	Chr11:14268083
			OsDUF50610.4	C	G	Chr11:14268298
			OsDUF50610.5	C	T	Chr11:14268593
			OsDUF50610.6	T	C	Chr11:14268614
II	*OsDUF50606*	3	OsDUF50606.1	G	A	Chr5:25778321
	OSDUF50601	9	OSDUF50601.1	A	G	Chr1:31272839
IIIa	*OSDUF50603*	12	OSDUF50603.1	G	A	Chr1:43017098
			OSDUF50603.2	C	A	Chr1:43018592
IIIb	*OsDUF50608*	7	OsDUF50608.1	G	C	Chr10:14658953
			OsDUF50608.2	G	A	Chr10:14658996
			OsDUF50608.3	C	G	Chr10:14659216
	OsDUF50604	3	OsDUF50604.1	T	C	Chr3:3381185
	OsDUF50607	3	OsDUF50607.1	G	C	Chr7:4305870
	OsDUF50605	7	OsDUF50605.1	T	A	Chr3:33171316

1

Figure 1

Figure 1. Phylogenetic tree of *DUF506* members identified in ten plant species.

The members were divided into I subfamilies shown in different colors. Circles in different colors represent different bootstrap values

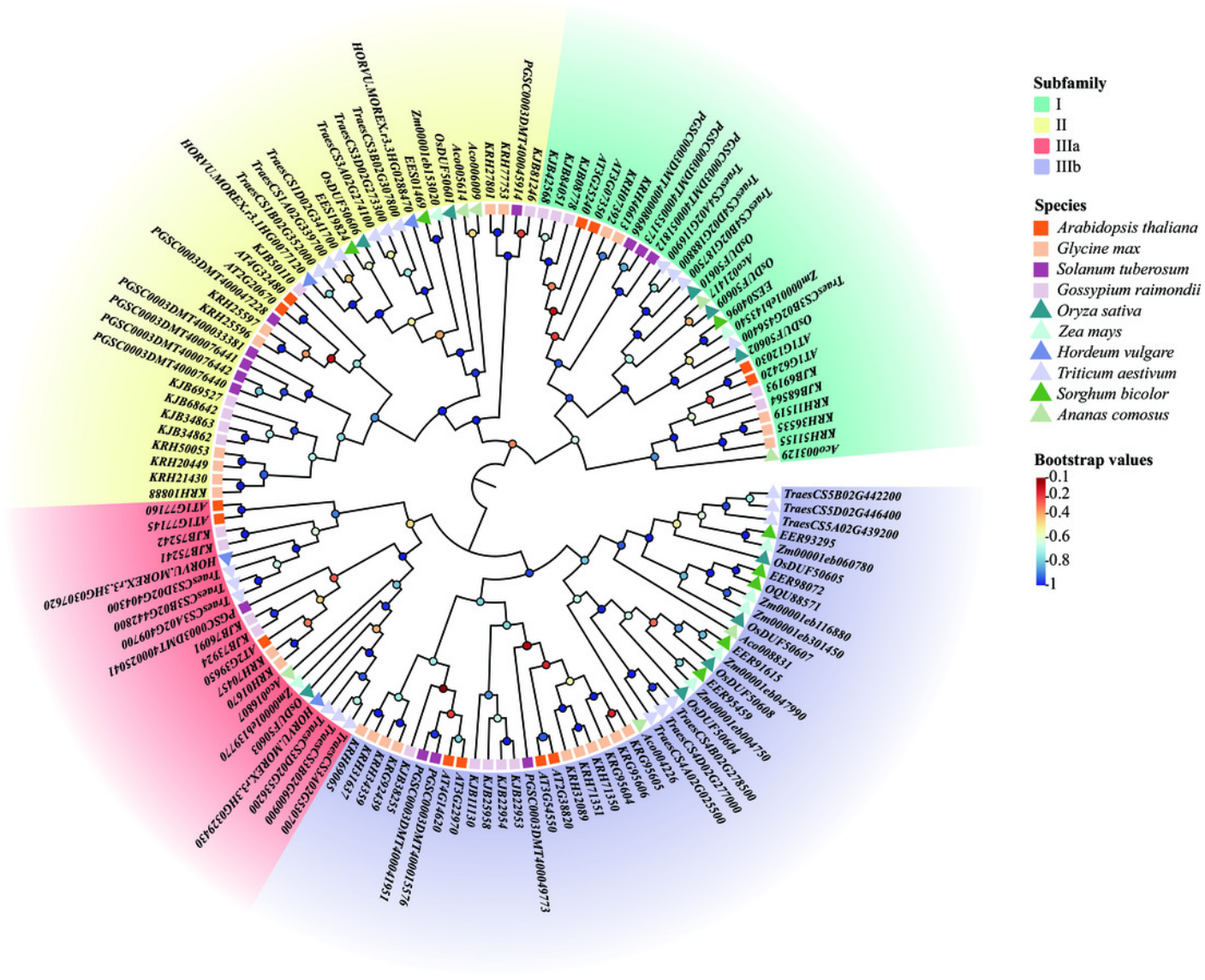


Figure 2

Figure 2. Phylogenetic tree, gene structure, and conserved motifs of *DUF506* in *Oryza sativa* and *Arabidopsis thaliana*

(A) Phylogenetic tree. Four background colors indicate four subfamilies. (B) Exon/intron structures. Yellow bars, green bars, and lines indicate exons, UTRs, and introns respectively. (C)Distributions of conserved motifs. Motifs are shown by 15 different color bars.

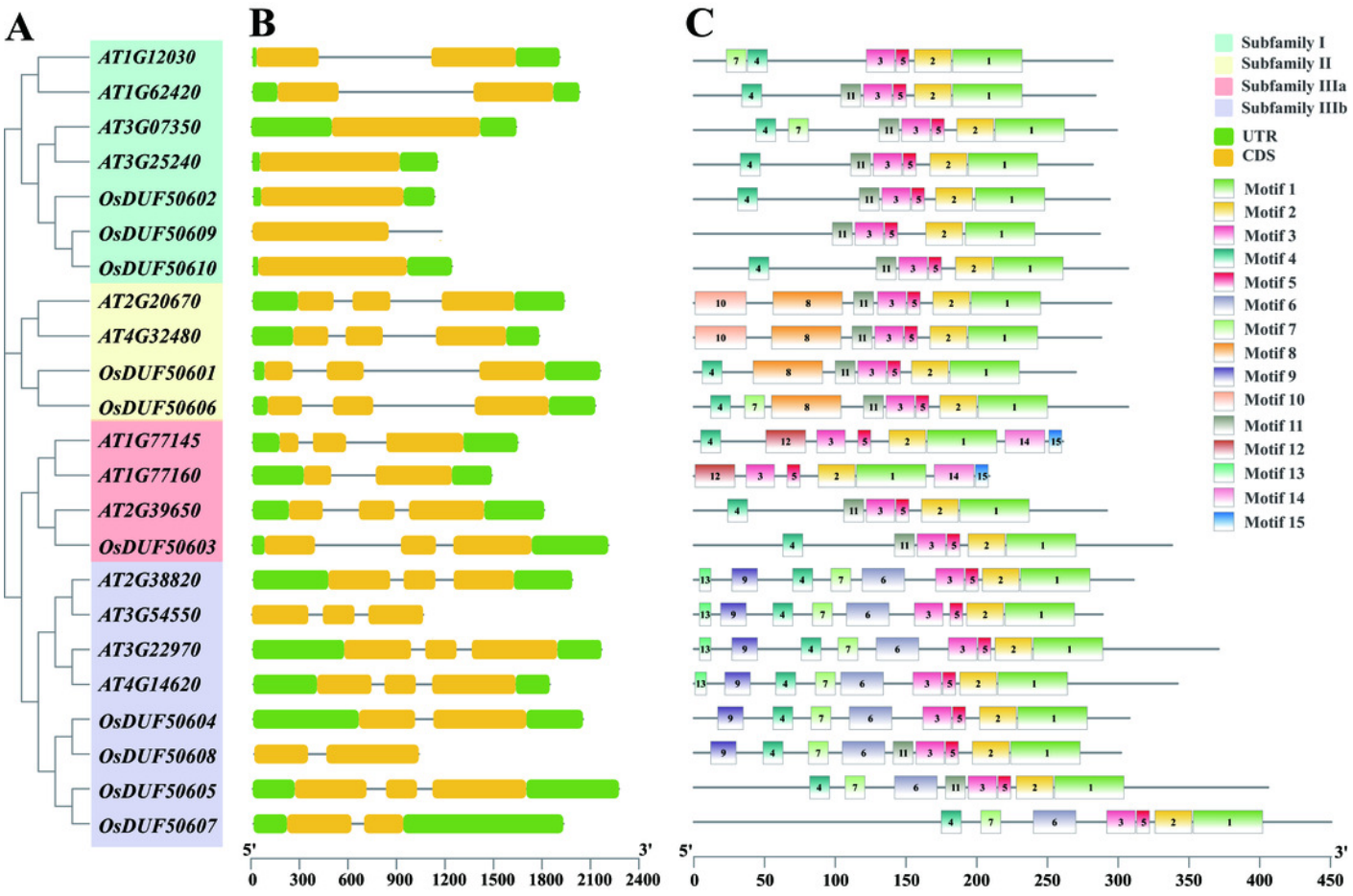


Figure 3

Figure 3. Distribution of CREs in *OsDUF506s* promoters

Predicted CREs with different functions are displayed with boxes in different colors. The light gray bars represent the 2000 bp sequence upstream of *OsDUF506s*.

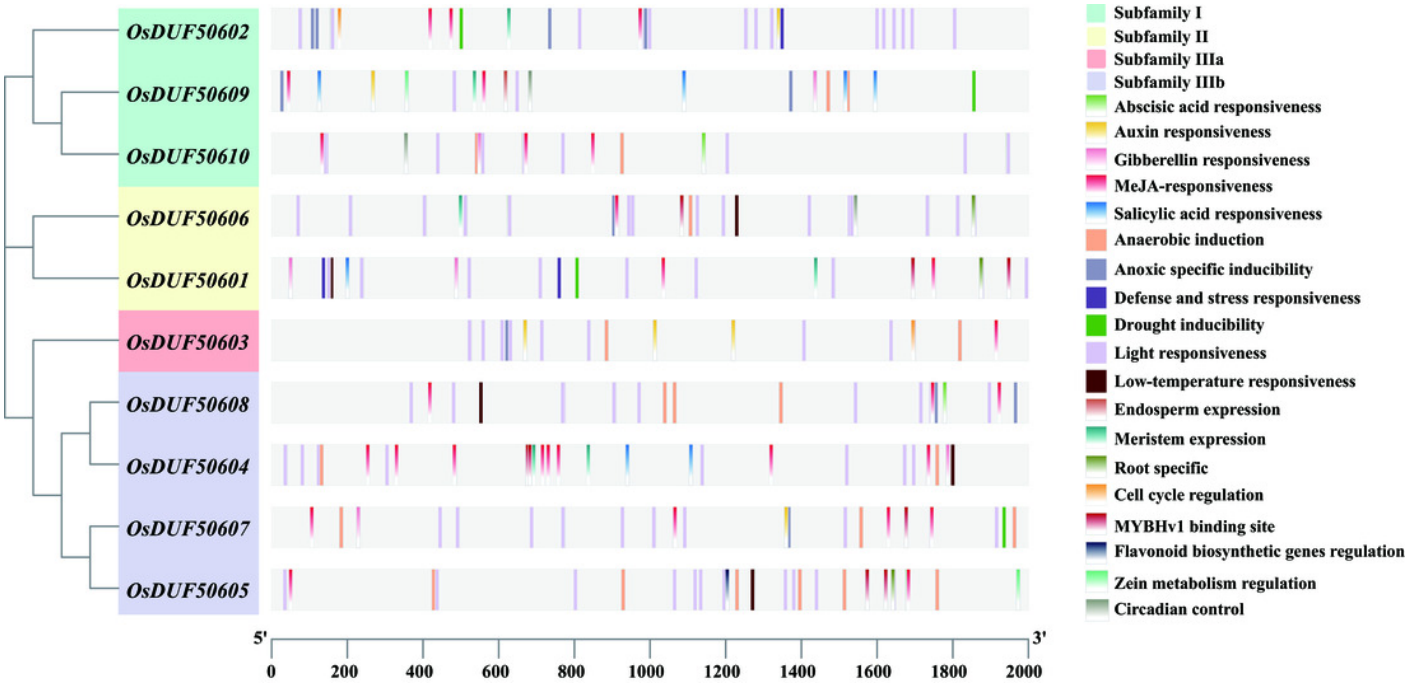


Figure 4

Figure 4. SNP_{GJ}-index values of *OsDUF506s* nsSNPs in 9 subpopulations

Circles in different colors and sizes represent different SNP_{GJ}-index values.

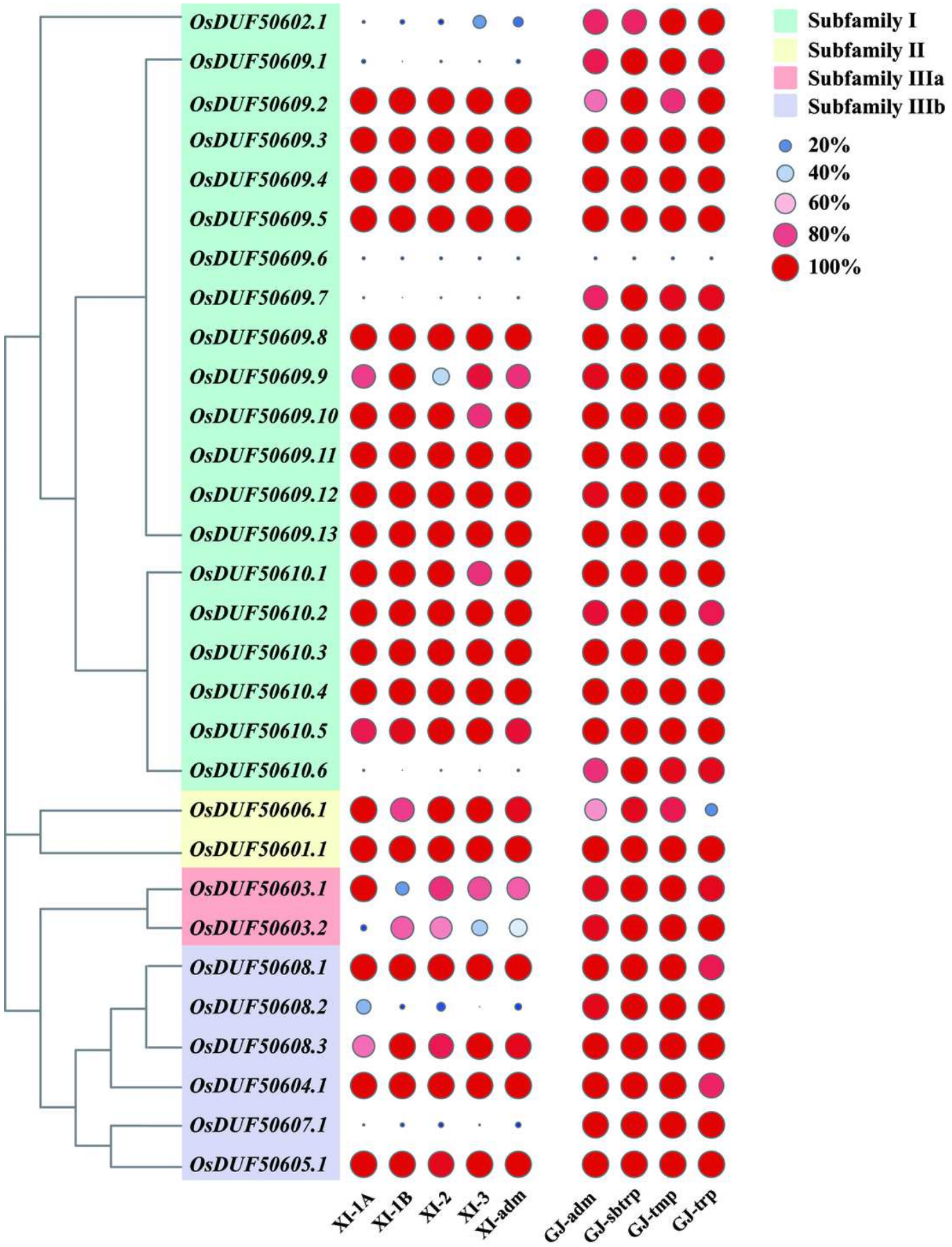

OsDUF50602.1
OsDUF50609.1
OsDUF50609.2
OsDUF50609.3
OsDUF50609.4
OsDUF50609.5
OsDUF50609.6
OsDUF50609.7
OsDUF50609.8
OsDUF50609.9
OsDUF50609.10
OsDUF50609.11
OsDUF50609.12
OsDUF50609.13
OsDUF50610.1
OsDUF50610.2
OsDUF50610.3
OsDUF50610.4
OsDUF50610.5
OsDUF50610.6
OsDUF50606.1
OsDUF50601.1
OsDUF50603.1
OsDUF50603.2
OsDUF50608.1
OsDUF50608.2
OsDUF50608.3
OsDUF50604.1
OsDUF50607.1
OsDUF50605.1
XI-1A
XI-1B
XI-2
XI-3
XI-adm
GJ-adm
GJ-sbtrp
GJ-tmp
GJ-trp
Subfamily I
Subfamily II
Subfamily IIIa
Subfamily IIIb
20%
40%
60%
80%
100%

Figure 5

Figure 5. Synteny analysis of *DUF506*s in rice and *Arabidopsis*

Pink and green bars represent the chromosomes of rice and *Arabidopsis* respectively. The black lines in the colored box show the gene density of the chromosomes. The green lines suggest duplicated gene pairs in *Oryza sativa*, the yellow lines indicate the collinearity between *Oryza sativa* and *Arabidopsis thaliana.*

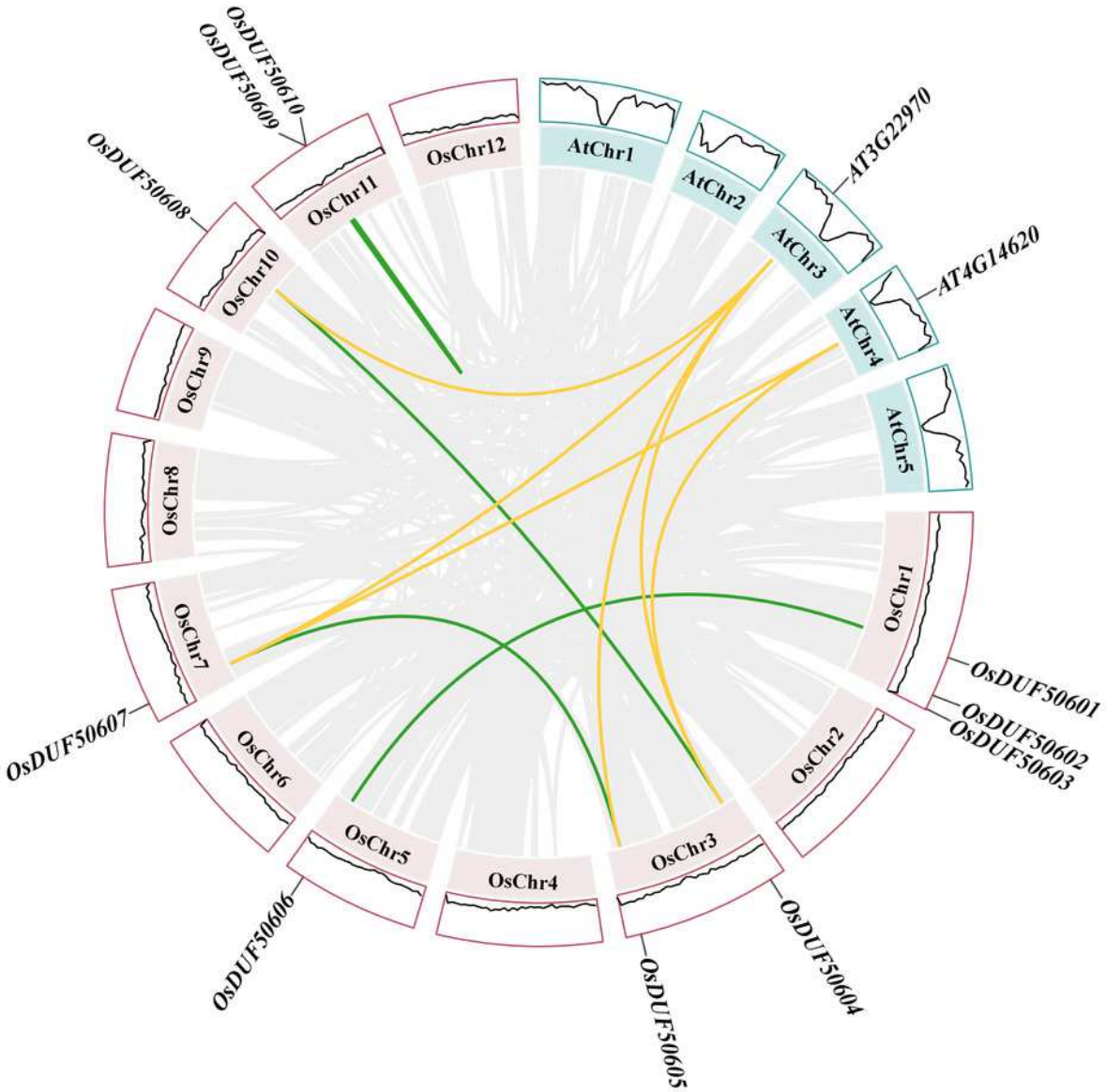

OsDUF50610
OsDUF50609
OsDUF50608
OsDUF50607
OsDUF50606
OsDUF50605
OsDUF50604
OsDUF50603
OsDUF50602
OsDUF50601
AT4G14620
AT3G22970
OsChr12
AtChr1
AtChr2
AtChr3
AtChr4
AtChr5
OsChr11
OsChr10
OsChr9
OsChr8
OsChr7
OsChr6
OsChr5
OsChr4
OsChr3
OsChr2
OsChr1

Figure 6

Figure 6. Synteny analysis of *DUF506s* between *Oryza sativa* and five monocot species

Red lines suggest syntenic gene pairs.

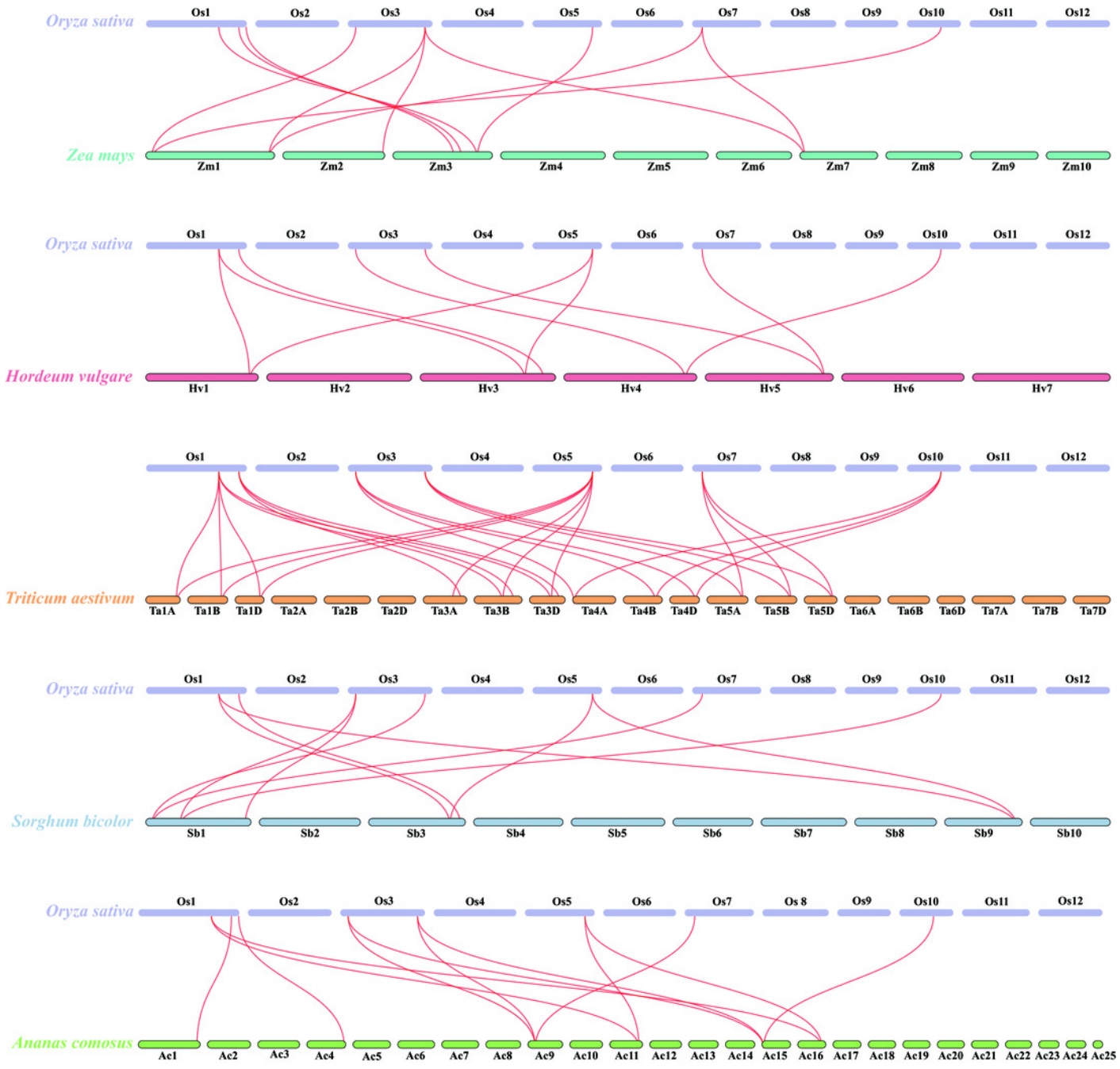


Figure 7

Figure 7. Analysis of predicted miRNA targeting *OsDUF506s*

(A) Identified miRNAs targeting *OsDUF506s*.Circles represent *OsDUF506s*, squares represent the related miRNAs. (B) Expressions of predicted miRNAs in different tissues and under abiotic stresses. The heatmap demonstrates the expression level, the color gradient from blue to red presents increasing expression values.

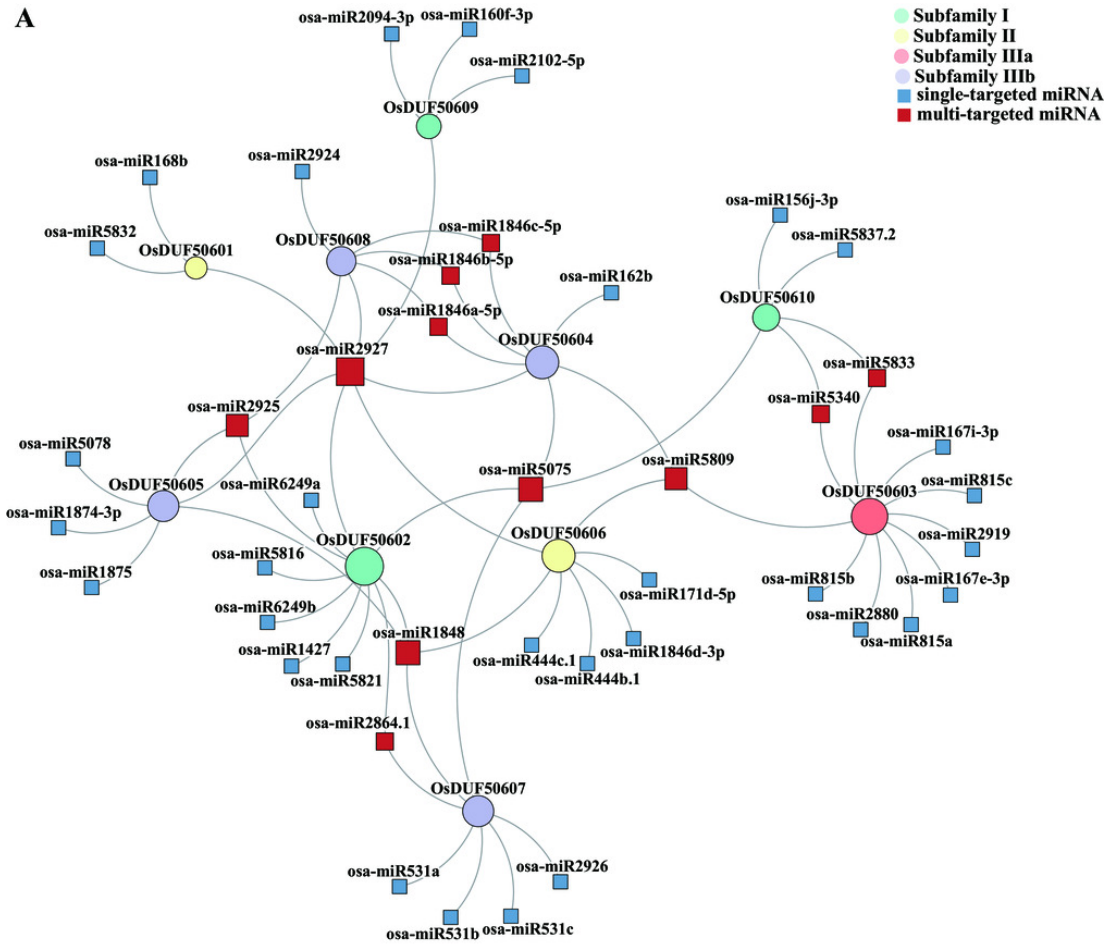

A
Subfamily I
Subfamily II
Subfamily IIIa
Subfamily IIIb
single-targeted miRNA
multi-targeted miRNA
osa-miR2094-3p
osa-miR160f-3p
osa-miR2102-5p
OsDUF50609
osa-miR168b
osa-miR2924
osa-miR5832
OsDUF50601
OsDUF50608
osa-miR1846c-5p
osa-miR1846b-5p
osa-miR162b
osa-miR1846a-5p
OsDUF50604
osa-miR2927
osa-miR156j-3p
osa-miR5837.2
OsDUF50610
osa-miR5833
osa-miR5340
osa-miR2925
osa-miR5078
osa-miR167i-3p
osa-miR5075
osa-miR5809
osa-miR815c
OsDUF50605
osa-miR6249a
OsDUF50603
osa-miR1874-3p
OsDUF50602
OsDUF50606
osa-miR2919
osa-miR5816
osa-miR1875
osa-miR171d-5p
osa-miR815b
osa-miR167e-3p
osa-miR6249b
osa-miR2880
osa-miR1848
osa-miR1427
osa-miR444c.1
osa-miR1846d-3p
osa-miR815a
osa-miR5821
osa-miR444b.1
osa-miR2864.1
OsDUF50607
osa-miR531a
osa-miR2926
osa-miR531b
osa-miR531c
B
Cold: seeding 14 days
Salt: seeding 14 days
Drought: seeding 14 days
Control: seeding 14 days
Endosperm
Embryo
Anther
Leaf
Root
Shoot
osa-miR6249a
osa-miR2094-3p
osa-miR2927
osa-miR5075
osa-miR1848
osa-miR2925
osa-miR5833
osa-miR815a
osa-miR160f-3p
osa-miR168b
osa-miR2864.1
osa-miR2880
osa-miR2924
osa-miR2926
osa-miR531a
osa-miR531b
osa-miR5340
osa-miR1427
osa-miR156j-3p
osa-miR162b
osa-miR167e-3p
osa-miR171d-5p
osa-miR1846a-5p
osa-miR1846d-3p
osa-miR1874-3p
osa-miR1875
osa-miR2102-5p
osa-miR2919
osa-miR444b.1
osa-miR5078
osa-miR5809
osa-miR5816
osa-miR5821
osa-miR5832
osa-miR5837.2
13
10
7
4
1
-2
-5

Figure 8

Figure 8. Expressions of *OsDUF506* members in different tissues

The heatmap demonstrates the expression level, the color gradient from blue to red presents increasing expression values. Grey presents missing data.

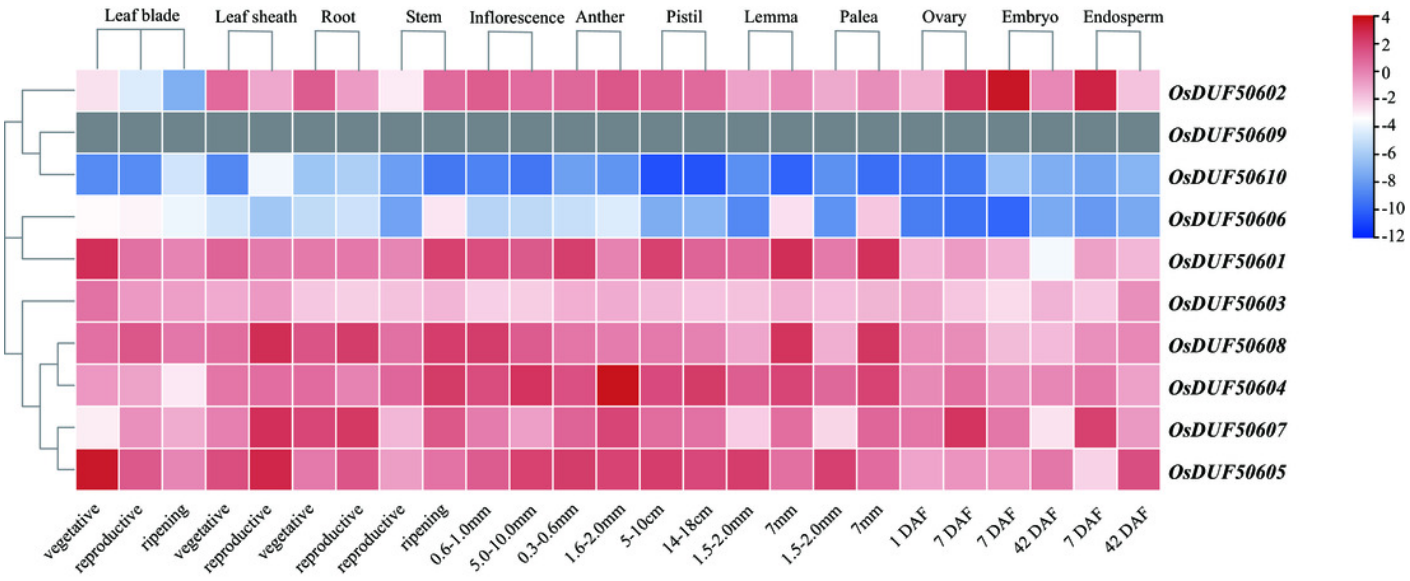


Figure 9

Figure 9. Expressions of *OsDUF506s* induced by plant hormones in root(A) and shoot(B).

The heatmap demonstrates the expression levels, the color gradient from blue to red presents increasing expression values. Grey presents missing data.

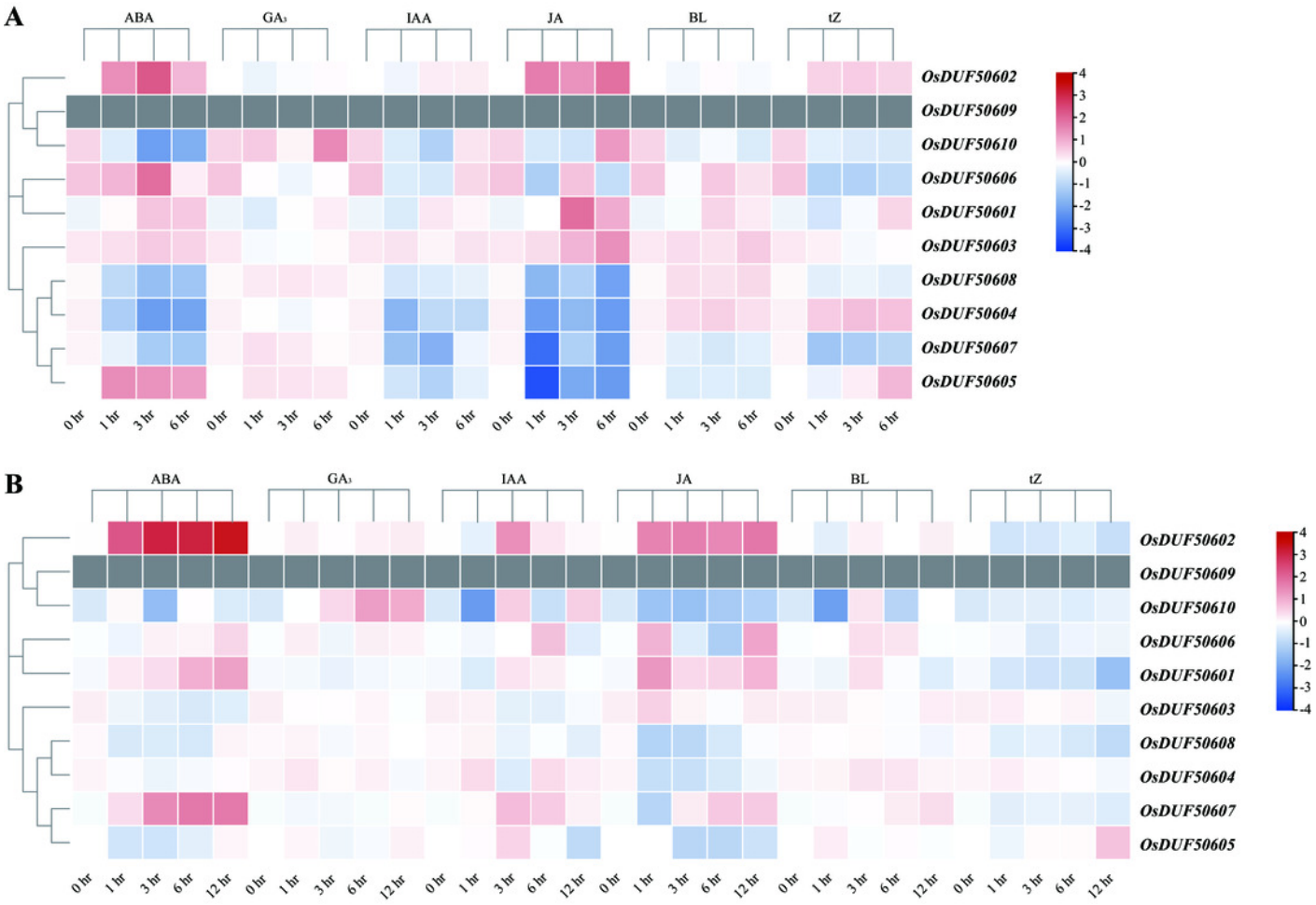
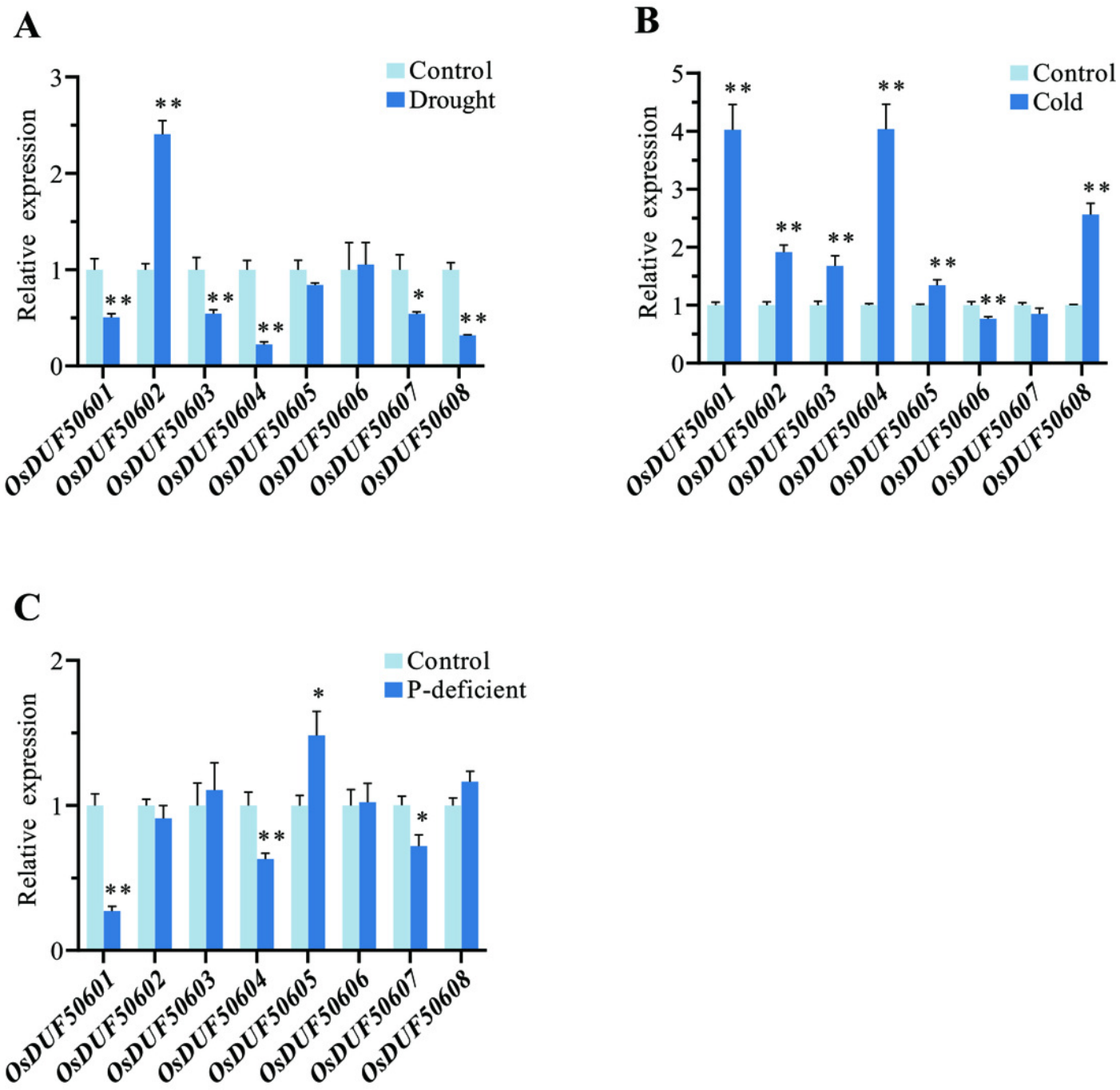


Figure 10

Figure 10. Relative gene expressions of *OsDUF506s* under drought(A), cold(B) and phosphorus-deficient(C) stresses

Data represent the mean±SE of three biological and technical replicates. The significant differences level was analyzed by unpaired t-test(*p□0.05, **p□0.01).


A
Control
Drought
Relative expression
0
1
2
3
OsDUF50601
OsDUF50602
OsDUF50603
OsDUF50604
OsDUF50605
OsDUF50606
OsDUF50607
OsDUF50608
B
Control
Cold
Relative expression
0
1
2
3
4
5
OsDUF50601
OsDUF50602
OsDUF50603
OsDUF50604
OsDUF50605
OsDUF50606
OsDUF50607
OsDUF50608
C
Control
P-deficient
Relative expression
0
1
2
OsDUF50601
OsDUF50602
OsDUF50603
OsDUF50604
OsDUF50605
OsDUF50606
OsDUF50607
OsDUF50608