Plant Tissue Cult. & Biotech. **36**(1): 237-257, 2026 (June) **ISSN 1817-3721, E-ISSN 1818-8745**
DOI: https://doi.org/10.3329/ptcb.v36i1.91182
©Bangladesh Assoc. for Plant Tissue Culture & Biotechnology
PTC&B

WHIRLY Gene Family in *Oryza*: Genome-Wide Identification, Conserved Organellar Roles and Paralog-Specific Abiotic Stress Responses in Rice

Abida Lotif Tamanna, Shajneen Jahan Shoily and Tahmina Islam*

Plant Breeding and Biotechnology Laboratory, Department of Botany, University of Dhaka, Dhaka-1000, Bangladesh

Key words: *WHIRLY* genes, *Oryza*, *Oryza sativa,* Genome-wide identification, Organellar Genome maintenance, Abiotic stress, *In silico* expression analysis

Abstract

The WHIRLY (WHY) protein family plays an indispensable role in regulation of mitochondrial and plastid genomes and stress adaptation across plant species. However, a comprehensive genome-wide characterization of *WHY* gene family among the members of *Oryza* has not yet been studied. Here, 14 *WHY* genes were identified throughout seven *Oryza* species, comprising a pair of members per species present on chromosomes 6 (WHY1-type) and 2 (WHY2-type). Phylogenetic evaluation clustered all WHY members into two principal clades associated with chloroplastic and mitochondrial targeting patterns. Gene architecture characterization identified a predominantly six-exon structural organization in *WHY1* genes, while *WHY2* members exhibited more intricate structure at intronic regions. Motif analysis revealed conserved pattern of six core motifs. Motif 7 was exclusively present in all WHY2 type, while Motif 5 and 10 were found within WHY1 type. All *WHY* genes evolved through segmental or whole genome duplication under purifying selection. Promoter analysis found 33 types of cis-acting regulatory elements; the DRE core element was conserved across all 14 *WHY* gene promoters, while ABA-responsive elements occurred in only a subset. Protein–protein interaction analysis suggested that OsWHY1 and OsWHY2 are involved in the maintenance of organelle genomes, DNA repair, and RNA regulation. *In silico* expression analysis in *O. sativa* revealed that *OsWHY1* is downregulated under drought and heat, while *OsWHY2* is particularly downregulated under ABA treatment, highlighting differential stress response between two genes. These findings present the first genus-wide genomic characterization of the *WHY* gene family across seven *Oryza* species, complemented by *in silico* expression analysis in *O. sativa*. This study provides a foundation for future experimental validation of *WHY* gene functions in rice stress adaptation and crop improvement.

*Author for correspondence: <t.islam@du.ac.bd>.

Introduction

Rice (*Oryza sativa* L.) is of great agricultural and nutritional importance among grain crops around the world, serving as the staple food for more than half of global population and is cultivated in more than 100 countries, with 90% of worldwide harvest originating from Asia (Fukagawa et al. 2019). Worldwide crop production and accumulation are estimated to be 60% greater than in 2005/07 by 2050, demanding immediate advancements in sustainable crop production (Alexandratos and Bruinsma 2012). As the climate is becoming worse day by day, crop production models predict that the yield of most staple crops will reduce significantly, including rice, driven by increased occurrence of drought, flood and rising temperatures, with serious consequences for reliable food supply (Mickelbart et al. 2015).

The WHIRLY (WHY) protein family denotes a unique group of plant-specific single-stranded DNA (ssDNA)-binding proteins that have drawn considerable scientific interest as a promising target in plant breeding programs for enhancing stress mitigation capacity and long-term agricultural productivity under evolving climatic conditions (Desveaux et al. 2005, Taylor et al. 2023). WHY proteins are predominantly localized in chloroplasts, mitochondria, and nuclei, where they perform essential functions in maintaining genomic integrity and regulating nuclear and organelle genomes (Desveaux et al. 2005, Maréchal et al. 2009). Structurally, WHY proteins are characterized by a conserved "KGKAAL" motif within the Whirly DNA-binding domain, which mediates binding to ssDNA (Desveaux et al. 2002, Cappadocia et al. 2012). WHY-like proteins lacking the "KGKAAL" motif have also been identified in green algae such as *Ostreococcus tauri*, suggesting duplication of the original *WHY* gene may have occurred as an ancestral eukaryotic element of chloroplast nucleoids (Krause and Krupinska 2009, Taylor et al. 2023).

In most higher plant species, two types of WHY proteins are present, namely WHY1 and WHY2, whereas members of the Brassicaceae family and *Arabidopsis thaliana* additionally possess a third member, WHY3 (Krause et al. 2005). Loss of *AtWHY1* and *AtWHY3* produces leaf variegation associated with aberrant plastid genome rearrangements, while the triple mutant *why1why3pol1b-1* displays severe yellow-variegated phenotypes with elevated reactive oxygen species accumulation (Lepage et al. 2013). *AtWHY2* regulates mitochondrial genome maintenance, and its overexpression causes reduction in mitochondrial transcripts and acceleration of senescence (Marechal et al. 2008). Beyond model species, WHY proteins have been studied in economically important crops. In maize, *ZmWHY1* is essential for chloroplast biogenesis, and its loss results in albino plants that die at the three to four-leaf stage (Prikryl et al. 2008). In barley, *WHY1* regulates senescence-associated gene expression and its overexpression delays drought-induced leaf senescence (Krupinska et al. 2014, Manh et al. 2023). In tomato (*Solanum lycopersicum*), *SlWHY1* enhances cold tolerance by activating the expression of the *RbcS1* gene (Zhuang et al. 2020) and promotes thermotolerance by

upregulating HSP21.5A under heat stress (Zhuang et al. 2020), while *SlWHY2* is induced by drought stress and its silencing reduces photosynthetic efficiency and disrupts reactive oxygen species homeostasis (Meng et al. 2020). In cassava, MeWHY proteins interact with MeCIPK23 to activate abscisic acid biosynthesis and enhance drought stress responses, while in sweet cherry, *PavWHY1/2* respond to low temperature and are associated with bud dormancy regulation (Yan et al. 2020, Wang et al. 2024).

Genome-wide characterization of the WHY gene family has been conducted in several plant species, including alfalfa, with ten members showing differential expression under abiotic stress (Ruan et al. 2022), wheat (*Triticum aestivum*) and Triticeae species where *TaWHY1-7D* was identified as an osmotic stress regulator (Liu et al. 2023), soybean, with seven *GmWHY* genes responding to low phosphorus stress (Li et al. 2024) and cotton (*Gossypium* spp.), where *GhWHY1-D* was functionally validated as a negative regulator of salt and drought stress responses (He et al. 2025).

Within the *Oryza* genus, *OsWHY1* has been shown to interact with *OsTRX z* and is essential for early chloroplast development, with the *Oswhy1* CRISPR-Cas9 knockout displaying an albino seedling-lethal phenotype (Qiu et al. 2022). However, a systematic genome-wide investigation of the WHY gene family across multiple *Oryza* species, encompassing both cultivated and wild relatives that harbor valuable genetic diversity for crop improvement, remains lacking. The present study addressed this gap by providing the first comprehensive genome-wide characterization of the WHY gene family across seven *Oryza* species, identifying conserved and divergent structural features of *OsWHY1* and *OsWHY2*. This was complemented by *in silico* expression analysis in *O. sativa* to reveal their differential transcriptional responses to abiotic stress.

In this study, we performed comprehensive genome-wide identification and characterization of *WHY* genes across seven *Oryza* species, including the cultivated species *O. sativa* and *O. glaberrima* and five wild relatives (*O. barthii*, *O. brachyantha*, *O. nivara*, *O. punctata*, and *O. rufipogon*) representing AA, BB, and FF genome types, through systematic analysis of physicochemical properties, chromosomal distribution, gene structure, conserved motifs, phylogenetic trees, cis-regulatory elements, collinearity, Ka/Ks ratios, protein interactions, miRNA regulation, and *in silico* expression profiling in *O sativa* under various stress treatments, providing fundamental understanding of functional validation and precision breeding approaches in future studies.

Materials and Methods

For determining all genes within WHY family throughout the seven *Oryza* species, the genome and proteome sequences of seven *Oryza* species were downloaded from Ensembl Plants database (Cunningham et al. 2022), namely *O. sativa* Japonica (IRGSP-1.0), *O. barthii*, *O. brachyantha*, *O. glaberrima*, *O. nivara* , *O. punctata*, and *O. rufipogon*. Protein sequences of three *Arabidopsis thaliana* WHY proteins from the TAIR database (https://www.arabidopsis.org; Reiser et al. 2022) were used as reference queries in BLASTp searches (NCBI BLAST + v2.17.0; E-value $\leq 1 \times 10^{-5}$) against each species

proteome. All candidates were validated using InterProScan (https://www.ebi.ac.uk/interpro/; Mistry et al. 2021) to confirm the conserved WHIRLY domain (PF08536). Confirmed members were named and classified as WHY1 or WHY2 based on their orthology with *Arabidopsis thaliana* reference genes (*AtWHY1* and *AtWHY2*), supported by phylogenetic placement, sequence similarity to reference proteins, chromosomal position, conserved motif composition, and predicted subcellular localization. Physicochemical properties were predicted using ExPASy ProtParam (https://web.expasy.org/protparam; Artimo et al. 2012), and subcellular localization was predicted using WoLF PSORT (https://wolfpsort.hgc.jp/; Horton et al. 2007) and TargetP 2.0 (https://services.healthtech.dtu.dk/services/TargetP-2.0/; Armenteros et al. 2019).

Chromosomal localization of all 14 *WHY* genes was visualized using TBtools-II (v2.482; Chen et al. 2023) with chromosome length and density data from Ensembl Plants GTF files (release 62). For gene structure analysis, GFF3 annotation files were obtained from RAP-DB (rapdb.dna.affrc.go.jp) for *O. sativa* Japonica and from Ensembl Plants for the remaining species. Exon/intron/UTR coordinates were extracted using custom bash scripts and visualized in TBtools-II v2.482.

Conserved protein motifs were identified using MEME Suite v5.5.9 (http://meme.nbcr.net/meme; Bailey et al. 2015)with the following parameters: site distribution, Zero or One Occurrence Per Sequence (ZOOPS); number of motifs, 10; minimum motif width, 6 amino acids; maximum motif width, 50 amino acids; E-value threshold for motif retention, 1×10^{-4}. Domain architecture was analyzed using the NCBI Batch CD-Search tool (E-value ≤0.01) and visualized in TBtools-II v2.482. For phylogenetic analysis, 58 full-length WHY sequences from 24 plant species spanning green algae to angiosperms were retrieved from Ensembl Plants using BLASTp and validated by HMMER v3.4 (E-value $\leq 1 \times 10^{-5}$). Multiple sequence alignment was performed with MAFFT v7 (Katoh et al. 2013), and a Maximum Likelihood tree was constructed using IQ-TREE2 v2.3.6 (Minh et al. 2020) with the Q.plant+F+G4 substitution model (Kalyaanamoorthy et al. 2017) and 1,000 ultrafast bootstrap replicates (Hoang et al. 2018). The tree was annotated using iTOL v6 (https://itol.embl.de/; Letunic and Bork 2021).

To investigate the selection pressure acting on the *WHY* gene pairs, synonymous (Ks) and non-synonymous (Ka) substitution rates were calculated using the Simple Ka/Ks Calculator (Nei-Gojobori method) implemented in TBtools-II v2.482. Whole-genome synteny analysis was performed using MCScanX (v1.1; Wang et al. 2012). Protein sequences from all seven *Oryza* species were subjected to an all-against-all BLASTp search using NCBI BLAST+ (v2.17.0; E-value $\leq 1 \times 10^{-5}$). The resulting BLAST output and genome annotation GFF files were used as input for MCScanX to identify pairwise syntenic blocks. Duplication mode of all *WHY* genes was determined using the MCScanX duplicate-gene classifier. Synteny visualization, including dual synteny ribbon plots and a circular Circos plot, was generated using TBtools-II (v2.482; Chen et al. 2023).

Protein-protein interaction networks for OsWHY1 (Os06g0145800; UniProt Q5VP52) and OsWHY2 (Os02g0158400; UniProt Q0JBW2) were constructed using STRING v12.0 (https://string-db.org; Szklarczyk et al. 2023) with a minimum interaction score of 0.150 (low confidence) to capture the broadest interaction space. Only interactions with a combined score ≥0.400 were retained for visualization and discussion.

Three-dimensional homology models of all 14 OsWHY proteins were built using SWISS-MODEL (https://swissmodel.expasy.org), using the *A. thaliana* WHY1 crystal structure (PDB: 4KOO; 1.88 Å) as template for WHY1 proteins and the *A. thaliana* WHY2 crystal structure (PDB: 4KOP; 1.75 Å) for WHY2 proteins. Model structures were visualized in UCSF ChimeraX v1.11.1.

Co-expression networks for *OsWHY1* and *OsWHY2* were constructed using the ATTED-II database (https://atted.jp/; Obayashi et al. 2018). The top 13 co-expressed genes for *OsWHY1* (Entrez ID: 4340109) and *OsWHY2* (Entrez ID: 4328365) in *Oryza sativa* were retrieved based on Mutual Rank (MR)-based Pearson correlation coefficients and visualized as circular co-expression network figures using the igraph package in R (.5.1). To predict miRNA-mediated regulation, CDS sequences of all 14 *WHY* genes were submitted to psRNATarget (Schema V2; https://www.zhaolab. org/psRNATarget/) against the *O. sativa* miRNA library from miRBase Release 21. The predicted miRNA–target interactions were visualized as a bubble plot using ggplot2 (v3.5.1) in R (v4.5.1).

Cis-acting regulatory elements (CAREs) in 2,000 bp upstream promoter sequences of all 14 *WHY* genes were identified using Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Lescot et al. 2002). The distribution of cis-acting elements across all 14 *WHY* ene promoters was visualized as a schematic diagram using TBtools-II v2.482, and the count matrix was visualized as a heatmap with functional grouping using TBtools-II v2.482.

Tissue-specific expression profiles of *OsWHY1* and *OsWHY2* were retrieved from the Rice Expression Database (RED; https://ngdc.cncb.ac.cn/red/; Xia et al. 2017) across nine tissue types and $\log_2$-transformed ($\log_2$(FPKM+1)). RNA-seq expression analysis under abiotic stress was performed using publicly available datasets from the NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/). Cold stress data (BioProject PRJNA827493; 4°C, 24 h; cv. Nipponbare; 3 replicates/condition), drought and heat data (GSE221542; mild/ severe drought and 30/60 min heat shock; cv. Nipponbare) were processed through a common pipeline: Raw reads were quality-assessed using FastQC (v0.11.9) and trimmed using Trim Galore (v0.6.7). Trimmed reads were aligned to the *O. sativa* IRGSP-1.0 reference genome (Ensembl Plants) using HISAT2 (v2.2.1), and gene-level read counts were quantified using feature Counts (v2.0.1). Differential expression analysis was performed using DESeq2 (v1.38.0) in R (v4.5.1) with Benjamini-Hochberg FDR correction; genes with padj <0.05 and $|\log_2 FC| \geq 1$ was considered significantly differentially expressed. Microarray expression data under ABA treatment were retrieved from GEO accession GSE39429 (Sato et al. 2013, GPL6864 platform, root tissue, 0 min vs. 6 hrs ABA) and analyzed using the limma package (v3.54.0) with empirical Bayes modeling (Ritchie et al. 2015).

Results and Discussion

A total of 14 *WHY* genes were identified across seven *Oryza* species, with exactly two members per species (Fig. 1). BLASTP searches using *Arabidopsis thaliana* WHY protein sequences identified a Chr6 gene as the top hit for *AtWHY1* and *AtWHY3* in each species, while *AtWHY2* matched a distinct Chr2 gene. This confirmed that all WHY1-type genes are located on chromosome 6 and all WHY2-type genes on chromosome 2. The uniform presence of two *WHY* genes in each diploid *Oryza* species is similar to the two-member WHY families reported in other diploid plant species including *Solanum lycopersicum* and *Prunus avium* (Akbudak et al. 2019, Wang et al. 2024). Larger families have also been reported in polyploid species such as *Brassica napus* (6 members), *Glycine max* (7 members), and *Gossypium* species (12 members), where successive genome duplications have expanded the family (Li et al. 2024, Wang et al. 2024, He et al. 2025). All 14 proteins had negative GRAVY values, indicating hydrophilic character along with their nucleic acid-binding functions. Subcellular localization predictions showed similarity between WoLF PSORT and TargetP 2.0 for WHY1-type proteins. Hence, most WHY1 proteins were predicted to be chloroplast-localized, the exception being *OpWHY1*, whose top WoLF PSORT prediction was nuclear (Table 1). However, WoLF PSORT and TargetP 2.0 gave inconsistent predictions for *OnWHY2*, *OpWHY1*, and *OrWHY2*. This may explain why WHY2 proteins can be found in both plastids and mitochondria. The result showed resemblance with observations found in sweet cherry (Wang et al. 2024).

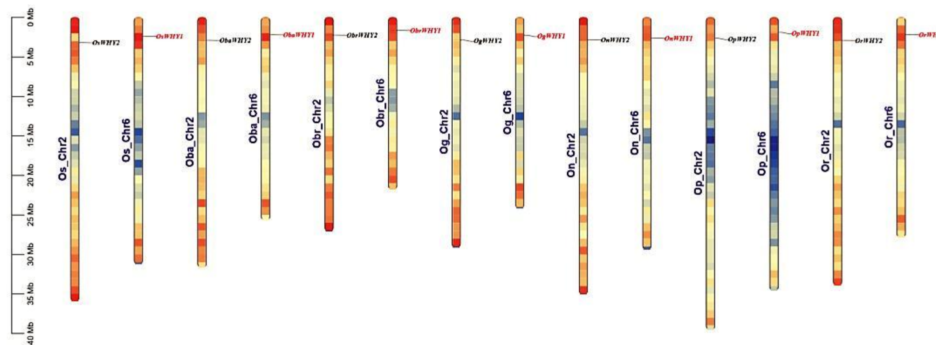


Fig. 1. Chromosomal localization of *WHY* genes across seven *Oryza* species. The physical positions of WHY1-type and WHY2-type genes are mapped on chromosomes 2 and 6 of *O. sativa* (Os), *O. barthii* (Oba), *O. brachyantha* (Obr), *O. glaberrima* (Og), *O. nivara* (On), *O. punctata* (Op), and *O. rufipogon* (Or). The scale is in megabase pairs (Mbp).

In *Arabidopsis*, *AtWHY1* and *AtWHY3* localize to chloroplasts while *AtWHY2* localizes to mitochondria (Maréchal et al. 2009). This pattern is broadly conserved across the OsWHY proteins identified here.

Chromosomal mapping revealed a conserved distribution of *WHY* genes across all seven *Oryza* species (Fig. 1). All *WHY1*-type genes (*OsWHY1*, *ObaWHY1*, *ObrWHY1*, *OgWHY1*, *OnWHY1*, *OpWHY1*, and *OrWHY1*) were localized exclusively on chromosome

6, while all *WHY2*-type genes were located on chromosome 2 without exception across both cultivated species and five wild relatives. The chromosomal positions of these genes are conserved across *Oryza* species. This suggests they were present in the common ancestor of the genus and points to their key roles in organellar genome stability. A similar trend was found in *Brassica*, where *WHY1* genes kept their chromosomal locations across *B. rapa*, *B. oleracea*, and *B. napus* with no major rearrangements during polyploidization (Wang et al. 2024). In contrast, the seven *GmWHY* genes in soybean were spread across five chromosomes, which reflects the higher genomic complexity of the polyploid soybean genome (Li et al. 2024).

Table 1. Genomic information and physicochemical properties of WHY proteins identified across seven *Oryza* species.

Species	Gene Name	Transcript ID	Chr	Genomic location	Strand	Size (aa)	MW (kDa)	GRAVY	pI	WoLF PSORT	TargetP
O. sativa	*OsWHY1*	Os06t0145800-00	Chr 6	2405409-2408049	+	272	30.16	-0.427	9.41	chlo: 13, nucl: 1	cTP
O. sativa	*OsWHY2*	Os02t0158400-01	Chr 2	3182926-3187016	-	228	25.19	-0.396	9.68	mito: 8, nucl: 3.5, cyto_nucl: 2.5	mTP
O. barthii	*ObaWHY1*	OBART06G03050.1	Chr 6	2186702-2189554	+	272	30.17	-0.405	9.3	chlo: 13, nucl: 1	cTP
O. barthii	*ObaWHY2*	OBART02G04450.1	Chr 2	2895803-2899642	-	319	34.66	-0.343	9.6	chlo: 7, nucl: 4, mito: 2	cTP
O. brachyantha	*ObrWHY1*	OB06G12750.1	Chr 6	1642026-1644939	+	269	30.25	-0.368	8.33	chlo: 12, mito: 2	cTP
O. brachyantha	*ObrWHY2*	OB02G13770.1	Chr 2	2233481-2238805	-	294	32.80	-0.474	10.73	chlo: 8, nucl: 3, mito: 3	OTHER
O. glaberrima	*OgWHY1*	ORGLA06G0029900.1	Chr 6	2254601-2257000	+	272	30.18	-0.415	9.3	chlo: 12, nucl: 2	cTP
O. glaberrima	*OgWHY2*	ORGLA02G0041200.1	Chr 2	2858526-2862160	-	228	25.19	-0.396	9.68	mito: 8, nucl: 3.5, cyto_nucl: 2.5	mTP
O. nivara	*OnWHY1*	ONIVA06G03980.1	Chr 6	2624508-2628992	+	271	30.01	-0.412	9.3	chlo: 13, nucl: 1	cTP
O. nivara	*OnWHY2*	ONIVA02G04450.1	Chr 2	2849758-2854252	-	334	36.19	-0.289	9.66	nucl: 7, chlo: 6, cysk: 1	cTP
O. punctata	*OpWHY1*	OPUNC06G02750.1	Chr 6	1886169-1889035	+	320	35.28	-0.38	8.76	nucl: 10, cyto: 3, vacu: 1	cTP
O. punctata	*OpWHY2*	OPUNC02G03720.1	Chr 2	2630851-2634587	-	317	33.86	-0.068	9.68	chlo: 6, vacu: 4, extr: 2	cTP
O. rufipogon	*OrWHY1*	ORUFI06G03010.1	Chr 6	2167393-2170248	+	272	30.16	-0.427	9.41	chlo: 13, nucl: 1	cTP
O. rufipogon	*OrWHY2*	ORUFI02G04590.1	Chr 2	2907531-2911377	-	304	33.51	-0.37	10.04	nucl: 6, chlo: 5, mito: 1	cTP

aa, amino acids; Da, Daltons; pI, isoelectric point; GRAVY, grand average of hydropathicity; WoLF PSORT, subcellular localization predicted by WoLF PSORT server; TargetP, subcellular localization predicted by TargetP 2.0 server; cTP, chloroplast transit peptide; mTP, mitochondrial targeting peptide; chlo, chloroplast; mito, mitochondria; nucl, nucleus; cyto, cytoplasm; vacu, vacuole; extr, extracellular.

To examine the structural features of the identified *WHY* genes, the exon-intron organization of all 14 *OsWHY* genes was analyzed using GFF3 annotation data. The results showed clear differences between the two WHY subfamilies (Fig. 2a, Table 2). WHY1 members showed a mostly six-exon structure across AA-genome and BB-genome species, with *OsWHY1, OgWHY1, ObaWHY1, OpWHY1,* and *OrWHY1* all possessing six exons. *OnWHY1* contains 8 exons and *ObrWHY1* (FF-genome) contains 9 exons. WHY2 members displayed more complex gene structures, with exon counts ranging from 7 (*OsWHY2*) to 10 (*OnWHY2*). This pattern is similar with observations in other plant species like soybean and *Brassica* where *WHY2* genes show more variable intron configurations than their WHY1 counterparts (Li et al. 2024, Wang et al. 2024).

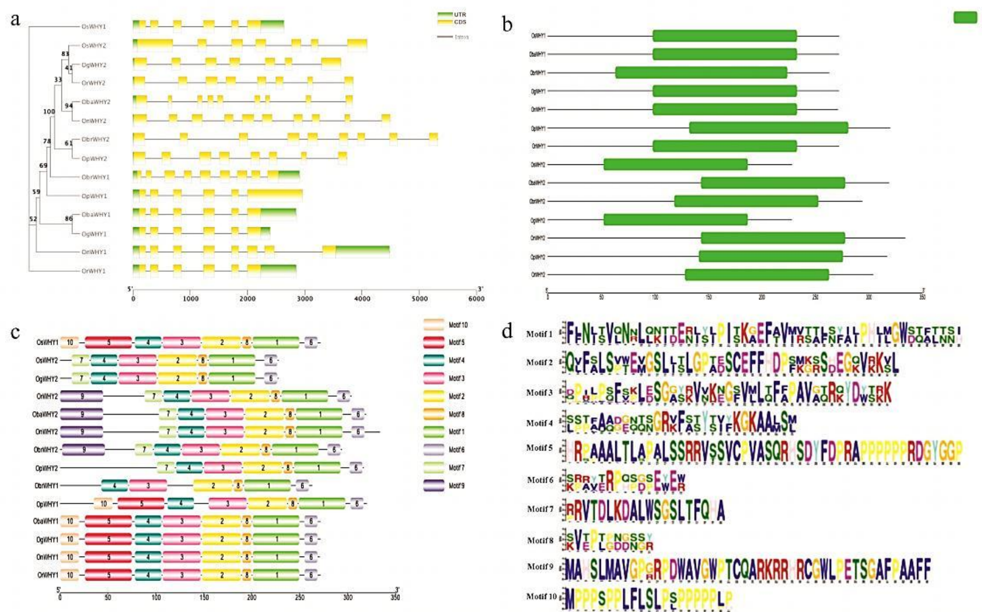


Fig. 2. Gene structure, domain architecture and conserved motif analysis of Whirly gene family of *O. sativa* (Os), *O. barthii* (Oba), *O. brachyantha* (Obr), *O. glaberrima* (Og), *O. nivara* (On), *O. punctata* (Op), and *O. rufipogon* (Or): (a) the exon–intron structure of family members, (b) domains of the Whirly gene family, (c) the 10 conserved motifs have different colors and sequences, and the different motifs correspond to one another in the structure, and (d) logo of the motif.

Domain analysis using the NCBI Batch CD-Search tool confirmed the universal presence of the Whirly domain (PF08536) in every family member (Fig. 2b). These results suggest that the core WHIRLY domain is conserved, while clade-specific functional differences exist. This is consistent with findings in soybean (Li et al. 2024), *Brassica napus* (Wang et al. 2024), and alfalfa (Ruan et al. 2022).

Conserved motif discovery using the MEME Suite identified 10 conserved motifs among all 14 OsWHY proteins (Fig. 2c, 2d). Six core motifs (Motifs 1, 2, 3, 4, 6, and 8) were shared universally across all 14 proteins in both the WHY1 (chloroplastic) and

WHY2 (mitochondrial) clades, constituting the structural and functional core of the WHIRLY domain (PF08536). Motifs 5 and 10 were present in six of the seven WHY1 members but absent in all WHY2 members and *ObrWHY1* (FF-genome). This identifies them as WHY1-associated motifs that may contribute to chloroplast-specific functions. Motif 7 showed a strictly complementary distribution, present in all seven WHY2 members but absent from all WHY1 members. This suggests its contribution to mitochondria-specific functions including mitochondrial DNA copy number regulation, as documented for *AtWHY2* (Golin et al. 2020).

Table 2. Exon counts of *OsWHY* genes across seven *Oryza* species.

Gene Name	Species	Clade	Exon Count	Genome Type
OsWHY1	*O. sativa*	WHY1	6	AA (japonica)
OsWHY2	*O. sativa*	WHY2	7	AA (japonica)
ObaWHY1	*O. barthii*	WHY1	6	AA (wild)
ObaWHY2	*O. barthii*	WHY2	9	AA (wild)
ObrWHY1	*O. brachyantha*	WHY1	9	FF (wild)
ObrWHY2	*O. brachyantha*	WHY2	9	FF (wild)
OgWHY1	*O. glaberrima*	WHY1	6	AA (glaberrima)
OgWHY2	*O. glaberrima*	WHY2	8	AA (glaberrima)
OnWHY1	*O. nivara*	WHY1	8	AA (wild)
OnWHY2	*O. nivara*	WHY2	10	AA (wild)
OpWHY1	*O. punctata*	WHY1	6	BB (wild)
OpWHY2	*O. punctata*	WHY2	9	BB (wild)
OrWHY1	*O. rufipogon*	WHY1	6	AA (wild)
OrWHY2	*O. rufipogon*	WHY2	8	AA (wild)

To understand the evolutionary origin of *Oryza* WHY proteins, a maximum likelihood tree was built using protein sequences from 24 plant species (Fig. 3). The phylogenetic tree grouped into two main clades. Group A comprised all WHY1-type chloroplast-targeted proteins, while Group B encompassed all WHY2-type mitochondria-targeted proteins. This clear division is consistent with findings across multiple plant species (Akbudak and Filiz 2019, Ruan et al. 2022, Li et al. 2024, Wang et al. 2024). Within Group A, Subgroup A2 contained all monocot WHY1 proteins, with the seven *Oryza* WHY1 sequences forming a tightly clustered clade alongside WHY1 proteins of *Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon*, *Triticum aestivum*, and *Musa acuminata*, confirming monocot WHY1 monophyly. Subgroup A1 comprised all dicot WHY1 proteins. *AtWHY3*, a Brassicaceae-specific member, also grouped here with *AtWHY1* indicating its known role in plastid genome stability (Maréchal et al. 2009). Within Group B, the seven *Oryza* WHY2 sequences formed a tight cluster in Subgroup B2. This occurrence is consistent with their conserved mitochondrial function. The *Ginkgo biloba* WHY sequences formed a separate clade between Groups A and B, supporting the idea that WHY1/WHY2 sub functionalization arose before or alongside the emergence of angiosperms.

Fig. 3. Phylogenetic analysis of WHIRLY (WHY) proteins from 24 plant species. Bootstrap values ≥70% are shown at nodes. Branch colors indicate phylogenetic groupings: red, Group A2 (monocot WHY1); orange, Group A1 (dicot WHY1); teal, Group B2 (monocot WHY2); blue, Group B1 (dicot WHY2); brown, gymnosperm (*Ginkgo biloba*); grey, outgroup. The pink and blue background shading denotes Group A (WHY1, chloroplast-targeted) and Group B (WHY2, mitochondria-targeted), respectively.

The Ka/Ks ratio values for the pairwise comparisons of *WHY1*-type genes were between 0 and 0.683, whereas those for pairwise comparison of *WHY2*-type genes were between 0 and 0.450. All calculable values were below 1.0 (two pairs were incalculable owing to zero substitution rates; Table 3), which suggests that the *WHY* genes have undergone purifying selection in *Oryza* lineage (Table 3). Similar observations have been made in soybean, wheat, and rapeseed (Liu et al. 2023, Li et al. 2024, Wang et al. 2024). According to MCScanX analysis, all 14 *WHY* gene loci were classified as WGD/segmental duplicates (Table 3). This indicates that WHY1 and WHY2 originated from an ancient whole-genome or segmental duplication event that predated *Oryza* diversification. However, the gene pairs in Table 3 are interspecific orthologs, not paralogs. Their Ka/Ks values reflect sequence divergence due to speciation. The original duplication event that gave rise to WHY1 and WHY2 as distinct paralogs was the ancestral split into chloroplast-targeted WHY1 and mitochondria-targeted WHY2 lineages, which occurred before the diversification of *Oryza* species.

Table 3. Ka/Ks ratios of homologous *WHY* gene pairs and duplication classification in *Oryza* species.

Seq 1	Seq 2	Ka	Ks	Ka/Ks	Type of Mutation/Evolution	Duplication
WHY1 Orthologous Pairs (*OsWHY1* vs other species WHY1)						
OsWHY1	*ObaWHY1*	0.003296	0.004827	0.6828	Purifying	WGD/Segmental
OsWHY1	*ObrWHY1*	0.046966	0.205058	0.2290	Purifying	WGD/Segmental
OsWHY1	*OgWHY1*	0.004949	0.009685	0.5110	Purifying	WGD/Segmental
OsWHY1	*OnWHY1*	0.001647	0.000000	N/A	Purifying*	WGD/Segmental
OsWHY1	*OpWHY1*	0.009963	0.191862	0.0519	Purifying	WGD/Segmental
OsWHY1	*OrWHY1*	0.000000	0.004819	0.0000	Purifying	WGD/Segmental
WHY2 Orthologous Pairs (*OsWHY2* vs other species WHY2)						
OsWHY2	*ObaWHY2*	0.025724	0.063984	0.4020	Purifying	WGD/Segmental
OsWHY2	*ObrWHY2*	0.031290	0.142398	0.2197	Purifying	WGD/Segmental
OsWHY2	*OgWHY2*	0.000000	0.012526	0.0000	Purifying	WGD/Segmental
OsWHY2	*OnWHY2*	0.025728	0.057191	0.4499	Purifying	WGD/Segmental
OsWHY2	*OpWHY2*	0.035589	0.098085	0.3628	Purifying	WGD/Segmental
OsWHY2	*OrWHY2*	0.000000	0.000000	NaN	Undetermined	WGD/Segmental

Ka = non-synonymous substitution rate; Ks = synonymous substitution rate. Purifying* = Ka/Ks incalculable but Ka = 0, Undetermined = Ka/Ks incalculable due to synonymous site saturation; Duplication type: WGD/Segmental = whole genome or segmental duplication.

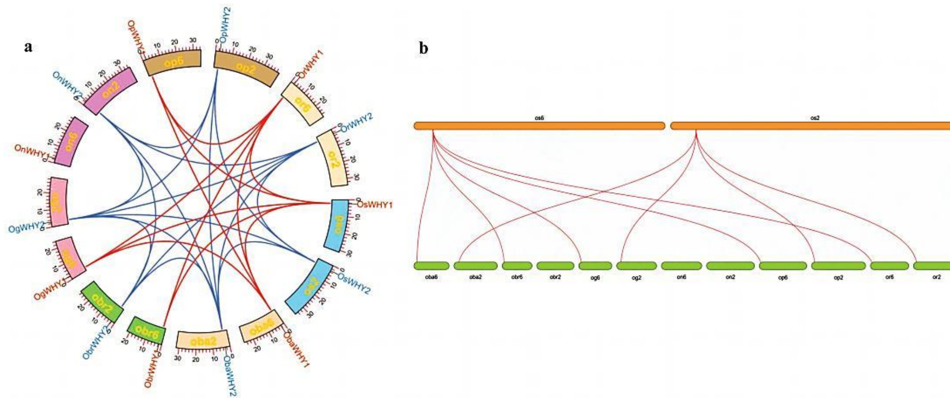


Fig. 4. Collinearity analysis of *WHY* genes in *Oryza* species: (a) circos plot showing collinear connections of WHY1 (red lines, Chr6) and WHY2 (blue lines, Chr2) genes across all seven *Oryza* species. Each colored arc represents chromosome 6 or 2 of one species. Species chromosome color codes: *O. sativa* (blue), *O. rufipogon* (tan/orange), *O. barthii* (light orange), *O. brachyantha* (green), *O. glaberrima* (pink), *O. nivara* (purple), *O. punctata* (peach), and (b) dual synteny ribbon plot comparing *O. sativa* chromosomes 6 and 2 with the corresponding chromosomes of six *Oryza* species. Red lines indicate conserved collinear positions of *WHY* genes; grey ribbons represent genome-wide syntenic gene pairs.

Whole-genome collinearity analysis identified 2,882 pairwise collinear blocks, covering 197,740 collinear gene pairs out of 240,729 genes analyzed (82.14%). The Circos plot showed that all *WHY1*-type genes were consistently located on chromosome 6 and

all *WHY2*-type genes on chromosome 2 across all seven *Oryza* species. This indicates complete conservation of chromosomal location throughout the *Oryza* genus (Fig. 4a). This conservation was further confirmed by the dual synteny ribbon plot where *OsWHY1* and *OsWHY2* were detected within conserved collinear blocks in all pairwise comparisons between *O. sativa* and the six other species (Fig. 4b). The exceptional conservation of *WHY* gene synteny across approximately 15 million years of *Oryza* evolution is consistent with the essential and non-redundant roles of WHY1 and WHY2 in organellar genome maintenance. Similar collinearity has been reported in *Brassica napus* and soybean (Li et al. 2024, Wang et al. 2024).

WHIRLY proteins are known to function within larger protein complexes. To understand which cellular processes OsWHY1 and OsWHY2 are involved in, a protein-protein interaction (PPI) network analysis was done (Fig. 5a). The strongest interaction was between OsWHY1 and OsWHY2. This pattern was also observed in *Arabidopsis* and maize, where WHY1 and WHY2 cooperate to maintain organelle genome stability (Maréchal and Brisson 2010; crystal structures of DNA-Whirly complexes have established their role as organizational scaffolds for genome maintenance complexes in organelles (Cappadocia et al. 2010), and WHY1, WHY2, and WHY3 are involved in the initial phase of homologous recombination in both plastids and mitochondria (Maréchal and Brisson 2010). Two RecA family proteins were also identified as interaction partners of OsWHY1 and OsWHY2. This is biologically relevant given that AtWHY1 and AtWHY3 are known to suppress illegitimate recombination in the plastid genome (Maréchal et al. 2009). The DNA helicase RTEL1 and a CRS1 homolog were also identified. CRS1 is associated with chloroplast intron splicing, suggesting that OsWHY1 may have a role in chloroplast gene expression beyond DNA binding. PTAC12, a component of the plastid RNA polymerase complex, was also identified as an interaction partner, further linking OsWHY proteins to chloroplast transcription. Collectively, the PPI network analysis positions OsWHY1 and OsWHY2 within a functional interaction network centered on organelle genome maintenance, DNA repair, and RNA processing, consistent with their essential roles in plastid and mitochondrial genome stability.

To assess whether the functional conservation of WHIRLY proteins is reflected at the structural level across *Oryza* species, homology modeling was performed for all 14 *Oryza* WHY proteins. The results yielded high-confidence structural models for WHY1 and WHY2 proteins. All 14 models adopted homo-tetrameric quaternary assemblies, consistent with the experimentally established oligomeric state of WHIRLY proteins (Fig. 5b). The characteristic beta-sheet-rich Whirly domain comprising eight antiparallel beta-strands arranged in two perpendicular groups connected by an alpha-helix was conserved across all models. Structural differences between the two subfamilies were confined to the variable N-terminal transit peptide and C-terminal regions, while the core Whirly domain was highly conserved across all 14 models. The high structural similarity across all seven *Oryza* species for both subfamilies support strong purifying selection acting on the WHIRLY fold, consistent with the Ka/Ks data.

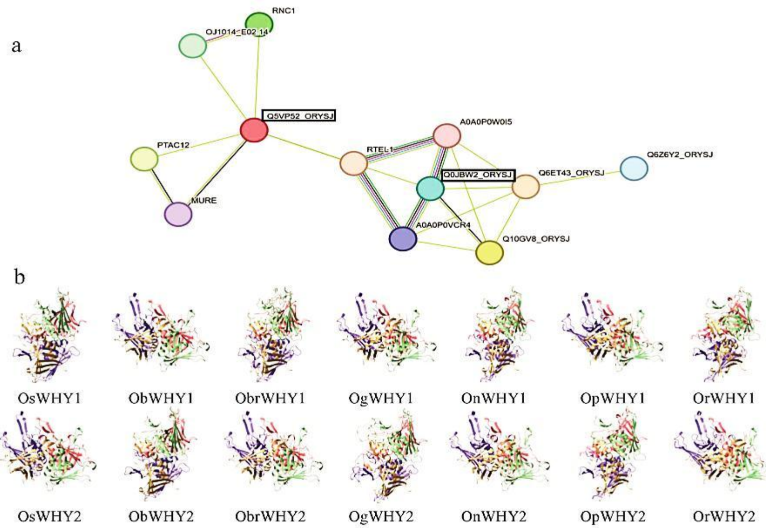
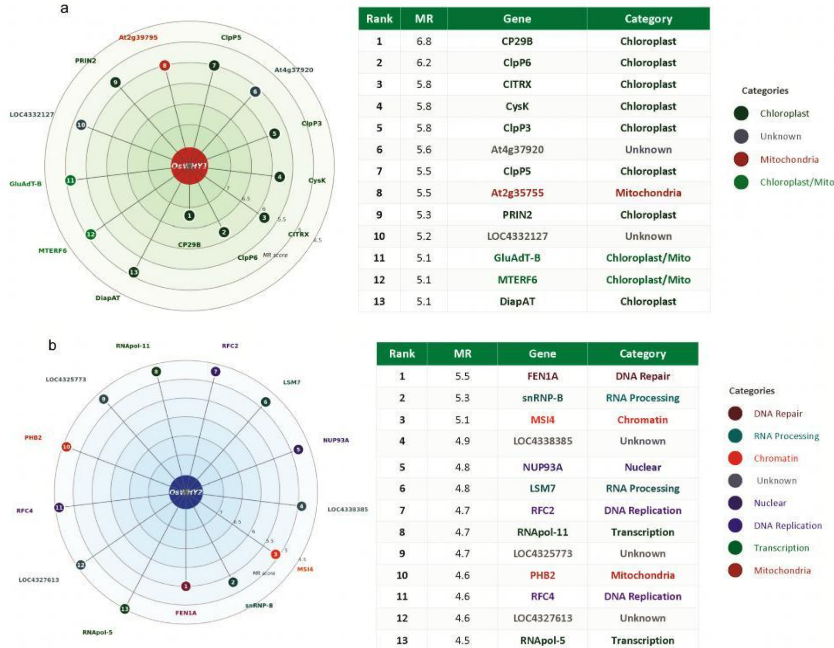


Fig. 5. Protein-protein interaction network and three-dimensional structural models of OsWHY proteins: (a) PPI network of OsWHY1 and OsWHY2 constructed using STRING v12.0. OsWHY1 (Q5VP52_ORYSJ) is shown in red (marked in box) and OsWHY2 (Q0JBW2_ORYSJ) in cyan (marked in box). Edge colors represent the type of interaction evidence, and (b) predicted three-dimensional structural models of all 14 *Oryza* WHY proteins across seven *Oryza* species generated by homology modeling. Upper row: WHY1-type (chloroplastic) members; lower row: WHY2-type (mitochondrial) members.



Rank	MR	Gene	Category
1	6.8	CP29B	Chloroplast
2	6.2	ClpP6	Chloroplast
3	5.8	CITRX	Chloroplast
4	5.8	CysK	Chloroplast
5	5.8	ClpP3	Chloroplast
6	5.6	At4g37920	Unknown
7	5.5	ClpP5	Chloroplast
8	5.5	At2g35755	Mitochondria
9	5.3	PRIN2	Chloroplast
10	5.2	LOC4332127	Unknown
11	5.1	GluAdT-B	Chloroplast/Mito
12	5.1	MTERF6	Chloroplast/Mito
13	5.1	DiapAT	Chloroplast

Rank	MR	Gene	Category
1	5.5	FEN1A	DNA Repair
2	5.3	snRNP-B	RNA Processing
3	5.1	MSI4	Chromatin
4	4.9	LOC4338385	Unknown
5	4.8	NUP93A	Nuclear
6	4.8	LSM7	RNA Processing
7	4.7	RFC2	DNA Replication
8	4.7	RNApol-11	Transcription
9	4.7	LOC4325773	Unknown
10	4.6	PHB2	Mitochondria
11	4.6	RFC4	DNA Replication
12	4.6	LOC4327613	Unknown
13	4.5	RNApol-5	Transcription

Fig. 6. Co-expression network analysis of *OsWHY1*: (a) and *OsWHY2*, and (b) in *Oryza sativa* using the ATTED-II database. The top 13 most correlated genes are shown with Mutual Rank (MR)-based Pearson correlation coefficients.

Co-expression analysis was carried out to explore the functional gene networks of *OsWHY1* and *OsWHY2*. This revealed distinct gene associations for each protein in *O. sativa*. *OsWHY1* was co-expressed mainly with chloroplast-localized proteins (Fig. 6a). These included RNA-binding proteins, subunits of the ATP-dependent Clp protease complex involved in plastid protein quality control under stress (Nishimura and van Wijk 2015), and thioredoxin-like proteins linked to redox regulation. This pattern places *OsWHY1* within plastid proteostasis networks. In contrast, *OsWHY2* was co-expressed with genes enriched for DNA replication and repair functions (Fig. 6b). Key co-expressed partners included flap endonuclease FEN1A, involved in Okazaki fragment processing and base excision repair (Liu et al. 2004), replication factor C subunits, and mitochondria-localized prohibitin-2 (Fig 9b). This supports a functional link between *OsWHY2* and active DNA replication machinery, consistent with co-expression patterns reported for WHY2 in soybean (Li et al. 2024). Collectively, these findings provide independent transcriptomic confirmation of the distinct and complementary roles of *OsWHY1* and *OsWHY2* in chloroplastic and mitochondrial genome maintenance, respectively.

To investigate whether miRNA-mediated post-transcriptional regulation also contributes to their differential control, a screening of all 14 WHY coding sequences against published *O. sativa* miRNAs using psRNATarget was done. This identified predicted miRNA-target interactions involving 24 distinct miRNA families across the seven species (Fig. 7). *WHY1* genes were predominantly targeted by stress-responsive miRNA families. Notably, osa-miR395 targeted *WHY1* genes in six of the seven species. This is consistent with its reported targeting of *WHY* genes in tomato (Akbudak and Filiz 2019). *WHY2* genes were targeted by a distinct set of miRNA families. Several of these interactions were predicted to operate through translation inhibition rather than cleavage. This suggests divergent post-transcriptional regulatory mechanisms between the two subfamilies. Notably, osa-miR2919 was the only miRNA family predicted to target both *WHY1* and *WHY2* genes across multiple species. This points to a conserved role in coordinating expression across the entire WHY family. Overall, the contrasting miRNA regulatory landscapes of the two subfamilies are consistent with their functional divergence and may contribute to the differential expression patterns observed between WHY1 and WHY2 under abiotic stress.

To understand how *WHY* gene expression is regulated, the 2-kb promoter regions of all 14 *OsWHY* genes were analyzed for cis-acting regulatory elements. The results identified diverse cis-acting regulatory elements (CAREs) associated with hormone responses (MeJA, ABA, auxin, gibberellin, salicylic acid), light responsiveness, stress responses (drought, defense, low temperature, anaerobic induction), and developmental regulation (meristem expression, seed-specific regulation), with differential distribution organization found between two types of genes (Fig. 8a). Heatmap analysis revealed a total 33 type cis-regulatory elements under the previous classification in the promoter region of *WHY* genes where light-responsive elements were the most abundantly distributed across all 14 promoters (Fig. 8b). This aligns with the known role of WHY1 in

chloroplast genome maintenance and its link to photosynthetic activity. A similar abundance of light-responsive elements has been reported in *Brassica napus WHY* gene promoters (Wang et al. 2024). Among hormonal elements, MeJA-responsive elements were found in several promoters, especially in *ObrWHY2* and *OpWHY2*, suggesting possible jasmonate-mediated regulation of organellar genome maintenance. ABA-responsive elements (ABRE) were also found in *ObrWHY1* and *OpWHY2* promoters. Similar hormonal elements have been reported in soybean *GmWHY* promoters (Li et al. 2024). The DRE core element, which is the binding site for DREB/CBF transcription factors and mediates responses to cold and drought (Stockinger et al. 1997), was present in all 14 *WHY* gene promoters. This suggests that *WHY* gene expression across *Oryza* species may be coordinately regulated under abiotic stress through the DREB/CBF pathway.

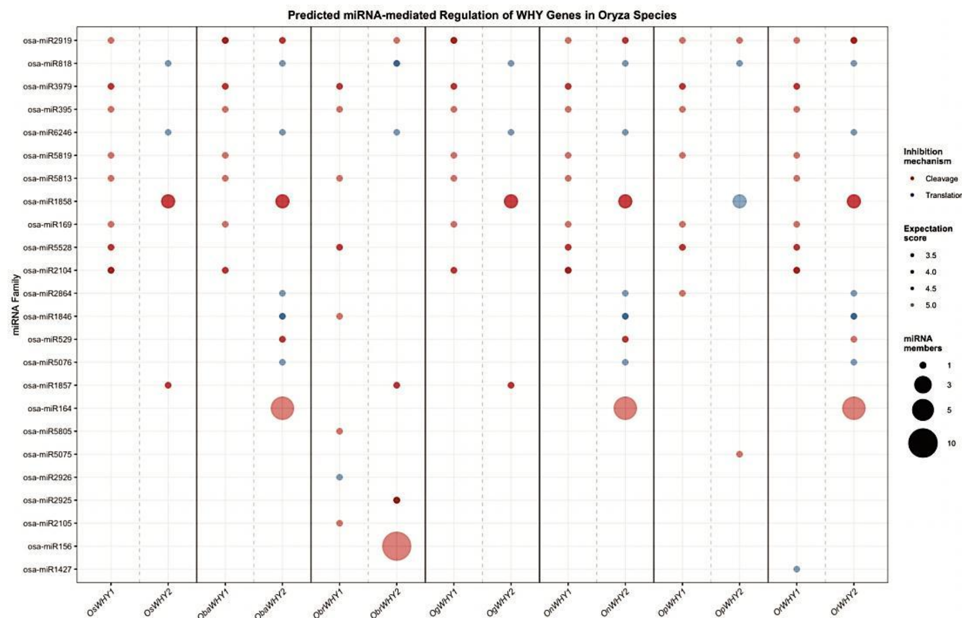


Fig. 7. Predicted miRNA-mediated regulation of *WHY* genes across seven *Oryza* species. Bubble plot showing predicted miRNA-target interactions identified using psRNATarget. The x-axis represents individual *WHY* genes grouped by species and the y-axis represents miRNA families. Bubble color indicates the inhibition mechanism (red, cleavage; blue, translation inhibition) and bubble size indicates the number of miRNA family members.

Tissue expression analysis using the RED database revealed distinct patterns between *OsWHY1* and *OsWHY2* across nine tissue types. *OsWHY1* showed moderate and uniform expression, with highest levels in seed (FPKM = 9.63), anther (8.71), and leaf (8.04). This is consistent with its chloroplast localization and role in plastid genome maintenance in photo synthetically active tissues (Fig. 9). *OsWHY2* showed higher overall expression, with peak levels in callus (FPKM = 68.01) and anther (67.45). This

reflects the continuous demand for mitochondrial genome maintenance during cell division, consistent with the role of WHY2 in mitochondrial DNA copy number regulation (Golin et al. 2020).

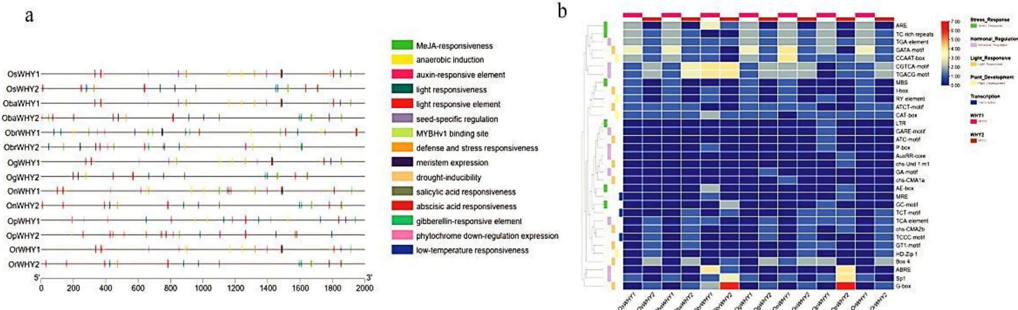


Fig. 8. Analysis of cis-acting regulatory elements in the 2000 bp upstream promoter regions of 14 *WHY* genes across seven *Oryza* species: (a) schematic diagram showing the positions of identified cis-elements along each promoter sequence (5′ to 3′) and classified into different functional categories, and (b) heatmap showing the number of each cis-element per gene promoter, with hierarchical clustering of cis-elements shown on the left. *WHY1*-type and *WHY2*-type genes are indicated by the color bar at the top.

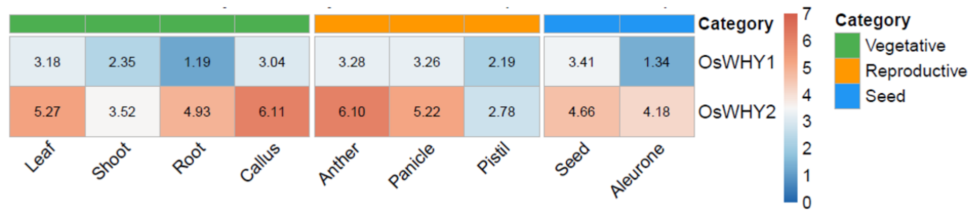


Fig. 9. Tissue expression patterns of the Whirly gene family in *Oryza sativa*. Heatmap showing tissue-specific expression profiles of *OsWHY1* and *OsWHY2* across nine tissue types retrieved from the Rice Expression Database (RED), with values shown as $\log_2$ (FPKM+1). Tissues are grouped into vegetative, reproductive, and seed categories.

To determine whether the two paralogs show different transcriptional responses to abiotic stress, *in silico* RNA-seq analysis was performed in *Oryza sativa*. This revealed a clear and biologically meaningful divergence in the stress-responsive expression of the two *OsWHY* paralogs. Under cold stress, neither *OsWHY1* nor *OsWHY2* showed significant transcriptional changes, suggesting that short-term cold exposure is insufficient to activate the *WHY* stress response pathway in rice (Fig. 10a). Under drought and heat, however, *OsWHY1* was consistently and progressively suppressed across all conditions tested, increasing with stress duration or severity (Fig. 10b, 10c). This duration-dependent downregulation, observed independently across two distinct stress types (drought and heat), suggests a conserved transcriptional stress response linked to sustained chloroplast impairment rather than a stress-type-specific effect. As these findings are based solely on *in silico* expression data, further experimental validation through qRT-PCR or loss-of-function approaches is needed to confirm their

biological significance. A comparable suppression of *WHY1* homologs under drought and salinity has been reported in cotton (He et al. 2025), supporting the idea that *WHY1* downregulation under sustained stress may be a conserved response in certain species. In contrast, *SlWHY1* in tomato promotes thermotolerance through upregulation of HSP21.5A expression under heat stress (Zhuang et al. 2020) and enhances cold tolerance by binding to the RbcS1 promoter (Zhuang et al. 2020), reflecting species-specific differences in WHIRLY1 transcriptional regulation across plant families. *OsWHY2* remained transcriptionally stable under all direct abiotic stress conditions. Microarray analysis under ABA treatment, however, identified significant and consistent downregulation of *OsWHY2* (Fig. 10d), confirmed across three independent probes, while *OsWHY1* showed no significant response.

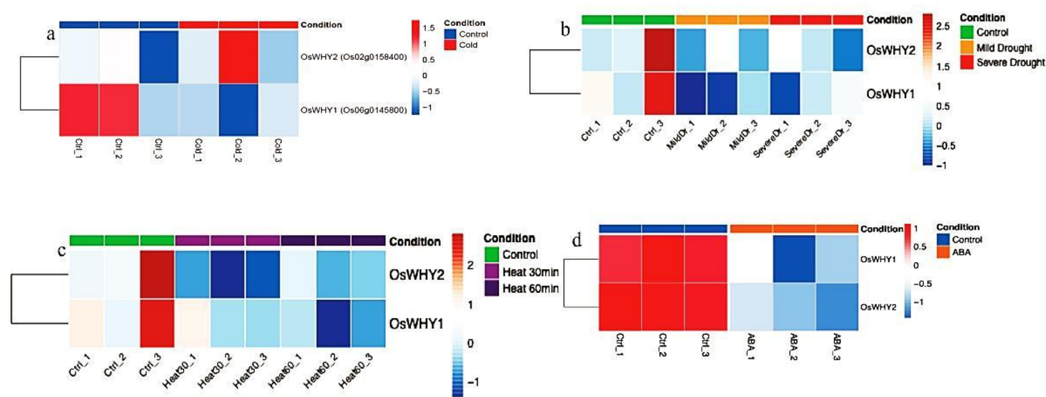


Fig. 10. Expression analysis of *OsWHY1* and *OsWHY2* under abiotic stress conditions in *Oryza sativa*: (a) heatmap showing expression levels under cold stress (4°C, 24 hrs) versus control across three biological replicates, (b) heatmap showing expression levels under mild and severe drought stress versus control across three biological replicates, (c) heatmap showing VST-normalized expression levels of *OsWHY1* and *OsWHY2* under heat stress (30 min and 60 min) versus control across three biological replicates, and (d) heatmap showing expression levels under ABA treatment versus control in root tissue across three biological replicates. Red indicates upregulation and blue indicates downregulation. All heatmaps were generated using the pheatmap package in R (v4.5.1).

This suggests a possible mechanistic distinction: *OsWHY2* is not regulated by the direct physical stress signal but specifically by the hormonal interpretation of stress through ABA (Cutler et al. 2010). These findings are based on *in silico* microarray data and require further experimental confirmation through qRT-PCR. This ABA-specific regulation might suggest that mitochondrial genome maintenance in rice is subject to hormone-mediated rather than stress-signal-mediated transcriptional control. The constitutive expression of *OsWHY2* under direct abiotic stress parallels observations in cotton and *Arabidopsis* (Taylor et al. 2023, He et al. 2025), supporting constitutive mitochondrial genome maintenance as a broadly conserved angiosperm strategy (Golin et al. 2020).

The results of this study reveal a clear functional divergence between the two *Oryza* WHIRLY paralogs. *OsWHY1* is a stress-responsive chloroplastic factor whose expression is consistently suppressed under drought and heat, with suppression increasing progressively with stress duration. This pattern is consistent with a role in plastid-to-nucleus retrograde signaling under sustained stress conditions (Lepage et al. 2013). *OsWHY2*, by contrast, maintains stable expression under direct abiotic stress but is specifically downregulated in response to ABA treatment, indicating that its regulation is hormonal rather than stress-signal driven. This difference is supported by the presence of ABRE elements in *Oryza WHY2*-type promoters and represents a novel finding for the mitochondrial WHIRLY subfamily in rice. Overall findings of this study shows that *OsWHY1* and *OsWHY2* have diverged into distinct stress-regulatory roles, with *OsWHY1* responding to direct abiotic stress signals and *OsWHY2* regulated through ABA-mediated hormonal signaling. This divergence is supported by multiple lines of evidence, including gene structure, motif composition, cis-regulatory elements, protein–protein interactions, and expression profiling. These findings provide a preliminary molecular framework for the WHY gene family across *Oryza* species. Future experimental studies involving qRT-PCR validation, overexpression lines, and loss-of-function mutants under abiotic stress conditions are recommended to confirm these *in silico* findings and evaluate their potential application in improving stress tolerance in rice.

Acknowledgments

We would like to express our sincere gratitude to the Centennial Research Grant of University of Dhaka for providing financial support. Authors also acknowledge Plant Breeding & Biotechnology Laboratory, Department of Botany, University of Dhaka for their generous support in conducting this research.

References

Akbudak MA and **Filiz E** (2019) Whirly (Why) transcription factors in tomato (*Solanum lycopersicum* L.): genome-wide identification and transcriptional profiling under drought and salt stresses. Mol. Biol. Rep. **46**: 4139-4150.

Alexandratos N and **Bruinsma J** (2012) World agriculture towards 2030/2050: the 2012 revision. FAO ESA Working Paper No. 12-03. Food and Agriculture Organization, Rome.

Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A and **Nielsen H** (2019) Detecting sequence signals in targeting peptides using deep learning. Life Sci. Allian. **2**(5): e201900429.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I and **Stockinger H** (2012) ExPASy: SIB bioinformatics resource portal. Nucl. Acids Res. **40**(W1): W597-W603.

Bailey TL, Johnson J, Grant CE and **Noble WS** (2015) The MEME suite. Nucl. Acids Res. **43**(W1): W39-W49.

Cappadocia L, Marechal A, Parent JS, Lepage E, Sygusch J and **Brisson N** (2010) Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. Plant Cell. **22**(6): 1849-1867.

Cappadocia L, Parent JS, Zampini E, Lepage E, Sygusch J and **Brisson N** (2012) A conserved lysine residue of plant Whirly proteins is necessary for higher order protein assembly and protection against DNA damage. Nucl. Acids Res. **40**(1): 258-269.

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y and **Xia R** (2023) TBtools-II: A 'one for all, all for one' bioinformatics platform for biological big-data mining. Mol. Plant. **16**(11): 1733-1742.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R and **Berry A** (2022) Ensembl 2022. Nucl. Acids Res. **50**(D1): D988-D995.

Cutler SR, Rodriguez PL, Finkelstein RR and **Abrams SR** (2010) Abscisic acid: emergence of a core signaling network. Ann. Rev. Plant Biol. **61**: 651-679.

Desveaux D, Allard J, Brisson N and **Sygusch J** (2002) A new family of plant transcription factors displays a novel ssDNA-binding surface. Nat. Struct. Biol. **9**(7): 512-517.

Desveaux D, Marechal A and **Brisson N** (2005) Whirly transcription factors: defense gene regulation and beyond. Trends Plant Sci. **10**(2): 95-102.

Fukagawa NK and **Ziska LH** (2019) Rice: importance for global nutrition. J. Nutr. Sci. Vitaminol. 65(Supplement): S2-S3.

Golin S, Negroni YL, Bennewitz B, Klosgen RB, Mulisch M, La Rocca N, Cantele F, Vigani G, Lo Schiavo F, Krupinska K and **Zottini M** (2020) WHIRLY2 plays a key role in mitochondria morphology, dynamics, and functionality in Arabidopsis thaliana. Plant Direct. **4**(5): e00229.

He S, Sun H, Li Y, Chen Q, Yang J and **Xiao F** (2025) Genome-wide characterization of WHIRLY genes in cotton identifies GhWHY1-D as a negative regulator of salt and drought stress responses. Gene. Res. Crop Evol. **72**(Suppl 1): 141-159.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ and **Vinh LS** (2018) UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. **35**(2): 518-522.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ and **Nakai K** (2007) WoLF PSORT: protein localization predictor. Nucl. Acids Res. **35**(suppl_2): W585-W587.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A and **Jermiin LS** (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. **14**(6): 587-589.

Katoh K and **Standley DM** (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**(4): 772-780.

Krause K and **Krupinska K** (2009) Nuclear regulators with a second home in organelles. Trends Plant Sci. **14**(4): 194-199.

Krause K, Kilbienski I, Mulisch M, Rodiger A, Schafer A and **Krupinska K** (2005) DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. FEBS Lett. **579**(17): 3707-3712.

Krupinska K, Oetke S, Desel C, Mulisch M, Schafer A, Hollmann J and **Hensel G** (2014) WHIRLY1 is a major organizer of chloroplast nucleoids. Front. Plant Sci. **5**: 432.

Lepage E, Zampini E and **Brisson N** (2013) Plastid genome instability leads to reactive oxygen species production and plastid-to-nucleus retrograde signaling in Arabidopsis. Plant Physiol. **163**(2): 867-881.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P and **Rombauts S** (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucl. Acids Res. **30**(1): 325-327.

Letunic I and Bork P (2021) Interactive Tree Of Life (iTOL) v6: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. **49**(W1): W293-W296.

Li Z, Zhai X, Zhang L, Yang Y, Zhu H, Lu H and **Hu D** (2024) Genome-wide identification of the WHIRLY gene family and its potential function in low phosphate stress in soybean (*Glycine max*). Genes. **15**(7): 833.

Liu H, Wang X, Yang W, Liu W, Wang Y, Wang Q and **Zhao Y** (2023) Identification of Whirly transcription factors in Triticeae species and functional analysis of *TaWHY1-7D* in response to osmotic stress. Front. Plant Sci. **14**: 1297228.

Liu Y, Kao HI and **Bambara RA** (2004) Flap endonuclease 1: a central component of DNA metabolism. Ann. Rev. Biochem. **73**: 589-615.

Manh MB, Ost C, Peiter E, Hause B, Krupinska K and **Humbeck K** (2023) WHIRLY1 acts upstream of ABA-related reprogramming of drought-induced gene expression in barley and affects stress-related histone modifications. Int. J. Mol. Sci. **24**(7): 6326.

Marechal A and **Brisson N** (2010) Recombination and the maintenance of plant organelle genome stability. New Phytol. **186**(2): 299-317.

Marechal A, Parent JS, Sabar M, Veronnneau-Lafortune F, Abou-Rached C and **Brisson N** (2008) Overexpression of mtDNA-associated *AtWhy2* compromises mitochondrial function. BMC Plant Biol. **8**(1): 42.

Marechal A, Parent JS, Veronnneau-Lafortune F, Joyeux A, Lang BF and **Brisson N** (2009) Whirly proteins maintain plastid genome stability in *Arabidopsis.* Proc. Natl. Acad. Sci. USA. **106**(34): 14693-14698.

Meng C, Yang M, Wang Y, Chen C, Sui N, Meng Q and **Lv W** (2020) *SlWHY2* interacts with *SlRECA2* to maintain mitochondrial function under drought stress in tomato. Plant Sci. **301**: 110674.

Mickelbart MV, Hasegawa PM and **Bailey-Serres J** (2015) Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nat. Rev. Gene. **16**(4): 237-251.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A and **Lanfear R** (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**(5): 1530-1534.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD and **Bateman A** (2021) Pfam: The protein families database in 2021. Nucl. Acids Res. **49**(D1): D412-D419.

Nishimura K and **van Wijk KJ** (2015) Organization, function and substrates of the essential Clp protease system in plastids. Biochim. Biophys. Acta. **1847**(9): 915-930.

Obayashi T, Aoki Y, Tadaka S, Kagaya Y and **Kinoshita K** (2018) ATTED-II in 2018: A plant coexpression database based on investigation of the statistical property of the mutual rank index. Plant Cell Physiol. **59**(1): e3.

Prikryl J, Watkins KP, Friso G, van Wijk KJ and **Barkan A** (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. Nucl. Acids Res. **36**(16): 5152-5165.

Qiu Z, Chen D, Teng L, Guan P, Yu G, Zhang P and **Zhu L** (2022) *OsWHY1* interacts with OsTRX z and is essential for early chloroplast development in rice. Rice **15**(1): 50.

Reiser L, Subramaniam S, Zhang P and **Berardini T** (2022) Using the *Arabidopsis* information resource (TAIR) to find information about *Arabidopsis* genes. Curr. Protoc. **2**(10): e574.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and **Smyth GK** (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43**(7): e47.

Ruan Q, Wang Y, Xu H, Wang B, Zhu X, Wei B and **Wei X** (2022) Genome-wide identification, phylogenetic, and expression analysis under abiotic stress conditions of Whirly (WHY) gene family in *Medicago sativa* L. Sci. Rep. **12**(1): 18676.

Sato Y, Takehisa H, Kamatsuki K, Minami H, Namiki N, Ikawa H, Ohyanagi H, Sugimoto K, Antonio BA and **Nagamura Y** (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. Nucl. Acids Res. **41**(D1): D1206-D1213.

Stockinger EJ, Gilmour SJ and **Thomashow MF** (1997) Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA. **94**(3): 1035-1040.

Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ and **von Mering C** (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucl. Acids Res. **51**(D1): D638-D646.

Taylor RE, West CE and **Foyer CH** (2023) WHIRLY protein functions in plants. Food Energy Secur. **12**(2): e379.

Wang L, Hou Q and **Qiao G** (2024) Genome-wide identification and expression analysis of the sweet cherry WHIRLY gene family. Curr. Issues Mol. Biol. **46**(8): 8015-8030.

Wang L, Zhao Z, Li H, Pei D, Ma Q, Huang Z, Wang H and **Xiao L** (2024) Genome-wide identification and molecular evolutionary history of the Whirly family genes in *Brassica napus*. Plants. **13**(16): 2243.

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. Nucl. Acids Res. **40**(7): e49.

Xia L, Zou D, Sang J, Xu X, Yin H, Li M and **Zhang Z** (2017) Rice Expression Database (RED): an integrated RNA-Seq-derived gene expression database for rice. J. Genet. Genomics. **44**(5): 235-241.

Yan Y, Liu W, Wei Y and **Shi H** (2020) MeCIPK23 interacts with Whirly transcription factors to activate abscisic acid biosynthesis and regulate drought resistance in cassava. Plant Biotechnol. J. **18**(7): 1504-1506.

Zhuang K, Gao Y, Liu Z, Diao P, Sui N, Meng Q, Meng C and **Kong F** (2020) WHIRLY1 regulates HSP21.5A expression to promote thermotolerance in tomato. Plant Cell Physiol. **61**(1): 169-177.

Zhuang K, Kong F, Zhang S, Meng C, Yang M, Liu Z and **Meng Q** (2019) Whirly1 enhances tolerance to chilling stress in tomato via protection of photosystem II and regulation of starch degradation. New Phytol. **221**(4): 1998-2012.

Zhuang K, Wang J, Jiao B, Chen C, Zhang J, Ma N and **Meng Q** (2020) WHIRLY1 maintains leaf photosynthetic capacity in tomato by regulating the expression of RbcS1 under chilling stress. J. Exp. Bot. **71**(12): 3653-3663.

(Manuscript received on 30 May, 2026; revised on 17 June, 2026)