International Journal of Advanced Engineering and Technology Research
ISSN: 3106-8677 (Print), ISSN: 3106-8685 (Online) | Vol. 2, No. 2, 2026

Genome-wide Identification and Bioinformatics Analysis of the Rice DnaJ Gene Family

Jiao Zhong [1], Qiong Ren [1], Zhenhui Kang [1, *], Xiaojun Ren [2, *]

[1] School of Sichuan University of Science & Engneering, Yibing 644005, China

[2] School of Chengdu Neusoft University, Chengdu 611844, China

[*] **Corresponding author**

Abstract: This study aims to analyze the structural characteristics of the DnaJ gene family in rice (Oryza sativa L.) and provide support for its molecular regulatory mechanism research. Using whole-genome data from the RAP-DB database, we identified 104 DnaJ family members, which are divided into three subfamilies (OsDjA, OsDjB, and OsDjC) based on phylogenetic relationships and conserved structural features. Most members contain a typical J domain, while some also have specific domai ns such as zinc finger, TPR, and PDI-related domains. Different subfamilies have highly conserved motif compositions but significant differences in arrangement, and their exon-intron structures also vary notably. The promoter regions of these DnaJ genes are rich in light-responsive, hormone-responsive, and stress-responsive elements, and extensive homologous collinearity between the rice DnaJ gene family and that of Arabidopsis reflects its high evolutionary conservation. This study provides an important theoretical basis and data support for further elucidating the functions of the rice DnaJ family in rice growth, development, and stress responses.

Keywords: Rice, DnaJ gene, genome-wide identification, bioinformatics analysis.

1. Introduction

Heat shock proteins (HSPs) are a class of molecular chaperone proteins widely present in prokaryotic and eukaryotic organisms. They were first named because they can be rapidly induced to express in large quantities under high-temperature stress conditions [1]. Studies have shown that the expression of HSP can be induced by various abiotic stresses such as drought, salt stress, oxidative stress, and high light, and it plays an important role in maintaining cellular protein homeostasis, promoting protein folding and assembly [2]. According to molecular weight and structural characteristics, HSPs are generally divided into six categories: Hsp100,Hsp90,Hsp70,Hsp60,Hsp40,and small heat shock proteins. Among them, the Hsp40 family, also known as the DnaJ protein family, is an important regulatory factor in the HSP molecular chaperone system, which can act alone or synergistically with molecular chaperones such as Hsp70 and Hsp100 to accurately regulate protein folding and the maintenance of homeostasis [3].

In plants, HSPs and their cofactors have important biological functions in cell development and abiotic stress. As the core organelle for photosynthesis and substance metabolism in plants, chloroplast development begins from proplastids and relies on the highly coordinated regulation of nuclear and plastid genes [4]. Research has shown that the nucleus-encoded 'chloroplast development factors' play a key role in chloroplast differentiation, thylakoid membrane structure formation, and photosystem complex assembly. Among them, molecular chaperone proteins containing DnaJ domains (also known as J proteins) are indispensable auxiliary factors in the maintenance of chloroplast protein homeostasis and functional completion [5].

In plants, DnaJ proteins are not only involved in basic protein quality control but also play important roles in chloroplast development, photosynthesis regulation, and responses to abiotic stress. As an important organelle for photosynthesis and various metabolic reactions in plants, the normal establishment of chloroplast structure and function relies on the coordinated expression of nuclear and plastid genes. Studies have shown that some DnaJ proteins are localized in chloroplasts and participate in the assembly and stabilization of photosystem protein complexes. For example, in Arabidopsis, the DnaJ-like protein PSA2 has been confirmed to be involved in the assembly process of photosystem I (PSI), and its functional loss leads to decreased photosynthetic efficiency and abnormal chloroplast development. In addition, some DnaJ proteins have been found to be involved in the folding and transport of thylakoid membrane proteins, further indicating their important role in maintaining the stability of the photosynthetic system.

Apart from being involved in chloroplast development, DnaJ proteins also play a key role in plant responses to abiotic stresses. In natural environments, plants are often exposed to various adverse factors such as high light, drought, salt stress, and temperature fluctuations. These stresses can lead to protein misfolding and the accumulation of reactive oxygen species (ROS), thereby causing cellular damage. As an important component of molecular chaperones, DnaJ proteins help alleviate stress-induced cellular damage to some extent by regulating protein folding and stability. For example, research has shown that certain DnaJ proteins are upregulated under high light conditions and help maintain the stability of photosystem structures, improving plant tolerance to light stress. Moreover, DnaJ proteins can indirectly reduce ROS accumulation by participating in the regulation of antioxidant systems, thus enhancing plant adaptability to stress. DnaJ proteins mainly activate the ATPase activity of Hsp70 through their conserved J domain, promoting the binding and release of substrate proteins with Hsp70 and forming an efficient molecular chaperone cycle system [6].

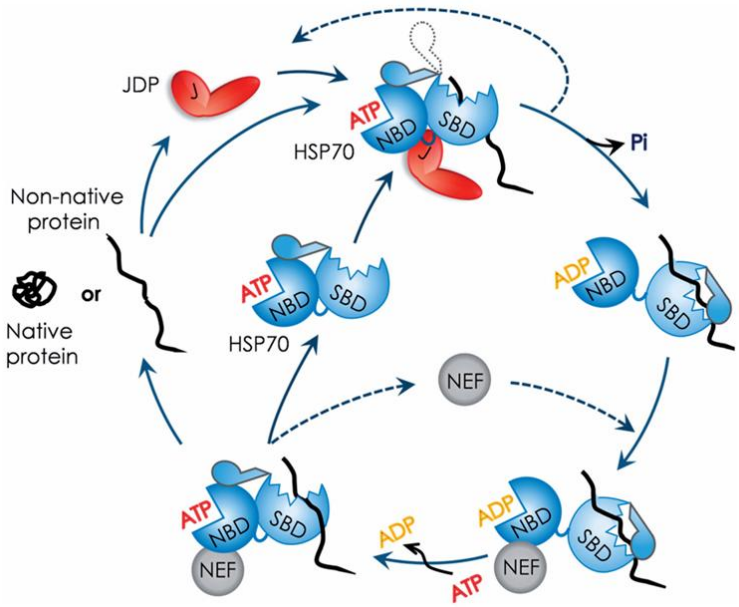

Figure 1. The Hsp70-JDP chaperone machinery

A typical DnaJ protein consists of four highly conserved functional domains, including the N-terminal 'J' domain, the Gly/Phe-rich G/F region, the CxxCxGxG-type zinc finger domain (ZF), and the C-terminal C-term domain. These domains together determine its recognition ability and molecular chaperone activity [7]. Based on differences in structural integrity and conserved domains, plant DnaJ proteins are generally classified into four types: A,B,C, and D. Among them, class A DnaJ proteins contain all typical domains and are usually considered to have relatively complete molecular chaperone functions [8]. Research by Liu and others indicates that the Class A DnaJ gene OsDNAJ15 is significantly induced under salt stress conditions. It interacts with OsBAG4,affecting the DNA binding activity of the transcription factor OsMYB106,thereby positively regulating the expression of OsHKT1,ultimately enhancing the tolerance of rice to salt stress [9].

This study uses bioinformatics methods to conduct a whole-genome analysis of the DnaJ gene family in rice. By analyzing the physicochemical properties of the genes, their chromosomal locations, conserved domain structures, promoter cis-acting elements, and collinearity relationships, it lays a foundation for in-depth research on the biological functions of the rice DnaJ family genes in rice growth, development, and stress responses.

2. Materials and Methods

2.1. Characteristic Analysis of the DnaJ Gene Family

Use the ExPASy Proteomics Server website (http://vweb.expasy.org/protparam/) to analyze the molecular weight and theoretical isoelectric point of rice DnaJ family genes online. Use the subcellular localization prediction website WOLF PSORT (http://psort1.hgc.jpfom.html) to perform protein localization analysis.

2.2. Chromosomal Localization Analysis of the Rice DnaJ Family

Download the rice genome files and annotation files from the Ensemble plant (https://asia.ensemhl.org) database, extract gene location files, gene collinearity files, and chromosome length files on TBtools, and visualize them on Advance Circos. Perform Ka/Ks analysis using Simple Ka/Ks Calculators.

2.3. Conserved motif analysis of the Rice DnaJ Gene Family

To identify the conserved motifs in the rice DnaJ gene family, the online analysis website MEME (http://meme-suite.org) was used to analyze the motifs of the rice DnaJ gene family, setting the number of motifs to 8, with the other parameters set to default.

2.4. Promoter Cis-acting Element Analysis

Based on rice genome data, the Tbtools 2.0 software was used to extract the 2000 bp upstream sequences of the DnaJ gene family (before the start codon ATG) as sequences for cis-element analysis. The online analysis website Plant CARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used for the analysis of promoter cis-acting elements.

2.5. Collinearity Analysis

To systematically analyze the collinearity of the rice DnaJ gene family, download the whole-genome protein sequences and gene location information (GFF files) of the target species and comparison species from the genome database Ensembl Plants (https://asia.ensembl.org). Perform whole-genome homologous comparison using BLASTP (usually setting E-value $\leq 1e^{-5}$) to screen potential homologous gene pairs. Then, input the comparison results along with gene location information into MCScanX software to identify collinear blocks within the genome and determine the types of gene duplication, thereby completing the intra-species collinearity analysis. Based on this, homologous gene pairs between different species are further used to construct cross-species collinearity relationships, identify conserved homologous segments, and achieve inter-species collinearity analysis. Finally, visualization tools such as TBtools 2.0 are used to process the MCScanX output results and separately plot intra-species and inter-species collinearity maps.

3. Results and Analysis

3.1. Genome-wide Information of Rice Dnaj Protein Family Members

Table 1. Genome-wide members information of the rice DnaJ protein family

Gene Name	Gene ID	pi	mw	Subcellular Localization Prediction
OsDjA1	LOC_Os02g43930	6.94	47232.41	nucl: 8, mito: 2.5, cyto_mito: 2,
OsDjC1	LOC_Os01g01160	9.84	21444.96	chlo: 14
OsDjC2	LOC_Os01g06454	9.99	12077.08	chlo: 13, extr: 1
OsDjB1	LOC_Os01g13760	9.05	38538.73	cysk: 13, pero: 1
OsDjC3	LOC_Os01g17030	10.13	15665.59	cyto: 3.5, E.R.: 3.5, mito: 3,
OsDjC4	LOC_Os01g17040	9.61	22952.69	E.R.: 3.5, E.R._plas: 3, chlo: 2,cyto: 2,
OsDjC5	LOC_Os01g25320	5.02	104277.81	nucl: 13, golg: 1
OsDjC6	LOC_Os01g27740	8.37	113412.41	nucl: 14
OsDjC7	LOC_Os01g32870	5.93	44959.46	nucl: 13, plas: 1
OsDjC8	LOC_Os01g33800	9.13	67633.89	nucl: 14
OsDjC9	LOC_Os01g37560	8.14	42749.81	E.R.: 5.5, E.R._plas: 4, cyto: 3,
OsDjC10	LOC_Os01g42190	5.32	21677.09	chlo: 8, nucl: 4,mito: 2
OsDjC11	LOC_Os01g44310	5.19	163187.68	nucl: 14
OsDjC12	LOC_Os01g50700	7.07	71043.97	nucl: 11, cyto: 2, plas: 1
OsDjC13	LOC_Os01g53020	8.59	38118.97	chlo: 14
OsDjB2	LOC_Os01g65480	9.25	36053.57	cyto: 8, chlo: 4, mito: 1, cysk: 1
OsDjC14	LOC_Os01g69930	8.96	83446.91	nucl: 14
OsDjC15	LOC_Os01g74580	6.45	52839.8	mito: 11, chlo: 2, cyto: 1
OsDjB3	LOC_Os02g03600	9	42996.57	cyto: 4.5, E.R.: 4,
OsDjC16	LOC_Os02g10180	6.66	53277.53	chlo: 4, nucl: 2.5, mito: 2, vacu: 2, E.R.: 2,
OsDjC17	LOC_Os02g10220	6.89	32454.47	nucl: 11, cyto: 2, chlo: 1
OsDjB4	LOC_Os02g20394	8.74	38539.9	cysk: 5, nucl: 4, cyto: 3, mito: 1, pero: 1
OsDjC18	LOC_Os02g30620	9.03	82390.14	nucl: 13, cyto: 1
OsDjC19	LOC_Os02g35000	5.42	42484.19	chlo: 4, cyto: 3, nucl: 2, mito: 2, cysk: 2, plas: 1
OsDjC20	LOC_Os02g46640	6.05	14209.99	cyto: 8, nucl: 2, mito: 2, extr: 1, cysk: 1
OsDjC21	LOC_Os02g50760	8.93	49325.46	nucl: 10, cyto: 3, chlo: 1
OsDjC22	LOC_Os02g52270	5.24	15548.45	nucl: 2, cysk: 2, golg: 2, E.R._vacu: 1.33333,
OsDjC23	LOC_Os02g54130	5.8	29255.13	nucl: 12, chlo: 1, mito: 1
OsDjA2	LOC_Os02g56040	9.1	52017.85	chlo: 11.5, chlo_mito: 6.83333,
OsDjC24	LOC_Os03g04400	9.02	32305.94	nucl: 13, cyto: 1
OsDjC25	LOC_Os03g10180	5.62	66473.23	nucl: 12.5, cyto_nucl: 7.5, cyto: 1.5
OsDjA3	LOC_Os03g12236	8.99	28113.41	chlo: 12, nucl: 1, mito: 1
OsDjC26	LOC_Os03g15480	9.52	34569.54	chlo: 11, nucl: 3
OsDjC27	LOC_Os03g18200	9.39	72123.74	vacu: 8, chlo: 2, golg: 2, nucl: 1, extr: 1
OsDjC28	LOC_Os03g18870	5.85	17994	chlo: 10, nucl: 3, mito: 1
OsDjC29	LOC_Os03g20730	10.67	17919.19	chlo: 14
OsDjC30	LOC_Os03g28310	7.54	83405.56	nucl: 7, mito: 6, vacu: 1
OsDjC31	LOC_Os03g36160	9.5	31715.18	mito: 10, cyto: 3, chlo: 1
OsDjA4	LOC_Os03g44620	6.01	46485.6	nucl: 8, cyto: 2, plas: 2, mito: 1, extr: 1
OsDjC32	LOC_Os03g51830	9.79	28067.8	chlo: 1, nucl: 1, mito: 1, plas: 1
OsDjC33	LOC_Os03g54150	5.01	67041.41	chlo: 6, nucl: 6, mito: 1, cysk: 1
OsDjC34	LOC_Os03g55360	7.93	57035.82	chlo: 6, nucl: 2, cyto: 2, mito: 2, plas: 1, cysk: 1
OsDjC35	LOC_Os03g56540	10.72	10934.75	chlo: 6, mito: 5, nucl: 2, cyto: 1
OsDjA5	LOC_Os03g57340	5.84	46690.67	nucl: 9, mito: 2, plas: 2, extr: 1
OsDjC36	LOC_Os03g60790	6.88	29652.77	chlo: 10, nucl: 2, mito: 2
OsDjC37	LOC_Os03g61550	10.27	28639.19	mito: 6, nucl: 3, chlo: 2, cyto: 2, plas: 1
OsDjC38	LOC_Os03g61730	9.21	80049.76	plas: 7, E.R.: 3, nucl: 2, mito: 1, pero: 1
OsDjC39	LOC_Os03g62120	9.77	50236.16	nucl: 11, cyto: 1, mito: 1, plas: 1
OsDjC40	LOC_Os03g62130	8.94	29784.85	nucl: 10, cyto: 3, mito: 1
OsDjC41	LOC_Os03g62140	8.79	30864.28	nucl: 5, mito: 4, chlo: 3, cyto: 2
OsDjC42	LOC_Os03g62150	8.94	28247.7	nucl: 9, cyto: 3, chlo: 2
OsDjC43	LOC_Os04g24180	5.74	76249.65	plas: 8, E.R.: 3, vacu: 2, nucl: 1

Table 1. (Continued)

OsDjC44	LOC_Os04g31940	9.02	81735.14	nucl: 12, chlo: 1, cysk: 1
OsDjA6b	LOC_Os04g46390	6.55	47147.56	nucl: 6, cyto: 4, chlo: 1, mito: 1, plas: 1, extr: 1
OsDjC45	LOC_Os04g57880	9.39	52581.36	chlo: 9, vacu: 2, nucl: 1, E.R.: 1, golg: 1
OsDjC46	LOC_Os04g59060	9.73	31357.15	nucl: 9, mito: 3, chlo: 2
OsDjC47	LOC_Os05g01590	9.4	26201.04	nucl: 6, chlo: 4, mito: 3, cysk: 1
OsDjB5	LOC_Os05g03630	9.34	35029.83	cysk: 7, mito: 4, nucl: 2, cyto: 1
OsDjB6	LOC_Os05g06440	6.43	39276.68	vacu: 8, extr: 2, nucl: 1, plas: 1, E.R.: 1, golg: 1
OsDjA7c	LOC_Os05g26902	9.66	48214.78	chlo: 10.5, chlo_mito: 6, nucl: 3
OsDjA8c	LOC_Os05g26926	9.66	48214.78	chlo: 10.5, chlo_mito: 6, nucl: 3
OsDjC48	LOC_Os05g30130	9.41	42595.8	vacu: 4, chlo: 3, cyto: 2, plas: 2,
OsDjC49	LOC_Os05g31062	8.04	44439.18	cyto: 5, chlo: 4, nucl: 2.5,
OsDjC50	LOC_Os05g45350	6.09	39142.07	chlo: 14
OsDjC51	LOC_Os05g46620	6.15	37903.07	cyto: 7, chlo: 4, mito: 2, plas: 1
OsDjB7	LOC_Os05g48810	8.96	38678.98	cysk: 10, cyto: 3, chlo: 1
OsDjC52	LOC_Os05g50370	5.25	157294.87	nucl: 13, chlo: 1
OsDjA9	LOC_Os06g02620	9.18	47377.34	chlo: 12, mito: 1, extr: 1
OsDjC53	LOC_Os06g09560	4.44	24538.04	nucl: 12, cyto: 1, mito: 1
OsDjC54	LOC_Os06g13060	8.66	49130.35	nucl: 11, cyto: 2, plas: 1
OsDjC55	LOC_Os06g34440	6.4	113289.35	nucl: 3.5, E.R.: 3, cysk_nucl: 2.5,
OsDjC56	LOC_Os06g44160	6.43	16090.95	chlo: 8, mito: 4, vacu: 2
OsDjC57	LOC_Os07g03270	9.33	28441.55	mito: 4, E.R.: 3, plas: 2, vacu: 2,
OsDjC58	LOC_Os07g09450	10.11	12029.98	chlo: 13, extr: 1
OsDjC59	LOC_Os07g28800	9.07	31411.56	nucl: 11, mito: 2, chlo: 1
OsDjC60	LOC_Os07g43330	9.42	31193.45	chlo: 12, mito: 2
OsDjC61	LOC_Os07g44310	10.28	14944.33	E.R.: 4.5, mito: 3, golg: 3,
OsDjB8	LOC_Os08g06460	9.41	38204.32	cyto: 12, nucl: 2
OsDjB9	LOC_Os08g28700	5.42	37056.34	chlo: 6, plas: 3, cyto: 2, vacu: 2, mito: 1
OsDjC62	LOC_Os08g35160	5.48	16734.51	extr: 6, chlo: 5, nucl: 1, cyto: 1, mito: 1
OsDjA10	LOC_Os08g36140	9.42	20029.47	chlo: 12, mito: 2
OsDjC64	LOC_Os08g36980	4.66	19290.48	chlo: 7, cyto: 6, nucl: 1
OsDjC65	LOC_Os08g37270	9.52	41570.27	nucl: 9, chlo: 3, cyto: 2
OsDjC66	LOC_Os08g41110	5.43	44232.93	nucl: 6, cyto: 4, plas: 2, chlo: 1, mito: 1
OsDjC67	LOC_Os08g43490	10.43	15732.93	chlo: 12, mito: 2
OsDjC68	LOC_Os09g20320	9.9	36242.61	chlo: 14
OsDjC69	LOC_Os09g28590	4.77	19321.68	chlo: 5, extr: 4, nucl: 3, cyto: 2
OsDjC70	LOC_Os09g28890	8.84	39497.07	mito: 10, cyto: 2, chlo: 1, plas: 1
OsDjC71	LOC_Os09g32050	5.12	44179.92	cyto: 4, chlo: 3, nucl: 3, mito: 2, plas: 2
OsDjC72	LOC_Os10g03610	10.27	27563.34	nucl: 6, mito: 5, cyto: 2, plas: 1
OsDjC73	LOC_Os10g11012	7.45	40062.49	cyto: 3.5, E.R.: 3.5, mito: 3,
OsDjC74	LOC_Os10g36370	8.63	60140.4	chlo: 9, nucl: 2.5, cysk_nucl: 2, mito: 1, vacu: 1
OsDjC75	LOC_Os10g42439	5.96	287691.27	plas: 9, E.R.: 3, vacu: 1, pero: 1
OsDjC76	LOC_Os11g36530	8.17	31818.09	chlo: 6, mito: 5, cyto: 2, plas: 1
OsDjC77	LOC_Os11g36960	6.38	119025.4	nucl: 13, cyto: 1
OsDjC78	LOC_Os11g37000	5.82	69490.91	nucl: 12, cyto: 1, cysk: 1
OsDjC79	LOC_Os11g43950	5.38	96862.64	nucl: 12, mito: 1, golg: 1
OsDjA11	LOC_Os12g07060	8.73	45697.62	chlo: 12, mito: 2
OsDjC80	LOC_Os12g15590	9.33	36162.79	plas: 7.5, golg_plas: 4.5, E.R.: 4, cyto: 1, extr: 1
OsDjC81	LOC_Os12g27070	8.83	29050.73	chlo: 12, nucl: 1, mito: 1
OsDjC82	LOC_Os12g31840	5.56	68218.86	nucl: 14
OsDjC83	LOC_Os12g36180	6.36	102039.08	nucl: 14
OsDjC84	LOC_Os12g41820	8.86	61408.97	plas: 8, nucl: 3, E.R.: 2, mito: 1
OsDjA12	LOC_Os12g42440	5.09	49391.19	cyto: 8, chlo: 2, nucl: 2

This study used the ExPASy Proteomics Server website to analyze the theoretical isoelectric point (pI) and relative molecular weight of the rice DnaJ protein family, and predicted the subcellular localization of the DnaJ protein family using the WOLF PSORT website. The results are shown in Table 1. The family contains a total of 104 members, with the OsDjA, OsDjB, and OsDjC subfamilies containing 12, 9, and 83 members, respectively, with the OsDjC subfamily having the highest proportion. Physicochemical property analysis revealed that the theoretical isoelectric points of the proteins ranged from 4.44 to 10.72, and the relative molecular weights ranged from 10.93 to 287.69 kDa,

indicating considerable differences in the charge properties and structural composition of the rice DnaJ protein family. The subcellular localization prediction results showed that rice DnaJ family members are mainly distributed in the nucleus and chloroplasts, while a few members are located in the cytoplasm, mitochondria, plasma membrane, endoplasmic reticulum, and vacuoles. In summary, some members of the rice DnaJ protein family may be involved in chloroplast-related biological processes, while others may play a role in nuclear regulation or the maintenance of intracellular protein homeostasis.

3.2. Chromosomal Localization of the DnaJ gene Family

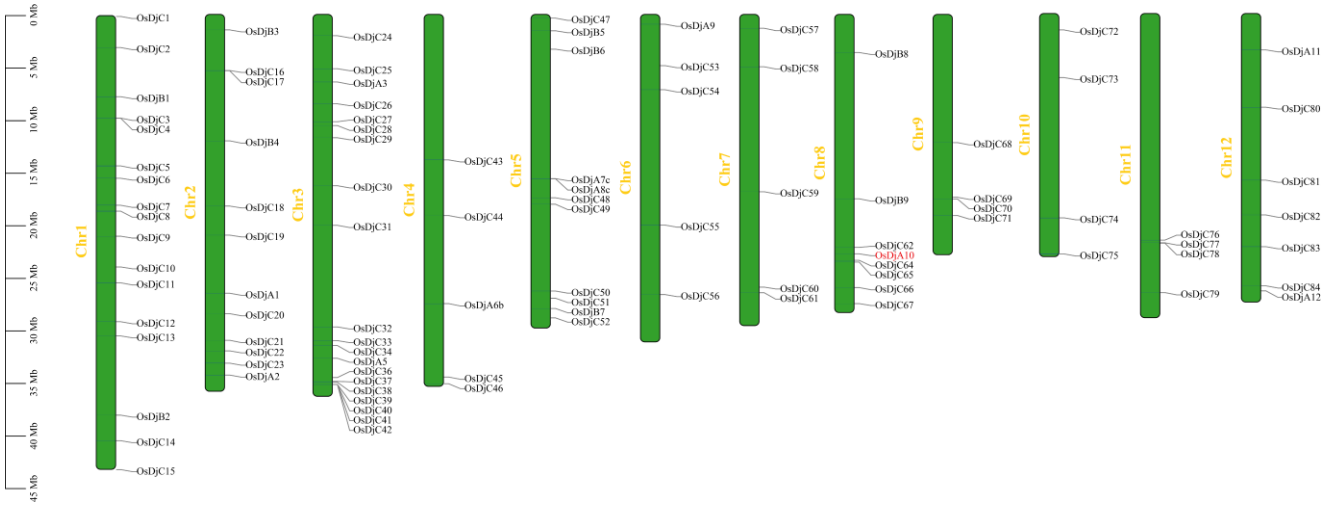

Figure 2. Chromosomal localization of the DnaJ gene family

The results of chromosome localization show that members of the rice DnaJ protein family are widely distributed across 12 chromosomes, but the distribution is uneven. The largest number of genes is located on Chr3, with a total of 20; Chr1, Chr2, and Chr5 contain 17, 12, and 11 members, respectively; Chr8 and Chr12 each have 8 and 7 members; Chr4, Chr6, and Chr7 contain 5 members; and Chr9, Chr10, and Chr11 each have 4 members. Overall, members of the OsDjC subfamily are widely distributed and dominant on various chromosomes, whereas members of the OsDjA and OsDjB subfamilies are relatively few and more sparsely distributed.

3.3. Analysis of Conserved Motifs in the DnaJ Protein Family

Conserved motif analysis shows that rice DnaJ family proteins exhibit both significant conservation and differentiation characteristics in motif composition and arrangement. Most members contain core motifs such as Motif1, Motif2 and Motif3, and display relatively consistent arrangement patterns within the same phylogenetic branch, indicating that these proteins have maintained a high degree of structural conservation during evolution. The motif composition differences among different subfamilies are more pronounced. The motif types and arrangements of the OsDjA and OsDjB subfamilies are relatively stable, while the OsDjC subfamily shows greater variation in motif number, arrangement order, and protein length, displaying obvious structural differentiation features. Domain analysis further indicates that members of OsDjA and OsDjB generally contain typical DnaJ conserved domains, whereas the OsDjC subfamily exhibits more diverse domain compositions; in addition to some members containing DnaJ-related domains, it also includes functional domains such as TPR, suggesting their potential involvement in more complex biological processes. Overall, the rice DnaJ family achieves functional differentiation through motif recombination and domain diversification while retaining core structural features. This result is consistent with the phylogenetic relationship analysis, and from a structural perspective, supports the functional division and evolutionary adaptability of different subfamilies.

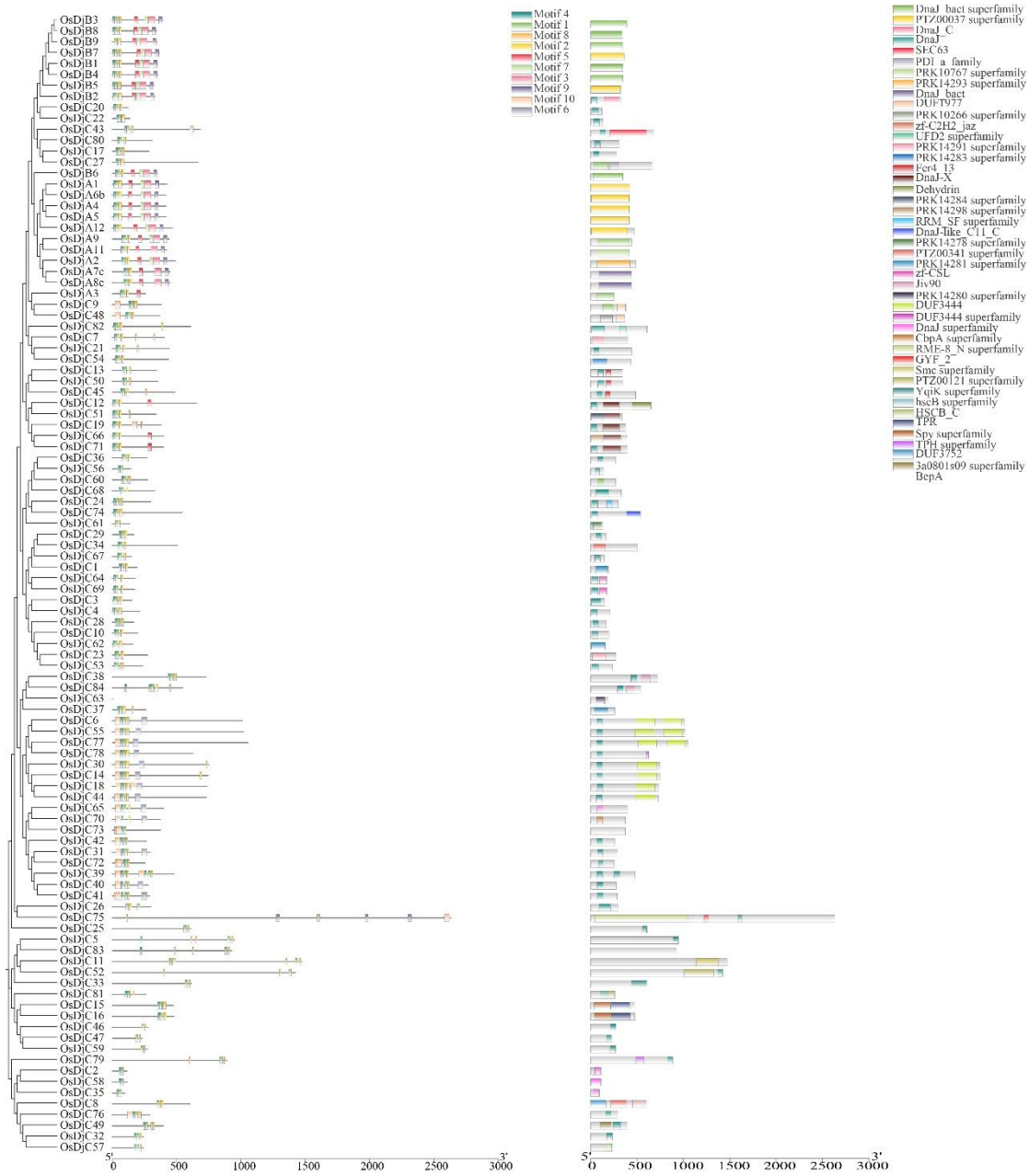

Figure 3. Analysis of Conserved Motifs in the DnaJ Protein Family

3.4. Promoter Cis-Element Analysis

Analysis of promoter cis-acting elements shows that the upstream regulatory regions of rice DnaJ family genes contain various cis-acting elements associated with plant hormone responses, stress responses, and growth and development regulation. Among the hormone response elements, ABRE, CGTCA-motif, and TGACG-motif are present in the promoters of most genes, with relatively high frequency, suggesting that members of the OsDj family may be involved in regulating ABA and jasmonic acid signaling pathways. In addition, some genes also contain P-box, GARE-motif, TATC-box, TCA-element, and auxin response-related elements, indicating that their expression may be co-regulated by multiple hormone signals. Regarding stress responses, elements such as LTR, MBS, and TC-rich repeats are detected in multiple members, indicating that this gene family may participate in cold, drought, and defense response processes. Compared with hormone and stress response elements, growth and development-related cis-elements are more diverse, including CAT-box, GCN4_motif, O2-site, RY-element, and light response-related elements, suggesting that some OsDj genes may be involved in meristem development, seed formation, and light signal regulation. It is noteworthy that the types and numbers of cis-elements in different gene promoters vary significantly, with some genes showing an enrichment of specific types of elements. Overall, OsDj family gene promoters generally have the characteristic of coexisting multiple types of cis-acting elements, indicating that their expression may be coordinately regulated by multiple signaling pathways, and providing cis-regulatory evidence for the functional differentiation of this family in growth, development, and stress adaptation. These results reveal the potential multifunctional regulatory nature of OsDj family genes at the transcriptional regulation level.

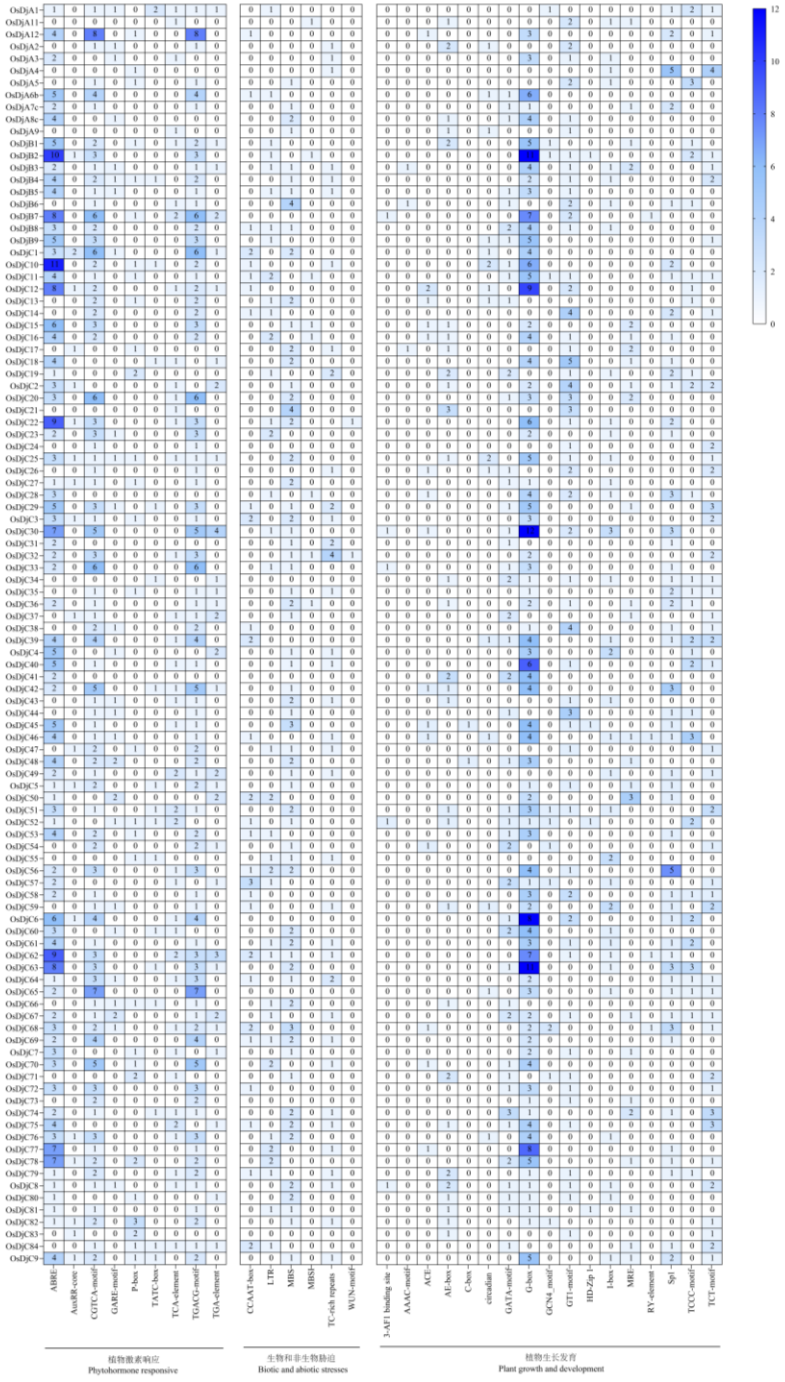

Figure 4. Cis-acting elements of rice DnaJ family promoters

3.5. Inter-species Collinearity Analysis

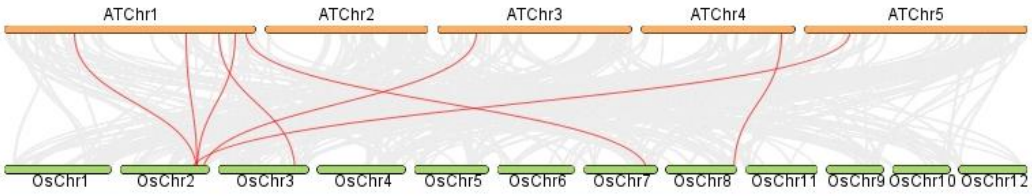

Figure 5. Collinearity between the rice DnaJ family and Arabidopsis

Inter-species collinearity analysis showed that there is extensive homology correspondence between the rice DnaJ family genes and the Arabidopsis genome. A large number of collinear gene pairs are distributed among different chromosomes, displaying a clear many-to-many correspondence pattern, indicating that this family underwent complex chromosomal rearrangements and gene redistributions after the divergence of monocots and dicots. Some collinearity relationships are concentrated in chromosomal regions such as OsChr2, OsChr3, ATChr1 and ATChr3, suggesting that these regions may be conserved evolutionary hotspots for this family. Overall, the collinearity between rice and Arabidopsis indicates that the DnaJ family is highly evolutionarily conserved, with its members possibly

originating before the divergence of monocots and dicots, and later expanding through segmental and whole-genome duplications during subsequent evolution.

3.6. Intraspecific Collinearity Analysis

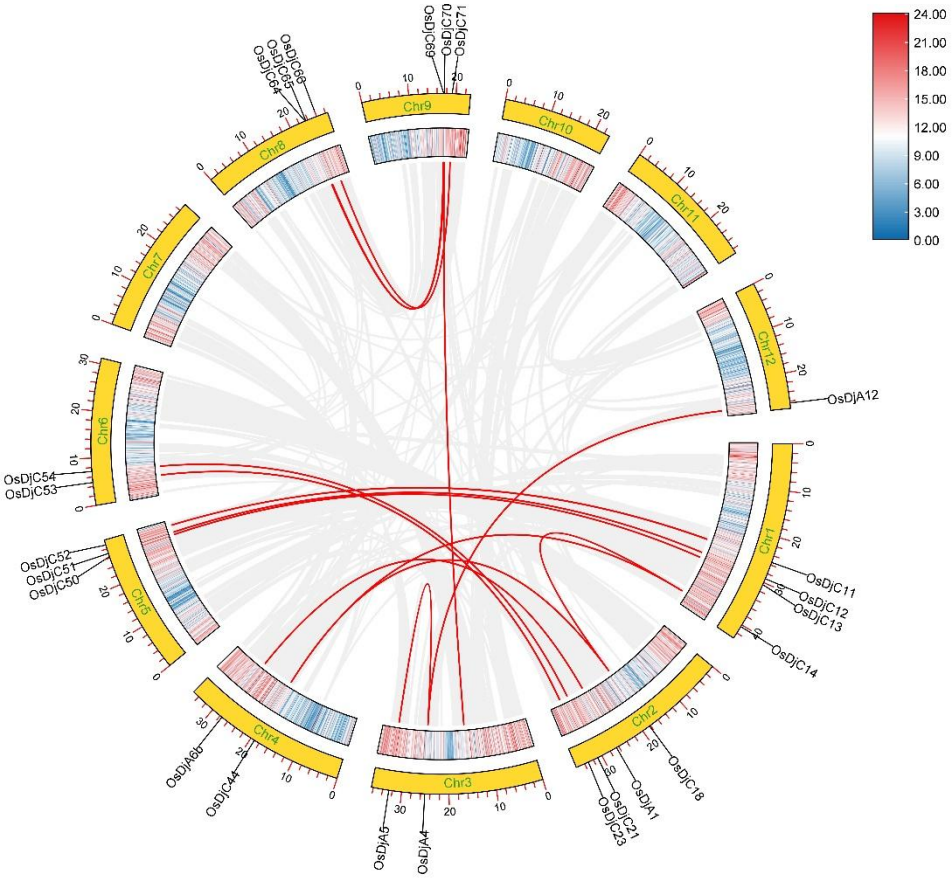

Figure 6. Intra-species collinearity of the rice DnaJ family

Intra-species collinearity analysis showed that rice DnaJ family genes have multiple collinear relationships within the genome, mainly distributed across different chromosomes, exhibiting obvious cross-chromosomal connection characteristics. This result indicates that the expansion of the OsDj family mainly relies on segmental duplication events rather than tandem repeats. Some collinear relationships are concentrated in chromosome regions such as Chr3, Chr6and Chr8–Chr9, suggesting that these regions may be key collinear segments for family expansion. In addition, most genes can form connections with multiple collinear genes, indicating that this family has undergone multiple rounds of duplication and rearrangement during evolution. Overall, the formation and expansion of the rice DnaJ family may be related to the combined effects of whole-genome duplication and segmental duplication.

4. Discussion

HSP proteins play an important role in the response of plants to abiotic stress, among which DnaJ proteins, as key co-chaperones of the Hsp70 system, are involved in various processes such as protein folding, stabilization, and degradation [10]. This study, based on bioinformatics methods, systematically analyzed the composition characteristics and potential functions of the rice DnaJ gene family. By combining results on gene number, domain composition, chromosomal distribution, and collinearity relationships, the family was comprehensively discussed from the aspects of evolutionary expansion patterns, structural differentiation, and regulatory characteristics.

From the perspective of family size and classification structure, this study identified a total of 102 DnaJ family members, among which the OsDjC subfamily accounted for the highest proportion (about 80%), significantly more than the OsDjA and OsDjB subfamilies. Similar family composition characteristics have also been observed in other plants; for example, in Arabidopsis and maize, the DnaJ family is also dominated by DjC members [11]., However, the results of this study show that this subfamily is more prominent in rice. Combined with intrafamily collinearity analysis, it can be seen that there are numerous inter chromosomal collinearity relationships among OsDj genes, while tandem duplications are rare, indicating that their expansion mainly relies on segmental duplication and whole-genome duplication events. This result is consistent with previous findings that the expansion of plant gene families is primarily derived from WGD and segmental duplication [12]. In addition, there is a significant collinear enrichment in the Chr3, Chr6, and Chr8–Chr9 regions, suggesting that these chromosomal regions may have played an important role in the expansion of gene families. It is worth noting that members of the OsDjC subfamily are widely distributed across various chromosomes. Combined with their high structural diversity, it is speculated that after expansion they experienced significant functional differentiation, which led to their long-term retention.

In terms of the physicochemical properties and structural characteristics of proteins, members of the rice DnaJ family exhibit considerable differences in molecular weight and isoelectric point, reflecting the diversity of their structural composition. Members of the OsDjA and OsDjB subfamilies

usually contain relatively complete J domains and related conserved regions, with relatively stable structures, mainly functioning as classical molecular chaperones involved in the protein folding process [13]. In contrast, the OsDjC subfamily shows significant structural differences, with not only a complex motif composition but also various function-related domains such as TPR and RRM. Previous studies have shown that the TPR domain can mediate protein-protein interactions, while the RRM domain is involved in RNA binding and processing [14]. This suggests that OsDjC members may be involved in transcriptional regulation and RNA metabolism processes, in addition to their traditional molecular chaperone functions. Therefore, the large-scale expansion of the OsDjC subfamily may be closely related to the demand for functional diversification.

Subcellular localization prediction results show that members of the OsDj family are mainly located in the nucleus and chloroplasts, which is consistent with the functional distribution of DnaJ proteins in plants. Previous studies have indicated that chloroplast-localized DnaJ proteins play a crucial role in the assembly of photosystems. For example, PSA2 in Arabidopsis, as a DnaJ-like zinc finger protein, participates in the assembly of the Photosystem I reaction center on the thylakoid membrane, and its loss of function significantly affects the accumulation of the PSI complex [15]. In addition, nuclear-localized DnaJ proteins may be involved in transcriptional regulation or signal transduction processes [16]. Analysis of promoter cis-acting elements in this study showed that the OsDj gene is generally rich in hormone- and stress-responsive elements, which is consistent with its potential multifunctional regulatory role.

In terms of regulatory mechanisms, hormone-responsive elements such as ABRE, CGTCA-motif, and TGACG-motif are commonly present in the promoter region, indicating that the OsDj family may be involved in the regulation of ABA and jasmonic acid signaling pathways. Previous studies have indicated that these hormones play a key role in plant responses to stresses such as drought, high light, and diseases [17]. At the same time, this study also detected stress-responsive elements such as MBS, LTR and TC-rich repeats, further supporting the potential role of this family in stress response. Studies have shown that DnaJ proteins can enhance plant stress tolerance by maintaining protein homeostasis and reducing the accumulation of misfolded proteins [18], The results of this study provide support for this function from the perspective of cis-regulatory regulation.

From an evolutionary perspective, interspecies collinearity results indicate that there are a large number of DnaJ homologous gene pairs between rice and Arabidopsis, showing a many-to-many correspondence, suggesting that this family was largely established before the differentiation between monocots and dicots. This phenomenon has been reported in multiple plant gene families [19]. At the same time, some chromosomal regions exhibit higher collinearity conservation, suggesting that these regions may have retained relatively ancient gene structures. Many-to-many collinearity relationships also indicate that after gene duplication, some members may have undergone functional differentiation or even acquired new functions, which is consistent with the common pattern of functional divergence during gene family evolution [20].

Overall, this study supports the following model: the rice DnaJ gene family originated from early plant ancestors and already had a basic structural framework before species differentiation. It subsequently expanded through whole-genome duplication and segmental duplication events, with the OsDjC subfamily showing significant expansion. During evolution, some members maintained the classic molecular chaperone functions, while others achieved functional diversification through domain recombination and changes in regulatory patterns, thereby participating in various biological processes such as chloroplast development, nuclear regulation, and stress response.

5. Conclusion

This study conducted a systematic identification and comprehensive analysis of the rice DnaJ gene family based on whole-genome data. The results showed that the family contains a total of 104 members, which can be divided into three subfamilies: OsDjA, OsDjB and OsDjC with OsDjC having the largest number of members. Analysis of protein physicochemical properties and conserved motifs indicated that members differ in molecular weight, isoelectric point, and motif composition, reflecting structural diversity on a conserved basis. Chromosomal localization results showed that family members are distributed across 12 chromosomes in an uneven manner, with certain regions exhibiting gene enrichment. Collinearity analysis revealed multiple cross-chromosomal collinearity relationships among family members, indicating that segmental duplication plays an important role in family expansion; meanwhile, widespread homologous collinearity with Arabidopsis suggests high evolutionary conservation. Subcellular localization prediction indicated that members are mainly distributed in the nucleus and chloroplast. Combined with promoter cis-element analysis, most genes were found to be rich in hormone response and stress-related elements, suggesting that their expression may be regulated by multiple signaling pathways. In summary, the rice DnaJ gene family has expanded through gene duplication, forming a structurally diverse and functionally rich gene set, providing a foundation for subsequent functional studies.

Acknowledgements

This research is jointly supported by the National Natural Science Foundation of China (31601291), the Wuliangye Group Co., Ltd. and Sichuan University of Light Industry cooperation project (CXY2021ZR010), the Talent Introduction Project of Sichuan University of Light Industry (2024RC073), and the Sichuan Provincial Department of Science and Technology project (2020YFN0023).

References

[1] Berka, M., Kopecká, R., Berková, V., et al. (2022). Regulation of heat shock proteins 70 and their role in plant immunity. Journal of Experimental Botany, 73(7), 1894–1909. https://doi.org/10.1093/jxb/erac025

[2] Yurina, N. P. (2023). Heat shock proteins in plant protection from oxidative stress. Molecular Biology, 57(6), 951–964. https://doi.org/10.1134/S0026893323060148

[3] Quan, Y., Wang, Z., Wei, H., et al. (2022). Transcription dynamics of heat shock proteins in response to thermal acclimation in Ostrinia furnacalis. Frontiers in Physiology, 13, 992293. https://doi.org/10.3389/fphys.2022.992293

[4] Daniell, H., Jin, S., Zhu, X. G., et al. (2021). Green giant—a tiny chloroplast genome with mighty power to produce high-

value proteins: History and phylogeny. Plant Biotechnology Journal, 19(3), 430–447. https://doi.org/10.1111/pbi.13526

[5] Stan, G., Lorimer, G. H., & Thirumalai, D. (2022). Friends in need: How chaperonins recognize and remodel proteins that require folding assistance. Frontiers in Molecular Biosciences, 9, 1071168. https://doi.org/10.3389/fmolb.2022.1071168

[6] Verma, A. K., Tamadaddi, C., Tak, Y., et al. (2019). The expanding world of plant J-domain proteins. Critical Reviews in Plant Sciences, 38(5–6), 382–400. https://doi.org/10.1080/07352689.2019.1658587

[7] Kampinga, H. H., Andreasson, C., Barducci, A., et al. (2019). Function, evolution, and structure of J-domain proteins. Cell Stress & Chaperones, 24(1), 7–15. https://doi.org/10.1007/s12192-018-00936-4

[8] Chen, T., Xu, T., Zhang, T., et al. (2021). Genome-wide identification and characterization of dnaJ gene family in grape (Vitis vinifera L.). Horticulturae, 7(12), 589. https://doi.org/10.3390/horticulturae7120589

[9] Liu, Y., Li, M., Yu, J., et al. (2023). Plasma membrane-localized Hsp40/DNAJ chaperone protein facilitates OsSUVH7-OsBAG4-OsMYB106 transcriptional complex formation for OsHKT1;5 activation. Journal of Integrative Plant Biology, 65(1), 265–279. https://doi.org/10.1111/jipb.13382

[10] Walsh, P., Bursać, D., Law, Y. C., et al. (2004). The J-protein family: Modulating protein assembly, disassembly and translocation. EMBO Reports, 5(6), 567–571. https://doi.org/10.1038/sj.embor.7400172

[11] Tamadaddi, C., Verma, A. K., Zambare, V., et al. (2022). J-like protein family of Arabidopsis thaliana: The enigmatic cousins of J-domain proteins. Plant Cell Reports, 41(6), 1343–1355. https://doi.org/10.1007/s00299-022-02856-2

[12] Wang, Y., Tang, H., DeBarry, J. D., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research, 40(7), e49. https://doi.org/10.1093/nar/gkr1293

[13] Kampinga, H. H., & Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nature Reviews Molecular Cell Biology, 11(8), 579–592. https://doi.org/10.1038/nrm2941

[14] Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS Journal, 272(9), 2118–2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x

[15] Fristedt, R., Williams-Carrier, R., Merchant, S. S., et al. (2014). A thylakoid membrane protein harboring a DnaJ-type zinc finger domain is required for photosystem I accumulation in plants. Journal of Biological Chemistry, 289(44), 30657–30667. https://doi.org/10.1074/jbc.M114.578149

[16] Liu, J. Z., & Whitham, S. A. (2013). Overexpression of a soybean nuclear localized type–III DnaJ domain-containing HSP40 reveals its roles in cell death and disease resistance. The Plant Journal, 74(1), 110–121. https://doi.org/10.1111/tpj.12108

[17] Yang, J., Duan, G., Li, C., et al. (2019). The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. Frontiers in Plant Science, 10, 1349. https://doi.org/10.3389/fpls.2019.01349

[18] Kong, F., Deng, Y., Zhou, B., et al. (2013). A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. Journal of Experimental Botany, 65(1), 143–158. https://doi.org/10.1093/jxb/ert353

[19] Heidari, P., Faraji, S., Ahmadizadeh, M., et al. (2021). New insights into structure and function of TIFY genes in Zea mays and Solanum lycopersicum: A genome-wide comprehensive analysis. Frontiers in Genetics, 12, 657970. https://doi.org/10.3389/fgene.2021.657970

[20] Qian, W., & Zhang, J. (2014). Genomic evidence for adaptation by gene duplication. Genome Research, 24(8), 1356–1362. https://doi.org/10.1101/gr.173743.114