Genome-wide identification and analysis of the regulation wheat DnaJ family genes following wheat yellow mosaic virus infection

LIU Ting-ting1, XU Miao-ze1, GAO Shi-qi1, ZHANG Yang2, HU Yang3, JIN Peng1, CAI Lin-na1, CHENG Ye1, CHEN Jian-ping1, YANG Jian1, ZHONG Kai-li1

1 State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products/Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Rural Affairs and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo 315211, P.R.China
2 Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, P.R.China
3 Institute of Forest Health, Zhejiang Academy of Forestry, Hangzhou 310021, P.R.China

Abstract
The co-chaperone DnaJ plays an important role in protein folding and regulation of various physiological activities, and participates in several pathological processes. DnaJ has been extensively studied in many species including humans, drosophila, mushrooms, tomatoes, and Arabidopsis. However, few studies have examined the role of DnaJ in wheat (Triticum aestivum), and the interaction mechanism between TaDnaJs and plant viruses. Here, we identified 236 TaDnaJs and performed a comprehensive genome-wide analysis of conserved domains, gene structure and protein motifs, chromosomal positions and duplication relationships, and cis-acting elements. We grouped these TaDnaJs according to their domains, and randomly selected six genes from the groups for tissue-specific analysis, and expression profiles analysis under hormone stress, and 17 genes for plant virus infection stress. In qRT-PCR, we found that among the 17 TaDnaJ genes tested, 16 genes were up-regulated after wheat yellow mosaic virus (WYMV) infection, indicating that the TaDnaJ family is involved in plant defense response. Subsequent yeast two-hybrid assays verified the WYMV NIa, NIb and 7KD proteins interacted with TaDJC (TraesCS7A02G506000), which had the most significant changes in gene expression levels after WYMV infection. Insights into the molecular mechanisms of TaDnaJ-mediated stress tolerance and sensitivity could inform different strategies designed to improve crop resistance to abiotic and biotic stress. This study provides a basis for future investigation of the TaDnaJ family and plant defense mechanisms.

Keywords: TaDnaJ, WYMV, wheat, genome-wide, expression, hormone, biotic stress, HSP40

1. Introduction
Heat shock proteins (HSPs), the largest of molecular chaperones, protect interacting surfaces of protein substrates from conformational damage by binding to them and facilitating the folding of unfolded or nascent polypeptides (Tutar and Tutar 2010). Since their first discovery in Drosophila melanogaster (Ritossa 1964),
HSPs have received widespread attention, and presently, studies on HSPs are quite common. In humans, the roles of HSPs during fertilization and pregnancy, tumor, cancer, and immunity have been studied (Hristova 2012a, b; Sterrenberg et al. 2011; Tan and Bernstein 2013; Lang et al. 2019). HSPs have also been researched in other species including zebrafish, yeast, rice, and Arabidopsis (Krone and Sass 1994; Yeh and Hsu 2002; Piechuch and Oblak 2013; Qian et al. 2014; Mertz-Henning et al. 2016; Kumar et al. 2018; Matsuoka et al. 2019). HSP families in plants mainly include HSP60, HSP70, HSP90, HSP100, HSP40, and sHSP (Muthusamy et al. 2017; Pulido and Leister 2018). HSP70 (also called DnaK) is one of the most abundant HSPs in eukaryotic cells and plays an important role in protein quality control and requires two co-chaperones to perform its function. The first is the nucleotide exchange factor (NEF), which exchanges HSP70-bound ADP for ATP to trigger the release of the folding substrate (Rampelt et al. 2018). The other is DnaJ (also called HSP40 or J-proteins), which recognizes an unfolded substrate and delivers it to DnaK, accelerates ATP hydrolysis of HSP70s, and induces conformational changes in stable chaperone proteins (Kampinga and Craig 2010). A specific factor of the molecular chaperone system, DnaJ can handle different substrates. In Escherichia coli, the presence of DnaJ and GrpE HSPs strongly stimulates the ATPase activity of HSP70. DnaJ or GrpE alone can slightly stimulate the ATPase activity of HSP70 (Liberek et al. 1991).

The DnaJ family is the most diverse family of co-chaperones whose members are characterized by the presence of conserved J-domains, which were first identified in the E. coli DnaJ, and a highly conserved HPD tripeptide signature motif. The J-domain is responsible for binding to the ATPase domain of HSP70 and is typically located at the N-terminus. DnaJ proteins (J-proteins) are classified into four subtypes according to the presence or absence of domains (Cheetham and Capian 1998; Walsh et al. 2004; Rajan and D’Silva 2009). For example, DnaJA proteins have all the domains found in DnaJ, such as the J-domain, the CxxCxxGxG zinc-finger domain (C, cysteine; G, glycine; X, other amino acid residues) and the C-terminal domain. The CxxCxxGxG zinc-finger domain is located in the central part of the J-protein and participates in protein–protein interactions (Miernyk 2001). The C-terminal domain facilitates DnaJ dimerization and participates in the interaction between multiple polypeptides. By comparison, DnaJB proteins lack only the CxxCxxGxG zinc-finger domain, while DnaJC proteins retain only the J-domain, which may appear in any region. DnaJD proteins have been classified as “DnaJ-like proteins”, lacking several important residues (including the HPD tripeptides), but they share structural similarities with the J-domain and function independently of HSP70. Recent studies have shown that domains other than the J-domain are also critical for the functions of J-proteins (Ajit Tamadaddi and Sahi 2016).

Previous studies have identified 129 DnaJ homologs in Arabidopsis (Zhang et al. 2018), 115 in rice (Luo et al. 2019), and 76 in pepper (Fan et al. 2020). However, there is limited research on wheat DnaJ. DnaJ has been implicated in diverse physiological roles including hormone regulation, but it has also been shown to be involved in resistance to disease resistance (Knox et al. 2011; Bekh-Ochir et al. 2013; Sporn and Hines 2015). For instance, the Arabidopsis AU1 is essential for abscisic acid (ABA) response (Park and Kim 2014) and the expression of the J-protein LeCDJ2 is triggered by salicylic acid (SA) (Wang et al. 2014). In addition, the tobacco mosaic virus movement protein interacts with a DnaJ-like protein (Hwang et al. 1998; Hofius et al. 2007; Shimizu et al. 2009; Cho et al. 2012; Verchot 2012), and the interaction between DnaJ and soybean mosaic virus coat protein leads to susceptibility (Zong et al. 2020). Furthermore, distinct DnaJ family members regulate the replication and pathogenesis of various viruses (Uran0 et al. 2013). Wheat yellow mosaic virus (WYMV), belonging to the genus Bymovirus in the family Potyviridae, (Li et al. 2017; Yang et al. 2018; Ohki et al. 2019; Zhang et al. 2019), is transmitted in the soil by Polymyxa graminis, which causes wheat yellow mosaic disease (Xu et al. 2018). The publication of the world’s first complete hexaploid wheat annotation has enabled more efficient and convenient analysis of the functions of wheat genes, and accelerated the research on wheat.

In this study, we performed a genome-wide analysis of the DnaJ family in wheat, and identified 236 DnaJ homologs, which were divided into three categories according to the features of the J-domains. The chromosomal loci and replication relationships, exon-intron structures, protein motifs and protein structures of these TaDnaJ were also analyzed. Subsequently, the expression profiles of TaDnaJ homologues in different tissues and under different hormone and WYMV infection conditions were analyzed. Yeast two-hybrid assay was performed to validate interactions between J-proteins and WYMV proteins. This research provides some clues to the defense mechanism of wheat.

2. Materials and methods

2.1. Genome-wide identification of TaDnaJ family genes

Two approaches were used to identify DnaJ family genes in wheat. In the first approach, DnaJ was used as a keyword to search and download protein sequences of wheat J-proteins from the plant genomics resource Phytozome v12.1 (https://
phytozome.jgi.doe.gov/pz/portal.html#} (Goodstein et al. 2012). In the second approach, the protein sequences of Arabidopsis DnaJ family members (Pulido et al. 2018; Zhang et al. 2018) were downloaded from TAIR (https://www.arabidopsis.org/) (Reiser et al. 2017). The protein sequences were used as queries for BLASTp searches against the wheat genome in the Ensembl Plants Database (http://plants.ensembl.org/Triticum_aestivum/Info/Index) (Bolser et al. 2015), to screen out the wheat homologs with a cut-off of Eval<10^-6 and ID%>70. All candidate proteins were further analyzed using the Pfam database (http://pfam.xfam.org/search#tabview=tab1) (Finn et al. 2014), NCBI Batch Web CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/bwropsb/bwropsb.cgi), or SMART database (http://smart.embl-heidelberg.de/). ExPaSy (https://web.expasy.org/compcomp_pl&group=programs&subgroup=proloc) (Wilkins et al. 1999; Artimo et al. 2012) was used to calculate the theoretical isoelectric point (pI) and molecular weight (MW) of each J-protein. SoftBerry (http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc) (Galas 2001) was used to predict the subcellular location of the J-proteins, and SignalP5.0 (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen et al. 1996; Bagos et al. 2010) was used to predict the signal peptides.

2.2. Analysis of gene domains and phylogenetic tree construction based on TaDnaJ

Gene domains data for TaDnaJ were obtained from Pfam (http://pfam.xfam.org/) and mapped with TBtools (Chen et al. 2020). The coding sequences (CDS) and amino acid sequences of wheat J-protein were downloaded from the Ensembl Plants BioMart (http://plants.ensembl.org/biomart/martview/85492d152b91fe26280cb79d80856174) and used for phylogenetic analysis. The default parameters of ClustalW were used to align all acquired sequences.

A phylogenetic tree based on unrooted neighbor-joining were constructed with 1 000 bootstrap tests with MEGA6 Software (Tamura et al. 2013), and TaDnaJs were classified according to the topology of the phylogenetic trees.

2.3. Chromosomal localization and gene duplication in TaDnaJ

To map the TaDnaJ genes to wheat chromosomes, their chromosomal locus information was retrieved from the Ensembl Plants Database (http://plants.ensembl.org/Triticum_aestivum/Location/Genome). After downloading the genome annotation file (ftp://ftp.ensemblgenomes.org/pub/release-47/plants/gtf/triticum_aestivum), duplication events of TaDnaJ genes were calculated using TBtools MCScanX with a threshold e-values1e-4 (Chen et al. 2020). The chromosomal locations and duplication relationships of TaDnaJs were illustrated using Circos-0.69 (http://circos.ca/software/download/) (Krzywinski et al. 2009).

2.4. TaDnaJ gene structure analysis and prediction of protein tertiary structures

The TaDnaJ motifs were analyzed online using MEME Suite 5.1.1 (http://alternate.meme-suite.org/tools/meme) (Bailey et al. 2009). Fifteen conserved motifs were predicted. The TaDnaJ coding and genome sequences were used to determine the exon-intron structures on the Gene Structure Display Server (http://gsds.gao-lab.org/) (Hu et al. 2015). All the results were rearranged TBtools (Chen et al. 2020). The SWISS-MODEL template library is a huge structure database, containing experimentally determined protein structures from the Protein Data Bank (Berman et al. 2002). The tertiary structure of TaDnaJ was predicted by homology modeling using SWISS-MODEL (https://swissmodel.expasy.org/) (Waterhouse et al. 2018). The quality of the model was assessed on GMQE range (Biasini et al. 2014) of 0 to 1, where the larger the value, the better the quality, and QMEAN range (Benkert et al. 2011) of –4 to 0, where the closer the value to 0, the better the match between the protein tested and the template. A prediction result up to 30% consistency indicated up to 80% accuracy of the model and a consistency of up to 50% indicated an accuracy of up to 95%. The best model was selected based on QMEAN and GMQE.

2.5. Tissue-specific expression of TaDnaJ

The expression of six randomly selected TaDnaJ genes was analyzed in eight different tissues. Three replicates were taken from each tissue of wheat and stored at –80°C until total RNA was extracted. Gene expression was detected using quantitative real-time PCR (qRT-PCR). A tissue-specific expression profile and a heat map of analysis results were displayed using TBtools (Chen et al. 2020).

2.6. Prediction cis-acting elements of TaDnaJ

Promoter sequences comprising 2 000 bp upstream of all TaDnaJ genes were downloaded from the wheat genome database, Ensembl Plants (Muthusamy et al. 2017). Conserved cis-acting regulatory elements present in the identified promoter regions were computationally predicted using the PlantCARE Database (http://bioinforatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al. 2002). A heat map of all analysis results was displayed using TBtools
(Chen et al. 2020), and promoter element analysis was performed for the hormone and oxygen response.

2.7. Expression profiles of TaDnaJ under hormone treatments

Wheat seedlings (YangMai 158) were sprayed with 100 μmol L⁻¹ abscisic acid (ABA), 100 μmol L⁻¹ gibberellin (GA), 100 μmol L⁻¹ methyl jasmonate (MeJA), and 100 μmol L⁻¹ salicylic acid (SA) (Yu et al. 2019). Distilled water was used as control. three biological replicates of leaf samples were collected at five different time points (0, 2, 4, 6, and 8 h), and stored at –80°C until total RNA was extracted. Then, gene expression was detected with qRT-PCR.

2.8. Expression of TaDnaJ during WYMV infection

Three biological replicates of two groups of wheat with the same growth conditions were selected for gene expression analysis. The wheat leaves were inoculated with WYMV following the previous study (Ohki et al. 2019) and the un-effected wheat leaves were used as a control. After 7 days, the test group was confirmed to be infected with WYMV and wheat leaves were collected for qPCR analysis. Seventeen TaDnaJ genes were randomly selected to detect changes in gene expression after virus infection.

2.9. TaDnaJ expression analysis by qRT-PCR

Total RNA in each sample was extracted using the HiPure Plant RNA Mini Kit (Magen, China, Guangdong) according to the manufacturer’s instructions, and stored at –80°C. First-strand cDNA was synthesized from ≤1 μg total RNA using the First-Strand cDNA Synthesis Kit (TOYOBO, Japan, Osaka) according to the manufacturer’s instructions (He et al. 2020). The gene expression level was assessed using qPCR SYBR green master mix on a real-time PCR machine (ABI). All primers used for qRT-PCR were designed using NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast). The relative expression levels of TaDnaJ were calculated and quantified using the $2^{-\Delta\Delta CT}$ method after normalization with CDC levels, respectively. The changes in gene expression were analyzed for significance using SPSS Software, and visualized on a bar chart using GraphPad Prisma Software.

2.10. Prediction protein-protein interactions between WYMV and TaDnaJ

Prediction models of protein–protein interactions (PPIs) between virus and their plant host was available at http://bclab.inha.ac.kr/VirusHostPPI (Zhou et al. 2018), these dataset could be network analyzed and visualized with the help of the Cytoscape (Doncheva et al. 2019).

2.11. Yeast two-hybrid assay

The coding sequence of TaDJC (accession number: TraesCS7A02G506000.1) and WYMV proteins were cloned to the Gal DNA-binding domain (vector: pGBKKT7) or Gal4 activation domain (vector: pGADT7), respectively, and using primers listed in Appendix A. Yeast two-hybrid assays were performed following the TaKaRa protocol handbook using strain yeast Y2H Gold. Yeast cells carrying the co-transformed plasmids were plated onto a low-stringency selective medium lacking leucine and tryptophan (SD/–Leu–Trp) to confirm the positive transformation and plated onto a high-stringency selective medium lacking leucine, tryptophan, histidine, and adenine (SD/–Leu–Trp–His–Ade) to analyze the interaction.

3. Results

3.1. Identification and analysis of TaDnaJ

In this study, we performed a genome-wide analysis to identify members of the DnaJ family in the wheat genome using two approaches to identify TaDnaJ family members. In the first approach, 109 TaDnaJs were identified in the Phytozome Database. In the second approach, a BlastP search was done in the wheat genome and wheat full length cDNA databases using known DnaJ sequences in Arabidopsis (Pulido et al. 2018; Zhang et al. 2018) and rice (Luo et al. 2019) as query sequences. The results were further confirmed by reciprocal BLAST. Finally, our analysis yielded 384 candidate genes in wheat. Further in silico analysis on Pfam and NCBI-CD network tools found that DnaJD proteins lacked the typical J-domain and were consequently, not considered as J-proteins. Therefore, J-proteins containing the J-domain were subsequently studied. Pseudogenes were further excluded in Pfam on the basis of the presence or absence of the J-domain (El-Gebali et al. 2019; Lamb et al. 2019). After removing redundant sequences, a total of 236 TaDnaJs were identified in wheat (Appendix B). All of the identified wheat proteins had conserved J-domains, which are essential for J-proteins to interact with and modulate the activity of their HSP70 partners. The gene IDs, clades, chromosomal loci, CDS lengths, protein sizes, molecular weights, isoelectric points (pI), sub-cellular localizations, and signal peptides (SP) are listed in Table S1. The average molecular weight of the TaDnaJ family was 44.6 kDa and the average isoelectric point was 8 (Appendix B). The molecular weights and
isoelectric points of the TaDnaJ family members showed large variations (Appendix B). Of the 236 predicted TaDnaJ family genes, 232 genes were mapped to 21 chromosomes of wheat (Appendix B). The TaDnaJs were predicted to localize to the extracellular, cytoplasmic, endoplasmic, plasma, mitochondrial, chloroplast and nuclear compartments (Appendix B). Of these genes, 181 localized to extracellular compartments, whereas only one gene localized to the chloroplast (Appendix B). An N-terminal secretion signal was predicted 11 genes (Appendix B).

3.2. Domain organization and phylogenetic analysis of TaDnaJ

J-proteins usually contain conserved J-domain, zinc-finger, and C-terminal domains. To comprehensively analyze the evolutionary relationship of the TaDnaJ family, TaDnaJ genes were classified according to the presence of the complete three domains into three types (DJA, DJB and DJC) containing 20, 24 and 190 members, respectively (Fig. 1-A). TaDnaJs show complex diversity in the amino acid sequence. Phylogenetic analysis of the TaDnaJ family indicated that these J-proteins could be classified by structural domains into three types and nine groups (I–IX) (Fig. 1-B), consistent with the DnaJ groups found in rice (Luo et al. 2019) and Arabidopsis (Zhang et al. 2018). Accordingly, TaDnaJs were divided into three major clades DJA, DJB and DJC. DJC was divided into nine clusters, each representing a structure other than J-domains, with a different function. Members of clade I contained a J-domain at the C-terminus and a domain of unknown function (DUF) domain at the N-terminus. In clade II genes, all J-domains were located at the C-terminus and contained the zinc-finger domain downstream of the J-domain (Fig. 1-B). Clade III contained the J-domain at the C-terminus and a Myb_DNA-binding domain downstream of the J-domain (Fig. 1-B). The J-domain of clade IV was located at the N-terminus, and multiple tandem tetratricopeptide repeat (TPR) domains were located upstream of the J-domain (Fig. 1-B). In clade V genes, all J-domains were located in the central region and contained Fer4 domains downstream of the J-domain (Fig. 1-B). The J-domain of members of clade VI was located at the C-terminus, and a DnaJ-X domain was included downstream of the J-domain (Fig. 1-B). In clade VII, the N-terminus contained the J-domain and a Jiv90 domain (Fig. 1-B). In clade VIII, the J-domain was located in the central region and a HSCB domain was included after the J domain (Fig. 1-B). Members of clade IX contained a GYF domain in the central region, followed by the J-domain (Fig. 1-B).

3.3. Analysis of gene structures, conserved motifs and tertiary structures of TaDnaJ

Since genes with similar structures may have evolved from a common ancestor, and the coding potential is most likely determined by the structure of the genes, the origin of its genes can also be traced. To identify common motifs among the different groups of J-proteins, we used the MEME motif search tool and mapped the results with TBtools. Fifteen conserved motifs were identified (Appendix C). The distribution of these motifs in J-proteins is illustrated in Fig. 2-A. J-proteins of the same group exhibited similar motif distribution patterns. Motifs 1 and 2 were present in every group, while only DJC had 5, 8, 11 and 13 motifs (Fig. 2-A). Exon-intron structural diversity within a gene family is an important clue about the evolutionary and functional analyses of the gene family members (Yan et al. 2014). To examine the structural features of TaDnaJ genes, the exon-intron structures of all the TaDnaJ genes were analyzed. The results revealed that genes in the same group shared a similar number of exons but had different exon and intron lengths (Fig. 2-B). To gain insight into the structural effects of the J-proteins, three-dimensional protein models were generated using homology modeling with SWISS-MODEL template library. One or two J-proteins from each group were randomly selected and displayed. Similar results were observed in the DJA and DJB group. However, in the DJC group, tertiary structures of J-proteins were complex and diverse (Fig. 2-C).

3.4. Analysis of Chromosomai Location and Duplication of TaDnaJ

Hexaploid wheat contains three sub-genomes (A, B, and D), and every wheat gene can have three homologs from the triad of homologous chromosomes (Panchy et al. 2016). Among the 236 predicted TaDnaJ family genes, chromosome 1, 5 and 7 were found to be enriched DnaJ genes (Appendix D). TaDnaJ genes in the different groups showed an uneven distribution across the A, B, and D sub-genomes and biased distribution among the seven chromosomes of each sub-genome (Appendix D). Chromosome 7 was the most frequent, with 19 genes located on chromosome 7A and 20 genes on chromosome 7B (Appendix D). Ancient duplication events and a high rate of retention of extant duplicate gene pairs contribute to abundant duplicate genes in plant genomes. To understand the situation of hexaploid wheat, we evaluated tandem replication events in the TaDnaJ family using Circos to determine the chromosomal locus and duplication relationship of all TaDnaJ genes. We found 308 duplication events in TaDnaJs, with maximum duplications
I, DUF; II, zinc-finger; III, Myb_DNA-binding; IV, TPR; V, Fer; VI, DnaJ-X; VII, Jiv90; VIII, HSCB_C; IX, GYF_2.

acid sequences of 236 wheat J-proteins aligned with ClustalW. DJA, DJB, and DJC proteins are displayed in red, blue and green.

Fig. 1 Schematic depiction of the characteristic domains and phylogenetic analysis of TaDnaJs. A, TaDnaJs were classified into three different types. DnaJA proteins mainly consist of three highly conserved domains: N-terminal J-domain (DnaJ) responsible for binding the ATPase domain of HSP70, the zinc-finger domain (CXXCXXGXG) and the C-terminal domain (DnaJC). DnaJB proteins lack the zinc-finger domain, whereas DnaJC proteins contain only the J-domain that may be situated anywhere along the length of the J-proteins, but there can be other domains besides the J-domain. The results were mapped by TBtools. B, a phylogenetic tree was constructed in MEGA6 using the neighbor-joining method with 1000 bootstrap replicates based on the full-length amino acid sequences of 236 wheat J-proteins aligned with ClustalW. DJA, DJB, and DJC proteins are displayed in red, blue and green dots, respectively. Clades DJC I–IX represent the nine different clades separated by low-score branches on the basis of domain. I, DUF; II, zinc-finger; III, Myb_DNA-binding; IV, TPR; V, Fer; VI, DnaJ-X; VII, Jiv90; VIII, HSCB_C; IX, GYF_2.
by TBtools. C, the SWISS-MODEL was used for structural prediction, and the model with the best results was selected based on QMEAN and GMQE.

Fig. 2 Gene structures, conserved motifs and tertiary structure of TaDnaJ. A, distribution of all motifs identified by MEME. Numbers 1–15 are displayed in different colors. The results were mapped by TBtools. B, exon-intron structure of TaDnaJ. Green stripes indicate untranslated 5'- and 3'-regions, yellow stripes indicate exons, and gray lines indicate introns. The results were mapped by TBtools. C, the SWISS-MODEL was used for structural prediction, and the model with the best results was selected based on QMEAN and GMQE.
events on chromosome 7D. In addition, all chromosomes, except chromosome 2, had tandemly duplicated gene pairs (Fig. 3). TaDnaJs were the largest HSP sub-family and accounted for 11 tandem duplication events (Fig. 3).

3.5. Tissue-specific analysis of TaDnaJ

Gene regulation plays a vital role in plant growth and development, and control of gene expression in a specific biological process or under certain pressures are aspects of major interest in plant research (Gautam et al. 2015). To comprehensively understand the functions of TaDnaJ, the expression of two randomly selected TaDnaJ genes from each group was analyzed in eight different tissues (root, stem, top leaf, two leaf, three leaf, four leaf, five leaf, and seed) by qRT-PCR. The results showed that TaDnaJ genes were differentially expressed in the different groups. The expression levels of genes in the DJA group tended to be higher in the top leaf tissues, while the expression levels of genes in the DJB group were highest in roots and seeds (Fig. 4). The expression of genes in the DJC group was highest in wheat leaves, followed by roots and seeds (Fig. 4).
Fig. 4 Differential tissue-specific expression of \textit{TaDnaJ}. The mean expression levels of \textit{TaDnaJ} in other tissues relative to the parietal lobe were calculated for three independent biological replicates, and the results were visualized in TBtools. Red represents a high expression level and blue represents a low expression level.
3.6. Prediction of TaDnaJ cis-acting elements

Recent advances in transcriptomic technologies and genome sequencing of plants have enabled routine large-scale prediction of promoter sequences and their contributing cis-acting elements (Hernandez-Garcia and Finer 2014). The analysis of promoter elements is essential for studying the regulation of natural genes and developing transgenic crops (Hernandez-Garcia and Finer 2016). Regulation of gene expression at the promoter level is mainly controlled by the cis-acting elements located upstream of the transcription start site (Hernandez-Garcia and Finer 2014). To decipher the roles of cis-regulatory elements of TaDnaJ genes under biotic and abiotic stress-specific responses, we analyzed the promoter regions (2,000 bp upstream of the translation start site) of the 236 TaDnaJ genes using the PlantCARE Database. The identified cis-regulatory elements were classified into seven functional groups: development, hormone, adversity stresses, oxygen, light responsive, binding site, and promoter (Fig. 5-A). The most abundant elements were hormone-responsive and oxygen-responsive elements. Five hormone-responsive elements including abscisic acid (ABA), gibberellin (GA), methyl jasmonate (MeJA), salicylic acid (SA), and auxin, were identified (Fig. 5-B). Nearly all TaDnaJ genes contained cis-acting regulatory elements related to hormone regulation, indicating that TaDnaJ gene family is significantly affected by hormone factors. In addition, most TaDnaJ genes encoded enhancer-like elements related to anaerobic and hypoxia-specific induction (Fig. 5-C).

3.7. Analysis of J-proteins expression under hormone stress

It has been established that plant hormones act as chemical messengers in the physiological processes regulating the plant life cycle from germination to senescence, and that plant hormones also coordinate the physiological responses to biotic and abiotic stresses (Muller and Munne-Bosch 2017). Since nearly all TaDnaJ genes contained hormone regulatory factors, we explored the role of the TaDnaJ family in hormone response, we selected two TaDnaJ genes with hormone (SA, ABA, GA, and Me-JA) regulatory sites in each group, and the effect of hormone stress on TaDnaJ was verified. The results revealed that the genes in the DJA group were more sensitive to ABA and GA, while the genes in the DJB group were more sensitive to ABA, and the genes in the DJC group are most sensitive to SA and GA (Fig. 6). All six genes tested had significant changes in the hormone response, confirming the presence of hormone-responsive cis-acting elements.

3.8. Analysis of J-proteins expression after WYMV infection

In a previous study, we found that DnaJ interacted with the tobacco mosaic virus (Shimizu et al. 2009). To explore the relationship between the TaDnaJ family and plant viruses, we analyzed the gene expression profile of wheat following WYMV infection. During evolution, regions other than the J-domain have become more important for the function of the J-protein. In addition to detecting the relative expression levels of TaDnaJ genes in the DJA and DJB groups, expression in genes with special domains in the DJC group was also analyzed in leaves after virus infection. At 7 days post virus infection, leaves were collected and the expression pattern of 17 TaDnaJ genes that selected from subfamily at random was determined by RT-qPCR. 16 TaDnaJ genes were upregulated (Fig. 7). DJB-1 (TraesCS1B02G423600), DJC-HSB (TraesCS1A02G349900), DJC-GYF (TraesCS1B02G208100) and DJC-DUF (TraesCS1B02G054900) approximately two-fold upregulated, while DJC (TraesCS1A02G506000) and DJC-Fer4 (TraesCS1A02G349900) three-fold increased (Fig. 7). These results showed that most TaDnaJ genes were substantially affected by WYMV and may therefore play a role in post-infection responses.

3.9. Analysis of protein–protein interactions between WYMV and TaDnaJ

The role of DnaJ family in viral replication had been described in numerous publications. From the previous results, we found that TaDnaJ played an important role in the process of virus infection. To investigate the WYMV proteins that interact with the TaDnaJ, we used the PPIs model to predict the 17 J-proteins mentioned above, and selected TaDJC with the most significant changes in qRT-PCR for yeast two-hybrid assay verification. The results of the PPIs network diagram showed that there was an interaction between J-proteins and one or more proteins of WYMV (Fig. 8-A), and further yeast two-hybrid results demonstrated TaDJC interacted with WYMV Nia, Nib and 7KD proteins (Fig. 8-B). These results indicated that TaDnaJ does play a role in the stress response after WYMV infection.

4. Discussion

DnaJ regulates various physiological activities in an organism and participates in various pathological processes. Positive-strand RNA viruses are one of the largest viruses that infect plants worldwide, and cellular chaperones play a central role in the establishment of positive-strand and negative-strand RNA virus infections (Verchot 2012; Newburn and White...
Fig. 5 Prediction of TaDnaJ gene cis-acting elements. A, several cis-acting elements were detected in the promoter region of each TaDnaJ gene. The green blocks represent the numbers of cis-element in TaDnaJ genes, and the darker colors indicate a greater number. The cis-acting elements were divided into seven types. B, the type, number and location of hormone-related elements in TaDnaJ genes. C, the type, number, and location of oxygen components in TaDnaJ genes.
However, few studies have examined the interaction mechanism between the TaDnaJ family and plant viruses. Recently, the completion of whole-genome sequencing and annotation of the allohexaploid wheat genome has facilitated the analysis of wheat gene families. Here, we performed a comprehensive genome-wide analysis of conserved domains, gene structure and protein motifs, chromosome locations, duplication relationships, and expression profiles of TaDnaJs under hormone stress and plant virus infection.

The exon-intron structure distribution pattern of most genes in the same cluster is similar. To respond to various pressures in time, genes must be activated quickly, and a compact gene structure with few introns helps to activate genes (Jeffares et al. 2008; Yu et al. 2016). Fifteen conserved motifs were identified in the TaDnaJ primary sequences, and J-proteins in the same group showed similar motif distribution patterns which were closely related to the catalytic core domain of the J-proteins. The distribution
Fig. 7 TaDnaJ gene expression levels after wheat yellow mosaic virus (WYMV) infection. Seven days after the wheat seedlings were infected with WYMV, the changes in gene expression were detected by RT-qPCR. The gene expression data were analyzed using SPSS software and visualized with GraphPad Software. Statistically significant differences (P<0.05 (Student’s t-test) are represented by an asterisk (*), and bars represents the standard deviation (±SD) calculated for three biological replicates.

Fig. 8 Protein–protein interactions between wheat yellow mosaic virus (WYMV) and TaDnaJ. A, the PPIs network between WYMV proteins and J-proteins. The pink circles represent WYMV proteins, the green hexagons represent J-proteins, and the black lines represent possible interactions. B, Nia, Nib and 7KD was fused to the DNA-binding domain and TaDJC was fused to the activation domain (AD-TaLIP), co-transformed into yeast cells, and then coated uniformly on selection plates of SD/–Trp–Leu–His–Ade solid medium. Positive and negative controls were co-transformed with AD-RECT/BD-53 and AD-RECT/BD-Lam, respectively.

of motifs thus provided a reference for the functional analysis of different genes. Plant gene families evolve mainly through tandem duplication and high rates of birth and death in clusters, but may also evolve through non-
repetitive polyploidy or large-scale segmental duplications and subsequent losses (Cannon et al. 2004; Magadum et al. 2013). During the evolution, TaDnaJs have undergone duplication events, which may have conferred certain selection advantages under certain conditions. Temporal and spatial expression analysis of TaDnaJ revealed tissuespecific expression patterns. DJA J-proteins were highly expressed in leaves, suggesting a role in leaf dynamics. However, DJB J-proteins were highly expressed in roots and species, indicating that they may be involved in the development of root tip meristems. In contrast, DJC J-proteins were expressed in various tissues, suggesting that these proteins are involved in multiple physiological activities in wheat.

Although genetic diversity increases during evolution, DnaJ has remained evolutionarily conserved. A total of 236 TaDnaJ homologs were identified, distributed on each chromosome. However, the classification of TaDnaJ based only on their domains simplifies the complexity of analysis. Therefore, we constructed a phylogenetic tree according to the domains other than the J-domains, to provide a novel perspective on the annotations of TaDnaJ. Accordingly, we identified the Fer4 domain, which was first identified in sulfate-reducing bacteria, and plays an important role in the biosynthesis of ferrichrome A (Sery et al. 1994; Winterberg et al. 2010; Luo et al. 2019). Iron is an essential element for almost all living things. We also identified the Jiv90 domain, the J-domain protein that interacts with viral proteins in the N-terminal region of the polyprotein (Muller et al. 2003; Guo et al. 2017). The J-domain was preceded by multiple TPR domains (tetra-tripeptide repeats in series), a structural motif found in several proteins. This motif mediates the interaction between proteins and can bind to various domains to perform multiple functions (Muller et al. 2003; Prasad et al. 2010; Luo et al. 2019). In addition, we identified the GYF domain, a small universal adapter domain that recognizes proline-rich sequences (PRS) (Kofer and Freund 2006; Matsui et al. 2017). The GYF domain protein reduces the susceptibility of Arabidopsis to the plantain asiatica mosaic virus (Hashimoto et al. 2016). J-proteins containing the GYF domain may play an important role in the resistance mechanism of the host plant to the virus (Albert et al. 2015). qRT-PCR showed that most TaDnaJ gene expression levels were up-regulated after WYMV infection, indicating that J-protein may be involved in plant defense mechanisms. In previous studies, it was found that the key factor of soybean susceptibility is the DnaJ protein that interacts with soybean mosaic virus coat protein (Zong et al. 2020). Further results of yeast two-hybrid assay showed that TaDJC interacted with WYMV Nla, Nib and 7KD proteins.

All in all, the genome-wide identification and basic functional analysis of TaDnaJ genes family in wheat were conducted in this study. These results provided novel insights for TaDnaJ homoeologs and provided with useful clues for further functional characterization. In subsequent research, we will be further validated the functions of TaDnaJ genes by gene overexpressing or gene silencing in wheat.

5. Conclusion

Understanding the molecular mechanisms of J-protein-mediated stress tolerance and sensitivity could facilitate the design of different strategies to improve crop resistance to abiotic and biotic stresses. In this study, we performed a comprehensive genome-wide analysis of conserved domains, gene structure and protein motifs, chromosome locations, duplication relationships, and expression profiles of TaDnaJs under hormone stress and plant virus infection. Results of qRT-PCR and yeast two-hybrid assay demonstrates that the TaDnaJ family is involved in plant defense mechanisms, and forms a basis for subsequent experiments.

Acknowledgements

This work was supported by the National Key R&D Plan of China (2018YFD0200507, 2017YFD-0201701, and 2018YFD0200408), the National Natural Science Foundation of China (31901954), the Natural Science Foundation of Ningbo City, China (2019A610415 and 2019A610410), the National Key Project for Research on Transgenic Biology, China (2016ZX08002-001), the China Modern Agricultural Industry Technology System (CARS-03), and the K.C. Wong Magna Funding in Ningbo University.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendices associated with this paper are available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

References


143, 2203–2214.