Phylogeny-dominant classification of J-proteins in Arabidopsis thaliana and Brassica oleracea

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0206.R2
Manuscript Type:	Article
Date Submitted by the Author:	03-Mar-2018
Complete List of Authors:	Zhang, Bin; Hunan Agricultural University Qiu, Hanlin; b State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University Qu, Donghai; Hunan Agricultural University Ruan, Ying; Hunan Agricultural University CHEN, Dong-Hong; State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	J-proteins, Arabidopsis, Brassica oleracea, phylogenetic analysis, domain organization, gene structure

SCHOLARONE™
Manuscripts

Phylogeny-dominant classification of J-proteins in *Arabidopsis thaliana* and *Brassica oleracea*

Bin Zhang ^{a,#}, Han-Lin Qiu ^{b,#}, Dong-Hai Qu ^a, Ying Ruan ^{a,*}, Dong-Hong CHEN ^{b,*}

^a Key Laboratory of Education Department of Hunan Province on Plant Genetics and Molecular Biology, Hunan Agricultural University, Changsha 410128, China

^b State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Lin'an, Zhejiang 311300, China

[#] The authors with equal contributions

^{*} Corresponding author: chendh212@163.com (D.H. Chen); yingruan@hotmail.com (Y. Ruan)

zhangbin27104@163.com (B. Zhang)

qiuhanlinyc@163.com (H. Qiu)

qudonghai1994@163.com (D. Qu)

yingruan@hotmail.com (Y. Ruan)

chendh212@163.com (D.H. Chen)

Abstract: HSP40s or DnaJ/J-proteins are evolutionarily conserved in all the organisms as cochaperones of molecular chaperone HSP70s that mainly participate in maintaining cellular protein homeostasis, such as protein folding, assembly, stabilization, and translocation under normal conditions as well as refolding and degradation under environmental stresses. It has been reported that Arabidopsis J-proteins are classified into four classes (type A~D) according to domain organization, but their phylogenetic relationships are unknown. Here, we identified 129 J-proteins in worldwidely popular vegetable *Brassica oleracea* as a close relative of model plant Arabidopsis and also revised the information of Arabidopsis J-proteins based on the latest online bioresources. According to phylogenetic analysis with domain organization and gene structure as references, the J-proteins from Arabidopsis and *B. oleracea* were classified into fifteen main clades (I~XV) separated by a number of undefined small branches with remote relationship. Based on the number of members, they respectively belong to multigene clades, oligo-gene clades, and mono-gene clades. The J-protein genes from different clades may function together or separately to constitute a complicated regulatory network. This study provides a constructive viewpoint for J-proteins classification and an informative platform for further functional dissection and resistant genes discovery related to genetically improvement of crop plants.

Key words: J-proteins, Arabidopsis, *Brassica oleracea*, phylogenetic analysis, domain organization, gene structure

Introduction

Prokaryotic and eukaryotic organisms respond and adapt to abiotic stress via accumulation of heat shock proteins (HSPs) as a defense mechanism. HSPs are divided into seven groups based on their molecular weight: the small HSPs (IbpA/B in *Escherichia coli*), HSP10 (GroES), HSP60 (GroEL), HSP40/J-proteins (DnaJ), HSP70 (DnaK), HSP90 (HtpG), HSP100 (ClpB/A/C) (Georgopoulos and Welch 1993; Lindquist and Craig 1988). HSP40/J-protein is the co-chaperone of molecular chaperone HSP70 and stimulates the ATPase activity of HSP70. HSP40 co-chaperones usually have more members than Hsp70 chaperones. For instance, there are 6 types of HSP40s and 3 HSP70s in *E. coli*, 22 HSP40s and 14 HSP70s in yeast, 41 HSP40s and 12 HSP40s in human, and even >100 HSP40s and 16 HSP70s in the model plant *Arabidopsis*, respectively (Qiu et al. 2006; Rajan and D'Silva 2009; Walsh et al. 2004). Therefore, the specific HSP70 may bind a diversity of J-proteins to perform multiple molecular functions and produce different biological effects.

J-protein is characteristic of the conserved DnaJ (J)-domain which is responsible for interaction with the nucleotide-binding domain of HSP70. According to the domain organization, J-proteins have been classified into four types (Cheetham and Caplan 1998; Walsh et al. 2004). Type A J-proteins share all the four domains with their *E. coli* prototype DnaJ, sequentially including a high conserved N-terminal DnaJ-domain (J-domain), a Gly/Phe-rich region, a zinc finger domain, and a C-terminal domain. Type B J-proteins just lack zinc-finger domain, while Type C J-proteins only carry a single J-domain. Type D J-proteins are identical to J-like proteins with the J-domain but absent of the extremely conserved HPD motif. For instance, there are 8 Type-A J-proteins DJAs, 16 Type-B DJBs, 82 Type-C DJCs, and 3 Type-D DJLDs in *Arabidopsis* (Chiu et al. 2013; Finka et al. 2011). However, the phylogenetic relationships among them are unclear.

J-proteins have widely localization not only in the cytoplasm and nucleus, but also in the endoplasmic reticulum (ER), chloroplast, and mitochondria (Miernyk 2001; Rajan and D'Silva 2009; Finka et al. 2011). For instance, *Arabidopsis* has been identified to have 6 and 19 J-proteins in ER and chloroplasts, respectively (Chiu et al. 2013; Ohta et al. 2013; Yamamoto et al. 2008). Moreover, J-proteins not only act as the co-chaperones participating

in diverse cellular events including protein folding, assembly, translocation, stabilization, refolding, and degradation (Du et al. 2013), but also perform the other roles as enzyme or epigenetic regulator (de Crouy-Chanel et al. 1995; Richly et al. 2010). In *Arabidopsis*, J-proteins are extensively involved in growth and development, gravity perception, hormone regulation and response to abiotic stress (Bekh-Ochir et al. 2013; Boonsirichai et al. 2003; Chen et al. 2011; Christensen et al. 2002; Feng et al. 2016; Harrison and Masson 2008; Kneissl et al. 2009; Kroczyńska et al. 1996; Ma et al. 2015; Park et al. 2011; Park and Kim 2014; Petti et al. 2014; Pulido et al. 2013; Tamura et al. 2007).

Brassica oleracea comprises a variety of widely popular vegetable cultivars displaying extreme morphological diversity, such as headed cabbage, broccoli, cauliflower, leaf cabbage (kale), Brussels sprouts, collard greens (collards), savoy cabbage, kohlrabi (turnip cabbage), and gai lan (Chinese kale). The leaf, stem or flowering head in specific cultivars are edible, closely related to human life. *B. oleracea* has the close genetic relationship with the model plant *Arabidopsis thaliana*, both belonging to the family *Brassicaceae*. Newly released genome of *B. oleracea* contributes to the study on polyploid genome evolution during *Brassica* speciation and on morphological diversification during domestication and breeding (Cheng et al. 2016; Liu et al. 2014), and also facilitates further investigation on specific gene(s), gene family, signal/metabolic pathway and network involved in *B. oleracea* growth and development, and stress responses. Furthermore, technique achievements of genetic transformation, plastid transformation and CRISPR/Cas9-mediated genome editing in *B. oleracea* pave the way to crop modification (Lawrenson et al. 2015; Liu et al. 2012; Tseng et al. 2014). Here we reevaluated *Arabidopsis* J-proteins and identified J-proteins in *B. oleracea* at the genome-wide level. We systematically analyzed their J-proteins on the aspects of gene classification, chromosomal localization, gene structure, domain organization, and phylogenetic relationship. This study may offer a primary panorama of J-proteins in *B. oleracea* and a novel classification of J-proteins, as well as provide some clues for potential resistant genes for accelerating genetic improvement in *B. oleracea* and other crop plants.

Materials and Methods

Identification and chromosome distribution of J-proteins in *B. oleracea*

J proteins in *Brassica oleracea capitata* v1.0 (Savoy cabbage) were identified from the plant genomics resource Phytozome v12.1

(https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Boleraceacapitata) via two approaches. ① ‘DnaJ’ was used as keyword in term of search type to identify J proteins in the website

(https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Boleraceacapitata). The hits were then subject to on-line SMART mode

(<http://smart.embl-heidelberg.de/>) with ticked option of PFAM domains or NCBI Batch Web CD-Search Tool (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) with Expect

Value threshold (set as 1) to ensure the presence of J-domains. ② In order to thoroughly identify the J proteins and avoid omission of the unannotated ones, we firstly collected all the genomic, CDS, and amino acid sequences of Arabidopsis J protein family genes, based on the published data (Rajan and D’Silva 2009; Finka et al. 2011; Chiu et al. 2013) and public TAIR Arabidopsis genomic database (www.arabidopsis.org), and then amino acid sequences of known Arabidopsis J proteins were used as queries to BLASTp against *Arabidopsis thaliana* proteome to obtain as many hits as possible in combination with domain detection. These hits were used to search *B. oleracea capitata* v1.0 database depositing in Phytozome. Reciprocal BLAST was used for further confirmation.

J-containing gene loci information of *B. oleracea capitata* in chromosomes were retrieved from Phytozome website. The genes were mapped to the chromosomes via MapChart program (Voorrips 2002). Tandem duplications indicated the tandemly arrayed genes with the close phylogenetic relationships are located at the same chromosomal location within ~100 kb (Kong et al. 2007).

Examination of gene structure, domain organization and phylogenetic relationship

Exon–intron gene structure schematics of J-domain-containing genes were analyzed by online program Gene Structure Display Server GSDS 2.0 (<http://gsds.cbi.pku.edu.cn/>) (Hu et al. 2015). The domain organization of J family proteins were analyzed using SMART and Pfam

databases. J proteins were aligned using Clustalw (Thompson et al. 1994), and phylogenetic analysis was performed using MEGA6 (Tamura et al. 2013). The full-length amino acid sequences of J-proteins were used for constructing neighbor-joining (NJ) trees with the following settings: pairwise deletion option for gaps/missing data treatment; *p*-distance method for Substitution model; and bootstrap test of 1000 replicates for internal branch reliability.

Results

Identification of J-proteins from *Arabidopsis thaliana* and *B. oleracea*

Systematic identification of J-proteins in *Arabidopsis* has been carried out several times as yet (Finka et al. 2011; Miernyk 2001; Rajan and D'Silva 2009). Most recently but still six years ago, Finka et al (2011) excluded some transposable elements and repeating items from *Arabidopsis* J/J-like proteins collected by Rajan and D'Silva (2009), corrected the total number from 120 to 108. Subsequently, Chiu et al (2013) found a new member, DJC82/At3G05345 during the investigation of chloroplast J proteins in *Arabidopsis*. Here, we further checked the published data and rescreened *Arabidopsis* J-protein members from Phytozome bioresource. We found that two type-C proteins (DJC58/At1g69050 and DJC49/At1g80120) and three type-D proteins (DJLD1/At5g11500, DJLD2/At3g59280 and DJLD3/At5g61880) should not be J-proteins due to absence of typical J-domain according to the detection of both online SMART and NCBI-CD webtools. Additionally, we also identified another 13 different members (Table S1) in which ARC6/DJLD14, CRRL/DJLD4 and EIP9/DJLD5 have been reported (Park et al. 2011; Vitha et al. 2003; Yamamoto et al. 2011), consequently giving rise to 117 *Arabidopsis* J-proteins in total. To identify the J proteins in *B. oleracea*, two search strategies were performed through both keyword search and BLASTp search against phytozome database (described in materials and methods). Finally, we obtained 129 J-proteins in *B. oleracea* (Table S2).

As conventional classification only based on the domain organization, J-proteins (including J-like proteins) can be simply divided into four types (Finka et al. 2011; Miernyk 2001; Rajan

and D'Silva 2009). However, the phylogenetic relationship among them are far from known. In fact, J-proteins characteristic of the conserved J-domain usually display greatly diversity in the rest of amino acid sequences. Thus, we performed phylogenetic analysis at the same time using domain organization as a reference to classify J-proteins. The results showed that J-proteins in *Arabidopsis* and *B. oleracea* are divided into 15 major clades (I~XV) often separately by low-score branches which might represent the unrelative mono-genes (Fig. 1). Clade I and II contain more than ten members in *Arabidopsis* and *B. oleracea*, named as multigene clades. Clade III to XV harbor 2 to 7 members, named as oligo-gene clades. The remaining separate clades have a single member at least in *Arabidopsis* or/and *B. oleracea*, named as mono-gene clades.

Based on the genomic information from Phytozome database, J-protein genes of both species have been mapped across all the chromosomes (Fig. 2). On the whole, J-protein genes along the chromosomes in *Arabidopsis* are denser than those in *B. oleracea*, possibly due to *Arabidopsis* with small genome and low number of chromosomes. Furthermore, the number of J-protein genes per chromosome varies in both species, such as chromosome 3 (chr3) and chr7 respectively contains 26 and 8 genes in *B. oleracea*. Chromosomal distribution of genes in each clade is usually uneven, such as *Arabidopsis* clade-I has 9 genes on chr5 but has only one member on chr1. In addition, tandem duplication events were identified in *Arabidopsis* including DJC68/DJC84 pairs (clade I), DJC53/DJC52 (clade I), and DJC6/DJC4 (clade XV) as well as *B. oleracea* including Bol015437/015438 (clade IX), Bol008689/008687 (clade XII), and Bol035735/035736 where Bol035735 belongs to clade II, and Bol35736 shares high sequence similarity with Bol35735 but only absent of N-terminal J-domain. These tandemly duplicated paralogs in both species didn't share homologs, indicating these events happened after the divergence of both species.

Multigene clades: Clade I~II

Multigene clade indicates the clade with more than ten members in each species, such as clade I and II (Fig. 3). Clade I contain 20 and 14 members in *Arabidopsis* and *B. oleracea*, respectively. The genes in this clade lack intron or only have few introns. Clade I can be

clustered into three subclades, I-1~I-3. All the members in subclade I-1 is characteristic of another one or double C-terminal DUF3444 domains with unknown function, compared with only one member from *Arabidopsis* and its ortholog from *B. oleracea* in subclade I-2 containing DUF3444 domain, which is thoroughly absent in subclade I-3. However, DJC56 in I-2 has another C-terminal WCC domain found in retrotransposons and Gemini viruses (Nakayashiki et al. 2001). The function of this clade was just slightly referred to in *Arabidopsis* so far. DJLD5/EIP9 is localized in both cytoplasm and nucleus, and interacts with EMF1 to implicate in flowering time control (Park et al. 2011).

Clade II respectively contain 17 and 22 members in *Arabidopsis* and *B. oleracea*. By means of the latest on-line Simple Module Architecture Tool (SMART), we found clade II usually have the second distinctive DnaJ C-terminal domain DnaJ_C, but which is absent in DJC20 and its ortholog Bol008420. Clade II can be divided into two subclades, subclade II-1 and II-2. The genes in subclade II-1 usually have more introns than those in subclade II-2. Additionally, subclade II-1 proteins usually possess the third characteristic DnaJ central domain DnaJ_CXXCXGXG which is imbedded in the N-terminus of DnaJ_C domain. DnaJ_CXXCXGXG domain contains four cysteine-rich repeats of the motif CXXCXGXG (X indicates any amino acid), where each set of two repeats associates with one unit of Zn^{2+} . Exceptionally, DJC20/Bol008420 branch uses CCHC-type zinc finger (ZnF_C2HC) domain instead. DJA4/Bol008420 branch carries a TM domain behind the Dna_C domain. Clade II proteins in *Arabidopsis* have been reported to play diverse roles. Mitochondrial-targeted DJC20/AtJ1 is involved in a wide range of growth and development, such as seedling growth, flowering time control, seed setting, and embryogenesis (Park and Kim 2014). Additionally, the expression of DJC20/AtJ1 was induced by salinity, dehydration and abscisic acid (ABA), and DJC20/AtJ1 also play a crucial role in stress response, such as thermotolerance, oxidative and osmotic stress (Park and Kim 2014; Wang et al. 2014b; Zhou et al. 2012). DJA3/GFA2 localized in the mitochondrial matrix is an ortholog of the yeast mitochondrial J protein Mdj1p, and is required for synergid cell death and polar nuclei fusion during megagametogenesis (Christensen et al. 2002). DJA1/J3 (At3g44110) mediates the integration of flowering signals via interacting with SVP during the floral transition, is involved in abiotic stress tolerance and activates plasma membrane H^+ -ATPase through interacting with

and suppressing protein kinase 5 (PKS5) activity under alkaline conditions (Salas-Munoz et al. 2016; Shen et al. 2011; Yang et al. 2010). DJA8/ERdj3b mediating ER-quality control is required for the activation of cell death and defense responses in *bir1-1* mutant with disrupted BAK1-INTERACTING RECEPTOR-LIKE KINASE 1 (BIR1) functions (Sun et al. 2014). Furthermore, DJA8/AtERdj3b associated with AtP58^{IPK} is involved in polar nuclear-membrane fusion (Maruyama et al. 2014).

In addition, we found our results mainly according to phylogenetic analysis were different from the classical domain-based J-protein classification (type A~D). In Arabidopsis, subclade II-1 consisting of seven type-A DJAs (DJA1~7) and one type-C DJC20 and subclade II-2 consisting of seven type-B DJBs (DJB 1~6, and DJB10), one type-A DJA8, and one type-C DJC51 are not completely consistent with previous type-A and type-B, respectively. So, phylogenetic-based classification in this study may reflect more truthful evolutionary relationship among J-proteins.

Oligo-gene clades: clade III~XV

Oligo-gene clade represents the clade with less than ten members in each species, such as clade III to XV (Fig. 4). Clade III contain 5 and 6 members in Arabidopsis and *B. oleracea*, respectively. The genes in this clade are composed of 7~10 exons. DnaJ-X domain (PF14308) with unknown function is the symbol of this clade.

Clade IV, XIV, and XV encompassing proteins only contain the signature J-domains in the C-terminus. The proteins in Clade IV encoded by one exon are short (~160 amino acid), whereas those in clade XIV encoded by six exons are medium (250~300 aa) and those in clade XV mostly encoded by eight exons (5~9) have the longest but diverse sizes (400~1500 aa). Clade IV protein DJC23/J11 is located in chloroplast stroma, involved in optimization of photosynthesis and stress response (Chen et al. 2010; Orme et al. 2001). Mitochondrial-localized DJC60/BIL2 in clade XIV promotes cell elongation downstream of BR signaling through improving ATP synthesis in mitochondria; BIL2 also participates in stress resistance against salinity and strong light (Bekh-Ochir et al. 2013). In clade XV, the cytosolic-targeted DJC15/JAC1 is required for phototropin-mediated chloroplast movement,

due to its function in regulating the biased chloroplast-actin filament distribution (Ichikawa et al. 2011; Suetsugu et al. 2005; Takano et al. 2010). DJC3/AUL1 as a auxilin-like protein might be related to protein translocation, because mammalian auxilin in conjunction with HSP70 is involved in removal of the lattices of clathrin-coated vesicles that mediate the vesicular transport of proteins between organelles (Quesada et al. 1999; Ungewickell et al. 1995)

Clade V respectively contain 3 and 2 members in Arabidopsis and *B. oleracea*. C-terminal DUF1977 domain (PF09320) with uncharacterized function is the representative domain of this clade except DJB15. In addition, DJC35 specific to Arabidopsis contains a transmembrane (TM) domain situating between J-domain and DUF1977. DJB14 and DJB15 independent of subclade II-2 containing most of previously classified DJB members were closer to DJC38 based on our study.

Clade VI ~ VIII and XII contain the members only carrying signature J-domains in N-terminus, except DJC48 with another TM domain in C-terminus. The genes in clade VI consist of eight exons encoding over 260 aa, those in clade VII consist of 4~6 exons encoding around 120 aa, those in clade VIII consist of 5~7 exons encoding over 160 aa, and those in clade XII consist of 9~11 exons encoding about 400 aa. No gene was functionally examined so far in clade VI and VII. clade VII gene *DJC47(atDjC53)* localized in the nucleus plays a negative role in thermotolerance (So et al. 2014). In clade XII, *DJB11/ALTERED RESPONSE TO GRAVITY1 (ARG1)* and *DJB13/ARG1-LIKE2 (ARL2)* but not *DJB12/ARL1* participate in the gravity signaling process required for root and hypocotyl gravitropism, and influence PIN3 relocalization and auxin redistribution upon gravistimulation (Boonsirichai et al. 2003; Guan et al. 2003; Harrison and Masson 2008; Sedbrook et al. 1999).

Clade IX was made up of DJC1/ZRF1a and DJC2/ZRF1b in Arabidopsis, each of which has two orthologs in *B. oleracea*. This clade genes usually are encoded by a single exon, except *Bol015437* with two exons. The ZRF1-like proteins usually contain a pair of closely juxtaposed SANT domains at the C terminus besides N-terminal J-domains. However, *Bol015437* and *Bol015438* as AtZRF1b orthologs lack SANT domains, which are found in transcriptional co-repressors and in the subunits of some chromatin remodeling complexes (Aasland et al. 1996). *ZRF1* orthologs are only existed in eukaryotes, not only retaining the

conventional function as co-chaperones but also obtaining diverse neofunctions (Chen et al. 2014). Arabidopsis *AtZRF1a/b* are involved in a wide range of growth and development, such as seed germination, organ size, meristem activity, flower development, gametophyte transmission and embryogenesis (Feng et al. 2016; Guzman-Lopez et al. 2016).

Clade X consists of DJC17/ATERDJ2A and DJC18/ATERDJ2B in Arabidopsis, as well as Bol021427 and Bol028398 in *B. oleracea*. *DJC17* and its ortholog *Bol021427* are composed of 10 exons, and *DJC18/Bol028398* are composed of 8 exons. J-domain in this clade situated between a pair of TM domains at the N-terminus and Sec63 domain at the C-terminus. Sec63 domain was named after the yeast Sec63p, which is a subunit of the protein translocation channel resident in ER membrane, involved in the biogenesis of secretory and transmembrane proteins and ER-associated Degradation (ERAD) of misfolded proteins (Romisch 2005; Servas and Romisch 2013). In Arabidopsis, AtERdj2A is necessary for plant growth and pollen germination; the *aterdj2a*^{-/-} mutant is lethal, but *aterdj2b*^{-/-} looks normal (Yamamoto et al. 2008).

Clade XI possesses three members each in Arabidopsis and *B. oleracea*, in which genes are composed by 9 exons. J-domain were localized between a batch of TM domains and Jiv90 domain, which indicates the region of a bovine J-protein Jiv interacting with viral polypeptide (Muller et al. 2003; Rinck et al. 2001).

Clade XIII can be divided into two subclades, XIII-1 and XIII-2. The genes in XIII-1 have three exons, much less than those in XIII-2 with around ten exons. The proteins in XIII-1 branch sometimes harbor another single TPR (tetratricopeptide repeat) domain, and those in XIII-2 branch contain multiple tandem TPR repeats prior to the J-domain. TPR domain is a structural motif of 34 amino acids residues and is often organized into tandem arrays of 3~16 repeats, which provide scaffolds to mediate protein-protein interactions and the assembly of multisubunit complexes (Blatch and Lassar 1999; D'Andrea and Regan 2003). In fact, TPR domains can couple with various domains to perform the diverse functions (Prasad et al. 2010).

Mono-gene clades: the other undefined clades

Mono-gene clades indicated the small disperse branches with distant relationships separated by above mentioned well-classified clades. As shown in Fig. 5, almost each Arabidopsis gene and its ortholog(s) in savoy cabbage here represent an individual clade, excluding some genes specific to Arabidopsis or savoy cabbage, such as *DJC72* and *Bol0115152*. The genes with closer relationship usually display similar gene structure and protein domain organization, which is a critical reference for identification of homologs, especially for the genes with higher number of introns and the proteins with recognizable domain(s). For instance, *DJC27/Bol039781* pairs and *DJC71/Bol032717* pairs should be the paralogs for each other. Due to too many clades present here, we would like focus on some special ones. *Bol023096* contains a single N-terminal IQ motif with conserved Ile and Gln residues. The IQ motif can be present in one or more copies, functioning as a binding site for calmodulin (CaM) and phosphorylation site of protein kinase C (PKC) (Baudier et al. 1991; Rhoads and Friedberg 1997). *Bol026594* harbors a N-terminal APO_RNA-bind domain, which resembles zinc fingers to bind zinc as well as serves as RNA binding domain (Watkins et al. 2011). *DJC14* and its orthologs *Bol015454* and *Bol043950* encompass a C-terminal oligomerisation domain HSCB_C, which is found in HscB/HSC20 as a co-chaperone to regulate the ATPase activity and peptide-binding specificity of the HSP70 chaperone HscA/HSC66 (Cupp-Vickery and Vickery 2000). *DJC10/Bol007713* branch contain a C-terminal zf-CSL motif which contains four conserved cysteine residues to chelate a single Zn^{2+} ion (Sun et al. 2005). *DJC76/Bol017088* and *DJC77/DJC82* branches cherish a Fer4_13 domain at the central region but behind the J-domain. Fer4_13 domain is a [4Fe-4S] single cluster domain of Ferredoxin I firstly identified in sulfate-reducing bacterium (Sery et al. 1994). J-proteins displaying a bacterial ferredoxin domain [4Fe-4S] cluster has been reported in *Chlamydomonas reinhardtii* (Dorn et al. 2010), likely acquired in horizontal gene transfer events (Petitjean et al. 2012). *DJC34/Bol017123* branch contain a C-terminal RNA recognition motif RRM, found in a variety of RNA binding proteins and a few single stranded DNA binding proteins (Birney et al. 1993). In addition, the branches of *DJC27/Bol027779*, *DJC57/Bol005191*, *DJC25 (OWL1)/Bol037782*, and *DJLD14 (ARC6)/Bol015212* respectively contain the domains of DUF4339, DUF3752, DUF3395, and DUF4101 all with unknown functions. In these mono-gene clades, the functions of some Arabidopsis J-proteins have been examined. *DJC27*

(*GRV2/KAM2*) is homolog to *Caenorhabditis elegans* *RECEPTOR-MEDIATED ENDOCYTOSIS-8* (*RME-8*), which plays an essential role in the endocytotic pathway in animal cell; Similarly, *GRV2* is localized in endosome, required for proper endosome formation related to vesicle trafficking from the multivesicular body (or pre-vacuolar compartment) to the lytic vacuole and also for shoot gravitropism (Silady et al. 2008; Silady et al. 2004; Tamura et al. 2007). *DJC25* (*OWL1*) localizes to the nucleus and cytoplasm, and light can induce its accumulation in the nucleus (Kneissl et al. 2009). *OWL1* participates in regulating plant growth and development, such as germination, cotyledon opening, hypocotyl elongation, and flowering time and perceiving very low light fluence essential for plant survival under unfavorable light conditions (Kneissl et al. 2009). *DJLD14/ARC6* (*ACCUMULATION AND REPLICATION OF CHLOROPLASTS 6*) homologous to the cyanobacterial cell division protein *Ftn2* is localized to the chloroplast envelope membrane, where it participates in plastid division (Vitha et al. 2003). *DJB9/AtERdj3A/TMS1* (*THERMOSENSITIVE MALE STERILE 1*) with disulfide isomerase activity, is required for thermotolerance of pollen tubes and other plant tissues, possibly functioning as a cochaperone of the ER-localized HSP70 protein BiP to participate in refolding or degradation of unfolded and misfolded proteins (Ma et al. 2015; Yang et al. 2009). *DJC22/J8* and *DJC26/J20*, like above-mentioned *DJC23/J11*, are small chloroplast-targeted J-proteins, involved in optimization of photosynthesis (Chen et al. 2010; Chen et al. 2011; Orme et al. 2001). Furthermore, *DJC26/J20* functions as an adaptor to recognize inactive forms of deoxyxylulose 5-phosphate synthase (*DXS*) and deliver them to HSP70 chaperone for proper folding in synergy with ClpB3 or degradation together with Clp protease (Pulido et al. 2016; Pulido et al. 2013). Thylakoid-targeted *DJC75/CRRJ/NdhT* and *DJLD4/CRRL* as the components of NAD(P)H dehydrogenase (*NDH*) subcomplex A are required for the activity of *NDH* which mediates photosystem I (*PSI*) cyclic electron flow (Ifuku et al. 2011; Yamamoto et al. 2011). *DJC76/atDjC17* (*At5g23240*) is involved in position dependent cell fate determination during root development (Petti et al. 2014).

Discussion

Phylogenetic-dominant classification contributes to overall understanding of J-proteins

The classical classification of J-proteins based on the domain organization simplifies the complexity of further analysis to a certain extent. However, sometimes the proteins belonging to the same types have remote evolutionary relationship. Here we pay more attention to phylogenetic relationship of J-proteins, providing a different viewing angle for classification and annotation of J-proteins.

With the constant updating of available genomic data and on-line public biotools, the identification and annotation of J-proteins in Arabidopsis has been greatly improved, which can be reflected by the three publications from the year 2001, 2009 to 2011 (Finka et al. 2011; Miernyk 2001; Rajan and D'Silva 2009). However, these publications each use similar nomenclature system but with reordered serial number, which prefers to cause confusion and even misunderstanding. For instance, At3g44110 was named as *atDjA3* by Miernyk (2001) but renamed as *atDjA1* by Rajan and D'Silva (2009) and *DJA1* by Finka et al (2011). In this study, we followed the J-protein nomenclature system by Finka et al (2011). We checked the published data by Finka et al (2011), finding that DJC49 and DJC58 (At1g80120 and At1g69050 belonging to Type-C before) as well as DJID1~3 (At5g11500, At3g59280, and At5g61880 belonging to Type-D before) lack J-domain according to the latest public SMART and Pfam databases. Later on, we searched the Arabidopsis genome again and found another 13 novel J-proteins though 11 of them lack conserved HPD motif in the J-domain, previously named J-like domain belonging to Type D. Additionally, DJLD14/ARC6 contain the so-called J-like domain with high E-value (30.136). Further phylogenetic analysis showed that these proteins with so-called J-like domains fail to cluster together but are grouped with the other different clades. So, the classification according to phylogenetic relationship with domain organization as a reference can provide a valuable clue for further dissection of J-proteins. By the similar means, we finally identified 129 J-proteins in *B. oleracea*, but 12 of them absent of HPD motif and the one (Bol015212) containing the J-domain with a little high E-value (0.2).

J-proteins can function in individual and coordinate manners

J-proteins can function in the cytosol, nucleus, and different organelles. The J-proteins with the similar subcellular localization maybe come from different clades and are involved in similar cellular events but with different roles. For instance, ER-resident J-proteins, AtP58^{IPK} (DJC16, mono-gene clades) respectively associated with AtERdj3A (DJB9, mono-gene clades) and AtERdj3B (DJA8, clade II), function at distinct steps of polar nuclear-membrane fusion (Maruyama et al. 2014). The chloroplast-targeted J8 (DJC22, mono-gene clades), J11 (DJC23, clade IV) and J20 (DJC26, mono-gene clades) are involved in optimization of various reactions of photosynthesis; correspondingly, *j8*, *j11* or *j20* single mutants display a series of common phenotype including decreased photosynthetic efficiency, destabilized PSII complexes and unbalanced redox reactions in chloroplasts (Chen et al. 2010; Chen et al. 2011; Orme et al. 2001). Additionally, J-protein paralogs usually have redundant functions. Clade XII genes *ARG1* (*DJB11*) and *ARL2* (*DJB13*) function in the same gravity signal transduction pathway required for normal root and hypocotyl gravitropism (Boonsirichai et al. 2003; Guan et al. 2003; Harrison and Masson 2008; Sedbrook et al. 1999). Loss of function of either *ZRF1a* (*DJC1*) or *ZRF1b* (*DJC2*) fails to influence plant growth and development, but knockout of both can produce pleiotropic phenotypes almost throughout the life cycle (Feng et al. 2016; Guzman-Lopez et al. 2016). *B. oleracea* also has the corresponding orthologs, which should have the similar functions.

J-proteins are widely involved in responses to biotic and abiotic stresses

In plants, J-proteins play critical roles in response to environmental stresses. *Atriplex nummularia* ANJ1 is a functional homolog of yeast cytoplasmic YDJ1/MAS5, which is required for mitochondrial protein import and protein translocation across the ER (Atencio and Yaffe 1992; Caplan et al. 1992). *ANJ1* can complements the yeast *mas5* temperature-sensitive mutation, and its expression is induced by heat shock and salt stress (Zhu et al. 1993), so does in *Salix gilgiana* *SGJ3* (Futamura et al. 1999). Tobacco DnaJ ortholog *NtDnaJ1* expression is obviously enhanced by drought stress. Over-expression of *NtDnaJ1* enhances tolerance to osmotic or drought stress in transgenic Arabidopsis (Xia et al. 2014). The tomato chloroplast-targeted J-protein (*LeCDJ1*) played an essential role in

maintaining photosystem II (PSII) activity under chilling or heat stress; *LeCDJ1* overexpression enhanced tolerance to chilling or heat in transgenic tomatoes (Kong et al. 2014a; Kong et al. 2014b). Ectopic expression of another tomato chloroplast-targeted DnaJ gene *LeCDJ2/SICDJ2* enhanced tolerance to drought stress and resistance to the pathogen *Pseudomonas solanacearum* in transgenic tobacco (Wang et al. 2014a). Overexpression of *SICDJ2* in tomato enhanced thermotolerance by protecting Rubisco activity and maintaining carbon assimilation capacity under heat stress (Wang et al. 2015). Furthermore, J-proteins have been found to participate in response to biotic stress even through different mechanisms. The nuclear-localized soybean J-protein *GmHSP40.1* was involved in hypersensitive response (HR)-like cell death and pathogen defense; Overexpression of *GmHSP40.1* caused HR-like cell death, and silencing of *GmHSP40.1* enhanced the susceptibility of soybean plants to *Soybean mosaic virus* (SMV) but did not influence R gene-mediated resistance to SMV (Liu and Whitham 2013). *Nicotiana benthamiana* J-proteins NbMIP1s can recognize both viral movement protein MP and resistance factor Tm-2², required for Tobacco mosaic virus (TMV) infection and plant innate immunity (Du et al. 2013). overexpression of *NbMIP1.1a* did not cause cell death, and silencing of *NbMIP1s* reduced both TMV infection and R-mediated resistance (Du et al. 2013). However, J-proteins also become the inevitable targets that viruses attack. The TMV movement protein MP of can bind to the host J-protein NtMIP1, leading to viral cell-to-cell spread (Shimizu et al. 2009). The capsid protein (CP) from potyviruses Potato virus Y (PVY) are able to interact with and hijack the tobacco host J-proteins NtCPIPs (NtCPIP1 and NtCPIP2a), which functions as essential susceptibility factors during PVY infection, possibly through recruiting HSP70 chaperones for viral assembly and intercellular spread (Hofius et al. 2007). Accordingly, overexpression of J-domain-disrupt versions of *NtCPIP1* and *NtCPIP2a* evidently increased the virus resistance of transgenic plants. Interestingly, the pathogenic *Pseudomonas syringae* can by its own produce J-protein HopI1 as a virulence effector to bind host HSP70 for suppressing plant immune responses (Jelenska et al. 2010). So, through investigation of the molecular mechanisms underlying J-protein-mediated stress tolerance and sensitiveness, different strategies may be designed to improve the resistance of crop plants to abiotic and biotic stresses. This study provides an informative platform for comprehensively exploiting the functions of J-proteins in

Arabidopsis and *B. oleracea*, therefore contributing to further genetic modification of cruciferous vegetables for biological traits and applicative qualities.

Acknowledgements This work was supported by the National Natural Science Foundation of China (30800595), Natural Science Foundation of Hunan Province of China (2015JJ3075), and Open project of Hunan Provincial Key Laboratory for Germplasm Innovation and Utilization of Crop (16KFXM02).

Conflict of interest All authors declare that they have no conflict of interest.

Ethical standards This article does not contain any studies with animals performed by any of the authors.

References

- Aasland, R., Stewart, A.F., and Gibson, T. 1996. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends in biochemical sciences* **21**(3): 87-88.
- Atencio, D.P., and Yaffe, M.P. 1992. MAS5, a yeast homolog of DnaJ involved in mitochondrial protein import. *Molecular and cellular biology* **12**(1): 283-291.
- Baudier, J., Deloulme, J.C., Van Dorsselaer, A., Black, D., and Matthes, H.W. 1991. Purification and characterization of a brain-specific protein kinase C substrate, neurogranin (p17). Identification of a consensus amino acid sequence between neurogranin and neuromodulin (GAP43) that corresponds to the protein kinase C phosphorylation site and the calmodulin-binding domain. *The Journal of biological chemistry* **266**(1): 229-237.
- Bekh-Ochir, D., Shimada, S., Yamagami, A., Kanda, S., Ogawa, K., Nakazawa, M., Matsui, M., Sakuta, M., Osada, H., Asami, T., and Nakano, T. 2013. A novel mitochondrial DnaJ/Hsp40 family protein BIL2 promotes plant growth and resistance against environmental stress in brassinosteroid signaling. *Planta* **237**(6): 1509-1525. doi: 10.1007/s00425-013-1859-3.
- Birney, E., Kumar, S., and Krainer, A.R. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic acids research* **21**(25): 5803-5816.
- Blatch, G.L., and Lassle, M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**(11): 932-939. doi: 10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N.
- Boonsirichai, K., Sedbrook, J.C., Chen, R., Gilroy, S., and Masson, P.H. 2003. ALTERED RESPONSE TO GRAVITY is a peripheral membrane protein that modulates gravity-induced cytoplasmic

- alkalinization and lateral auxin transport in plant statocytes. *The Plant cell* **15**(11): 2612-2625. doi: 10.1105/tpc.015560.
- Caplan, A.J., Cyr, D.M., and Douglas, M.G. 1992. YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism. *Cell* **71**(7): 1143-1155.
- Cheetham, M.E., and Caplan, A.J. 1998. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell stress & chaperones* **3**(1): 28-36.
- Chen, D.H., Huang, Y., Liu, C., Ruan, Y., and Shen, W.H. 2014. Functional conservation and divergence of J-domain-containing ZUO1/ZRF orthologs throughout evolution. *Planta* **239**(6): 1159-1173. doi: 10.1007/s00425-014-2058-6.
- Chen, K.M., Holmstrom, M., Raksajit, W., Suorsa, M., Piippo, M., and Aro, E.M. 2010. Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in *Arabidopsis thaliana*. *BMC plant biology* **10**: 43. doi: 10.1186/1471-2229-10-43.
- Chen, K.M., Piippo, M., Holmstrom, M., Nurmi, M., Pakula, E., Suorsa, M., and Aro, E.M. 2011. A chloroplast-targeted DnaJ protein AtJ8 is negatively regulated by light and has rapid turnover in darkness. *Journal of plant physiology* **168**(15): 1780-1783. doi: 10.1016/j.jplph.2011.04.001.
- Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., Liu, B., Liang, J., Zhuang, M., Liu, Y., Liu, D., Wang, X., Li, P., Liu, Y., Lin, K., Bucher, J., Zhang, N., Wang, Y., Wang, H., Deng, J., Liao, Y., Wei, K., Zhang, X., Fu, L., Hu, Y., Liu, J., Cai, C., Zhang, S., Zhang, S., Li, F., Zhang, H., Zhang, J., Guo, N., Liu, Z., Liu, J., Sun, C., Ma, Y., Zhang, H., Cui, Y., Freeling, M.R., Borm, T., Bonnema, G., Wu, J., and Wang, X. 2016. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nature genetics* **48**(10): 1218-1224. doi: 10.1038/ng.3634.
- Chiu, C.C., Chen, L.J., Su, P.H., and Li, H.M. 2013. Evolution of chloroplast J proteins. *PloS one* **8**(7): e70384. doi: 10.1371/journal.pone.0070384.
- Christensen, C.A., Gorsich, S.W., Brown, R.H., Jones, L.G., Brown, J., Shaw, J.M., and Drews, G.N. 2002. Mitochondrial GFA2 is required for synergid cell death in *Arabidopsis*. *The Plant cell* **14**(9): 2215-2232.
- Cupp-Vickery, J.R., and Vickery, L.E. 2000. Crystal structure of Hsc20, a J-type Co-chaperone from *Escherichia coli*. *Journal of molecular biology* **304**(5): 835-845. doi: 10.1006/jmbi.2000.4252.
- D'Andrea, L.D., and Regan, L. 2003. TPR proteins: the versatile helix. *Trends in biochemical sciences* **28**(12): 655-662. doi: 10.1016/j.tibs.2003.10.007.
- de Crouy-Chanel, A., Kohiyama, M., and Richarme, G. 1995. A novel function of *Escherichia coli* chaperone DnaJ. Protein-disulfide isomerase. *The Journal of biological chemistry* **270**(39): 22669-22672.
- Dorn, K.V., Willmund, F., Schwarz, C., Henselmann, C., Pohl, T., Hess, B., Veyel, D., Usadel, B., Friedrich, T., Nickelsen, J., and Schroda, M. 2010. Chloroplast DnaJ-like proteins 3 and 4 (CDJ3/4) from *Chlamydomonas reinhardtii* contain redox-active Fe-S clusters and interact with stromal HSP70B. *The Biochemical journal* **427**(2): 205-215. doi: 10.1042/BJ20091412.
- Du, Y., Zhao, J., Chen, T., Liu, Q., Zhang, H., Wang, Y., Hong, Y., Xiao, F., Zhang, L., Shen, Q., and Liu, Y. 2013. Type I J-domain NbMIP1 proteins are required for both Tobacco mosaic virus infection and plant innate immunity. *PLoS pathogens* **9**(10): e1003659. doi: 10.1371/journal.ppat.1003659.
- Feng, J., Chen, D., Berr, A., and Shen, W.H. 2016. ZRF1 Chromatin Regulators Have Polycomb Silencing and Independent Roles in Development. *Plant physiology* **172**(3): 1746-1759. doi: 10.1104/pp.16.00193.

- Finka, A., Mattoo, R.U., and Goloubinoff, P. 2011. Meta-analysis of heat- and chemically upregulated chaperone genes in plant and human cells. *Cell stress & chaperones* **16**(1): 15-31. doi: 10.1007/s12192-010-0216-8.
- Futamura, N., Ishiiminami, N., Hayashida, N., and Shinohara, K. 1999. Expression of DnaJ homologs and Hsp70 in the Japanese willow (*Salix gilgiana* Seemen). *Plant & cell physiology* **40**(5): 524-531.
- Georgopoulos, C., and Welch, W.J. 1993. Role of the major heat shock proteins as molecular chaperones. *Annual review of cell biology* **9**: 601-634. doi: 10.1146/annurev.cb.09.110193.003125.
- Guan, C., Rosen, E.S., Boonsirichai, K., Poff, K.L., and Masson, P.H. 2003. The ARG1-LIKE2 gene of *Arabidopsis* functions in a gravity signal transduction pathway that is genetically distinct from the PGM pathway. *Plant physiology* **133**(1): 100-112.
- Guzman-Lopez, J.A., Abraham-Juarez, M.J., Lozano-Sotomayor, P., de Folter, S., and Simpson, J. 2016. *Arabidopsis thaliana* gonidialess A/Zuotin related factors (GlsA/ZRF) are essential for maintenance of meristem integrity. *Plant molecular biology* **91**(1-2): 37-51. doi: 10.1007/s11103-016-0439-x.
- Harrison, B.R., and Masson, P.H. 2008. ARL2, ARG1 and PIN3 define a gravity signal transduction pathway in root statocytes. *The Plant journal : for cell and molecular biology* **53**(2): 380-392. doi: 10.1111/j.1365-313X.2007.03351.x.
- Hofius, D., Maier, A.T., Dietrich, C., Jungkunz, I., Bornke, F., Maiss, E., and Sonnewald, U. 2007. Capsid protein-mediated recruitment of host DnaJ-like proteins is required for Potato virus Y infection in tobacco plants. *Journal of virology* **81**(21): 11870-11880. doi: 10.1128/JVI.01525-07.
- Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J., and Gao, G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**(8): 1296-1297. doi: 10.1093/bioinformatics/btu817.
- Ichikawa, S., Yamada, N., Suetsugu, N., Wada, M., and Kadota, A. 2011. Red light, Phot1 and JAC1 modulate Phot2-dependent reorganization of chloroplast actin filaments and chloroplast avoidance movement. *Plant & cell physiology* **52**(8): 1422-1432. doi: 10.1093/pcp/pcr087.
- Ifuku, K., Endo, T., Shikanai, T., and Aro, E.M. 2011. Structure of the chloroplast NADH dehydrogenase-like complex: nomenclature for nuclear-encoded subunits. *Plant & cell physiology* **52**(9): 1560-1568. doi: 10.1093/pcp/pcr098.
- Jelenska, J., van Hal, J.A., and Greenberg, J.T. 2010. *Pseudomonas syringae* hijacks plant stress chaperone machinery for virulence. *Proceedings of the National Academy of Sciences of the United States of America* **107**(29): 13177-13182. doi: 10.1073/pnas.0910943107.
- Kneissl, J., Wachtler, V., Chua, N.H., and Bolle, C. 2009. OWL1: an *Arabidopsis* J-domain protein involved in perception of very low light fluences. *The Plant cell* **21**(10): 3212-3225. doi: 10.1105/tpc.109.066472.
- Kong, F., Deng, Y., Wang, G., Wang, J., Liang, X., and Meng, Q. 2014a. LeCDJ1, a chloroplast DnaJ protein, facilitates heat tolerance in transgenic tomatoes. *Journal of integrative plant biology* **56**(1): 63-74. doi: 10.1111/jipb.12119.
- Kong, F., Deng, Y., Zhou, B., Wang, G., Wang, Y., and Meng, Q. 2014b. A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. *Journal of experimental botany* **65**(1): 143-158. doi: 10.1093/jxb/ert357.
- Kong, H., Landherr, L.L., Frohlich, M.W., Leebens-Mack, J., Ma, H., and dePamphilis, C.W. 2007. Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *The Plant journal : for cell and molecular biology* **50**(5): 873-885. doi: 10.1111/j.1365-313X.2007.03097.x.
- Kroczyńska, B., Zhou, R., Wood, C., and Miernyk, J.A. 1996. AtJ1, a mitochondrial homologue of the

- Escherichia coli* DnaJ protein. *Plant molecular biology* **31**(3): 619-629.
- Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Ostergaard, L., Patron, N., Uauy, C., and Harwood, W. 2015. Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome biology* **16**: 258. doi: 10.1186/s13059-015-0826-7.
- Lindquist, S., and Craig, E.A. 1988. The heat-shock proteins. *Annual review of genetics* **22**: 631-677. doi: 10.1146/annurev.ge.22.120188.003215.
- Liu, J.Z., and 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 : for cell and molecular biology* **74**(1): 110-121. doi: 10.1111/tpj.12108.
- Liu, M.S., Chen, L.F., Lin, C.H., Lai, Y.M., Huang, J.Y., and Sung, Z.R. 2012. Molecular and functional characterization of broccoli EMBRYONIC FLOWER 2 genes. *Plant & cell physiology* **53**(7): 1217-1231. doi: 10.1093/pcp/pcs063.
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X., Wang, J., Lu, K., Fang, Z., Bancroft, I., Yang, T.J., Hu, Q., Wang, X., Yue, Z., Li, H., Yang, L., Wu, J., Zhou, Q., Wang, W., King, G.J., Pires, J.C., Lu, C., Wu, Z., Sampath, P., Wang, Z., Guo, H., Pan, S., Yang, L., Min, J., Zhang, D., Jin, D., Li, W., Belcram, H., Tu, J., Guan, M., Qi, C., Du, D., Li, J., Jiang, L., Batley, J., Sharpe, A.G., Park, B.S., Ruperao, P., Cheng, F., Waminal, N.E., Huang, Y., Dong, C., Wang, L., Li, J., Hu, Z., Zhuang, M., Huang, Y., Huang, J., Shi, J., Mei, D., Liu, J., Lee, T.H., Wang, J., Jin, H., Li, Z., Li, X., Zhang, J., Xiao, L., Zhou, Y., Liu, Z., Liu, X., Qin, R., Tang, X., Liu, W., Wang, Y., Zhang, Y., Lee, J., Kim, H.H., Denoeud, F., Xu, X., Liang, X., Hua, W., Wang, X., Wang, J., Chalhoub, B., and Paterson, A.H. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature communications* **5**: 3930. doi: 10.1038/ncomms4930.
- Ma, Z.X., Leng, Y.J., Chen, G.X., Zhou, P.M., Ye, D., and Chen, L.Q. 2015. The THERMOSENSITIVE MALE STERILE 1 Interacts with the BiPs via DnaJ Domain and Stimulates Their ATPase Enzyme Activities in Arabidopsis. *PloS one* **10**(7): e0132500. doi: 10.1371/journal.pone.0132500.
- Maruyama, D., Yamamoto, M., Endo, T., and Nishikawa, S. 2014. Different sets of ER-resident J-proteins regulate distinct polar nuclear-membrane fusion events in *Arabidopsis thaliana*. *Plant & cell physiology* **55**(11): 1937-1944. doi: 10.1093/pcp/pcu120.
- Miernyk, J.A. 2001. The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. *Cell stress & chaperones* **6**(3): 209-218.
- Muller, A., Rinck, G., Thiel, H.J., and Tautz, N. 2003. Cell-derived sequences in the N-terminal region of the polyprotein of a cytopathogenic pestivirus. *Journal of virology* **77**(19): 10663-10669.
- Nakayashiki, H., Matsuo, H., Chuma, I., Ikeda, K., Betsuyaku, S., Kusaba, M., Tosa, Y., and Mayama, S. 2001. Pyret, a Ty3/Gypsy retrotransposon in *Magnaporthe grisea* contains an extra domain between the nucleocapsid and protease domains. *Nucleic acids research* **29**(20): 4106-4113.
- Ohta, M., Wakasa, Y., Takahashi, H., Hayashi, S., Kudo, K., and Takaiwa, F. 2013. Analysis of rice ER-resident J-proteins reveals diversity and functional differentiation of the ER-resident Hsp70 system in plants. *Journal of experimental botany* **64**(18): 5429-5441. doi: 10.1093/jxb/ert312.
- Orme, W., Walker, A.R., Gupta, R., and Gray, J.C. 2001. A novel plastid-targeted J-domain protein in *Arabidopsis thaliana*. *Plant molecular biology* **46**(5): 615-626.
- Park, H.Y., Lee, S.Y., Seok, H.Y., Kim, S.H., Sung, Z.R., and Moon, Y.H. 2011. EMF1 interacts with EIP1, EIP6 or EIP9 involved in the regulation of flowering time in *Arabidopsis*. *Plant & cell physiology* **52**(8): 1376-1388. doi: 10.1093/pcp/pcr084.

- Park, M.Y., and Kim, S.Y. 2014. The Arabidopsis J protein AtJ1 is essential for seedling growth, flowering time control and ABA response. *Plant & cell physiology* **55**(12): 2152-2163. doi: 10.1093/pcp/pcu145.
- Petitjean, C., Moreira, D., Lopez-Garcia, P., and Brochier-Armanet, C. 2012. Horizontal gene transfer of a chloroplast DnaJ-Fer protein to Thaumarchaeota and the evolutionary history of the DnaK chaperone system in Archaea. *BMC evolutionary biology* **12**: 226. doi: 10.1186/1471-2148-12-226.
- Petti, C., Nair, M., and DeBolt, S. 2014. The involvement of J-protein AtDjC17 in root development in Arabidopsis. *Frontiers in plant science* **5**: 532. doi: 10.3389/fpls.2014.00532.
- Prasad, B.D., Goel, S., and Krishna, P. 2010. In silico identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. *PloS one* **5**(9): e12761. doi: 10.1371/journal.pone.0012761.
- Pulido, P., Llamas, E., Llorente, B., Ventura, S., Wright, L.P., and Rodriguez-Concepcion, M. 2016. Specific Hsp100 Chaperones Determine the Fate of the First Enzyme of the Plastidial Isoprenoid Pathway for Either Refolding or Degradation by the Stromal Clp Protease in Arabidopsis. *PLoS genetics* **12**(1): e1005824. doi: 10.1371/journal.pgen.1005824.
- Pulido, P., Toledo-Ortiz, G., Phillips, M.A., Wright, L.P., and Rodriguez-Concepcion, M. 2013. Arabidopsis J-protein J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. *The Plant cell* **25**(10): 4183-4194. doi: 10.1105/tpc.113.113001.
- Qiu, X.B., Shao, Y.M., Miao, S., and Wang, L. 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and molecular life sciences : CMLS* **63**(22): 2560-2570. doi: 10.1007/s00018-006-6192-6.
- Quesada, V., Ponce, M.R., and Micol, J.L. 1999. OTC and AUL1, two convergent and overlapping genes in the nuclear genome of Arabidopsis thaliana. *FEBS letters* **461**(1-2): 101-106.
- Rajan, V.B., and D'Silva, P. 2009. Arabidopsis thaliana J-class heat shock proteins: cellular stress sensors. *Functional & integrative genomics* **9**(4): 433-446. doi: 10.1007/s10142-009-0132-0.
- Rhoads, A.R., and Friedberg, F. 1997. Sequence motifs for calmodulin recognition. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **11**(5): 331-340.
- Richly, H., Rocha-Viegas, L., Ribeiro, J.D., Demajo, S., Gundem, G., Lopez-Bigas, N., Nakagawa, T., Rospert, S., Ito, T., and Di Croce, L. 2010. Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**(7327): 1124-1128. doi: 10.1038/nature09574.
- Rinck, G., Birghan, C., Harada, T., Meyers, G., Thiel, H.J., and Tautz, N. 2001. A cellular J-domain protein modulates polyprotein processing and cytopathogenicity of a pestivirus. *Journal of virology* **75**(19): 9470-9482. doi: 10.1128/JVI.75.19.9470-9482.2001.
- Romisch, K. 2005. Endoplasmic reticulum-associated degradation. *Annual review of cell and developmental biology* **21**: 435-456. doi: 10.1146/annurev.cellbio.21.012704.133250.
- Salas-Munoz, S., Rodriguez-Hernandez, A.A., Ortega-Amaro, M.A., Salazar-Badillo, F.B., and Jimenez-Bremont, J.F. 2016. Arabidopsis AtDjA3 Null Mutant Shows Increased Sensitivity to Absciscic Acid, Salt, and Osmotic Stress in Germination and Post-germination Stages. *Frontiers in plant science* **7**: 220. doi: 10.3389/fpls.2016.00220.
- Sedbrook, J.C., Chen, R., and Masson, P.H. 1999. ARG1 (altered response to gravity) encodes a DnaJ-like protein that potentially interacts with the cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America* **96**(3): 1140-1145.
- Servas, C., and Romisch, K. 2013. The Sec63p J-domain is required for ERAD of soluble proteins in yeast. *PloS one* **8**(12): e82058. doi: 10.1371/journal.pone.0082058.

- Sery, A., Housset, D., Serre, L., Bonicel, J., Hatchikian, C., Frey, M., and Roth, M. 1994. Crystal structure of the ferredoxin I from *Desulfovibrio africanus* at 2.3 Å resolution. *Biochemistry* **33**(51): 15408-15417.
- Shen, L., Kang, Y.G., Liu, L., and Yu, H. 2011. The J-domain protein J3 mediates the integration of flowering signals in *Arabidopsis*. *The Plant cell* **23**(2): 499-514. doi: 10.1105/tpc.111.083048.
- Shimizu, T., Yoshii, A., Sakurai, K., Hamada, K., Yamaji, Y., Suzuki, M., Namba, S., and Hibi, T. 2009. Identification of a novel tobacco DnaJ-like protein that interacts with the movement protein of tobacco mosaic virus. *Archives of virology* **154**(6): 959-967. doi: 10.1007/s00705-009-0397-6.
- Silady, R.A., Ehrhardt, D.W., Jackson, K., Faulkner, C., Oparka, K., and Somerville, C.R. 2008. The GRV2/RME-8 protein of *Arabidopsis* functions in the late endocytic pathway and is required for vacuolar membrane flow. *The Plant journal : for cell and molecular biology* **53**(1): 29-41. doi: 10.1111/j.1365-3113X.2007.03314.x.
- Silady, R.A., Kato, T., Lukowitz, W., Sieber, P., Tasaka, M., and Somerville, C.R. 2004. The gravitropism defective 2 mutants of *Arabidopsis* are deficient in a protein implicated in endocytosis in *Caenorhabditis elegans*. *Plant physiology* **136**(2): 3095-3103; discussion 3002. doi: 10.1104/pp.104.050583.
- So, H.A., Chung, E., and Lee, J.H. 2014. *Arabidopsis* atDjC53 encoding a type III J-protein plays a negative role in heat shock tolerance. *Genes & Genomics* **36**(6): 733-744.
- Suetsugu, N., Kagawa, T., and Wada, M. 2005. An auxilin-like J-domain protein, JAC1, regulates phototropin-mediated chloroplast movement in *Arabidopsis*. *Plant physiology* **139**(1): 151-162. doi: 10.1104/pp.105.067371.
- Sun, J., Zhang, J., Wu, F., Xu, C., Li, S., Zhao, W., Wu, Z., Wu, J., Zhou, C.Z., and Shi, Y. 2005. Solution structure of Kti11p from *Saccharomyces cerevisiae* reveals a novel zinc-binding module. *Biochemistry* **44**(24): 8801-8809. doi: 10.1021/bi0504714.
- Sun, T., Zhang, Q., Gao, M., and Zhang, Y. 2014. Regulation of SOBIR1 accumulation and activation of defense responses in *bir1-1* by specific components of ER quality control. *The Plant journal : for cell and molecular biology* **77**(5): 748-756. doi: 10.1111/tpj.12425.
- Takano, A., Suetsugu, N., Wada, M., and Kohda, D. 2010. Crystallographic and functional analyses of J-domain of JAC1 essential for chloroplast photorelocation movement in *Arabidopsis thaliana*. *Plant & cell physiology* **51**(8): 1372-1376. doi: 10.1093/pcp/pcq089.
- Tamura, K., Stecher, G., Peterson, D., Filipinski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**(12): 2725-2729. doi: 10.1093/molbev/mst197.
- Tamura, K., Takahashi, H., Kunieda, T., Fuji, K., Shimada, T., and Hara-Nishimura, I. 2007. *Arabidopsis* KAM2/GRV2 is required for proper endosome formation and functions in vacuolar sorting and determination of the embryo growth axis. *The Plant cell* **19**(1): 320-332. doi: 10.1105/tpc.106.046631.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**(22): 4673-4680.
- Tseng, M.J., Yang, M.T., Chu, W.R., and Liu, C.W. 2014. Plastid transformation in cabbage (*Brassica oleracea* L. var. *capitata* L.) by the biolistic process. *Methods in molecular biology* **1132**: 355-366. doi: 10.1007/978-1-62703-995-6_23.
- Ungewickell, E., Ungewickell, H., Holstein, S.E., Lindner, R., Prasad, K., Barouch, W., Martin, B.,

- Greene, L.E., and Eisenberg, E. 1995. Role of auxilin in uncoating clathrin-coated vesicles. *Nature* **378**(6557): 632-635. doi: 10.1038/378632a0.
- Vitha, S., Froehlich, J.E., Koksharova, O., Pyke, K.A., van Erp, H., and Osteryoung, K.W. 2003. ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *The Plant cell* **15**(8): 1918-1933.
- Voorrips, R.E. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of heredity* **93**(1): 77-78.
- Walsh, P., Bursac, D., Law, Y.C., Cyr, D., and Lithgow, T. 2004. The J-protein family: modulating protein assembly, disassembly and translocation. *EMBO reports* **5**(6): 567-571. doi: 10.1038/sj.embor.7400172.
- Wang, G., Cai, G., Kong, F., Deng, Y., Ma, N., and Meng, Q. 2014a. Overexpression of tomato chloroplast-targeted DnaJ protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. *Plant physiology and biochemistry : PPB* **82**: 95-104. doi: 10.1016/j.plaphy.2014.05.011.
- Wang, G., Kong, F., Zhang, S., Meng, X., Wang, Y., and Meng, Q. 2015. A tomato chloroplast-targeted DnaJ protein protects Rubisco activity under heat stress. *Journal of experimental botany* **66**(11): 3027-3040. doi: 10.1093/jxb/erv102.
- Wang, X., Jia, N., Zhao, C., Fang, Y., Lv, T., Zhou, W., Sun, Y., and Li, B. 2014b. Knockout of AtDjB1, a J-domain protein from *Arabidopsis thaliana*, alters plant responses to osmotic stress and abscisic acid. *Physiologia plantarum* **152**(2): 286-300. doi: 10.1111/ppl.12169.
- Watkins, K.P., Rojas, M., Friso, G., van Wijk, K.J., Meurer, J., and Barkan, A. 2011. APO1 promotes the splicing of chloroplast group II introns and harbors a plant-specific zinc-dependent RNA binding domain. *The Plant cell* **23**(3): 1082-1092. doi: 10.1105/tpc.111.084335.
- Xia, Z., Zhang, X., Li, J., Su, X., and Liu, J. 2014. Overexpression of a tobacco J-domain protein enhances drought tolerance in transgenic *Arabidopsis*. *Plant physiology and biochemistry : PPB* **83**: 100-106. doi: 10.1016/j.plaphy.2014.07.023.
- Yamamoto, H., Peng, L., Fukao, Y., and Shikanai, T. 2011. An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *The Plant cell* **23**(4): 1480-1493. doi: 10.1105/tpc.110.080291.
- Yamamoto, M., Maruyama, D., Endo, T., and Nishikawa, S. 2008. *Arabidopsis thaliana* has a set of J proteins in the endoplasmic reticulum that are conserved from yeast to animals and plants. *Plant & cell physiology* **49**(10): 1547-1562. doi: 10.1093/pcp/pcn119.
- Yang, K.Z., Xia, C., Liu, X.L., Dou, X.Y., Wang, W., Chen, L.Q., Zhang, X.Q., Xie, L.F., He, L., Ma, X., and Ye, D. 2009. A mutation in Thermosensitive Male Sterile 1, encoding a heat shock protein with DnaJ and PDI domains, leads to thermosensitive gametophytic male sterility in *Arabidopsis*. *The Plant journal : for cell and molecular biology* **57**(5): 870-882. doi: 10.1111/j.1365-313X.2008.03732.x.
- Yang, Y., Qin, Y., Xie, C., Zhao, F., Zhao, J., Liu, D., Chen, S., Fuglsang, A.T., Palmgren, M.G., Schumaker, K.S., Deng, X.W., and Guo, Y. 2010. The *Arabidopsis* chaperone J3 regulates the plasma membrane H⁺-ATPase through interaction with the PKS5 kinase. *The Plant cell* **22**(4): 1313-1332. doi: 10.1105/tpc.109.069609.
- Zhou, W., Zhou, T., Li, M.X., Zhao, C.L., Jia, N., Wang, X.X., Sun, Y.Z., Li, G.L., Xu, M., Zhou, R.G., and Li, B. 2012. The *Arabidopsis* J-protein AtDjB1 facilitates thermotolerance by protecting cells against heat-induced oxidative damage. *The New phytologist* **194**(2): 364-378. doi: 10.1111/j.1469-8137.2012.04070.x.

Zhu, J.K., Shi, J., Bressan, R.A., and Hasegawa, P.M. 1993. Expression of an *Atriplex nummularia* gene encoding a protein homologous to the bacterial molecular chaperone DnaJ. *The Plant cell* **5**(3): 341-349. doi: 10.1105/tpc.5.3.341.

Draft

Figure captions

Fig. 1 Phylogenetic analysis of J-proteins in Arabidopsis and *B. oleracea*. Full-length amino acid sequences of 117 *A. thaliana* and 129 *B. oleracea* J-proteins were aligned via ClustalW, and the unrooted NJ trees were constructed using MEGA6 software with the following settings: pairwise deletion option for gaps/missing data treatment; *p*-distance method for Substitution model; and Bootstrap test of 1000 replicates for internal branch reliability. Colored Roman numerals I~XV represent fifteen major clades with high scores, separated by the low-score branches. The J-proteins from *A. thaliana* and *B. oleracea* are indicated by red and green dots, respectively.

Fig. 2 Chromosomal distribution of J-protein genes in Arabidopsis (At) and *B. oleracea* (Bol). Tandemly duplicated genes are highlighted by pink. The scale on the left of each chromosome represents megabases (Mb) and the chromosome number is indicated at the top of each chromosome. Roman numerals or “others” in brackets indicate the corresponding gene clades based on this study.

Fig. 3 Phylogenetic analysis, domain architecture and gene structure in multigene Clades. The unrooted NJ tree was generated using MEGA6 with parameter settings as Fig. 1 based on full-length amino acid sequences of J-proteins of Clade I and Clade II in *A. thaliana* and *B. oleracea*. Brackets indicate the other typical domains in addition to J-domain. Gene structures of J-proteins in *A. thaliana* and *B. oleracea* were indicated in blue and red, respectively. The solid boxes represent exons and black lines represent introns. Asterisk indicated Bol35736 that shares high sequence similarity with Bol35735 but absent of N-terminal J-domain.

Fig. 4 Phylogenetic analysis, domain architecture and gene structure in oligo-gene Clades. The unrooted NJ tree was generated using MEGA6 with parameter settings as Fig. 1 based on full-length amino acid sequences of J-proteins of Clade III to Clade XV in *A. thaliana* and *B. oleracea*. Brackets indicate the other typical domains in addition to J-domain. Gene structures in *A. thaliana* and *B. oleracea* were indicated in blue and red, respectively. The solid boxes represent exons and black lines represent introns.

Fig. 5 Phylogenetic analysis, domain architecture and gene structure in mono-gene clades.

The unrooted NJ tree was generated using MEGA6 with parameter settings as Fig. 1 based on full-length amino acid sequences of J-proteins of the other undefined clades in *A. thaliana* and *B. oleracea*. Brackets indicate the other typical domains in addition to J-domain. The solid boxes represent exons and black lines represent introns.

Supplementary material

gen-2017-0206. R2Suppla: Table S1 Summary of J-proteins in model plant Arabidopsis.

gen-2017-0206. R2Supplb: Table S2 Summary of J-proteins in *B. oleracea*.

Draft

129x122mm (300 x 300 DPI)

366x175mm (300 x 300 DPI)

364x232mm (300 x 300 DPI)

318x325mm (300 x 300 DPI)