Article

Genome-Wide Identification and Transcriptional Analysis of Arabidopsis *DUF506* Gene Family

Sheng Ying †

Noble Research Institute LLC, Ardmore, OK 73401, USA; yingshen@msu.edu
† Current address: Department of Biochemistry and Molecular Biology, Michigan State University,
East Lansing, MI 48823, USA.

Abstract: The Domain of unknown function 506 (*DUF506*) family, which belongs to the PD-(D/E)XK nuclease superfamily, has not been functionally characterized. In this study, 266 *DUF506* domain-containing genes were identified from algae, mosses, and land plants showing their wide occurrence in photosynthetic organisms. Bioinformatics analysis identified 211 high-confidence *DUF506* genes across 17 representative land plant species. Phylogenetic modeling classified three groups of plant *DUF506* genes that suggested functional preservation among the groups based on conserved gene structure and motifs. Gene duplication and Ka/Ks evolutionary rates revealed that *DUF506* genes are under purifying positive selection pressure. Subcellular protein localization analysis revealed that *DUF506* proteins were present in different organelles. Transcript analyses showed that 13 of the Arabidopsis *DUF506* genes are ubiquitously expressed in various tissues and respond to different abiotic stresses and ABA treatment. Protein-protein interaction network analysis using the STRING-DB, AtPIN (*Arabidopsis thaliana* Protein Interaction Network), and AI-1 (Arabidopsis Interactome-1) tools indicated that AtDUF506s potentially interact with iron-deficiency response proteins, salt-inducible transcription factors, or calcium sensors (calmodulins), implying that *DUF506* genes have distinct biological functions including responses to environmental stimuli, nutrient-deficiencies, and participate in Ca(2+) signaling. Current results provide insightful information regarding the molecular features of the *DUF506* family in plants, to support further functional characterizations.

Keywords: abiotic stress; arabidopsis; domain of unknown function (DUF); domain-structure; evolutionary analysis; expression profiling

1. Introduction

Domain of unknown function (DUF) is a conserved protein domain that has not been extensively characterized [1]. Designation as a DUF domain was derived from two bacterial motifs (DUF1 and 2), which were, respectively, renamed as GGDEF and EAL domains, based on their featured peptides [1,2]. The availability of whole genome sequence information has enabled the generation of a list of DUF-classified proteins in a wide range of organisms with a considerable portion (31%, 5889 DUFs) found in the 19,179 entries in the Pfam database (34.0, https://pfam.xfam.org/, accessed on August 2021) [3]. Numerous *DUF* genes have been studied in different species of plants. In Arabidopsis, *DUF6* [4], *DUF26* [5], *DUF538* [6], *DUF579* [7], *DUF581* [8], *DUF617/MIZ* [9], *DUF642* [10], *DUF647* [11], *DUF724* [12], *DUF784* [13], *DUF827* [14], *DUF1117* [15] and *DUF4228* [16] have been functionally characterized, and these studies have revealed the involvement of DUF-domain containing proteins in diverse biological processes. For example, *At5g06700/TRICHOME BIREFRINGENCE/TBR* and its homolog *At5g01360/TRICHOME BIREFRINGENCE-LIKE 3/TBL3*, which belong to the plant-specific DUF231 family, function in maintaining homeostasis of secondary wall cellulose synthesis in Arabidopsis [17]. Furthermore, another DUF231 member, At3g55990/ESK1/TBL29, was shown to be a negative regulator of freeze tolerance [18]. A recent evolutionary study revealed that *DUF231* genes are found in all land plants and conduct conserved functions [19].

The *DUF506* family belongs to the PD-(D/E)XK nuclease superfamily [20], which exists in all organisms. Most of the studies of *DUF506* domain-containing proteins involve biochemical, mutagenesis, and structural biological approaches [21,22]. Currently, this superfamily includes approximately 22,900 proteins classified into 121 family branches [23]. However, due to low protein sequence similarity, *DUF506* genes display extremely distant homologs among families [20] making it difficult to elucidate the molecular and biological functions of individual *DUF506* genes. Recently, a phosphate-stress response *DUF506* gene (*At3g25240/REPRESSOR OF EXCESSIVE ROOT HAIR GROWTH 1/RXR1*) in Arabidopsis was experimentally demonstrated to repress root hair elongation via interaction with RAB GTPase regardless of P supplied status [24]. Furthermore, overexpression of the *RXR1* homolog (*Bradi2g58590*) in the monocot model plant, Brachypodium, exhibited the identical short root hair phenotype, suggesting the conservative function of *RXR1* in regulation of plant cell development. Thus, the aim of this work was to conduct an extensive bioinformatics, expression profiling and subcellular analyses of plant *DUF506* genes, to uncover their additional biological functions.

In this study, 211 *DUF506* genes from 17 land plant species were identified with high confidence. These *DUF506* genes were subjected to evolutionary analysis. In addition, subcellular localization, expression profiling analysis, and putative interactor(s) predictions of *DUF506* family members not only extend the current knowledge of the PD-(D/E)XK superfamily in plants, but also provide beneficial insights for further studies into the distinct functions, such as abiotic stress responses, or regulation of calcium signaling.

2. Results

2.1. Genome-Wide Survey and Phylogenetic Clustering of DUF506 Genes in Plants

To extend current understanding of the *DUF506* gene family, 323 putative *DUF506* candidate gene sequences were extracted from the NCBI and Phytozome databases. Gene sequences encoding *DUF506* domain-containing proteins were identified from 23 representative species, including Chlorophyta, Bryophyta, Lycophyta, Gymnospermae and Angiospermae (basal angiosperm, monocots and eudicots). Then, comprehensive protein sequences alignment [25] and domain alignment (HMM, Hidden Markov Model) [26] were respectively conducted to eliminate candidates with low integrity or incomplete domains. The remaining 266 *DUF506* genes were then subjected to evolutionary analysis (Figure 1). Notably, no *DUF506* genes were identified from *Volvox carteri*, and only one candidate was discovered from another chlorophyte (*Chlamydomona reinhardtii*). Generally, the dicot species have more *DUF506* genes compared to monocots. The number of *DUF506* gene sequences was highest in soybean (24), followed by two other legumes.

A phylogenetic tree was constructed to examine the evolutionary relationships among the 211 *DUF506* candidates derived from 17 species in Angiosperms, using full-length protein sequence alignments. These *DUF506* proteins were classified into three distinct groups (I,II, and III), consisting of 66, 51, and 94 members, respectively (Figure 2a, Table S1). Group III was further divided into two subgroups, named IIIa and IIIb, with 33 and 61 members, respectively. *DUF506* proteins were found to be evenly distributed in each group (Table S2). For example, in Arabidopsis, there were four, two and seven *DUF506* members in group I, group II, and group III, respectively. The lengths of Arabidopsis *DUF506* protein ranged from 261 to 371 amino acid residues, while their predicted theoretical pI (Isoelectric point, the pH value at which the total charge on the protein is zero) varied from 5.02 to 8.96 (Table 1). Importantly, monocot and eudicot *DUF506* proteins clustered into different branches (Figure 2a). These results indicate that classification of *DUF506* members is largely concordant with the species tree (Figure 1).

2.2. Motif Conservation within Diverse DUF506 Family Groups

Within the *DUF506* family, ten putative conserved motifs between five and one hundred amino acid residues were predicted using MEME [27]. The corresponding motif sequence logos and their distribution along the sequence are illustrated in Figure 2b,c. Gen-

erally, group IIIa *DUF506* members had the highest motif abundance. The most ubiquitous was motif 4 and was present in all (211/211) *DUF506* proteins, while motif 1 (204/211), motif 2 (184/211), and motif 5 (203/211) were present in most *DUF506* proteins. In contrast, motifs 7 and 9 were exclusively found in group I and IIIa, respectively. Motifs 3, 6, 8, and 10 were present only in group IIIb. Overall, *DUF506* members from the same groups generally share distinctive motif compositions, indicating that functional similarities within groups and the diversification among groups.

2.3. Characterization of Chromosomal Distribution, Gene Synteny, and Sequence Structure of Arabidopsis DUF506 Genes

The localization of *DUF506* genes on Arabidopsis chromosomes was visualized through ePlant website (https://bar.utoronto.ca/eplant/, accessed on May 2021). The thirteen Arabidopsis *DUF506* genes were found on chromosomes 1–4 (Figure S1).

Gene synteny analysis revealed three segmental duplication events involving six Arabidopsis *DUF506* genes (Figure S1, Table 2), whereas no tandem duplication was found. These results implied that the Arabidopsis *DUF506* gene family expanded during evolution through gene segmental duplication. The ratio of nonsynonymous nucleotide substitutions (Ka), which led to amino acid altering, versus synonymous nucleotide substitutions (Ks), which did not change the amino acid sequence, was used to determine the selective pressure during genome evolution after gene duplication events [28]. Ka/Ks ratios for all three duplicated gene pairs are below 1, indicating that gene functional conservation during evolution, and suggesting that the Arabidopsis *DUF506* gene family underwent purifying selection after the duplication events. Purifying selection of *DUF506* genes was also found in monocot species, such as Brachypodium, rice, maize and sorghum (Table S4).

In addition, exon-intron distributions were examined to determine structural diversity among *DUF506* genes by using GSDS 2.0. The results showed almost 58% (122/211) of *DUF506* genes contained more than two exons (Figure S2, and Table S3). For example, nine of the Arabidopsis *DUF506* genes included three exons (Figure 3), and three duplicated gene pairs kept the identical exon-intron structure.
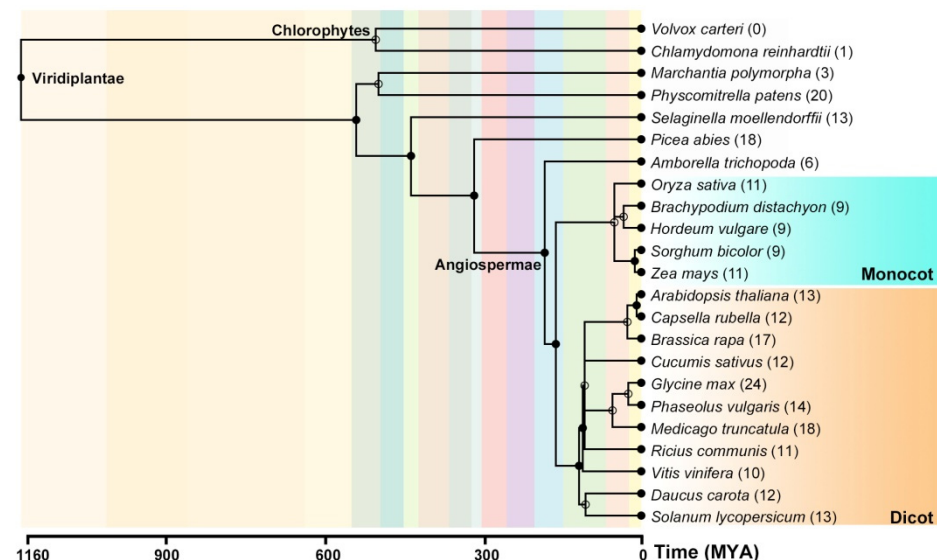

Figure 1. Phylogeny distribution of *DUF506* genes across 23 species used in this study. The number of high confidence *DUF506* genes is listed in parentheses. The order of tree branches and divergence time are derived from the TimeTree database (http://www.timetree.org/, accessed on August 2021). Monocot species are highlighted in the rectangle of teal color, whereas the dicot species are masked by rectangle in yellow color.
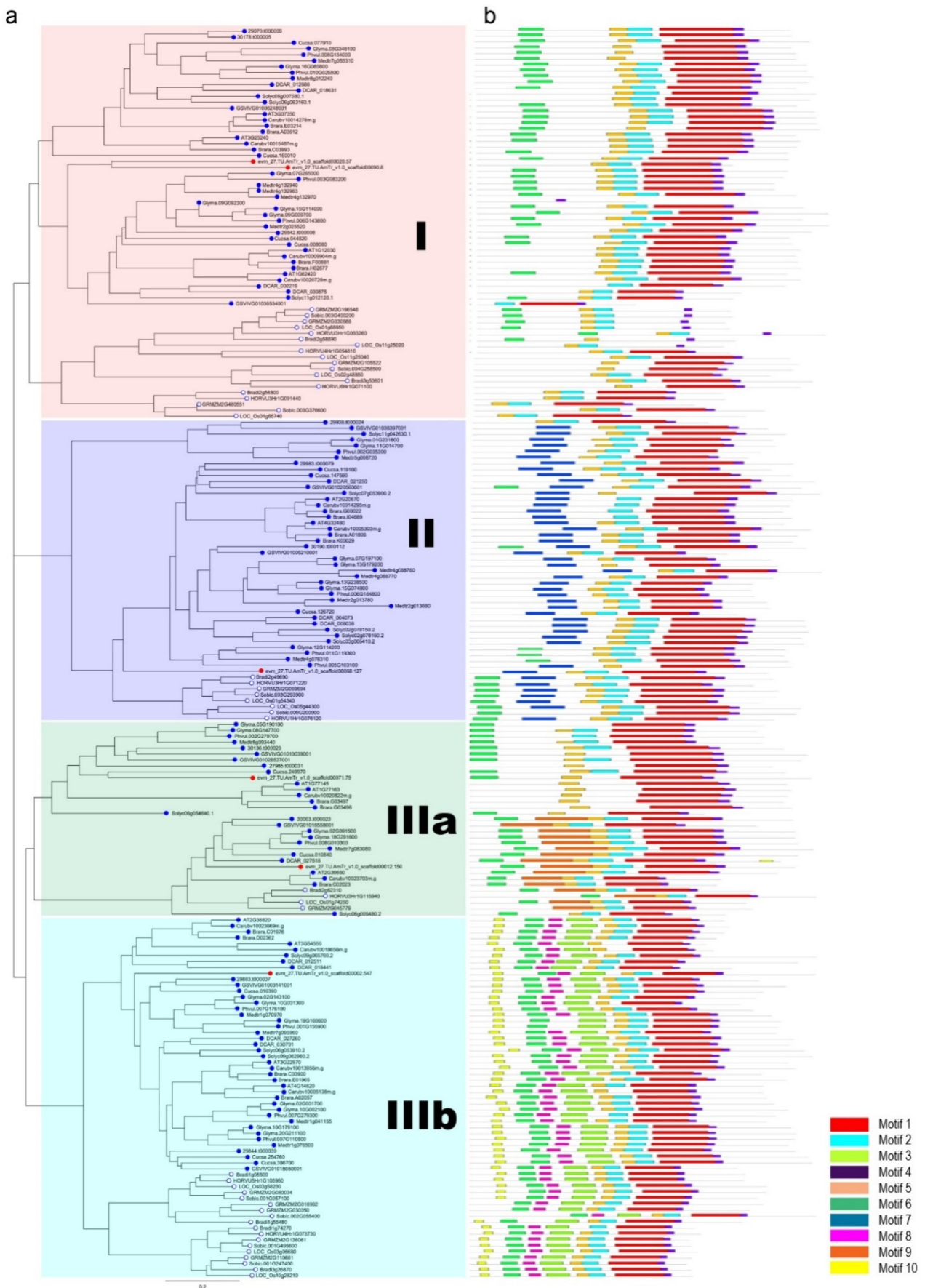
Figure 2. *Cont.*
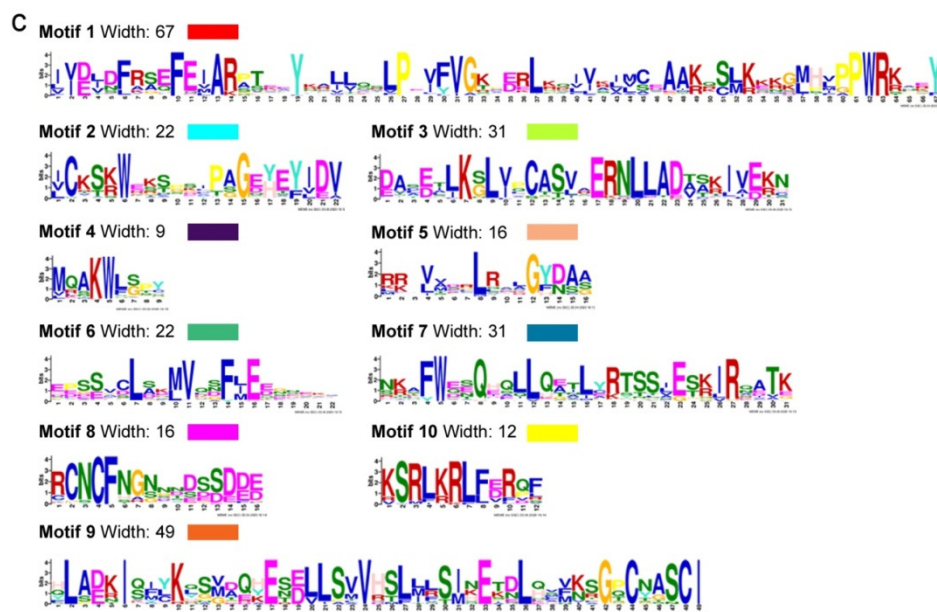
Figure 2. Phylogenetic representation and distribution of the conserved motifs in *DUF506* among angiosperms. (**a**) The phylogenetic tree was constructed based on full-length *DUF506* amino acid sequence alignments from 17 species. *DUF506* from monocots are represented by blue hollow dots, while those from dicots are represented by blue solid dots. *DUF506* from *Amborella trichopoda* (basal angiosperm) are represented by red solid dots. The distribution of *DUF506* conserved motifs (**b**) and the corresponding sequence logos (**c**) in plants are shown. Note that, the conserved β-sheet structures are mainly localized in Motif 2.

Table 1. Physicochemical properties of *DUF506* family members in Arabidopsis.

GeneID	# of A.A	p.I.	Subcellular Localization	Annotation (Phytozome v12.0)
At1g12030	296	5.85	cyto	phosphoenolpyruvate carboxylase
At1g62420	284	7.46	chlo; cyto	*DUF506* family protein
At1g77145	261	8.96	nucl	transmembrane protein
At1g77160	264	8.26	nucl	hypothetical protein
At2g20670	295	6.37	nucl	sugar phosphate exchanger
At2g38820	311	8.53	nucl	DNA-directed RNA polymerase subunit beta-beta protein
At2g39650	292	6.79	nucl	cruciferin (*DUF506*)
At3g07350	299	5.47	nucl	sulfate/thiosulfate import ATP-binding protein
At3g22970	371	5.02	extr	hypothetical protein
At3g25240	282	7.61	nucl	sulfate/thiosulfate import ATP-binding protein
At3g54550	289	8.63	chlo	DNA-directed RNA polymerase subunit beta-beta protein
At4g14620	342	8.28	cyto; cytoplas	hypothetical protein
At4g32480	288	7.66	nucl	sugar phosphate exchanger

Table 2. Synteny analysis of *DUF506* genes in Arabidopsis.

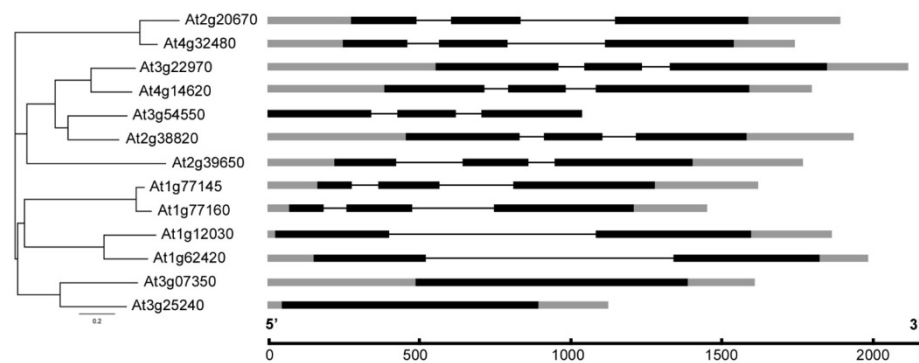| Duplicated Gene Pairs	Ka	Ks	Ka/Ks	Duplication Type	Type of Selection
At1g12030/At1g62420	0.28	1.03	0.2718	Segmental	Purify selection
At3g07350/At3g25240	0.46	1.76	0.2614	Segmental	Purify selection
At3g22970/At4g14620	0.21	1.05	0.2000	Segmental	Purify selection

Figure 3. Phylogenetic analysis and exon-intron distribution of Arabidopsis *DUF506* genes. Exon, intron and UTR are represented by black rectangle boxes, black lines and grey rectangle boxes, respectively.

2.4. Wide Spread Localization of Arabidopsis DUF506 Proteins

To study the subcellular localization of *DUF506* proteins in Arabidopsis, six representative Arabidopsis *DUF506* genes were cloned and transferred into the pEarleyGate 103 plasmid [29] to generate AtDUF506-GFP fused proteins driven by the CaMV35S promoter. Constructs were individually introduced into Arabidopsis mesophyll protoplasts using polyethylene glycol (PEG)-mediated transfection [30]. The resulting transient expression patterns are shown in Figure 4. The duplicated gene pairs were consistently expressed in identical organelles. For instance, At1g12030 and At1g62420 were localized in both the chloroplast and cytoplasm, while At3g22970 and At4g14620 were mainly in the plasma membrane, and slightly in the cytoplasm. Conversely, At3g25240 was concentrated in the nucleus and At3g07350 was present in the nucleus and cytoplasm. These results suggest that *DUF506* proteins play diverse physiological roles based on their distinct patterns of organelle localization.

2.5. Tissue Expression Profiles of Arabidopsis DUF506 Genes

To characterize the expression patterns of the Arabidopsis *DUF506* gene family, their transcript levels were examined in root, rosette leaves, stems, flowers, and siliques, using quantitative, real-time PCR (qRT-PCR). Arabidopsis *DUF506* gene transcripts were detected in all tested tissues and exhibited different expression patterns (Figure 5). The *At3g22970* transcript was abundant in all examined tissues, whereas the *At3g25240* transcript had a comparably low abundance. A few *DUF506* transcripts had much higher abundance in specific tissues or organs. For example, the *At3g54550* transcript was at least 50 times more abundant in siliques than in flowers or roots. Interestingly, in adult plants the *At1g62420* transcript was much more abundant in roots than in leaves, while in young seedlings, abundance was much higher in shoots than roots (Figure 6a). Such results suggest that *DUF506* genes might perform different functions during development of Arabidopsis plants.
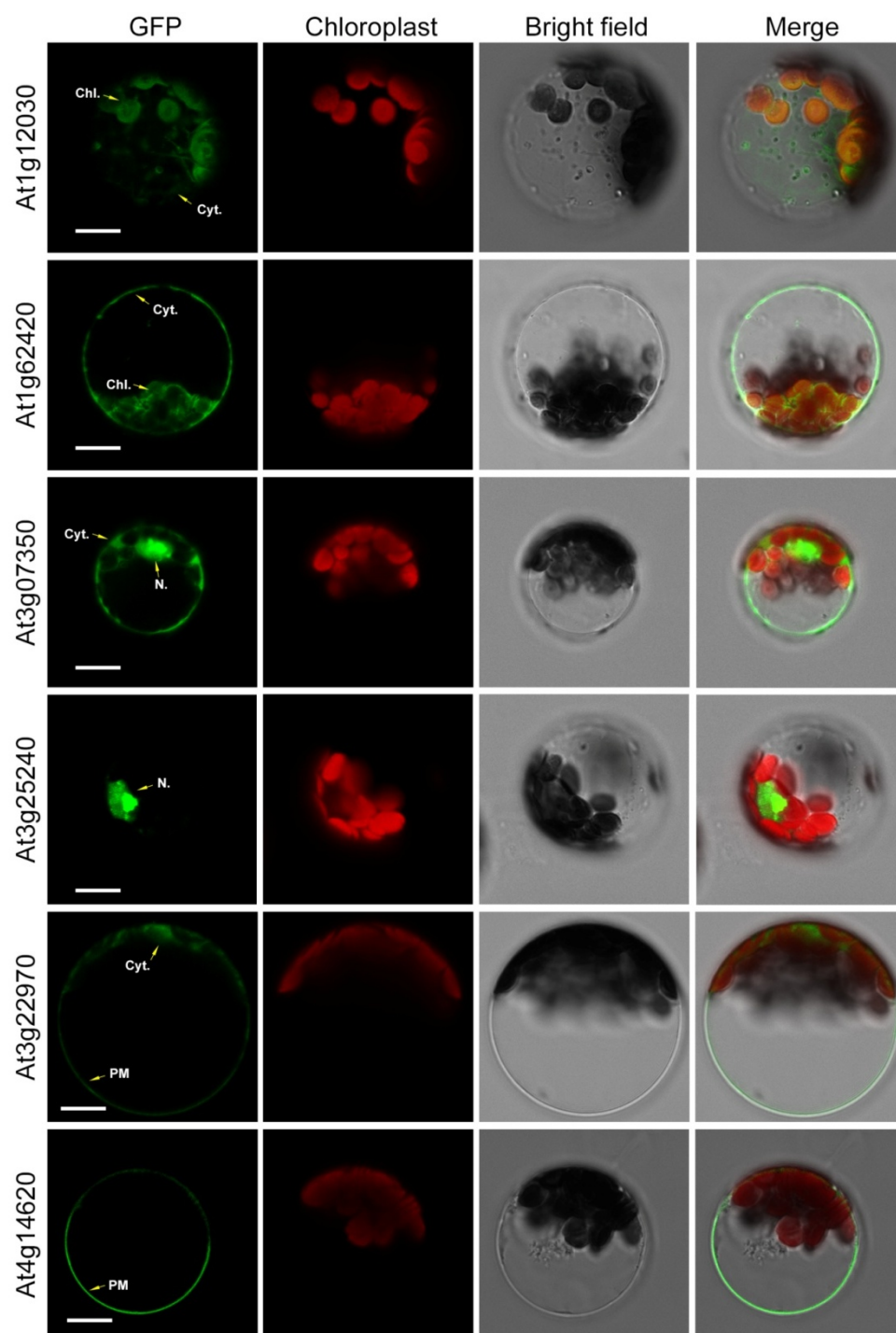
Figure 4. Subcellular localization of Arabidopsis *DUF506* proteins visualized by transient expression of GFP fusion constructs in Arabidopsis mesophyll protoplasts. Chl., chloroplast; Cyt., cytoplasm; N., nucleus; PM, plasma membrane. Scale bar, 10 μm.
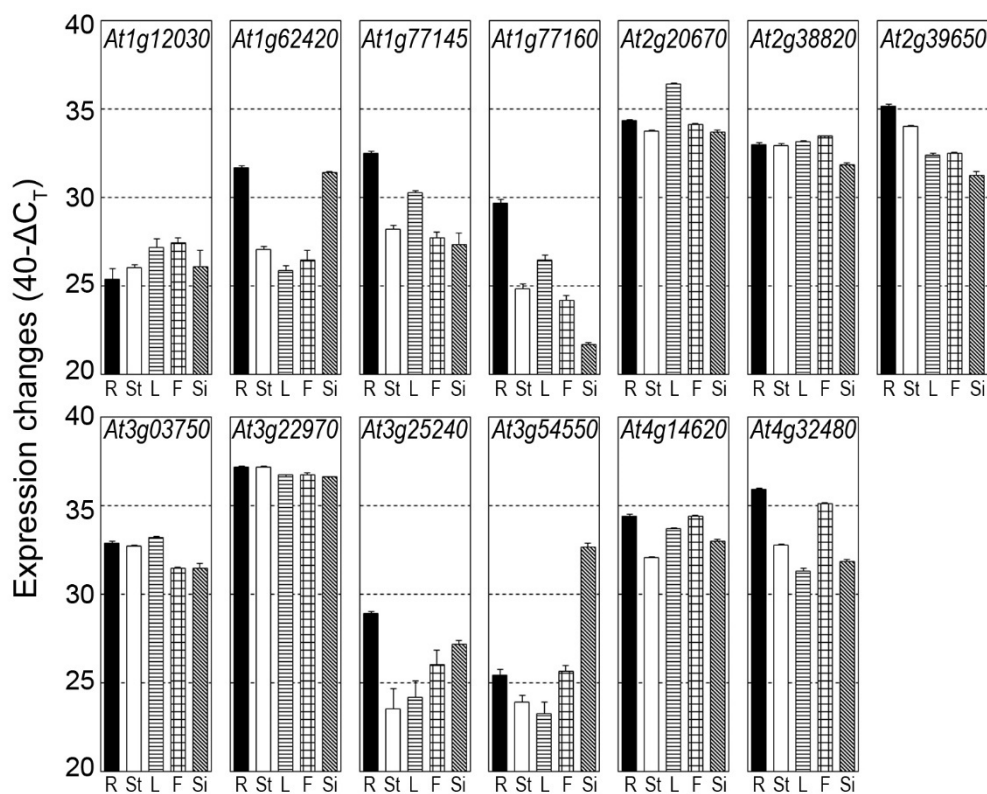
Figure 5. Expression profile of Arabidopsis *DUF506* genes in different tissues. R, root; St, stem; L, leave; F, flower; Si, silique. Changes of gene expression level between tissues are interpreted as described in [31]. Expression levels are given on a log scale expressed as $40^{-\Delta CT}$, where ΔCT is the difference in qRT-PCR threshold cycle number between the respective gene and the reference gene (*GAPDH*); 40 therefore equals the expression level of *GAPDH*; the number 40 was chosen because the PCR run stops after 40 cycles. The fold difference in expression is $2^{\Delta\Delta CT}$ when PCR efficiency is 2 (e.g., an ordinate value of 34 represents 16-fold lower expression than a value of 38).

2.6. Response of Arabidopsis DUF506 Gene Transcripts to Abiotic Stress and Abscisic Acid Treatments

Expression changes of Arabidopsis *DUF506* genes in young seedlings, under salt, osmotic, and Abscisic acid (ABA) treatment were investigated (Figure 6a). Due to the low expression level in seedlings, the *At3g54550* transcript was undetectable under all conditions. Under salt stress, *At1g12030*, *At1g62420*, *At1g77160*, *At2g38820*, *At2g39650*, and *At3g25240* were highly induced in shoot and root tissues (fold change [FC] ≥ 2), whereas *At1g77145* and *At3g07350* were slightly up-regulated (1 ≤ FC ≤ 2). By contrast, *At2g20670*, *At3g22970* and *At4g14620* were significantly down-regulated in shoots (FC ≥ 2). Except for *At1g62420* and *At3g25240*, no transcript changes were observed during osmotic stress.

Abscisic acid (ABA) has been shown to play an important role in stress responses and tolerance to abiotic stresses [32]. Most of the salt stress responsive Arabidopsis *DUF506* genes, including *At1g77145*, *At1g77160*, *At2g20670*, *At2g38820*, *At3g07350*, and *At4g14620*, positively responded to ABA treatment, while *At1g62420*, *At2g39650*, and *At3g25240* did not (Figure 6b). Meanwhile, expression of Arabidopsis *DUF506* genes were explored under heat and cold stress available through online public transcriptome databases (GENEVESTIGATOR® and EMBL-EBI/ArrayExpress). *At3g25240* and *At1g62420* were highly induced, but *At2g20670* and *At3g07350* were significantly repressed by both temperature stress conditions (FC ≥ 2).

Stress-related *cis*-elements in the promoter regions (2 kb upstream of start codon) of the Arabidopsis *DUF506* genes were also examined and identified (Figure S3). At least two copies of ABRE (ABA-responsive element) were found upstream of Arabidopsis *DUF506* genes with the exception of *At2g20670*. The transcripts changes of the majority

of Arabidopsis *DUF506* genes (8/13, Figure 6b) under ABA treatment suggest that some of them might be involved in integrating ABA-related signaling pathways with abiotic stress responses.
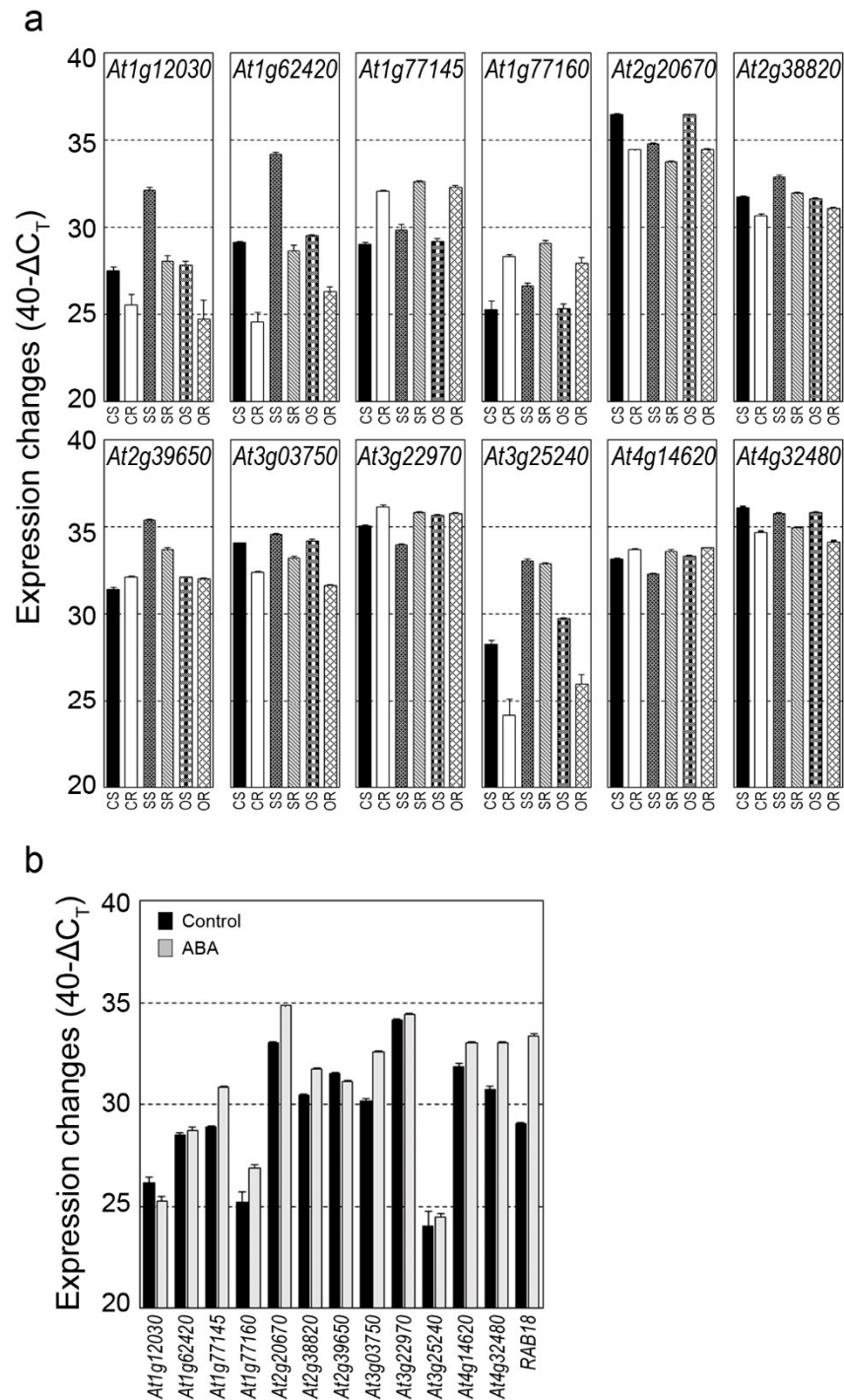

Figure 6. Expression profile of Arabidopsis *DUF506* genes in response to salt or osmotic stress (**a**) and Abscisic acid (ABA) treatment (**b**). CS, control shoot tissue; CR, control root tissue; SS, salt stress shoot tissue; SR, salt stress root tissue; OS, osmotic stress (mannitol) shoot tissue; OR, osmotic stress root tissue. *At5g66400/RAB18*, which is strongly induced by ABA, was used as positive external control (**b**).

2.7. Putative Interactors of Arabidopsis DUF506 Proteins

To further uncover potential biological functions of *DUF506* in Arabidopsis, protein-protein interaction analysis was conducted using three different online search tools (Table S6). STRING-DB [33] (v11.0, https://string-db.org/, accessed on May 2021) identified seventy-five potential interactors for nine Arabidopsis *DUF506* proteins, mainly based on coexpression profiles and text mining. Notably, several iron-deficiency related proteins (e.g., At1g47400/IRONMAM1, At5g67370/CGLD27, and At5g03570/IREG2) interacted with At1g12030, suggested its regulatory role in Fe signaling. Moreover, At2g39650 exhibited tight relationships with several salt stress regulators, such as salt-inducible ZINC Finger transcription factors (e.g., At3g55980/SZF1, At2g40140/SZF2, and At1g27730/ZAT10). On the other hand, the AtPIN and AI-1 programs only predicted two interactors of Arabidopsis *DUF506* proteins. Remarkably, At1g77145 was predicted and has been experimentally proven to extensively interact with different calmodulins (CaMs, Figure 7, [34]), which are small calcium-binding proteins that function as second messengers in plant Ca^{2+}-signaling [35]. The CaMs were identified from both databases, suggesting that At1g77145 might participate in Ca^{2+}/CaM mediating signal transduction.
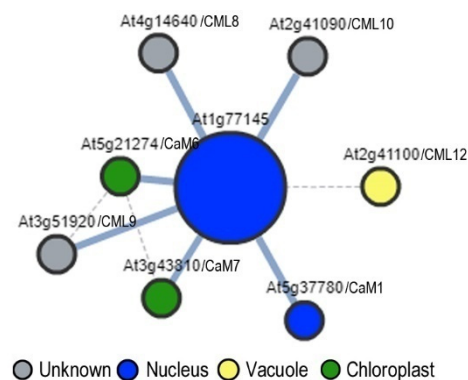

Figure 7. Predicted protein-protein interactions of At1g77145 and calmodulin/-like (CaM/CML). The interaction was predicted using AI-1 and the graph was generated by Arabidopsis interactions viewer (https://bar.utoronto.ca/interactions/cgi-bin/arabidopsis_interactions_viewer.cgi, accessed on June 2021). The color of nodes represents the subcellular localization predicted by SUBA (http://suba.live/, accessed on June 2021), and the thickness of lines represents the confident values of interactions. At2g41090, calmodulin like 10/CML10; At2g41100, calmodulin like 12/CML12; At3g43810, calmodulin 7/CaM7; At3g51920, calmodulin like 9/CML9; At4g14640, calmodulin like 8/CML8; At5g21274, calmodulin 6/CaM6; At5g37780, calmodulin 1/CaM1.

3. Discussion

Although the genome draft of Arabidopsis was released over two decades ago, more than 12,800 genes (46% of 27,655 protein-coding genes, Araport11, www.arabidopsis.org, accessed on March 2021) remain functionally uncharacterized. Intensive efforts have been devoted to identifying and characterizing these unknown or unannotated genes, such as *DUFs*. Several genes that contained various DUF domains have been studied in plants (e.g., *TBL/DUF231*). However, lack of genome-wide analysis of genes containing DUF domains has prevented a full understanding of their evolutionary history and biological functions.

3.1. Evolution, Expansion and Conservation of Plant DUF506 Genes

In the present study, examination of 23 representative species suggests that the green algae *Chlamydomonas reinhardtii* is the last common ancestor of *DUF506* prior to plant colonization of the land (Figure 1). In bryophytes (e.g., *M. polymorpha* and *P. patens*), the number of *DUF506* genes has largely increased, implying that they might have assisted with plant adaption to terrestrial conditions. Contrary to a recent report [16], the number of *DUF506* genes in different species is closely coordinated with the total number of protein-coding transcripts, regardless of genome size (Table S5). Moreover, the species

that belonged to the same taxa exhibited a similar percentage of *DUF506* genes among the total genes (*DUF506‰*). For instance, between Arabidopsis and soybean, the difference of *DUF506* (0.47‰ versus 0.43‰) was significantly smaller when considering the genome size (135 Mb versus 978 Mb). Additionally, phylogenetic analysis found that in every subgroup, *DUF506* proteins from monocots or eudicots tended to cluster in the same clades, suggesting the differentiation might be subjected before lineages diverged.

In the process of evolution, genome polyploidization has occurred frequently in Eukaryotes, concomitant with gene duplication events [36]. Three types of gene duplication events have been reported, namely tandem duplication, segmental duplication, and whole-genome duplication [37]. Here, three segmental duplicated gene pairs were identified from Arabidopsis, as well as in several monocot species (Table 2 and Table S4), suggesting that the expansion of the *DUF506* gene family might have been derived from genome polyploidy events to diversify gene function. Noteworthy is that considering the chromosome positioning, phylogenetic relationship, identity of CDS and protein sequences, *At1g77145* and *At1g77160* could be an undiscovered tandem duplicated gene pair (Figures 1, 2a and 3; Table 1). Previous studies have demonstrated essential and multifaceted genes, such as transcription factors and protein kinases, were retained during natural selection, whereas genes with a single specialized biological function, such as transferases, had been preferentially lost [38]. *At3g25240/RXR1*, and its duplicated pair, *At3g07350*, both strongly responded to P-stress [24]. However, they showed different expression patterns when exposed to abiotic stresses or ABA treatment (Figure 6). Taken together, the present results show that *DUF506* genes from same species have experienced strong purifying positive selection after the duplication event (Ka/Ks ratio < 1, Table 2), implicating their functional redundancy under certain conditions.

Structural-modeling (secondary and tertiary structures) is increasingly considered more effective at predicting protein function than protein sequencing (primary structure) alone, because the conservation of functional foldable motifs is likely higher than the sequence [8,39]. A previous study predicted *DUF506* proteins contain a conserved ($\alpha\alpha\alpha\beta\beta\beta\alpha\beta\alpha\alpha$) secondary structure pattern [20]. Comparative structural analysis found the three tandem β-sheet structures were mainly presented in motif 2 (Figure 2c), which is found in most *DUF506* proteins, whereas the α–coil structures were scattered in the neighborhood. Considering the ubiquitous presence of these secondary structures, *DUF506*s might perform conserved biological functions across species.

3.2. Distinct Roles of DUF506 Genes in Biological Processes

Expression profiling of Arabidopsis *DUF506* genes in different tissues, developmental stages, and under abiotic stress provided new data that could help in understanding potential biological functions. In general, transcript levels of *DUF506* genes fluctuated in various organs (Figures 5 and 6a). The hypothetical tandem duplicated gene pair, *At1g77145* and *At1g77160*, displayed identical expression patterns. These results supported the hypothesis of positive selection of *DUF506* genes based on the evolutionary analysis. However, expression changes of the segmental duplicated pair diverged from that of its homolog. For example, the *At1g62420* transcript was highly abundant in roots and siliques relative to other tissues, whereas the *At1g12030* transcript was steadily expressed.

The Arabidopsis *DUF506* transcripts responded more to salt stress than to osmotic stress. As shown in Figure 6a, most Arabidopsis *DUF506* genes (10/12, except *At3g54550*) responded to salt stress, whereas only two genes (*At1g62420* and *At3g25240*) were induced by osmotic treatment. Additionally, about 75% (8/12) of the genes responded to ABA treatment, an observation that was consistent with the presence of ABRE *cis*-elements in their promoters (Figure S3). *At1g12030*, *At1g77160*, *At2g20670*, and *At2g38820*, significantly responded to both salinity and ABA treatment. *At1g62420* and *At3g25240* transcripts were strongly induced by distinct abiotic stresses (e.g., salt, osmotic, cold, and heat), but were irresponsive to ABA treatment despite the presence of ABREs in their promoter regions (Figure 5, Figure 6 and Figure S3). Overall, these results suggest that the expression of

Arabidopsis *DUF506* genes are mediated through both ABA-dependent (e.g., *At1g12030* and *At1g77160*) and independent (e.g., *At1g62420* and *At3g25240*) pathways in response to environmental stimuli.

DUF506 genes in other plant species actively responded to abiotic stress (Figures S4–S9). For example, in Brachypodium and rice, *Bradi2g58590* and *LOC_Os01g68650*, the closest homologs of *At3g25240*, were strongly up-regulated by salinity and cold stress. A similar expression pattern was observed in *At2g39650* homologs (*Bradi2g62310* and *LOC_Os01g74250*). Due to limited data availability on soybean and Medicago, the expression profiles of *DUF506* genes responding to drought stress were analyzed. *Medtr1g070970*, *Medtr2g013780*, *Medtr4g088770*, *Medtr8g093440*, and their corresponding homologs in Soybean, *Glyma.02G143100*, *Glyma.15G074800*, *Glyma.13G179200*, *Glyma.05G190100* were significantly induced by drought stress.

Nowadays, protein interaction analysis, through yeast two hybridization (Y2H) or prediction based on public inter-proteomic data, is a useful tool to provide additional insights deciphering gene functions. *At1g12030*, which is the only *DUF506* gene that clearly responds to iron starvation (Fold Change [FC] $\geq$ 6 in shoot, and FC $\geq$ 2.5 in root, GENEVESTIGATOR®), was predicted to interact with several iron (Fe)-deficiency response proteins (Table S6). One example is At1g47400/FEP3/IRONMAN1, which encodes a 50 amino acid long FE-UPTAKE-INDUCING PEPTIDE, predominantly expressed in the vascular tissue (e.g., phloem) of leaves and roots. Overexpression of *FEP3* significantly accumulated Fe and Mn in Arabidopsis through activation of Fe uptake genes in roots [40,41]. Additional Fe-deficiency related proteins that interactin with At1g12030, including chloroplast-localized At5g67370/CONSERVED IN THE GREEN LINEAGE AND DIATOMS 27/CGLD27, 14-3-3 protein At1g34760/ROOT HAIR SPECIFIC 5/RHS5, tonoplast-localized iron efflux transporter At5g03570/IRON-REGULATED PROTEIN 2/IREG2, and phloem-based iron transporter At1g56430/NICOTIANAMINE SYNTHASE 4/NAS4, have been demonstrated to function effectively in maintaining endogenous iron homeostasis [42–45]. Lately, a novel basic helix–loop–helix (bHLH) transcription factor, At3g19860/Upstream Regulator of IRT1/URI, was demonstrated to act as a central regulator of the iron deficient signaling pathway in Arabidopsis [46]. Chromatin immunoprecipitation followed by a sequencing (ChIP-seq) experiment revealed that URI directly bound to promoters of many iron-regulated genes, including At1g12030 and its interactors, At5g05250/IRP6, IMA1, CGLD27, and NAS4 [46]. These results suggest that At1g12030 acts in an essential URI-dependent role in responding to iron deficiency.

In addition to responding nutrient-deficiency, another Arabidopsis *DUF506* member, At2g39650, intensively interacted with four well-known salt-stress-related transcription factors, including SZF1, SZF2, ZAT10, and At3g49530/NAC062 (Table S6). Three of the ZINC finger domain-containing proteins have been thoroughly characterized and were proven to be regulators under multiple stresses (e.g., salt, drought, cold, hypoxia, and H_2O_2) by activating the expression of stress-inducible genes [47–49]. Moreover, the membrane-bound transcription factor NAC062 was reported not only to promptly activate transcriptional responses upon exposure to abiotic/biotic stress, but also to mediate the unfolded protein response (UPR) under endoplasmic reticulum (ER) stress [50,51]. Thus, considering the high induction of At2g39650 under salt stress (Figure 6a), the present results suggest its involvement in transcriptional regulation of stress-response genes by interacting with different transcription factors.

Unlike the aforementioned interactions predicted based on text mining and co-expression data, At1g77145′s interaction with at least five Calmodulin/-like proteins (CaM/CML), such as CML8, CML9, CML10, CaM6, and CaM7, was determined through protein microarrays [34]. In plants, when intracellular levels of Ca^{2+} increase, CaM/CML directly binds to Ca^{2+} *via* symmetrical EF-hand motifs, then functions as an essential calcium secondary messenger by altering the activity of target proteins to contribute to signaling during developmental processes and adaptation to environmental stimuli [52]. CML8, CaM6, and CaM7 have been demonstrated to interact with the BRASSINOSTEROID-INSENSITIVE 1

(BRI1) receptor kinase [53]. The interactions inhibited the activity of BRI1, suggesting the participation of Ca^{2+} in BR signaling. Phylogenic analysis indicated that CML8, CML9, and CML10 belong to the same CML subgroup in Arabidopsis, and actively responded to various environmental stimuli [54]. The *CML8* transcript has been reported to be induced by salicylic acid (SA) and salt treatment [55]. *CML9* mutants exhibited hypersensitivity to ABA, and elevated salt and drought tolerance through up-regulation of stress-response genes, suggesting that AtCML9 is involved in the ABA-dependent pathway responding to abiotic stress [56]. Differently, through interacting with the Ascorbic acid (AsA) biosynthesis-related enzyme phosphomannomutase (PMM), CML10 augmented the activity of PMM and consequently modulated oxidative stress responses [57]. Moreover, CML8 and CML9 have been reported to play roles in plant immune response [58,59]. Searching the Calmodulin Target Database [60], a classic 1-12 motif ([FILVW]xxxxxxxxxx[FILVW], two bulky hydrophobic residues spaced by 10 amino acids residues) was identified in motif 1 of At1g77145. Further analysis showed different types of CaM binding motifs were scattered in twelve of the Arabidopsis *DUF506* proteins (Table S7). Presumably, At1g77145, and perhaps other *DUF506* members, might specifically bind to certain CaMs/CMLs in the presence of Ca^{2+}, leading to a structural conformation change or a shift in enzymatic activity, and ultimately influencing downstream stress-response gene expression.

The silique-specific At3g54550 potentially interacts with At2g38820 (Table S6), which is constitutively expressed in vivo, implying a possible chimeric complex consisting of different *DUF506* proteins, might perform regulatory functions during the reproductive stage, especially during seed development.

Overall, the broad response of *DUF506* genes to various abiotic stresses and the predicted interaction with various functional proteins suggest that Arabidopsis *DUF506* proteins have a diverse range of regulatory functions in response to abiotic stress (e.g., At2g39650 and At1g77145) or nutrient deficiency (e.g., At1g12030 and At3g25240).

4. Materials and Methods

4.1. Identification and Phylogenetic Analysis of DUF506 Protein in Plants

In this study, 211 *DUF506* candidate genes (Pfam accession: PF04720, ID: PDDEXK_6, Description: PDDEXK-like family of unknown function, https://pfam.xfam.org/) were identified from 17 plant species. The protein sequences, cDNA sequences, DNA sequences, upstream 2 kb genomic DNA sequences, and protein-coding sequences (CDS) of *DUF506* were downloaded from Phytozome v12.1.6 (*Amborella trichopoda* v1.0, *Arabidopsis thaliana* Araport11, *Capsella rubella* v1.0, *Brassica rapa* FPsc v1.3, *Cucumis sativus* v1.0, *Glycine max* Wm82.a2.v1, *Phaseolus vulgaris* v2.1, *Medicago truncatula* Mt4.0 v1, *Ricinus communis* v0.1, *Vitis vinifera* Genoscope.12X, *Daucus carota* v2.0, *Solanum lycopersicum* iTAG2.4, *Oryza sativa* v7_JGI, *Brachypodium distachyon* v3.1, *Hordeum vulgare* r1, *Sorghum bicolor* v3.1.1, and *Zea mays* PH207 v1.1). In genes undergoing alternative splicing, the longest transcript was used for downstream analysis. Predicted partial proteins carrying the signature domain were filtered out. The theoretical isoelectric point (pI) of *DUF506* proteins were calculated by Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/, Sequence analysis tool—Protein Isoelectric Point) [61]. The subcellular localization was predicted by using WoLF PSORT program (https://wolfpsort.hgc.jp/).

The amino acid sequence alignment of *DUF506* genes was performed by Clustal Omega1.2.2 (https://www.ebi.ac.uk/Tools/msa/clustalo/). The alignment was used to construct an unrooted phylogenic tree using the neighbor-joining method with 1000 bootstrap replications using Fast Tree. Both programs were carried out within Geneious Prime® software (v.2020.2.3.).

4.2. Gene Structure and Conserved Motif Analysis

The exon-intron distributions of *DUF506* genes were graphically generated by Gene Structure Display Server 2.0 [62] (http://gsds.gao-lab.org/) using coding sequences and their corresponding genomic DNA sequences under default settings. The conserved

motifs of *DUF506* protein were predicted by using the MEME web server [27] (v5.1.1, http://meme-suite.org/tools/meme) with the maximum number of motifs set at 10, and optimum width of motifs between 5 and 100 amino acids. The secondary structure of *DUF506* protein was predicted by using the JPred4 web server [63] (A Protein Secondary Structure Prediction Server, http://www.compbio.dundee.ac.uk/jpred4/index_up.html).

4.3. Chromosomal Localization, Gene Duplication and Syntenic Analysis

Chromosomal locations of Arabidopsis *DUF506* proteins were visualized by using ePlant web tool (https://bar.utoronto.ca/eplant/). Gene duplication events were analyzed by using the Multiple Collinearity Scan toolkit [64]. The precalculated nonsynonymous (Ka) and synonymous (Ks) substitutions of each duplicated Arabidopsis *DUF506* genes were obtained from Plant Genome Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/).

4.4. Subcellular Localization of Arabidopsis DUF506 Protein

The coding regions, except for the terminator codon of selected Arabidopsis *DUF506* genes, were amplified and then ligated into the destination vector, pEarleyGate 103 [29], to generate a AtDUF506-GFP fusion construct under the control of the CaMV35S promoter. Constructs were transformed into Arabidopsis (Col-0) mesophyll protoplast as described previously [30]. Twelve hours after transformation, fluorescence was detected using a Leica TCS SP8 confocal laser-scanning microscope. All experiments were repeated at least three times and representative images are displayed in figures.

4.5. Stress-Related Cis-Elements Analysis

The PlantCARE program [65] (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was employed to analyze the 2 kb upstream genomic sequences of plant *DUF506* genes.

4.6. Expression Analysis of Arabidopsis DUF506 Genes via qRT-PCR

Arabidopsis thaliana (ecotype Col-0) seeds were sterilized in 70% ethanol and 20% bleach as described previously [66]. After stratification for 3 days at 4 °C, sterilized seeds were directly sown in half strength (1/2) Murashige and Skoog (MS) medium (Caisson Labs) supplemented with 1% sucrose (pH 5.7) and 0.05% MES (w/v). Seeds were germinated and grown in a growth chamber maintained at 22°C (16-h-light and 8-h-dark cycle, 120 μmol^{-2}s^{-1} light intensity). The roots and shoots of 10-day–old seedlings, and the roots, stems, rosette leaves, flowers and siliques of mature Arabidopsis plants were used for tissue-specific expression analysis. For salt stress, osmotic stress, or ABA treatment, 7-day-old Arabidopsis seedling were transferred into 1/2 MS liquid medium supplemented with 150 mM NaCl, 300 mM mannitol or 100 μM ($\pm$)-*cis, trans*-ABA (Sigma), and the stress treatments were applied for 24 hr. A fresh medium-only control was conducted in parallel. Samples were rinsed with demineralized water, shoot and root tissues were harvested separately, snap-frozen in liquid N_2 and stored at -80 °C until further use.

The qRT-PCR experiments were performed according to previous studies [66]. All the experiments were repeated at least three times using cDNA prepared from two biological replicates. Primers used in the study are listed in Supplemental Table S8.

4.7. Protein-Protein Interaction Analysis

STRING-DB [33] (v11.0, https://string-db.org/), AtPIN [67] (Arabidopsis thaliana protein interaction network, https://atpin.bioinfoguy.net/cgi-bin/atpin.pl), and AI-1 [68] (Arabidopsis Interactome-1, http://interactome.dfci.harvard.edu/A_thaliana/index.php) online search tools were used to predict the putative protein-protein interaction networks with candidate Arabidopsis *DUF506* proteins using default settings.

5. Conclusions

In this study, 211 *DUF506* genes in 17 land species were identified and were conducted to estimate evolutionary analysis. Gene structure and conserve motif analysis indicated the conservation and dispersal of *DUF506* during evolution. Furthermore, expression profiling and protein interaction predictions suggest that Arabidopsis *DUF506* genes might be involved in plant resilience to environmental stresses.

References

1. Bateman, A.; Coggill, P.; Finn, R. DUFs: Families in search of function. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66*, 1148–1152. [CrossRef]
2. Simm, R.; Morr, M.; Kader, A.; Nimtz, M.; Römling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* **2004**, *53*, 1123–1134. [CrossRef]
3. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2020**, *49*, D412–D419. [CrossRef] [PubMed]
4. Ranocha, P.; Denancé, N.; Vanholme, R.; Freydier, A.; Martinez, Y.; Hoffmann, L.; Köhler, L.; Pouzet, C.; Renou, J.-P.; Sundberg, B.; et al. Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* **2010**, *63*, 469–483. [CrossRef] [PubMed]
5. Vaattovaara, A.; Brandt, B.; Rajaraman, S.; Safronov, O.; Veidenberg, A.; Luklová, M.; Kangasjärvi, J.; Löytynoja, A.; Ho-thorn, M.; Salojärvi, J.; et al. Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Commun. Biol.* **2019**, *2*, 56. [CrossRef]
6. Yu, C.-Y.; Sharma, O.; Nguyen, P.H.T.; Hartono, C.D.; Kanehara, K. A pair of DUF538 domain-containing proteins modulates plant growth and trichome development through the transcriptional regulation of GLABRA1 in Arabidopsis thaliana. *Plant J.* **2021**. Accepted Author Manuscript. [CrossRef] [PubMed]
7. Urbanowicz, B.R.; Peña, M.J.; Ratnaparkhe, S.; Avci, U.; Backe, J.; Steet, H.F.; Foston, M.; Li, H.; O'Neill, M.A.; Ragauskas, A.J.; et al. 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14253–14258. [CrossRef] [PubMed]
8. Laxmi, A. DUF581 Is Plant Specific FCS-Like Zinc Finger Involved in Protein-Protein Interaction. *PLoS ONE* **2014**, *9*, e99074. [CrossRef]
9. Moriwaki, T.; Miyazawa, Y.; Kobayashi, A.; Uchida, M.; Watanabe, C.; Fujii, N.; Takahashi, H. Hormonal Regulation of Lateral Root Development in Arabidopsis Modulated by MIZ1 and Requirement of GNOM Activity for MIZ1 Function. *Plant Physiol.* **2011**, *157*, 1209–1220. [CrossRef] [PubMed]
10. Zúñiga-Sánchez, E.; Soriano, D.; Martínez-Barajas, E.; Orozco-Segovia, A.; Gamboa-deBuen, A. *BIIDXI*, the *At4g32460* DUF642 gene, is involved in pectin methyl esterase regulation during Arabidopsis thaliana seed germination and plant development. *BMC Plant Biol.* **2014**, *14*, 338. [CrossRef]
11. Tong, H.; Leasure, C.D.; Hou, X.; Yuen, G.; Briggs, W.; He, Z.-H. Role of root UV-B sensing in Arabidopsis early seedling development. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 21039–21044. [CrossRef] [PubMed]
12. Cao, X.; Yang, K.-Z.; Xia, C.; Zhang, X.-Q.; Chen, L.-Q.; Ye, D. Characterization of DUF724 gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **2009**, *72*, 61. [CrossRef]
13. Jones-Rhoades, M.W.; Borevitz, J.O.; Preuss, D. Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS Genet.* **2007**, *3*, 1848–1861. [CrossRef]
14. Kodama, Y.; Suetsugu, N.; Wada, M. Novel protein-protein interaction family proteins involved in chloroplast movement response. *Plant Signal. Behav.* **2011**, *6*, 483–490. [CrossRef]

15. Kim, S.J.; Ryu, M.Y.; Kim, W.T. Suppression of Arabidopsis RING-DUF1117 E3 ubiquitin ligases, AtRDUF1 and AtRDUF2, reduces tolerance to ABA-mediated drought stress. *Biochem. Biophys. Res. Commun.* **2012**, *420*, 141–147. [CrossRef]

16. Yang, Q.; Niu, X.; Tian, X.; Zhang, X.; Cong, J.; Wang, R.; Zhang, G.; Li, G. Comprehensive genomic analysis of the DUF4228 gene family in land plants and expression profiling of ATDUF4228 under abiotic stresses. *BMC Genom.* **2020**, *21*, 12–16. [CrossRef]

17. Bischoff, V.; Nita, S.; Neumetzler, L.; Schindelasch, D.; Urbain, A.; Eshed, R.; Persson, S.; Delmer, D.; Scheible, W.-R. TRI-CHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Bi-osynthesis in Arabidopsis. *Plant Physiol.* **2010**, *153*, 590–602. [CrossRef] [PubMed]

18. Xin, Z.; Mandaokar, A.; Chen, J.; Last, R.L.; Browse, J. Arabidopsis ESK1 encodes a novel regulator of freezing tolerance. *Plant J.* **2007**, *49*, 786–799. [CrossRef]

19. Zhong, R.; Cui, D.; Ye, Z. Evolutionary origin of *O*-acetyltransferases responsible for glucomannan acetylation in land plants. *New Phytol.* **2019**, *224*, 466–479. [CrossRef] [PubMed]

20. Knizewski, L.; Kinch, L.N.; Grishin, N.V.; Rychlewski, L.; Ginalski, K. Realm of PD-(D/E)XK nuclease superfamily revisited: Detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* **2007**, *7*, 40. [CrossRef] [PubMed]

21. Kosinski, J.; Feder, M.; Bujnicki, J.M. The PD-(D/E)XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinform.* **2005**, *6*, 172. [CrossRef]

22. Zhao, L.; Bonocora, R.P.; Shub, D.A.; Stoddard, B.L. The restriction fold turns to the dark side: A bacterial homing endonu-clease with a PD-(D/E)-XK motif. *EMBO J.* **2007**, *26*, 2432–2442. [CrossRef]

23. Steczkiewicz, K.; Muszewska, A.; Knizewski, L.; Rychlewski, L.; Ginalski, K. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* **2012**, *40*, 7016–7045. [CrossRef]

24. Ying, S.; Blancaflor, E.; Liao, F.; Scheible, W.R. A Phosphorus-Limitation Induced, Functionally Conserved *DUF506* Protein is a Repressor of Root Hair Elongation in Arabidopsis thaliana. *bioRxiv* **2021**. [CrossRef]

25. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]

26. Mount, D.W. Using Hidden Markov Models to Align Multiple Sequences. *Cold Spring Harb. Protoc.* **2009**, *2009*, pdb.top41. [CrossRef] [PubMed]

27. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, w202–w208. [CrossRef]

28. Fay, J.C.; Wu, C.-I. Sequence Divergence, Functional Constraint, and Selection in Protein Evolution. *Annu. Rev. Genom. Hum. Genet.* **2003**, *4*, 213–235. [CrossRef] [PubMed]

29. Earley, K.W.; Haag, J.R.; Pontes, O.; Opper, K.; Juehne, T.; Song, K.; Pikaard, C.S. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **2006**, *45*, 616–629. [CrossRef] [PubMed]

30. Yoo, S.-D.; Cho, Y.-H.; Sheen, J. Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2007**, *2*, 1565–1572. [CrossRef]

31. Bari, R.; Pant, B.D.; Stitt, M.; Scheible, W.-R. PHO2, MicroRNA399, and PHR1 Define a Phosphate-Signaling Pathway in Plants. *Plant Physiol.* **2006**, *141*, 988–999. [CrossRef] [PubMed]

32. Zhu, J.-K. Abiotic Stress Signaling and Responses in Plants. *Cell* **2016**, *167*, 313–324. [CrossRef] [PubMed]

33. Von Mering, C.; Jensen, L.J.; Snel, B.; Hooper, S.D.; Krupp, M.; Foglierini, M.; Jouffre, N.; Huynen, M.A.; Bork, P. STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **2005**, *33*, D433–D437. [CrossRef]

34. Popescu, S.C.; Popescu, G.V.; Bachan, S.; Zhang, Z.; Seay, M.; Gerstein, M.; Snyder, M.; Dinesh-Kumar, S.P. Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4730–4735. [CrossRef]

35. Zielinski, R.E. Camodulin and calmodulin-binding proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 697–725. [CrossRef]

36. Adams, K.L.; Wendel, J.F. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **2005**, *8*, 135–141. [CrossRef]

37. Moore, R.C.; Purugganan, M.D. The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* **2005**, *8*, 122–128. [CrossRef] [PubMed]

38. Freeling, M. Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [CrossRef]

39. Mudgal, R.; Sandhya, S.; Chandra, N.; Srinivasan, N. De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. *Biol. Direct* **2015**, *10*, 38. [CrossRef]

40. Grillet, L.; Lan, P.; Li, W.; Mokkapati, G.; Schmidt, W. IRON MAN is a ubiquitous family of peptides that control iron transport in plants. *Nat. Plants* **2018**, *4*, 953–963. [CrossRef] [PubMed]

41. Hirayama, T.; Lei, G.J.; Yamaji, N.; Nakagawa, N.; Ma, J.F. The Putative Peptide Gene FEP1 Regulates Iron Deficiency Re-sponse in Arabidopsis. *Plant Cell Physiol.* **2018**, *59*, 1739–1752. [CrossRef]

42. Rodríguez-Celma, J.; Pan, I.C.; Li, W.D.; Lan, P.D.; Buckhout, T.J.; Schmidt, W. The transcriptional response of Arabidopsis leaves to Fe deficiency. *Front. Plant Sci.* **2013**, *4*, 276. [CrossRef] [PubMed]

43. Schaaf, G.; Honsbein, A.; Meda, A.R.; Kirchner, S.; Wipf, D.; von Wirén, N. AtIREG2 Encodes a Tonoplast Transport Protein Involved in Iron-dependent Nickel Detoxification in Arabidopsis thaliana Roots. *J. Biol. Chem.* **2006**, *281*, 25532–25540. [CrossRef] [PubMed]

44. Schuler, M.; Rellán-Álvarez, R.; Fink-Straube, C.; Abadia, J.; Bauer, P. Nicotianamine Functions in the Phloem-Based Transport of Iron to Sink Organs, in Pollen Development and Pollen Tube Growth in Arabidopsis. *Plant Cell* **2012**, *24*, 2380–2400. [CrossRef]

45. Yang, J.L.; Chen, W.W.; Chen, L.Q.; Qin, C.; Jin, C.W.; Chong, W.J.; Shi, Y.Z.; Zheng, S.J. The 14-3-3 protein GENERAL REGULATORY FACTOR11 (GRF11) acts downstream of nitric oxide to regulate iron acquisition in Arabidopsis thaliana. *New Phytol.* **2013**, *197*, 815–824. [CrossRef]

46. Kim, S.A.; LaCroix, I.S.; Gerber, S.; Guerinot, M.L. The iron deficiency response in Arabidopsis thaliana requires the phosphory-lated transcription factor URI. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24933–24942. [CrossRef] [PubMed]

47. Davletova, S.; Schlauch, K.; Coutu, J.; Mittler, R. The Zinc-Finger Protein Zat12 Plays a Central Role in Reactive Oxygen and Abiotic Stress Signaling in Arabidopsis. *Plant Physiol.* **2005**, *139*, 847–856. [CrossRef]

48. Mittler, R.; Kim, Y.; Song, L.; Coutu, J.; Coutu, A.; Ciftci-Yilmaz, S.; Lee, H.; Stevenson, B.; Zhu, J.-K. Gain- and loss-of-function mutations inZat10enhance the tolerance of plants to abiotic stress. *FEBS Lett.* **2006**, *580*, 6537–6542. [CrossRef]

49. Sun, J.; Jiang, H.; Xu, Y.; Li, H.; Wu, X.; Xie, Q.; Li, C. The CCCH-Type Zinc Finger Proteins AtSZF1 and AtSZF2 Regulate Salt Stress Responses in Arabidopsis. *Plant Cell Physiol.* **2007**, *48*, 1148–1158. [CrossRef] [PubMed]

50. Seo, P.J.; Kim, M.J.; Park, J.-Y.; Kim, S.-Y.; Jeon, J.; Lee, Y.-H.; Kim, J.; Park, C.-M. Cold activation of a plasma mem-brane-tethered NAC transcription factor induces a pathogen resistance response in Arabidopsis. *Plant J.* **2010**, *61*, 661–671. [CrossRef]

51. Yang, Z.-T.; Lu, S.-J.; Wang, M.-J.; Bi, D.-L.; Sun, L.; Zhou, S.-F.; Song, Z.-T.; Liu, J.-X. A plasma membrane-tethered transcription factor, NAC062/ANAC062/NTL6, mediates the unfolded protein response in Arabidopsis. *Plant J.* **2014**, *79*, 1033–1043. [CrossRef]

52. Bouché, N.; Yellin, A.; Snedden, W.A.; Fromm, H. Plant-specific calmodulin-binding proteins. *Annu. Rev. Plant Biol.* **2005**, *56*, 435–466. [CrossRef] [PubMed]

53. Oh, M.-H.; Kim, H.S.; Wu, X.; Clouse, S.D.; Zielinski, R.E.; Huber, S.C. Calcium/calmodulin inhibi-tion of the Arabidopsis BRASSINOSTEROID-INSENSITIVE 1 receptor kinase provides a possible link between calcium and brassinosteroid signalling. *Biochem. J.* **2012**, *443*, 515–523. [CrossRef]

54. Zhu, X.; Dunand, C.; Snedden, W.; Galaud, J.-P. CaM and CML emergence in the green lineage. *Trends Plant Sci.* **2015**, *20*, 483–489. [CrossRef] [PubMed]

55. Park, H.C.; Park, C.Y.; Koo, S.C.; Cheong, M.S.; Kim, K.E.; Kim, M.C.; Lim, C.O.; Lee, S.Y.; Yun, D.-J.; Chung, W.S. AtCML8, a calmodulin-like protein, differentially activating CaM-dependent enzymes in Arabidopsis thaliana. *Plant Cell Rep.* **2010**, *29*, 1297–1304. [CrossRef] [PubMed]

56. Magnan, F.; Ranty, B.; Charpenteau, M.; Sotta, B.; Galaud, J.-P.; Aldon, D. Mutations in AtCML9, a calmodulin-like protein fromArabidopsis thaliana, alter plant responses to abiotic stress and abscisic acid. *Plant J.* **2008**, *56*, 575–589. [CrossRef] [PubMed]

57. Cho, K.; Nguyen, H.T.K.; Kim, S.Y.; Shin, J.S.; Cho, D.H.; Hong, S.B.; Shin, J.S.; Ok, S.H. CML 10, a variant of calmodulin, modulates ascorbic acid synthesis. *New Phytol.* **2016**, *209*, 664–678. [CrossRef]

58. Leba, L.-J.; Cheval, C.; Ortiz-Martín, I.; Ranty, B.; Beuzón, C.R.; Galaud, J.-P.; Aldon, D. CML9, an Arabidopsis calmodulin-like protein, contributes to plant innate immunity through a flagellin-dependent signalling pathway. *Plant J.* **2012**, *71*, 976–989. [CrossRef] [PubMed]

59. Zhu, X.; Robe, E.; Jomat, L.; Aldon, D.; Mazars, C.; Galaud, J.-P. CML8, an Arabidopsis Calmodulin-Like Protein, Plays a Role inPseudomonas syringaePlant Immunity. *Plant Cell Physiol.* **2016**, *58*, 307–319. [CrossRef]

60. Yap, K.L.; Kim, J.; Truong, K.; Sherman, M.; Yuan, T.; Ikura, M. Calmodulin Target Database. *J. Struct. Funct. Genom.* **2000**, *1*, 8–14. [CrossRef]

61. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Se-quences. *BioTechniques* **2000**, *28*, 1102–1104. [CrossRef]

62. Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2014**, *31*, 1296–1297. [CrossRef]

63. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. [CrossRef]

64. Wang, Y.; Tang, H.; DeBarry, J.; Tan, X.; Li, J.; Wang, X.; Lee, T.-H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]

65. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]

66. Ying, S.; Zhang, D.-F.; Fu, J.; Shi, Y.-S.; Song, Y.-C.; Wang, T.-Y.; Li, Y. Cloning and characterization of a maize bZIP tran-scription factor, ZmbZIP72, confers drought and salt tolerance in transgenic Arabidopsis. *Planta* **2012**, *235*, 253–266. [CrossRef] [PubMed]

67. Brandão, M.M.; Dantas, L.L.; Silva-Filho, M.C. AtPIN: Arabidopsis thaliana Protein Interaction Network. *BMC Bioinform.* **2009**, *10*, 454. [CrossRef] [PubMed]

68. Consortium, A.I.M. Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science* **2011**, *333*, 601–607.