Article

Genome-Wide Identification and Characterization of the Soybean DEAD-Box Gene Family and Expression Response to Rhizobia

Yongliang Wang (ORCID)**, Junwen Liao, Jun Wu, Huimei Huang, Zhanxin Yuan, Wei Yang, Xinying Wu and Xia Li ***

National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory,
College of Plant Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Road,
Hongshan District, Wuhan 430070, China; wangyonglaing@webmail.hzau.edu.cn (Y.W.);
liaojunwenn@163.com (J.L.); wujun6662022@163.com (J.W.); huanghuimei163241@163.com (H.H.);
yuanzhanxin1218@163.com (Z.Y.); yangwei1051730475@163.com (W.Y.); 15927351341@163.com (X.W.)
* Correspondence: xli@mail.hzau.edu.cn

Abstract: DEAD-box proteins are a large family of RNA helicases that play important roles in almost all cellular RNA processes in model plants. However, little is known about this family of proteins in crops such as soybean. Here, we identified 80 DEAD-box family genes in the *Glycine max* (soybean) genome. These *DEAD-box* genes were distributed on 19 chromosomes, and some genes were clustered together. The majority of DEAD-box family proteins were highly conserved in Arabidopsis and soybean, but *Glyma.08G231300* and *Glyma.14G115100* were specific to soybean. The promoters of these *DEAD-box* genes share *cis*-acting elements involved in plant responses to MeJA, salicylic acid (SA), low temperature and biotic as well as abiotic stresses; interestingly, half of the genes contain nodulation-related *cis* elements in their promoters. Microarray data analysis revealed that the *DEAD-box* genes were differentially expressed in the root and nodule. Notably, 31 genes were induced by rhizobia and/or were highly expressed in the nodule. Real-time quantitative PCR analysis validated the expression patterns of some *DEAD-box* genes, and among them, *Glyma.08G231300* and *Glyma.14G115100* were induced by rhizobia in root hair. Thus, we provide a comprehensive view of the DEAD-box family genes in soybean and highlight the crucial role of these genes in symbiotic nodulation.

Keywords: *Glycine max*; DEAD-box RNA helicases; phylogenetic analysis; nodulation; rhizobia

1. Introduction

Helicases are essential enzymes found in all prokaryotes and eukaryotes and play important roles in essentially all nucleic acid metabolism processes. These helicase proteins can be divided into different families based on different conserved motifs [1]. The monomeric helicases have two RecA-like spherical domains with a linker, while hexameric helicases contain six domains arranged in a ring. The core structures of helicases also include the motifs involved in NTP binding and hydrolysis [2]. Depending on the type of target nucleic acids, helicases are classified as DNA or RNA helicases, which are involved in DNA and RNA metabolism, respectively.

The majority of RNA helicases are DExD/H-box proteins, which share conserved motifs and domain structures [3]. These RNA helicases have more than seven conserved motifs, which form two domains. Domain one is located at the N-terminal of DExD/H-box proteins and usually consists of motifs Q, I, Ia, Ib, II and III, while domain two is located at the C-terminal of these proteins and consists of motifs IV, V and VI. Both domains are interlinked by α-helices and β-folds, and the number of helices and folds varies from protein to protein [4]. Thus, the proteins form a cleft with an ATP binding site on one side and an RNA binding site on the other. This core structure of RNA helicases ensures that RNA helicases have the ability to simultaneously bind ATP and RNA to remodel RNA and RNA–protein complexes [5]. RNA helicases are classified into several subgroups, and the

DEAD-box (DDX) family is the largest protein subgroup of RNA helicases among them. The DDX family contains 12 highly conserved motifs and can be easily identified by the inclusion of special DEAD (ASP-Glu-Ala-ASP) motifs. These highly conserved DDX RNA helicases are necessary for almost all aspects of the RNA metabolism process, from RNA transcription to degradation [6–9]. Recently, emerging evidence has showed that DDX helicases can act as sensors, regulators and/or effectors during viral infections [10,11], implicating the diverse roles of these DDX helicases.

In plants, *DDX* genes are involved in various biological processes of plant growth and plant response to various abiotic stresses [12]. For example, the *Arabidopsis* DDX protein low expression osmotic response gene 4 (LOS4) is required to export mRNA and modulates plant development and stress responses [13]. Several DDX proteins, AtRH3, AtRH36, AtRH7, AtRH9 and AtRH25, regulate plant growth or female gamete development through regulating RNA splicing or rRNA biogenesis in Arabidopsis [14,15], in addition to plant responses to abiotic stresses, such as cold and drought [16]. ATRH7 was also involved in ribosomal assembly [17]. DDX17-like RH30 helicase is an effective antiviral limiting factor that can block tomabusvirus replication in plants [18]. In rice, the DDX RNA helicase OsBIRH1 plays important roles in defense as well as biotic and abiotic responses [19]. OsRH42 promotes plant growth and cold tolerance through the pre-mRNA splicing pathway, and precise control of OsRH42 homeostasis is crucial for the adaption of rice to environmental temperature changes [20]. These studies suggest crucial roles of DDX RNA helicases in plant development and adaptation to changing environmental conditions, such as biotic and abiotic stresses. Studies on these model plants helped us recognize the important roles of DDX RNA helicases in plants; however, it remains unknown how these DDX RNA helicases are evolved and whether these proteins play important roles in the growth and plant–microbial interactions of crops.

Soybean (*Glycine max* [L.] Merr.) is one of most important crops worldwide. It is the largest source of protein and the second largest source (contributing about 30% of total vegetable oil) of vegetable oil [21,22]. In contrast to other staple crops, soybean is associated with nitrogen-fixing bacteria, rhizobia, and forms root nodules where bacteroides can convert atmospheric N_2 into ammonia for plant growth [23]. Under low-soil-nitrogen conditions soybean plants produce and release flavonoid compounds into the rhizosphere to attract rhizobia to the root surface and to activate the production of signaling molecules, lipo-chitooligosaccharidic Nod factors (NFs) [24]. The perception of NFs by NF receptors in root hair cells of soybean plants leads to the activation of the nodulation signaling pathway, allowing rhizobia to infect root hairs and nodule organogenesis in the root cortex [25]. Several core nodulation signaling components have been identified, and the NMN module, which consists of GmNINa-miRNA172c-NNC1, plays a key role in soybean nodulation [26–28]. However, these studies have mainly focused on the transcriptional regulation of nodulation. In addition, soybean is one of most sensitive crops to abiotic stresses (e.g., drought), and rhizobial and mycorrhizal inoculation can enhance the stress tolerance as well as improve the seed quality and yield of soybean [29,30]. One can predict that there are more complex regulatory networks in soybean and that RNA helicases should have roles in modulating the post-transcriptional regulation of soybean nodulation and nitrogen fixation. Although many genes have been identified that are involved in stress response, flowering and nodulation of soybean in the past twenty years [31], there are only a couple of reports on RNA helicases in soybean. One study has shown that there are 213 putative RNA helicases in the soybean genome, which are widely distributed in chromosomes and differentially expressed during development under normal conditions [32]. Another study identified *GmRH* as the stress-responsive gene [33]. Our understanding of RNA helicases in soybean is very limited, and there are many questions that need to be answered. These include whether RNA helicases regulate soybean growth and stress responses, in particular plant symbiosis with nitrogen-fixing bacteria, rhizobia.

To gain insight into the *DDX* gene family and elucidate their roles in nodulation, we performed genome-wide identification and characterization of the subfamily of DDX RNA

helicases in soybean. We identified a total of 80 *DDX* genes in the soybean genome, which are distributed on almost all chromosomes, except chromosome 12. The phylogenetic tree analyses identified 78 conserved homologs of DDX proteins in *Arabidopsis* and two DDX proteins unique to soybean. Additionally, many of these *DDX* genes were expressed in different tissues of soybean; intriguingly, some of them were responsive to nitrogen and rhizobia. Finally, real-time quantitative PCR analysis validated the expression patterns of 10 *DDX*-RNA-helicase-coding genes in response to nitrogen and rhizobia treatments. To our knowledge this is the first report of a systemic, genome-wide analysis of the *DDX* gene family and their response to symbiotic nodulation in soybean. These results provide a strong foundation for understanding the functions of the *DDX* family genes and deciphering the molecular mechanisms of RNA-mediated nitrogen responses and symbiotic nitrogen fixation.

2. Results

2.1. Identification and Characterization of the DEAD-Box Family in Glycine Max

To identify the *DDX* genes in soybean we first searched for the DDX family members in the soybean genome using the Pfam and SMART databases, based on conserved domains of DEXDc and HELICc. A total of 258 genes were obtained and their amino acid sequences were subjected to motif analysis using MEME databases for further confirmation. Thus, we identified 80 soybean *DDX* family genes that encoded 118 DDX proteins, according to the 12 conserved motifs of the DDX family (Table S1 and Figure S1). We then analyzed the lengths of the protein sequences, the molecular weights and the gene lengths of all the *DDX* genes (Table S2). The lengths of the CDS of these *DDX* genes varied between 1089 and 3606, and accordingly the lengths of the DDX family proteins ranged from 362 to 1201 amino acids with an average length of 625 amino acids.

2.2. The Chromosome Localization of the DEAD-Box Family

Using the TBtools software and CorelDRAW drawing software tools, the *DDX* genes were then mapped onto the chromosomes of soybean and named with their own gene ID (Figure 1). As shown in Figure 1, the members of the *DDX* family are distributed on all the chromosomes, except chromosome 12. However, these *DDX* genes were not uniformly distributed on these chromosomes and the numbers of the *DDX* genes on these chromosomes varied, ranging from two (chromosomes 1, 4, 6 and 16) to eight (chromosomes 3 and 8) (Figure 1). Thus, the density of the *DDX* genes on 20 chromosomes ranged from 0% (chromosome 12) to 10% (chromosome 3 and 8) (Figure 1).

2.3. Phylogenetic and Classification Analyses of DEAD-Box Family Proteins in Soybean

In order to explore the evolution of soybean DDX proteins, we conducted a comparative study of soybean and *Arabidopsis* DDX family members. In *Arabidopsis* we identified 42 genes, which encode 57 proteins containing 12 conserved motifs (Table S1). The evolutionary relationships between the DDX family members in *Arabidopsis* and soybean were analyzed by the maximum likelihood method (Figure 2). Phylogenetic tree analysis showed that the DDX family members of soybean and *Arabidopsis* were divided into 15 subclades (Figure 2). Among them, nearly all the soybean and *Arabidopsis* DDX proteins were closely related. Interestingly, two DDX proteins (Glyma.14G115100 and Glyma.08G231300) fell into a separate subclade, XIV (Figure 2), indicating that these DDX family proteins might have evolved new functions in soybean. To confirm the phylogenetic relationship between the DDX family members in *Arabidopsis* and soybean, we performed a synteny analysis on these DDX homologs of *Arabidopsis* and soybean. The results showed that only 39 *DDX* genes in soybean were the orthologs of 16 members in *Arabidopsis* (Figure S2).
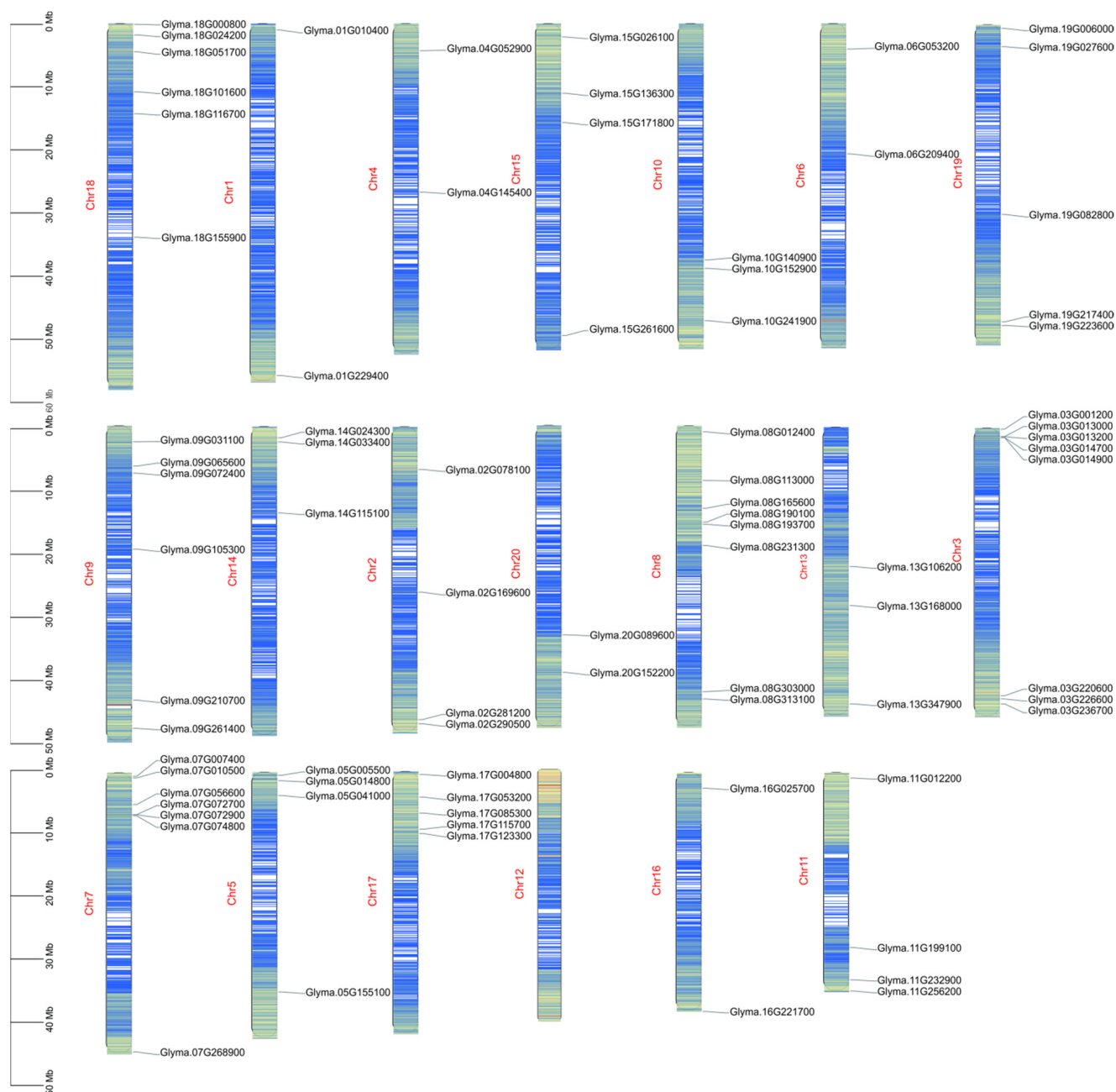
Figure 1. The chromosome distribution of the *DDX* genes.

2.4. Gene Structure and Conserved Motif Analyses of the DEAD-Box Family

To gain insights into the structural evolution of the DDX family in soybean we compared the conserved motifs and the compositions of introns/exons based on a new neighbor-joining (NJ) tree (Figure 3). According to their gene structures, they were divided into 13 subclades and named I to XIII (Figure 3). The structures of these soybean *DDX* genes varied greatly among the subclades, but genes within the same subclade had similar gene structures (Figure 3). The exon numbers of *DDX* gene family members varied from one to eighteen. Six genes had only one exon (7.5%), while others contained multiple exons. Interestingly, we noticed that two soybean unique *DDX* genes had the longest upstream untranscribed sequences, indicating a diverse regulation of gene transcription. Next, we analyzed the motifs in these DDX proteins on the online analysis site MEME and found that all the soybean DDX proteins contained 12 conserved motifs involved in ATP binding, ATP hydrolysis, nucleic acid binding and RNA helicase in DDX RNA helicases (Figure 3).

The order of these motifs was the same in these DDX proteins, but the N-terminal and C-terminal domains, as well as the amino acid sequences between the conserved motifs of these proteins, differed greatly (Figure 3). Collectively, these results suggest that soybean DDX proteins may have typical roles in RNA metabolism, but they may have diverse functions through different targets and partners.
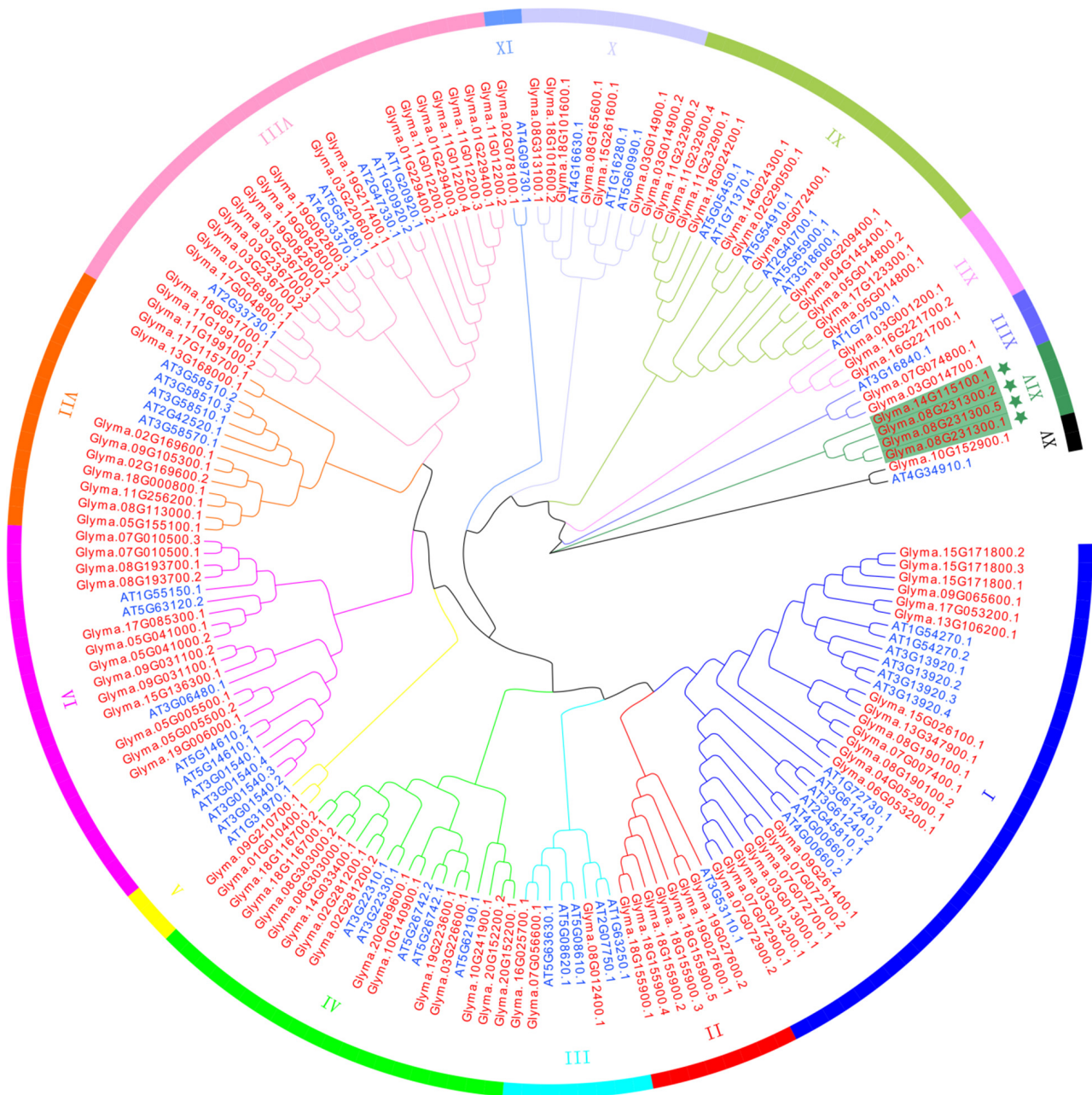

Figure 2. Phylogenetic tree of DDX proteins in Arabidopsis thaliana and soybean. Tree constructed with 1000 bootstrap replications. DDX proteins from Arabidopsis and soybean distinguished using blue and red, respectively. Proteins from different clusters (I–XV) were indicated with different color zones. Green stars at XIV indicate the soybean-specific DDX proteins.
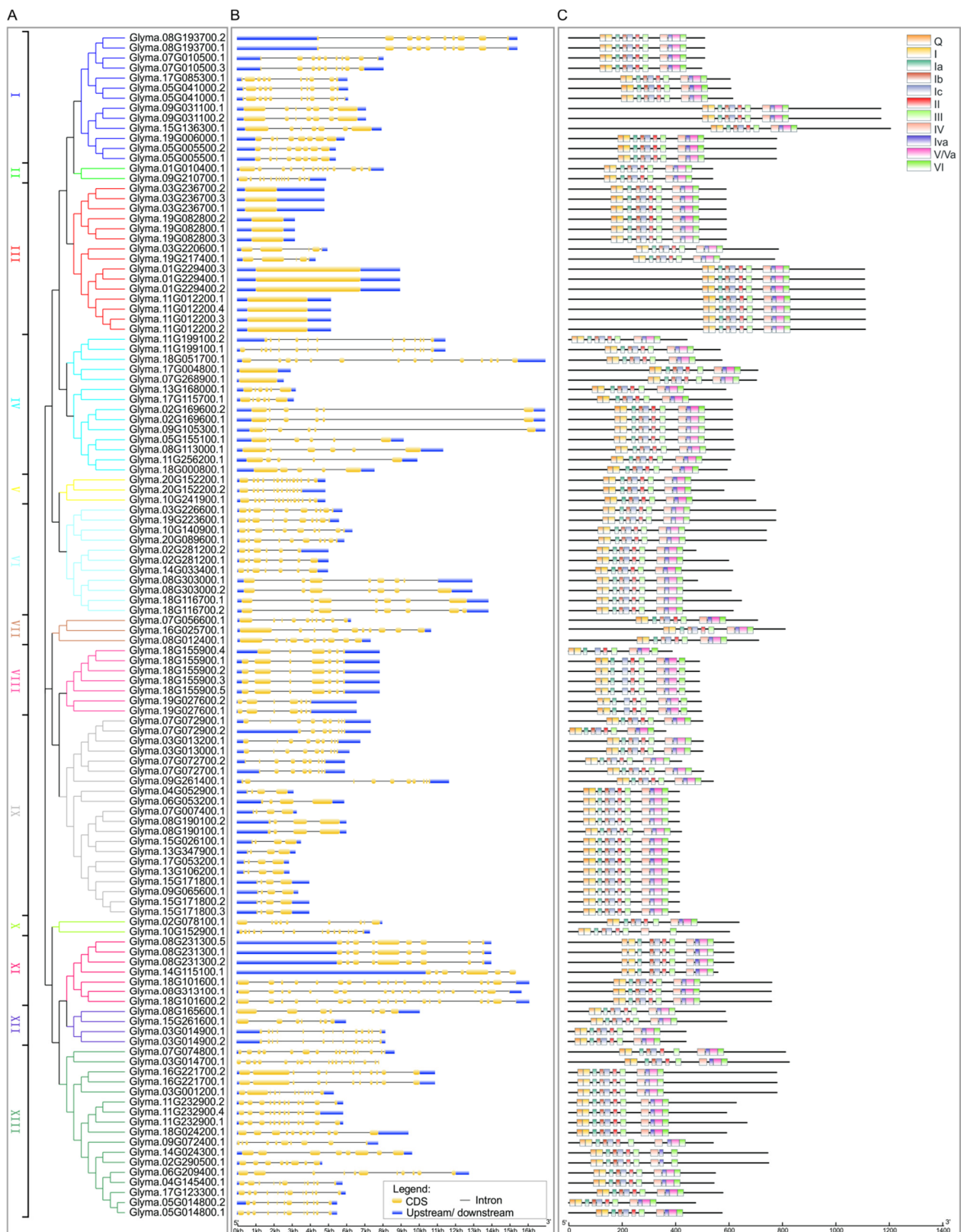
Figure 3. The structures and motif compositions of *DDX* genes and proteins. (**A**) Phylogenetic tree and classification of DDX proteins was constructed by MEGA. (**B**) Gene structures of *DDX* genes. The yellow boxes, black lines and blue boxes represent the CDS, intron and upstream/downstream. (**C**) Twelve conserved motif compositions of DDX proteins were identified using MEME. Each color represents a specific motif.

2.5. Cis Element Analysis of the DEAD-Box Gene Promoters

The transcription of a gene is regulated by the binding of transcription factors to the *cis* elements in the promoter regions [34,35]. To learn more about the transcription regulation of these soybean *DDX* genes, we analyzed the 2000 bp upstream sequences of *DDX* genes via the PlantCARE database (Figure 4A and Table S3). In addition to the abundant amount of core promoter (TATA-box), enhancer elements (CAAT-box) and light-responsive elements, twenty-six types of *cis*-regulatory elements frequently appeared in 80 promoters. These include the *cis* elements related to plant responses to phytohormones (e.g., abscisic acid, gibberellin, auxin, jasmonic acid and salicylic acid), low temperature, pathogens, wounding, drought, circadian, meristematic activity and flavonoid biosynthesis (Figure 4A). In addition, 13 types of elements are related to development and metabolism regulation. Notably, the total number of each type of *cis* element present in *DDX* promoters varied greatly, and the *cis* elements related to jasmonic acid, abscisic acid responsiveness and anaerobic were the most common in the soybean *DDX* gene family (Figures 4A and S3). Most importantly, 80 *cis* elements are related to nodule specificity in 49 *DDX* genes (Figures 4B and S3). The two genes specific to soybean (*Glyma.14G115100* and *Glyma.08G231300*) contain eleven hormone-responsive elements (e.g., abscisic acid, gibberellin, jasmonic acid and salicylic acid), three drought-responsive elements, one low-temperature-responsive element, five anaerobic-inducible elements, four metabolism regulation elements (e.g., meristem, endosperm), one AT-rich DNA-binding protein element and two nodule specificity elements (Figures 4C and S3). These results suggest that the expression of these *DDX* genes is regulated by phytohormones as well as biotic and abiotic stresses, and that these *DDX* genes might modulate plant responses to these internal or external stimuli.

2.6. Expression Profiles of the DEAD-Box Gene Family in Symbiotic Nitrogen Fixation

To investigate the possible functions of these *DDX* genes in soybean, we retrieved their expression profiles from the online transcriptome database *Glycine max* eFP Browser (http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi, accessed on 18 November 2021) and the plant comparative genomics portal Phytozome (https://phytozome-next.jgi.doe.gov/, accessed on 12 October 2021) to analyze the expression of *DDX* genes. The expression profiles of 76 *DDX* genes in the root and nodule were mined, but the remaining four members, including the two soybean-specific *DDX* genes, were not found in the database. The expression profiles of 80 *DDX* genes were mined in root hair 12 h and 24 h after inoculation. As shown in Figure 5A, eighty *DDX* genes were differentially expressed in the root and nodule: 18.75% genes were highly expressed in the root (*Glyma.13G168000*, *Glyma.20G089600*, *Glyma.04G052900*, *Glyma.07G007400*, *Glyma.07G072700*, *Glyma.10G241900*, *Glyma.15G171800*, *Glyma.08G190100*, *Glyma.14G115100*, *Glyma.03G014700*, *Glyma.08G190100*, *Glyma.17G053200*, *Glyma.13G106200*, *Glyma.15G026100*, *Glyma.03G013000*, *Glyma.13G347900*, *Glyma.18G155900*, *Glyma.06G053200*, *Glyma.08G193700* and *Glyma.09G065600*) and the nodule (*Glyma.13G168000*, *Glyma.20G089600*, *Glyma.04G052900*, *Glyma.07G007400*, *Glyma.07G072700*, *Glyma.10G241900*, *Glyma.15G171800*, *Glyma.20G152200*, *Glyma.08G190100*, *Glyma.14G115100*, *Glyma.03G014700*, *Glyma.08G190100*, *Glyma.17G053200*, *Glyma.13G106200*, *Glyma.15G026100*, *Glyma.13G347900*, *Glyma.18G155900*, *Glyma.07G072900*, *Glyma.03G013200* and *Glyma.09G065600*), respectively, 66% of which were highly expressed in the roots and nodule. The results suggest the potential role of these genes in nodulation and symbiotic nitrogen fixation.

Next, we analyzed the expression of the *DDX* genes in root hair during rhizobial infection. The majority of these *DDX* genes were expressed at very low levels in uninfected or infected root hairs 12 h after inoculation (HAI); only seven and six genes were up-regulated and down-regulated, respectively (Figure 5B). In contrast, more than 50% of the *DDX* genes were up-regulated at 24 HAI; very few genes were down-regulated in infected root hair (Figure 5B). Interestingly, the expression levels of almost all the genes were greatly increased in uninfected root hair at 24 HAI (Figure 5B), suggesting that these genes may be involved in root hair maturation. Taken together, these data suggest that *DDX* genes play important roles in different processes of symbiotic nitrogen fixation.
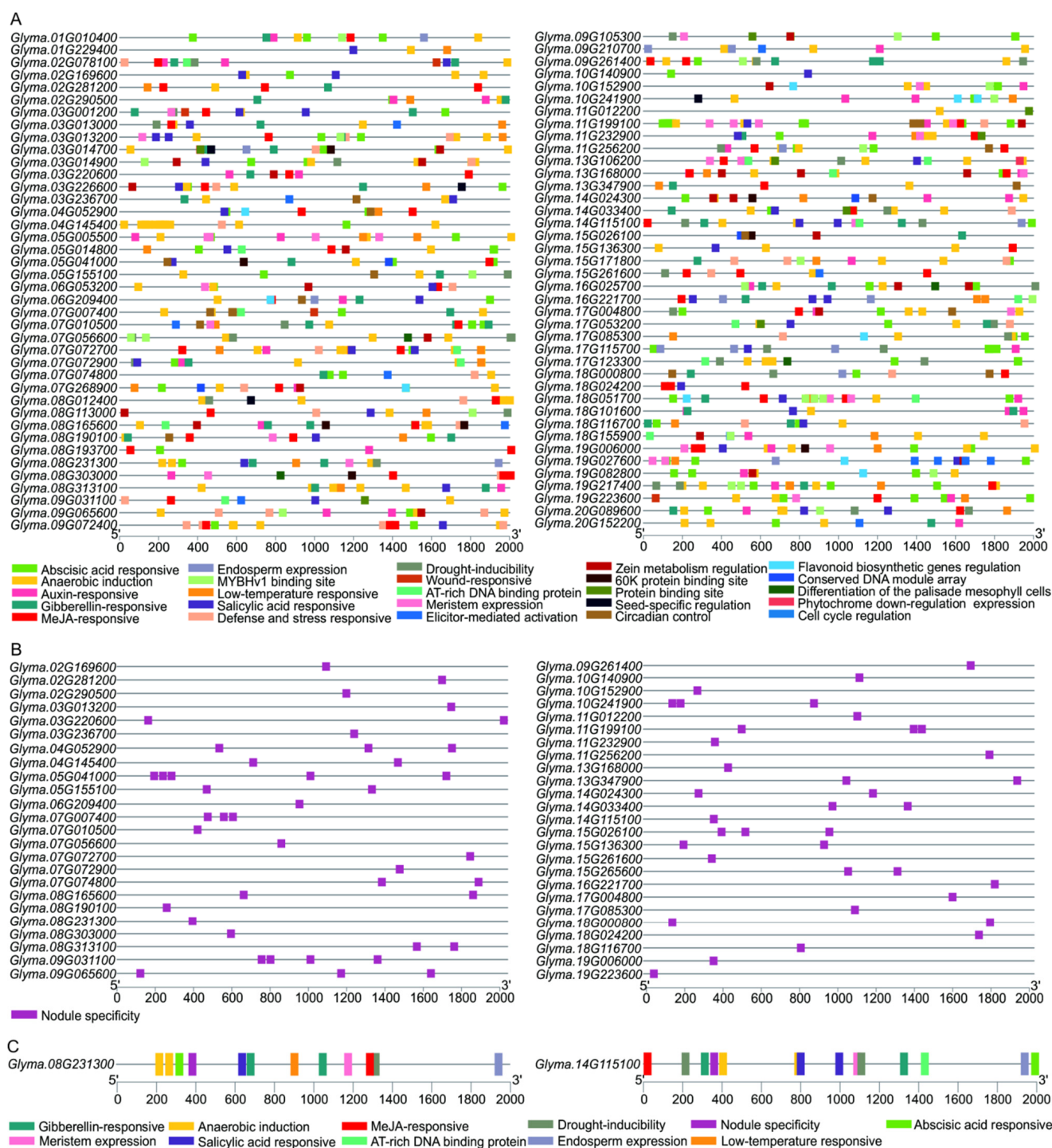
Figure 4. Predicted *cis* elements in the promoter regions of *DDX* genes. (**A**) Predicted *cis* elements of all *DDX* genes. (**B**) Predicted nodule-specific *cis* elements of 49 *DDX* genes. (**C**) Predicted *cis* elements of two soybean specific *DDX* genes. The scale bars at the bottom indicate the lengths of the promoter sequences. Color boxes indicate *cis* elements in the promoter regions.
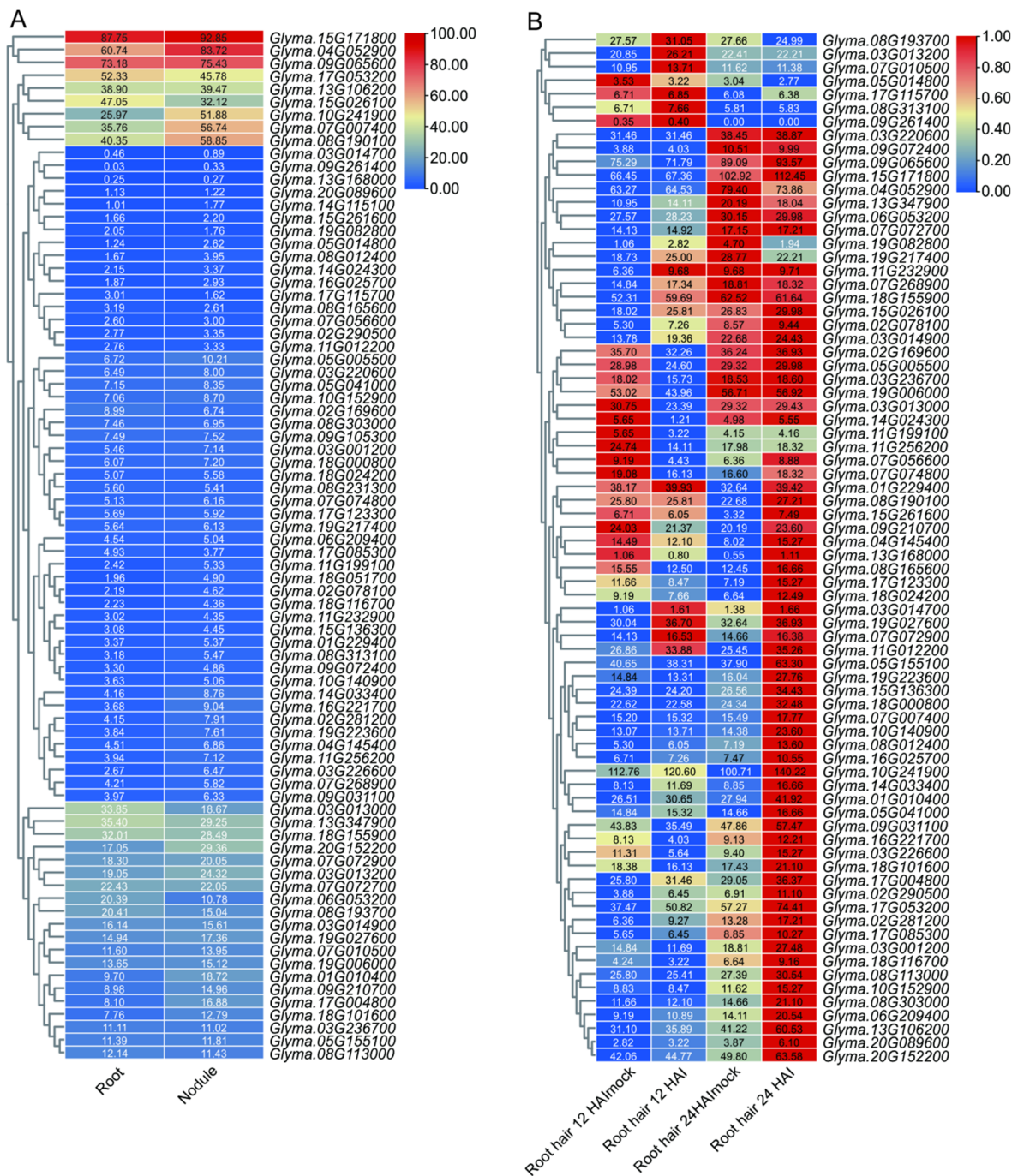
Figure 5. Expression analysis of *DDX* genes in the root and nodule, and their response to rhizobium. A heat map with clustering was created based on the RPKM value of 80 *DDX* genes of soybean. (**A**) Expression patterns of *DDX* genes in the root and nodule. (**B**) Expression patterns of *DDX* genes' response to rhizobium in root hair at 12 and 24 h after inoculation (HAI). The colored scale varies from green to red, which indicates low or high levels of gene expression. Root hair 12 HAImock/Root hair 24 HAImock and Root hair 12 HAI/Root hair 24 HAI indicate root hairs at 12 h/24 h after inoculation without and with rhizobia, respectively.

2.7. qPCR Analysis of DDX Gene Expression in Nodules and Root Hairs

To confirm the expression patterns of *DDX* genes, we performed qPCR assays to detect the expression of 18 highly expressed *DDX* genes (Figure S4) and three mildly expressed *DDX* genes (*Glyma.13G168000*, *Glyma.03G014700* and *Glyma.20G089600*) in roots

and mature nodules at 28 days after inoculation (DAI). The results showed that all the tested genes were expressed in the root and nodule, and that fifteen of them were expressed at higher levels in the root and mature nodules (Figures 6A and S4). Next, we analyzed the expression of the 13 *DDX* genes which were induced by rhizobia at 24 HAI in root hair (Figure S4) and the three marker genes of nodulation initiation (*GmNIN1a*, *GmENOD40-1* and *GmENOD40-2*). The qRT-PCR results showed that three marker genes were up-regulated, only six *DDX* genes were up-regulated, one *DDX* gene was unchanged and six *DDX* genes were down-regulated (Figure 6B). The inconsistent results between microarray data and our experimental analysis potentially resulted from plant culture and rhizobial infection artifacts.
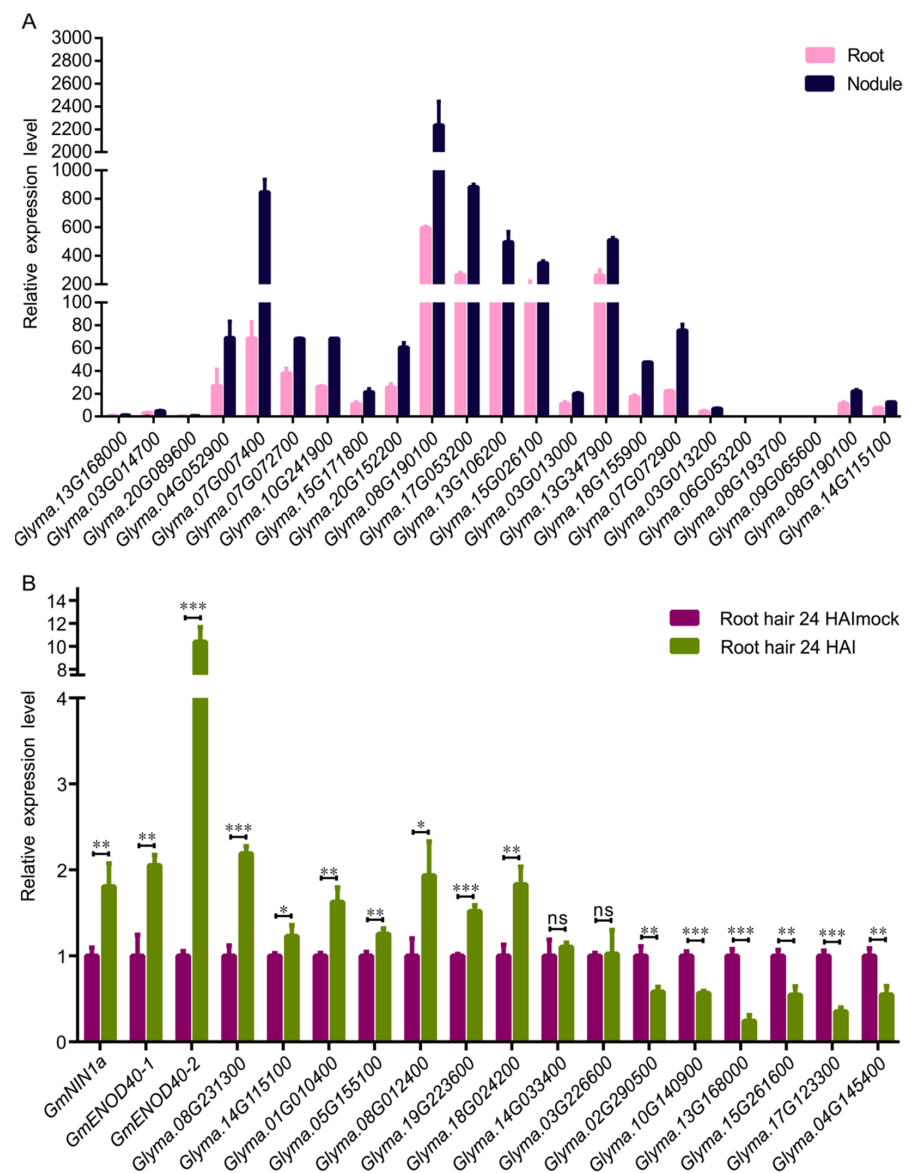

Figure 6. Expression profiles of *DDX* genes in the root and nodule, and their response to rhizobium. (**A**) Expression patterns of selected *DDX* genes in the root and nodule were examined by a qPCR assay. (**B**) Expression patterns of selected *DDX* genes in response to rhizobium in root hair were examined by a qPCR assay at 24 HAI. Data were the most representative of three biological replicates. The *GmELF1b* gene was used as an internal control. Root hair 24 HAImock and Root hair 24 HAI indicate root hairs at 24 h after inoculation without and with rhizobia, respectively. A Student's *t*-test tested the significance of the difference between two groups. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

To learn more about the two soybean specific *DDX* genes we also detected the expression of *Glyma.08G231300* and *Glyma.14G115100* in mature nodules and infected root hair. The expression patterns of *Glyma.08G231300* and *Glyma.14G115100* mRNAs were similar in both mature nodules and infected root hair (Figure 6). They were expressed at low levels in root nodules, but both genes were induced by rhizobial infection at 24 HAI. These results suggest that these two *DDX* genes may be involved in common processes contributing to nodulation and symbiotic nitrogen fixation.

3. Discussion

The DDX family of RNA helicases is widely present in all of life [36,37]. These highly conserved RNA helicases are necessary for processes from RNA transcription to degradation metabolism, and are therefore involved in almost all aspects of living organisms [6]. In the past, studies on the DDX family in plants have mainly focused on the dicot and monocot model plants, Arabidopsis and rice [38,39]. These DDX family proteins play an indispensable role in plant development, stress adaptation and plant–microbe interactions [12]. Soybean is one of the most important crops and has an important trait, plant–rhizobia symbiosis, but the characteristics of the *DDX* gene family and their potential functions in plant growth and symbiosis, in particular, are poorly understood. In this work, we presented a comprehensive analysis of the DDX family proteins in soybean, including gene classification, chromosomal locations, phylogenetic trees and expression profiles in different tissues as well as during symbiotic nodulation. Our results provide detailed information about DDX proteins and highlight the crucial roles of DDXs in soybean–rhizobia symbiosis.

The soybean genome is a partially diploidized tetraploid [40] and has more *DDX* genes than diploid species. In this study, we identified a total of 42 and 80 DDX RNA helicases in Arabidopsis and soybean, respectively, using stringent structural criteria and updated soybean genome information (Table S1). The marked difference in the number of DDX helicases between soybean and rice was also observed in a previous study. For example, they found 50 and 87 DDX RNA helicases in Arabidopsis and soybean, respectively [24]. However, the exact number of the DDX helicases in soybean and Arabidopsis is different between the mentioned study and ours. We speculate that this difference is mainly due to the criteria used for the identification of DDX RNA helicases. We identified the DEAD-box family members by 12 conserved motifs, while they mainly defined the DDX proteins based on the DEAD motif. The fact that the number of DDX RNA helicases almost doubled suggests that they may have evolved new functions and modulated new traits in soybean. Phylogenetic analysis reveals that 78 soybean DDX helicases have close relatives in Arabidopsis (Figure 2). Intriguingly, we found two soybean specific DDX RNA helicases with different gene and protein structures. Thus, these DDX RNA helicases may be essential not only for plant growth and environmental adaptation but also for soybean-specific traits, such as soybean–rhizobia symbiosis.

These 80 soybean *DDX* genes were separated into 13 subclades (Figure 3). The number of the DDX genes varied from two (subclades II and X) to 15 (subclade IX). These *DDX* genes contained different numbers of exons, varied lengths of introns and untranscribed regions. Interestingly, six genes in subclades III and IV contained only one exon, while others contained up to eighteen exons. Notably, the two soybean-specific genes have longer 5′ untranscribed regions than the others. The differences in gene structures may lead to different mRNA abundances and the proteins they produce. Indeed, the lengths of DDX family proteins vary greatly, ranging from 362 to 1201 amino acids. Importantly, they all contain 12 conserved ATP binding, ATP hydrolysis, nucleic acid binding and RNA helicase motifs of DDX family proteins all through the rest of the proteins, implicating the roles of these proteins in RNA metabolism. Highly divergent N-terminals, C-terminals and amino acid sequences between the conserved motifs of these DDX proteins may lead to different protein structures and functions during development and adaptation to their growth environment.

Indeed, we found that almost all the *DDX* promoters contained *cis*-regulatory elements related to plant development and responses to the environment, including light, phytohormones (e.g., abscisic acid, jasmonic acid) and abiotic stresses (e.g., temperature, drought, etc.). It is worth noting that the numbers of *cis*-regulatory elements related to abscisic acid and jasmonic acid phytohormones are the largest (Figure 4). It is well-known that abscisic acid and jasmonic acid are responsible for plant responses to abiotic and biotic stresses, respectively [41,42]. The large number of these *cis*-regulatory elements in soybean *DDX* gene promoters suggests crucial roles of the genes in plant adaptation to changing environmental conditions. Thus, we have identified potential novel candidate genes that regulate the stress tolerance of soybean through RNA metabolism. Considering the economic importance of soybean globally and continuing challenges to soybean cultivation due to global warming, the identification of stress-related genes is of great significance in facilitating basic research of soybean stress tolerance and in accelerating breeding for stress-tolerant soybean.

Soybean is able to associate with nitrogen-fixing rhizobia to form nodules where nitrogen is fixed for plant growth [26]. Symbiotic nitrogen fixation is determined by effective rhizobial infection, nodule organogenesis and the nitrogen fixation of mature nodules [27]. One can expect that these complex morphogenic and host–rhizobia interaction processes also undergo highly active RNA metabolism. Intriguingly, we provide evidence that many *DDX* genes are potentially involved in rhizobial infection and nodule development or functionality. Firstly, the promoters of 49 *DDX* genes have nodule-related *cis*-regulatory elements (Figure 4B); secondly, more than 18% of the *DDX* genes were expressed in the root and nodule in soybean (Figure 5); and finally, many genes were responsive to rhizobia infection in root hair, especially the two soybean-specific *DDX* genes (Figure 6). Our results implicate indispensable roles of these DDX RNA helicases in the process of soybean–rhizobia interactions and symbiotic nitrogen fixation. These DDX RNA helicases are a potential resource for the elucidation of the molecular mechanisms of symbiotic nodulation and for the genetic improvement of soybean nitrogen fixation efficiency.

In summary, we systemically analyzed the DDX family genes and proteins in soybean. Our findings lay a foundation for understanding RNA dynamics and functions in soybean development and environmental adaptation. To the best of our knowledge we revealed, for the first time, the crucial roles of DDX RNA helicases in symbiotic nitrogen fixation. The future functional characterization of these symbiosis related DDX RNA helicases will provide novel insights into the post-transcriptional regulation and genetic optimization of symbiotic nitrogen fixation in soybean.

4. Materials and Methods

4.1. The Identification of the DEAD-Box Genes in Arabidopsis and Soybean

To identify all members of the DDX family members of Arabidopsis and soybean, we searched for these family members in two ways, given that all DEAD-box proteins contain two conserved domains: DEXDc and HELICc. Firstly, DEXDc and HELICc were entered into the Domain Selection dialog of the SMART website (http://smart.embl-heidelberg.de/, accessed on 16 August 2021), and then Arabidopsis and soybean were entered into the Taxonomic selection dialog. We clicked on Architecture Query to obtain the results. Secondly, we clicked on BioMart under the Tools menu of the Phytozome website (https://phytozome-next.jgi.doe.gov/, accessed on 12 October 2021), selected crops and families under the SELECT MART sub-menu and clicked on Phytozome 12 Genomes. On the page, we clicked on attributes—we clicked to specify, selected Features and checked Gene Name, Gene Start and Gene End, as well as Transcript Name. We clicked filters—clicked to specify the species name, Arabidopsis and soybean, respectively. Then, we input PF00270 and PF00271 into the PROTEIN Limit to Genes with PFAM ID and clicked Results to output the results. The sequences of all the found amino acids were input into the online conservative motifs analysis MEME (https://meme-suite.org/meme/tools/meme, accessed on 18 October 2021) to retrieve the conserved motifs. All the amino acids were

analyzed one by one to confirm that they contained the 12 unique motifs of the DEAD-box family.

4.2. The Chromosomal Location of the DEAD-Box Genes

The GTF file of the soybean genome was downloaded from the Phytozome website. According to the GTF file and the gene IDs of the soybean DDX family members, the gene density and chromosome mapping maps were drawn by the TBtools software (College of Horticulture, South China Agricultural University, Guangzhou, China) [43].

4.3. Phylogenetic Trees of the DEAD-Box RNA Helicases

The amino acid sequences of all the soybean and Arabidopsis DEAD-box proteins were compared using the neighbor-joining (NJ) method of the MEGA7 program (Japan Department of Biological Sciences, Tokyo Metropolitan University, Tokyo, Japan) [44]. The images of the phylogenetic trees were drawn using MEGA7. Evolutionary tree files were imported from the Evoview online website (https://evolgenius.info//evolview-v2, accessed on 15 November 2021) on the evolutionary tree editor beautification.

4.4. Synteny Analysis of the DEAD-Box Genes in Soybean

The amino acid sequences of the main transcripts of *Arabidopsis* and soybean were compared with each other by the TBtools software to obtain comparison files. The synteny analysis was carried out by the MCscanX software (Department of Crop and Soil Sciences, University of Georgia, Athens, GA, USA) [45] to obtain the gene pairs of *DDX* family members with a synteny relationship between *Arabidopsis* and soybean; pictures were then drawn.

4.5. Expression Analyses of the DEAD-Box Genes in Soybean

The expression data of the *DDX* genes in soybean were obtained using the online website *Glycine max* eFP Browser (http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi, accessed on 18 November 2021). All the *DDX* gene expression data were then imported into TBtools (College of Horticulture, South China Agricultural University, Guangzhou, China) to draw the expression heat map. Next, the *DDX* genes, which are highly expressed in the roots, nodules and infected root hairs at 12 and 24 HAI were analyzed to draw an up-set Wayne map using TBtools.

4.6. Plant Materials and Growth Conditions

Glycine max var Williams 82 (Wm 82) was used as the plant material. Soybean seeds were germinated in sterile water for two days and then planted in solid media containing a low-nitrogen nutrient solution (25*25 disposable Petri dish). The pots were put in the plant growth room at 26 °C, with a light intensity of 200 μmol/m^2/s, under light for 16 h and in the dark for 8 h for 3 days, allowing seed germination. The germinating seedlings were inoculated with *Bradyrhizobium diazoefficens* rhizobia strain USDA110 (OD$_{600}$ = 0.08), suspended in sterile water. At 24 HAI the roots of inoculated and uninoculated plants were carefully removed and cleaned quickly in liquid nitrogen. We needed to quickly and gently brush the infected and uninfected root hair into a mortar. After the liquid nitrogen evaporated the samples were then quickly put into 2 mL centrifuge tubes and frozen in liquid nitrogen.

4.7. RNA Isolation and Real-Time Quantitative RT-PCR Expression Analysis

The total RNA of root materials was extracted using TRIpure Reagent (Aidlab Biotechnologies Co., Ltd., Beijing, China). A RT Supermix (No. R223-01, Vazyme Biotech Co., Ltd., Nanjing, China) kit was used for mRNA reverse transcription. The quantitative reagent for gene expression detection was Vazyme SYBR Green Mix (No. Q111-02, Vazyme Biotech Co., Ltd., Nanjing, China), and the quantitative PCR instrument was BIO-RAD CFX96/384 (Bio-Rad, Hercules, CA, USA). Each PCR reaction contained 10 μL 2 × STBR Green Mix,

0.5 µL of each primer, and appropriately diluted cDNA. The thermal cycling conditions were 95 °C for 5 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. The *GmELF1b* gene was used as an internal reference for all the qRT-PCR analyses. Each treatment was repeated three times independently. The primers used are described in Table S2.

5. Conclusions

In this work, we identified the DEAD-box family genes and proteins in soybean in a genome-wide manner. These *DDX* genes were widely distributed in nearly all chromosomes. The majority of the DDX proteins were highly homologous with Arabidopsis DDX proteins; two soybean specific DDX proteins also had evolved. These *DDX* genes contained multiple *cis*-regulatory elements in their promoters and were differentially expressed during development and in response to rhizobia. As a sessile organism, soybean has to deal with changing environments, low-nitrogen conditions in particular; it is conceivable that a large family of DDX helicases enables them to survive in their growth environments through the modulation of RNA metabolism from biogenesis to decay. This study lay a solid foundation for a further in-depth functional analysis of the *DDX* genes, and for deciphering their implications on soybean symbiotic nitrogen fixation and a wide range of agricultural traits.

References

1. Linder, P.; Jankowsky, E. From unwinding to clamping—The DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 505–516. [CrossRef] [PubMed]
2. Tanner, N.K.; Linder, P. DExD/H box RNA helicases: From generic motors to specific dissociation functions. *Mol. Cell* **2001**, *8*, 251–262. [CrossRef]
3. Fuller-Pace, F.V. DExD/H box RNA helicases: Multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.* **2006**, *15*, 4206–4215. [CrossRef]
4. Cruz, J.; Kressler, D.; Linder, P. Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem. Sci.* **1999**, *24*, 192–198. [CrossRef]
5. Fairman, M.E.; Maroney, P.A.; Wang, W.; Bowers, H.A.; Gollnick, P.; Nilsen, T.W.; Jankowsky, E. Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* **2004**, *304*, 730–734. [CrossRef] [PubMed]
6. Jankowsky, E. RNA helicases at work: Binding and rearranging. *Trends Biochem. Sci.* **2011**, *36*, 19–29. [CrossRef] [PubMed]

7.   Pryszlak, M.; Wiggans, M.; Chen, X.; Jaramillo, J.E.; Burns, S.E.; Richards, L.M.; Pugh, T.J.; Kaplan, D.R.; Huang, X.; Dirks, P.B. The DEAD-box helicase DDX56 is a conserved stemness regulator in normal and cancer stem cells. *Cell Rep.* **2021**, *34*, 108903. [CrossRef] [PubMed]

8.   Nawaz, G.; Sai, T.; Lee, K.; Su, J.P.; Dinh, S.N.; Kang, H. BrRH37, a cabbage (*Brassica rapa*) DEAD-Box RNA helicase, confers drought tolerance and ABA response in transgenic *Arabidopsis* plants. *J. Plant Biol.* **2021**, *64*, 327–336. [CrossRef]

9.   Nidumukkala, S.; Tayi, L.; Chittela, R.K.; Vudem, D.R.; Khareedu, V.R. DEAD box helicases as promising molecular tools for engineering abiotic stress tolerance in plants. *Crit. Rev. Biotechnol.* **2019**, *39*, 395–407. [CrossRef]

10. Parvatiyar, K.; Zhang, Z.; Teles, R.M.; Ouyang, S.; Jiang, Y.; Iyer, S.S.; Zaver, S.A.; Schenk, M.; Zeng, S.; Zhong, W.; et al. The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nat. Immunol.* **2012**, *13*, 1155–1162. [CrossRef]

11. Jiang, Y.; Zhu, Y.; Liu, Z.J.; Ouyang, S. The emerging roles of the DDX41 protein in immunity and diseases. *Protein Cell* **2017**, *8*, 83–89. [CrossRef]

12. Chen, D.; Peng, X.B. DEAD-box: Function on growth and development in *Arabidopsis thaliana*. *Plant Sci. J.* **2016**, *34*, 941–948. [CrossRef]

13. Gong, Z.; Dong, C.H.; Lee, H.; Zhu, J.H.; Xiong, L.M.; Gong, D.M.; Stevenson, B.; Zhu, J.K. A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in *Arabidopsis*. *Plant Cell* **2005**, *17*, 256–267. [CrossRef] [PubMed]

14. Huang, C.K.; Huang, L.F.; Huang, J.J.; Wu, S.J.; Ye, C.H.; Lu, C.A. A DEAD-box protein, AtRH36, is essential for female gametophyte development and is involved in rRNA biogenesis in *Arabidopsis*. *Plant Cell Physiol.* **2010**, *51*, 694–706. [CrossRef]

15. Gu, L.; Tao, X.; Lee, K.; Kang, H. A chloroplast-localized DEAD-box RNA helicase AtRH3 is essential for intron splicing and plays an important role in the growth and stress response in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **2014**, *82*, 309–318. [CrossRef]

16. Kim, J.S.; Kim, K.A.; Oh, T.R.; Park, C.M.; Kang, H. Functional characterization of DEAD-Box RNA helicases in *Arabidopsis thaliana* under abiotic stress conditions. *Plant Cell Physiol.* **2008**, *49*, 1563–1571. [CrossRef] [PubMed]

17. Huang, C.A.; Li, F.; Chun, K.; Shen, Y.L.; Yeh, S.J.; Chin, H. The DEAD-box RNA helicase AtRH7/PRH75 participates in pre-rRNA processing, plant development and cold tolerance in *Arabidopsis*. *Plant Cell Physiol.* **2016**, *57*, 174–191. [CrossRef] [PubMed]

18. Wu, C.Y.; Nagy, P.D. Blocking tombusvirus replication through the antiviral functions of DDX17-like RH30 DEAD-box helicase. *PLoS Pathog.* **2019**, *15*, e1007771. [CrossRef]

19. Li, D.Y.; Liu, H.; Zhang, H.J.; Wang, X.; Song, F.M. OsBIRH1, a DEAD-box RNA helicase with functions in modulating defence responses against pathogen infection and oxidative stress. *J. Exp. Bot.* **2008**, *59*, 2133–2146. [CrossRef] [PubMed]

20. Lu, C.A.; Huang, C.K.; Huang, W.S.; Huang, T.S.; Liu, H.Y.; Chen, Y.F. DEAD-box RNA helicase 42 plays a critical role in pre-mRNA splicing under cold stress. *Plant Physiol.* **2019**, *182*, 255–271. [CrossRef]

21. Philis, G.; Gracey, E.O.; Gansel, L.C.; Fet, A.M.; Rebours, C. Comparing the primary energy and phosphorus consumption of soybean and seaweed-based aquafeed proteins a material and substance flow analysis. *J. Clean Prod.* **2018**, *200*, 1142–1153. [CrossRef]

22. Zhan, J.; Twardowska, I.; Wang, S.; Wei, S.H.; Chen, Y.Q.; Ljupco, M. Prospective sustainable production of safe food for growing population based on the soybean (*Glycine max L. Merr*) crops under Cd soil contamination stress. *J. Clean Prod.* **2019**, *212*, 22–36. [CrossRef]

23. Collino, D.J.; Salvagiotti, F.; Perticari, A.; Piccinetti, C.; Ovando, G.; Urquiaga, S.; Racca, R.W. Biological nitrogen fixation in soybean in Argentina: Relationships with crop, soil, and meteorological factors. *Plant Soil* **2015**, *392*, 239–252. [CrossRef]

24. Cai, J.; Zhang, L.Y.; Liu, W.; Tian, Y.; Xiong, J.S.; Wang, Y.H.; Li, R.J.; Li, H.M.; Wen, J.; Mysore, K.S.; et al. Role of the Nod factor hydrolase MtNFH1 in regulating Nod factor levels during rhizobial infection and in mature nodules of *Medicago truncatula*. *Plant Cell* **2018**, *30*, 397–414. [CrossRef]

25. Malolepszy, A.; Kelly, S.; Srensen, K.K.; James, E.K.; Radutoiu, S. A plant chitinase controls cortical infection thread progression and nitrogen-fixing symbiosis. *eLife* **2018**, *7*, e38874. [CrossRef]

26. Wang, Z.J.; Wang, L.X.; Wang, Y.L.; Li, X. The NMN module conducts nodule number orchestra. *iScience* **2020**, *23*, 100825. [CrossRef]

27. Wang, L.X.; Sun, Z.X.; Su, C.; Wang, Y.L.; Yan, Q.Q.; Chen, J.H.; Ott, T.; Li, X. A GmNINa-miR172c-NNC1 regulatory network coordinates the nodulation and autoregulation of nodulation pathways in soybean. *Mol. Plant* **2019**, *12*, 1211–1226. [CrossRef] [PubMed]

28. Liu, H.; Zhang, C.; Yang, J.; Yu, N.; Wang, E.T. Hormone modulation of legume-rhizobial symbiosis. *J. Integr. Plant Biol.* **2018**, *8*, 632–648. [CrossRef]

29. Sheteiwy, M.S.; Fathi, D.; Xiong, Y.C.; Brestic, M.; Skalicky, M.; Hamoud, Y.A.; Ulhassan, Z.; Shaghaleh, H.; Elgawad, H.A.; Farooq, M.; et al. Physiological and biochemical responses of soybean plants inoculated with Arbuscular mycorrhizal fungi and Bradyrhizobium under drought stress. *BMC Plant Biol.* **2021**, *21*, 195–216. [CrossRef] [PubMed]

30. Sheteiwy, M.S.; Elgawad, H.A.; Xiong, Y.C.; Macovei, A.; Brestic, M.; Skalicky, M.; Shaghaleh, H.; Hamoud, Y.A.; El-Sawah, A.M. Inoculation with bacillus amyloliquefaciens and mycorrhiza confers tolerance to drought stress and improve seed yield and quality of soybean plant. *Physiol. Plant.* **2021**, *172*, 2153–2169. [CrossRef]

31. Zhang, M.; Liu, S.L.; Wang, Z.; Yuan, Y.Q.; Zhang, Z.F.; Liang, Q.J.; Yang, X.; Duan, Z.B.; Liu, Y.C.; Kong, F.J.; et al. Progress in soybean functional genomics over the past decade. *Plant Biotechnol. J.* **2021**, *20*, 256–282. [CrossRef] [PubMed]

32. Xu, R.; Zhang, S.; Huang, J.; Zheng, C.C. Correction: Genome-Wide comparative in silico analysis of the RNA helicase gene family in *Zea mays* and *Glycine max*: A Comparison with *Arabidopsis* and *Oryza sativa*. *PLoS ONE* **2013**, *8*, e78982. [CrossRef] [PubMed]

33. Chung, E.; Cho, C.W.; Yun, B.H.; Choi, H.K.; So, H.A.; Lee, S.W.; Lee, J.H. Molecular cloning and characterization of the soybean DEAD-box RNA helicase gene induced by low temperature and high salinity stress. *Gene* **2009**, *443*, 91–99. [CrossRef] [PubMed]

34. Yano, K.; Yoshida, S.; Müller, J.; Singh, S.; Banba, M.; Vickers, K.; Markmann, K.; White, C.; Schuller, B.; Sato, S.; et al. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20540–20545. [CrossRef] [PubMed]

35. Hirsch, S.; Kim, J.; Muñoz, A.; Heckmann, A.B.; Downie, J.A.; Oldroyd, G.E. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* **2009**, *21*, 545–557. [CrossRef]

36. Hashimoto, Y.; Lawrence, P.O. Comparative analysis of selected genes from *diachasmimorpha longicaudata* entomopoxvirus and other poxviruses. *J. Insect Physiol.* **2005**, *51*, 207–220. [CrossRef]

37. Fischer, M.G.; Allen, M.J.; Wilson, W.H.; Suttle, C.A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19508–19513. [CrossRef]

38. Dong, W.; Qin, B.; Li, X.; Ding, T.; Zhang, Y.; Cheng, Z.; Xue, Y.; Lin, H.X. Nucleolar DEAD-Box RNA helicase togr1 regulates thermotolerant growth as a pre-rRNA chaperone in rice. *PLoS Genet.* **2016**, *12*, e1005844. [CrossRef]

39. Ghazala, N.; Hunseung, K. Chloroplast- or mitochondria-targeted DEAD-Box RNA helicases play essential roles in organellar RNA metabolism and abiotic stress responses. *Front. Plant Sci.* **2017**, *8*, 871–880. [CrossRef]

40. Shoemaker, R.C.; Guffy, R.D.; Lorenzen, L.L.; Specht, J.E. Molecular genetic mapping of soybean: Map utilization. *Crop Sci.* **1992**, *32*, 1091–1098. [CrossRef]

41. Mauch-Mani, B.; Flors, V. The ATAF1 transcription factor: At the convergence point of ABA-dependent plant defense against biotic and abiotic stresses. *Cell Res.* **2009**, *19*, 1322–1323. [CrossRef] [PubMed]

42. Ra, C.; Je, M. Jasmonic acid distribution and action in plants-regulation during development and response to biotic and abiotic stress. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 4114–4119. [CrossRef]

43. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]

44. Sudhir, K.; Glen, S.; Koichiro, T. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]

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