Biochemical Genetics (2019) 57:487–506
https://doi.org/10.1007/s10528-019-09903-7

ORIGINAL ARTICLE

Genome-Wide Identification and Characterization of DIR Genes in *Medicago truncatula*

Min Song[1] · **Xiangyong Peng**[1]

Received: 6 August 2018 / Accepted: 4 January 2019 / Published online: 16 January 2019

Abstract

Dirigent proteins (DIRs) are critically involved in the formation of lignans, a diverse and widely distributed class of secondary plant metabolites exhibiting interesting pharmacological activities and implicated in natural plant defense. However, no detailed information is available about DIR gene family in *Medicago truncatula*. In this study, a total of 45 DIR genes were identified in *M. truncatula*. DIR proteins have variability in sequence. Most MtDIR genes have no intron. All MtDIR proteins contain single dirigent domain. A large number of MtDIR genes were expanded via gene duplication, and 37 MtDIR genes were duplicated in tandem. Digital expression data showed that 40% MtDIR genes had a higher expression level in the root. Analysis of RNA-seq and microarray data indicated that more than 30% MtDIR genes were responsive to biotic and/or abiotic treatments. This study will facilitate further studies on DIR family and provide useful clues for functional validation of DIR genes in higher plants.

Keywords *Medicago truncatula* · Dirigent domain · Gene family · Expression pattern

Introduction

The DIR protein containing the dirigent domain was first found in *Forsythia intermedia* (Davin et al. 1997). The word Dirigent comes from Latin dirigere, which means guide. The presence of the DIR protein has recently been found in mosses, ferns, gymnosperms, and angiosperms (Ralph et al. 2006, 2007; Wu et al. 2009; Li

Research area: Plant molecular genetics.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10528-019-09903-7) contains supplementary material, which is available to authorized users.

✉ Min Song
smin163@163.com

1 College of Life Science, Qufu Normal University, Qufu 273165, People's Republic of China

et al. 2014), but there is no report of DIR protein existing in primitive aquatic plants. The DIR gene emerged as a result of the evolution of vascular plants in terrestrial processes (Shi et al. 2012). As DIR protein is involved in the synthesis of lignin and lignans, it has attracted interest in research community in response to pests and certain abiotic stress induction (Ralph et al. 2006, 2007; Davin et al. 1997; Xia et al. 2000; Kim et al. 2002a; Hosmani et al. 2013).

Ralph et al. (2006, 2007) identified the DIR gene of 35 North American spruce (*Picea sitchensis*) and constructed a phylogenetic tree of 150 proteins including rice, *Arabidopsis thaliana*, barley, wheat, sugarcane, and pea. The DIR family is divided into six subfamilies: DIR-a, DIR-b/d, DIR-c, DIR-e, DIR-f, and DIR-g. The sequence similarity between these subfamilies is very low, and biochemical experiments show that only the DIR-a family of proteins can direct the correct stereo structure of lignin synthesis, and the biochemical function of other subfamily proteins is unknown, which are referred to as DIR-like proteins (Davin et al. 1997; Xia et al. 2000; Kim et al. 2002b).

The protein encoded by the DIR gene is known to have a signal sequence associated with the secretory pathway (Davin and Lewis 2000). Pickel et al. (2012) modeled the tertiary structure of Arabidopsis AOC proteins with high homology to AtDIR6 by homology modeling. The results showed that the DIR protein consisted of eight antiparallel chains of β-barrels, forming a hydrophobic cavity, which bound the substrate. It suggests that the DIR protein may have evolved from a hydrophobic bond-binding protein. Each subunit of the homodimer captures a substrate-free radical, directs them to react correctly, and forms an optically pure coupling product.

Lignin, surrounded by tracheal cells such as tracheids, ducts, wood fibers, and thick-walled cells, is a polymer that is randomly polymerized from highly substituted phenylpropane units. Lignin is filled in the cellulose framework, which enhances the mechanical strength of the plant, facilitates transport of tissue-transported water, resists invasion by undesirable external environments, and is resistant to pests and diseases (Boudet 2000; Zhou et al. 2009). Studies have shown that the DIR protein plays a role in the synthesis of lignin, which directs the coupling of lignan monomers to form dimers. In this process, the DIR protein recognizes specific substrates and controls the stereoselectivity of the coupling, resulting in the preferential production of optically active products (Davin et al. 1997; Burlat et al. 2001). DIR itself has no catalytic activity. *Pdh1*, which contains a dirigent domain, was highly expressed in the lignin-rich inner sclerenchyma of pod walls, especially at the stage of initiation in lignin deposition. *Pdh1* promotes pod dehiscence by increasing the torsion of dried pod walls, which serves as a driving force for pod dehiscence under low humidity (Funatsuki et al. 2014). Overexpression of *GhDIR1* can increase the content of lignin in transgenic cotton and increase the degree of lignification of epidermis and vascular bundles (Shi et al. 2012).

The DIR protein is also involved in the synthesis of lignans. Lignans are widely distributed plant secondary metabolites, have antibacterial properties, participate in plant disease resistance, constitutive or induced expression (Davin et al. 1997; Lewis and Davin 1999; Burlat et al. 2001). Studies have shown that *Forsythia suspensa*'s DIR protein directs the coupling of E-coniferyl alcohol to produce lignans (Davin et al. 1997). The mediation of stereoselective coupling with DIR proteins has also been reported in moco cotton (Liu et al. 2008) and Arabidopsis (Pickel et al. 2010; Kim et al. 2012). It was shown that TaDIR13 could effectively direct coniferyl alcohol coupling into (+)-pinoresinol. Overexpression of TaDIR13 in transgenic tobacco increased total lignan accumulation (Ma and Liu 2015). The GmDIR22 recombinant protein purified from Escherichia coli could effectively direct E-coniferyl alcohol coupling into lignan (+)-pinoresinol. Accordingly, the overexpression of GmDIR22 in transgenic soybean increased total lignan accumulation (Li et al. 2017).

In addition, Hosmani et al. (2013) reported the key role of a dirigent domain-containing protein, ESB1, in the construction of the Casparian strip in root epidermal cells. The Casparian strip are on the upper, lower, and radial walls of the cell and have a thickened structure of lignification and suppository, which wraps around the cells in a band and controls nutrients and moisture into the vascular column. ESB1 guides the Casparian domain complex to be located in the cell wall forming the Casparian strip. The lack of ESB1 results in the failure to form normal Casparian strip.

More than a dozen DIR genes in different plant species have been functionally characterized, and most are associated with disease resistance, insect damage, abiotic stress tolerance, or even responses to multiple stresses. For example, *PsDIR1* in pea (Daniels et al. 1987; Hadwiger et al. 1992), *GhDIR1* (Shi et al. 2012), six DIR genes in pine trees (Ralph et al. 2006), *BrDIR2* and four homologous genes in *Brassica rapa* (Thamil et al. 2013), *Hfr1*, *TaDIR13* and Dirigent domain-containing JRL genes in wheat (Subramanyam et al. 2006, 2008; Song et al. 2013; Ma and Liu 2015), *GmDIR22* in soybean (Li et al. 2017). *BhDIR1* of *Boea hygrometrica* (Wu et al. 2009), *Tadir* gene in *Tamarix androssowii* (Gao et al. 2010), *ScDir* in *Saccharum* spp. (Guo et al. 2012).

Some of the DIR genes also respond to the induction of disease-associated hormones, such as *SHDIR11* and *SHDIR16* expressions in sugarcane induced by SA, JA, and MeJA (Damaj et al. 2010). In the *Brassica rapa*, 12 DIR genes responded to ABA (Thamil Arasan et al. 2013).

Understanding the functions of the dirigent genes in plant biological and physiological processes would be a possible practicable approach to analyze and improve the defense response of crops against biotic and abiotic stresses. However, no study has been conducted on the DIR gene family in *Medicago truncatula* in detail until now. Hence, in the present study, we performed a genome-wide analysis of the DIR family in *M. truncatula*, an excellent legume model plant, based on complete genome sequence and annotation about several aspects, including genome composition, structure, and phylogenetic analyses, chromosomal localization, and gene duplication analysis. Meanwhile, we also investigated their expression patterns by public data, RNA-seq and microarray data, and explored their functions in plant development and response to stresses. These findings would be valuable for understanding DIR function and promoting their utilization in legumes' genetic improvement.

Materials and Methods

Genome-Wide Identification of *M. truncatula* Dirigent genes

Two methods were used to search Dirigent genes in *M. truncatula.* First, Dirigent (PF03018) hidden Markov model seed files were downloaded from the Pfam database (https://pfam.sanger.ac.uk/) and as a query sequence using HmmerWeb version 2.14.0 (https://www.ebi.ac.uk/Tools/hmmer/) searched for protein databases, parameter defaults. Secondly, all 26 Arabidopsis Dirigent protein sequences were downloaded from The Arabidopsis Information Resource (TAIR), and used as BLAST queries for NCBI BLAST 2.3.0+ against the Mt4.0v1 proteome in Phytozome12 (https://phytozome.jgi.doe.gov/), parameter defaults (Altschul et al. 1990; Lamesch et al. 2012). Redundant sequences were discarded manually and check the conserved domain of the protein using database of CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) in automatic mode (threshold = 0.01, maximum hits = 500).

Analysis of Gene Encoding Protein Features

The relative molecular mass and isoelectric point information of each protein was obtained through the ProtParm program (Gasteiger et al. 2005). Genomic sequences and CDS sequences were downloaded from the Ensembl database (https://plants.ensembl.org/), and the gene structure was analyzed using GSDS software (https://gsds.cbi.pku.edu.cn/, Guo et al. 2007). WoLF PSORT II (https://www.genscript.com/wolf-psort.html) was used to predict the subcellular locations (Horton et al. 2007). N-glycosylation sites (Asn) of the MtDIRs were searched online using NetNGlyc 1.0 server (https://www.cbs.dtu.dk/services/NetNGlyc/, Gupta et al. 2004).

We generated an alignment of all identified DIR and DIR-like amino acid sequences using the alignment program Clustal W (Larkin et al. 2007). Neighbor joining (NJ) was used for cluster analysis, and a phylogenetic tree was constructed using MEGA 5.1 with Model set to p-distance, the Bootstrap value was set to 1000, and the rest of the parameters were default (Tamura et al. 2011).

Chromosomal Localization and Gene Duplication Analysis of the DIR Genes

Positional information about DIR genes was obtained from the *M. truncatula* Genome Browser (https://www.medicagogenome.org/), and diagrams of their chromosome locations in *M. truncatula* were drawn using MapChart (Voorrips 2002). Homologous gene pairs were defined as having protein similarity of more than 70% and coverage greater than 75%. If two homologous genes were separated by five or fewer genes, they were identified as tandem duplications (TD), while if two genes were separated by more than five genes or distributed in different chromosomes, they were referred to as segmental duplications (SD). In addition, duplications between the DIR genes were also identified and complemented based on public data in the Plant Genome Duplication Database (https://chibba.agtec.uga.edu/duplication/; Lee et al. 2013).

Sequence Similarity

Sequence similarity between DIR protein sequences in *A. thaliana* and *M. truncatula* was visualized using Circoletto, a web interface for comparing two sequence libraries via Circos (Darzentas 2010). *M. truncatula* genes were used as the query against a database of *A. thaliana* DIR genes. An E-value of 1e−10 was used, and only the best match between the subject (MtDIR) and query (AtDIR) sequences was considered, with all local alignments per best-hit diagrammed.

All protein sequences for Arabidopsis were downloaded from the TAIR v10 (Lamesch et al. 2012), and converted to a BLAST-compatible protein database using the makeblastdb function of NCBI BLAST+2.2.28 (Altschul et al. 1990; Lamesch et al. 2012). This process was repeated using the Mt4.0v1 protein sequences downloaded from JCVI (Tang et al. 2014). BLASTp analysis was performed using the MtDIR and AtDIR protein sequences as queries against their respective protein database, and best-hits were collected based first on E-value, and in the event of ties, by bit score. Best-hits between the two families were then compared, and reciprocal best-hits were recorded as potential orthologous genes.

Motif Analysis and Domain Identification of MtDIR Proteins

MtDIR protein sequences were analyzed for novel domains using the Multiple EM for Motif Elicitation (MEME) web server (Bailey et al. 2009). The motif sites set at 10 and other default. The TBtool software shows the MEME results (Chen et al. 2018).

Cis-Acting Element Prediction in the Promoters of MtDIRs

The 1500bp sequences of the MtDIRs transcription start site (ATG) were downloaded from the *M. truncatula* Genome Browser (https://www.medicagogenome.org/) using the online tool PLANTCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al. 2002) to predict cis-elements.

Expression Analysis of the DIR Genes

The expression levels of DIR genes from *M. truncatula* in different tissues, including root, nodule, blade, bud, seedpod, and flower tissues, were downloaded from the *Medicago truncatula* Gene Expression Atlas Project (MtGEA, https://

mtgea.noble.org/v3/, Tang et al. 2009). Genome-wide transcriptome data from *M. truncatula* under different stresses, including cold, freezing, drought, salt and ABA, were downloaded from the NCBI short read archive database (SRA database; https://www.ncbi.nlm.nih.gov, Accession numbers SRX1056987–92, Shu et al. 2016), and gene expression was analyzed using TranslatomeDB (https://www.translatomedb.net/, Liu et al. 2018). Heatmaps were drawn with ClustVis tool (https://biit.cs.ut.ee/clustvis/, Metsalu and Vilo 2015).

Results

Genome-Wide Identification and Structural Analysis of DIR Genes in *M. truncatula*

Using PF03018 as a query, we identified 45 DIR genes in *M. truncatula*. Most of these genes contained one complete dirigent domain. We named them based on their distribution and linear order on the respective chromosomes (Table 1). The peptide length of MtDIRs varied ranging from 99 amino acids (MtDIR27) to 389 amino acids (MtDIR32). Nearly 35 (77.8%) MtDIRs were less than 200 amino acids. The predicted PI values ranged from 4.48 to 10.00 and MW ranged from 11,161.26 to 40,573.47 Da (Table 1). There were 43 MtDIRs containing Asn sites and 35 of them had signal peptides, indicating that these MtDIRs are probably secreted proteins. More than half MtDIRs were predicted final localization in the chloroplast. The rest located in the extra cellular region, vacuolar, nucleus, and cytoplasmic (Table S1). Sequences of MtDIR CDSs and proteins are shown in Text S1.

Bioinformatics Analysis, Exon/Intron Structure Determination, and Phylogenetic Tree Construction

To better understand the similarities and differences in MtDIRs between *M. truncatula* and other plants, an unrooted phylogenetic tree was generated using 165 DIR and DIR-like protein sequences from *M. truncatula,* Arabidopsis, rice, and poplar. DIRs were clearly separated into six distinct groups (Fig. 1). MtDIRs mainly were clustered in group 6 (26 MtDIRs) and group 3 (13 MtDIRs). DIRs in groups 2 and 5 were all from rice. Groups 1, 3 and 6 contain DIRs from 4 species. According to the classification of Ralph et al. (2007), group 1 belongs to the DIR-a subfamily, which includes MtDIR41/42, 7 rice, 5 Arabidopsis, and 9 poplar DIR proteins. It is speculated that the proteins of this group are all or partially involved in the synthesis of lignin. To elucidate the topological structures and evolutionary relationships of the MtDIRs, multiple sequence alignments of amino acid sequences were used to build a Neighbor-Joining (NJ) tree (Fig. S1A). The exon/intron structure showed that most MtDIR genes (95.6%) were intronless while two members (MtDIR18 and MtDIR43) contained just one intron (Fig. S1B).

Table 1 Name, ID, and various features of DIR genes in *M. truncatula*

Gene symbol	Gene ID	CDS	Peptide	Number of exons	Molecular weight/Da	Theoretical PI
MtDIR1	Medtr1g046490	711	236	1	24492.86	5.06
MtDIR2	Medtr1g046500	753	250	1	25879.73	5.28
MtDIR3	Medtr1g046800	753	250	1	25868.75	5.28
MtDIR4	Medtr1g054525	579	192	1	20983.02	9.56
MtDIR5	Medtr1g056370	561	186	1	20590.54	6.65
MtDIR6	Medtr1g115510	1062	353	1	37259.66	4.71
MtDIR7	Medtr1g115515	711	236	1	25942.59	6.3
MtDIR8	Medtr3g034030	561	186	1	20788.61	6.13
MtDIR9	Medtr3g105630	570	189	1	20941.87	8.91
MtDIR10	Medtr3g105640	567	188	1	20878.88	8.93
MtDIR11	Medtr4g013310	594	197	1	22441.68	9.1
MtDIR12	Medtr4g013320	600	199	1	22506.96	9.28
MtDIR13	Medtr4g013325	600	199	1	22361.66	9.06
MtDIR14	Medtr4g013330	600	199	1	22315.61	8.89
MtDIR15	Medtr4g013335	591	196	1	21880.16	8.94
MtDIR16	Medtr4g013345	591	196	1	21898.16	9.06
MtDIR17	Medtr4g013350	584	197	1	22015.22	8.89
MtDIR18	Medtr4g013355	543	180	2	19970.7	6.72
MtDIR19	Medtr4g013385	576	191	1	21382.41	6.83
MtDIR20	Medtr4g013770	579	192	1	21341.4	8.83
MtDIR21	Medtr4g049550	540	179	1	19175.76	8.09
MtDIR22	Medtr4g049570	537	178	1	19338.14	8.12
MtDIR23	Medtr4g062460	774	257	1	28487.36	6.15
MtDIR24	Medtr4g062470	915	304	1	32427.15	4.62
MtDIR25	Medtr4g062520	912	303	1	32422.17	4.62
MtDIR26	Medtr4g073950	579	192	1	21064.29	8.58
MtDIR27	Medtr4g074020	300	99	1	11161.26	10
MtDIR28	Medtr4g078885	561	186	1	20788.57	6.13
MtDIR29	Medtr4g122110	543	180	1	19993.81	5.5
MtDIR30	Medtr4g122130	543	180	1	20047.88	5.73
MtDIR31	Medtr5g096120	570	189	1	21110.85	6.03
MtDIR32	Medtr7g021300	1170	389	1	40573.47	4.48
MtDIR33	Medtr7g070390	579	192	1	21597.93	9.27
MtDIR34	Medtr7g093790	576	191	1	20851.97	5.14
MtDIR35	Medtr7g093820	585	194	1	21412.67	5.51
MtDIR36	Medtr7g093830	582	193	1	21372.73	6.19
MtDIR37	Medtr7g093850	585	194	1	21461.79	5.92
MtDIR38	Medtr7g093870	657	218	1	23950.49	6.15
MtDIR39	Medtr8g073770	573	190	1	20752.62	9.46
MtDIR40	Medtr8g073850	573	190	1	20789.68	9.46
MtDIR41	Medtr8g099115	570	189	1	21489.46	8.91
MtDIR42	Medtr8g099135	558	185	1	20624.75	6.95

Table 1 (continued)

Gene symbol	Gene ID	CDS	Peptide	Number of exons	Molecular weight/Da	Theoretical PI
MtDIR43	Medtr8g106405	513	170	2	19045.48	6.04
MtDIR44	Medtr8g106450	543	180	1	19963.7	6.29
MtDIR45	Medtr0433s0040	576	191	1	21382.41	6.83

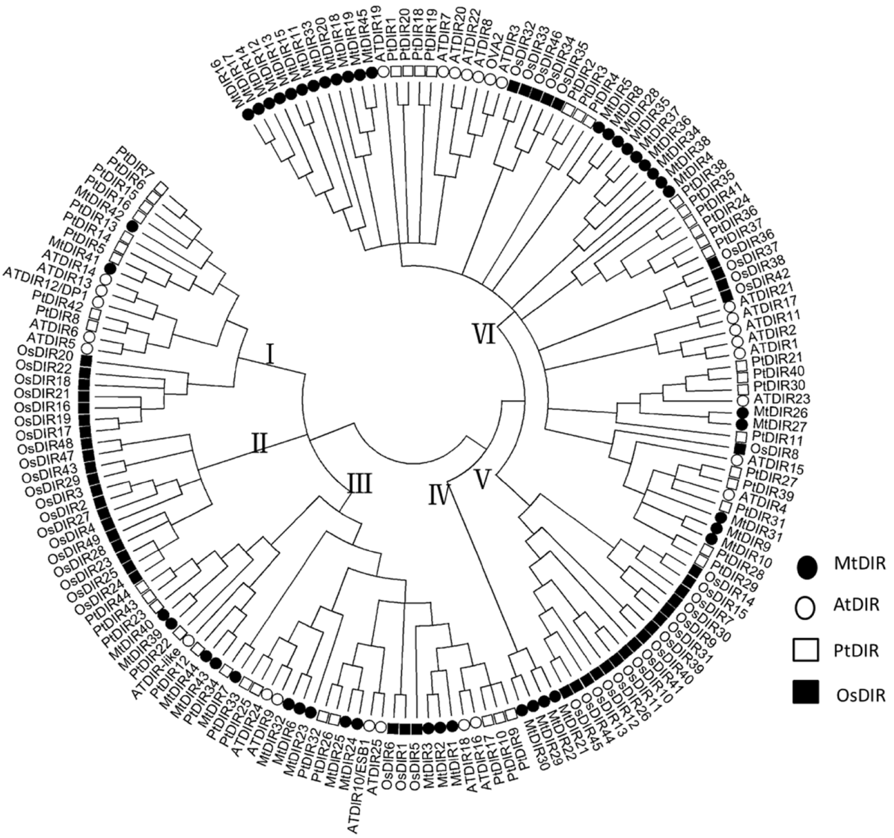


Fig. 1 Phylogenetic relationships of MtDIR genes. Phylogenetic analysis of the dirigent domain sequences of the DIR proteins in *M. truncatula*, rice, poplar, and Arabidopsis. The tree was constructed using MEGA5.1 by the neighbor-joining method with 1000 bootstrap replicates

DIR Gene Distribution on the Chromosome

Genome chromosomal location analysis revealed that MtDIRs are non-randomly located on chromosomes 1, 3, 4, 5, 7, and 8, respectively (Fig. 2). Among these chromosomes, chromosome 4 contained the largest number of 20 DIR genes followed by chromosomes 1 and 7, with 7 genes distributed, and only one occurred on chromosome 5. One DIR gene was found to have no chromosomal location information. The MtDIR genes were arranged in clusters on chromosomes. A total of 35 (77.8%) of the MtDIR genes belong to the tandem repeats. Chromosome 4 contains 10 DIR genes (MtDIR11–20) in a short chromosome region (~ 105 kb), and chromosomes 1, 3, 7, and 8 contain similar gene clusters.

Motif Analysis of MtDIRs

The MEME suite was used to identify motifs in each of the 45 MtDIR proteins as shown in Fig. 3. Each motif sequence was shown in Table S2. Motifs 1–4 were found to be similar to the dirigent domain region. Most DIRs contained 3–7 motifs. Not less than 40 DIRs contained motif 1, motif 2, and motif 3. Proteins in the same group contained similar motifs, while the motifs were divergent among different groups. For example, DIRs in group VI contained motifs 1–5 except MtDIR18 and MtDIR27, which lack motif 4 and only contained motif 3, respectively. Similarly, motif 8 and motif 10 were shared in group III. The differential motifs that DIRs contain suggest that the DIRs may have diverse biochemical features and biological functions.

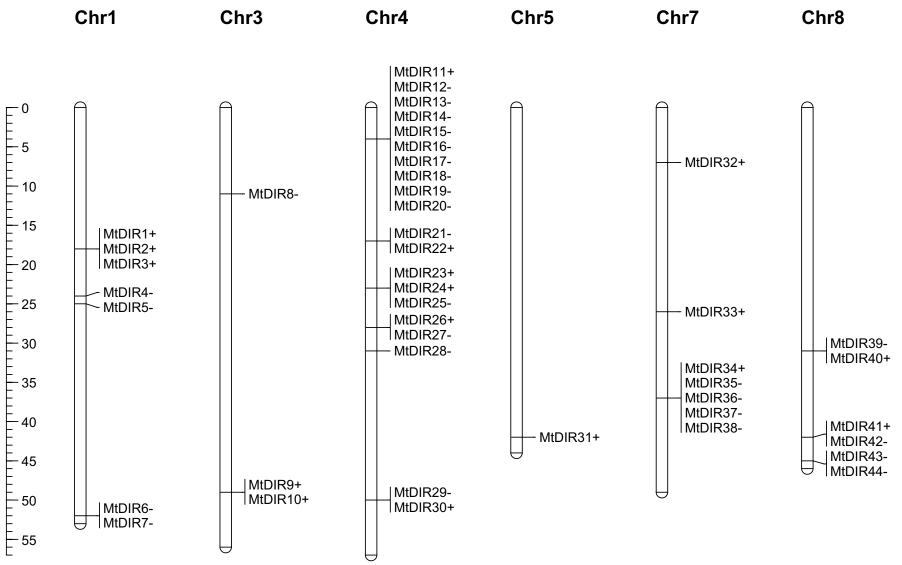


Fig. 2 Chromosomal location map of the 44 MtDIR genes. Arrow before a gene name indicates the coding polarity of the gene

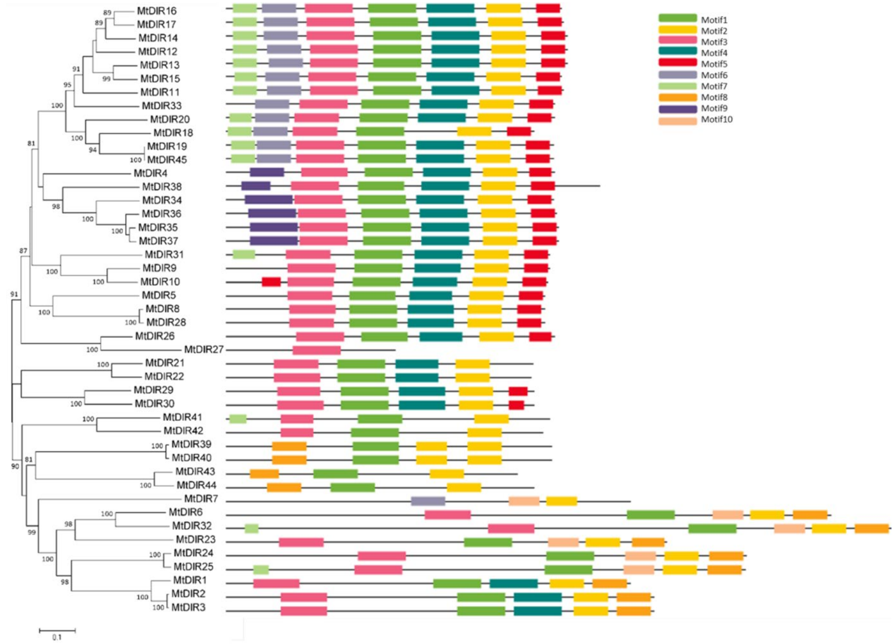


Fig. 3 Phylogenetic tree (left) and motif distributions (right) of the MtDIR proteins

Relationships Between AtDIRs and MtDIRs

Circoletto was used to identify and visualize the *A. thaliana* DIR protein sequences and their closest MtDIR orthologs. AtDIR2 produced the highest number of best-matches with six of the MtDIR sequences (Fig. 4). Reciprocal best-hit analysis was also performed to identify orthologous DIRs between both the species. With AtDIRs as queries and by performing two-way BLAST, a total of 8 reciprocal best blast hits were identified between *M. truncatula* and *A. thaliana*, which were distributed on chromosomes 1, 3, 4, and 8 (Table 2). These results correspond to the bands generated by Circoletto (Fig. 4).

The Promoter of the DIR Genes

The cis-acting elements of the predicted DIR genes were subsequently predicted by searching the promoter sequences from the PlantCARE database. As shown in Fig. 5, the DIR promoter contained a large number of light response elements, tissue-specific regulatory elements, hormone responsive elements, stress responsive elements, and other cis-acting elements associated with plant growth and development in addition to TATA-box, CAAT-box, and unknown function elements. Element number was from 18 (MtDIR44) to 58 (MtDIR39), and type number was from 15 (MtDIR44) to 31 (MtDIR37). All 45 MtDIR genes contained tissue-specific

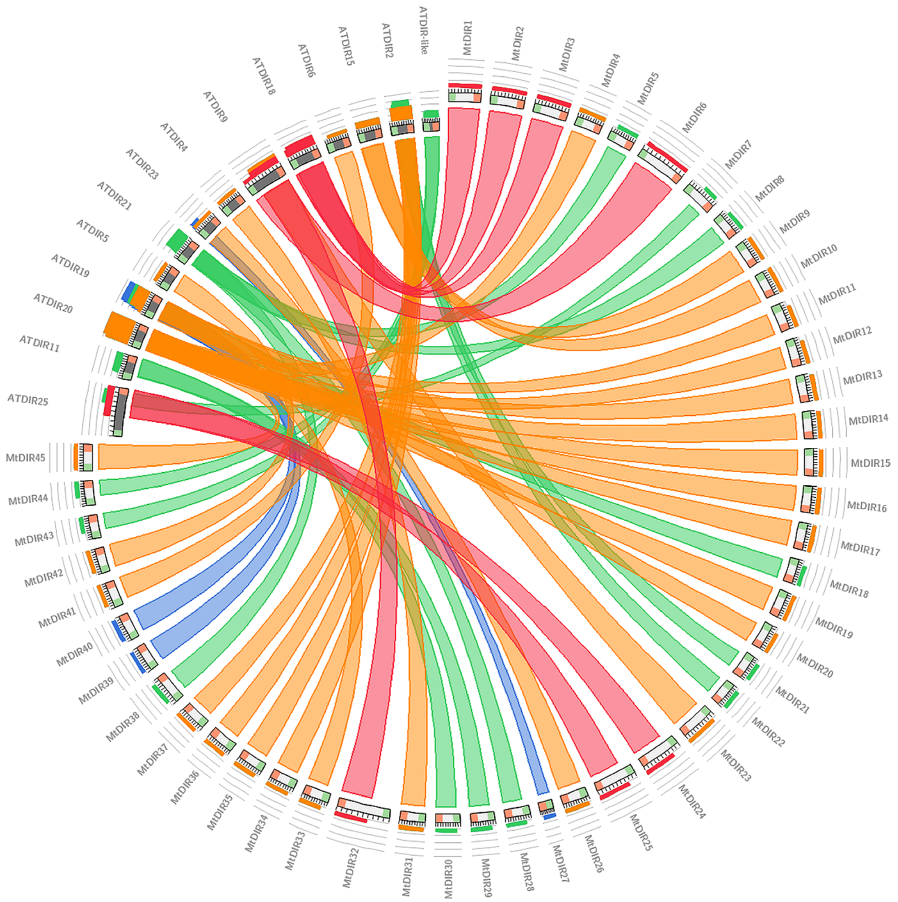


Fig. 4 Circoletto radial diagram linking the *M. truncatula* and *A. thaliana* DIR orthologs with ribbons. Colors of the ribbons are relative to the best BLAST alignment score, with matches within 25% of the best match as red, within 50% as orange, and within 75% as green. White (MtDIR) and black (AtDIR) bands on the periphery of the diagram represent the protein sequences, with the start and end of the sequence shown as green and red blocks, respectively. The terminal ribbon width corresponds to the portion of sequence aligned by BLAST (shown as sequence band) (Color figure online)

regulatory elements, light responsive elements, and stress responsive elements. All except MtDIR20 had hormone responsive elements. The names and functions of each elements included in the type were shown in Table S3.

Tissue Distribution of DIR Gene Transcripts

We investigated the expression profiles of MtDIR genes in various tissues using RNA-seq data from NCBI, including root, nodule, blade, bud, seedpod, and flower tissues. Thirty-seven MtDIR genes were expressed and FPKM>1 in at least one of

Table 2 Reciprocated best-hits between the *M. truncatula* and *A. thaliana* DIR genes using BLASTp analysis and their respective protein databases

MtDIR	Highest expressed tissue	ATDIR	Highest expressed tissue
MtDIR3	Root	ATDIR18	Root
MtDIR6	Root	ATDIR9	Root
MtDIR10	Flower	ATDIR15	Low expression levels in all organs
MtDIR16	Root	ATDIR20	Root
MtDIR24	Root	ATDIR25	Root
MtDIR26	Root	ATDIR23	Root
MtDIR41	Seedpod	ATDIR6	Root
MtDIR42	Seedpod	ATDIR5	Root

The tissue with the highest expression values of individual AtDIRs as determined by the eFP browser tool (Winter et al. 2007) and Genevestigator (Hruz et al. 2008)

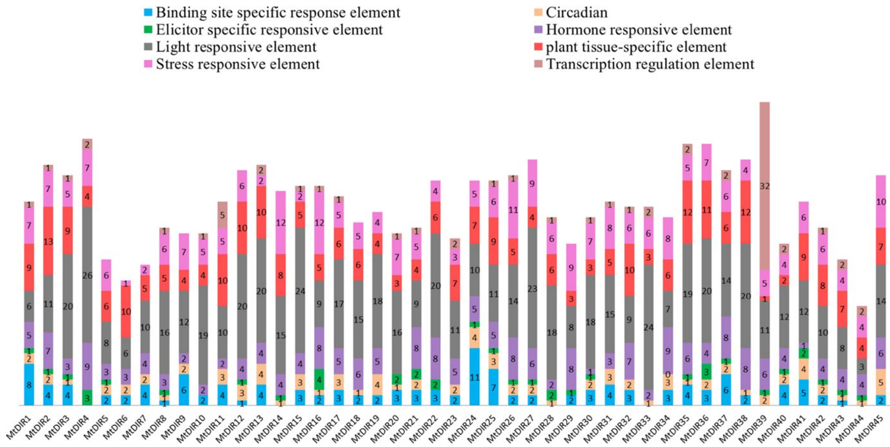


Fig. 5 Cis-acting elements in the promoter regions of MtDIR genes

six tissues, except *MtDIR18, -22*, *-27*, *-36*, *-40*, *-43, -44* and *-45*. Of these, 33 genes were expressed in nodule tissue, 30 genes in root tissue, 28 genes in flower tissue and 21, 23, and 24 genes in blade tissue, seedpod tissue, and bud, respectively (see Table S4). The heatmap also revealed that the majority of MtDIRs showed preferential expression and were clustered into five groups (Fig. 6). Among these MtDIR genes, group A (4 genes) were highly expressed in nodules, group B (18 genes) were highly expressed in root tissues and less in other tissues, group C (2 genes) and group D (6 genes) highly expressed in flowers and seedpods, respectively. Some genes exhibited tissue-specific profiles. For example, *MtDIR10* was major expressed in flower, *MtDIR28* and *MtDIR31* in nodules, *MtDIR25* and *MtDIR35* in root, and *MtDIR21* and *MtDIR11* in seedpods.

The expression patterns of the paralogous pairs were also revealed by line chart (Fig. 7), paralogous pairs with high sequence similarity had similar expression patterns. The best examples of this include *MtDIR1/2*, *MtDIR6/7*, and *MtDIR23/24*, which were strongly expressed in roots. *MtDIR11/33* and *MtDIR41/42* were strongly expressed in seedpods, with little or no expression in other tissues. Expression divergence was also found in paralogous pairs. For example, *MtDIR9* was highly expressed in seedpods, while its paralog, *MtDIR10*, was highly expressed in flowers. In addition, there were five pairs of orthologous between MtDIRs and AtDIRs with highest expression level in root (Table 2).

Expression Responses of MtDIR Genes to Abiotic Stress

To investigate the molecular functions of MtDIR genes in response to abiotic stress, we analyzed RNA-seq data from the NCBI SRA database (Accession numbers: SRX1056987–92) to detect the expression levels of DIR genes under different stresses, including cold, freezing, drought, salt, and ABA. The expressions of 29 MtDIR genes were detected and RPKM more than one in at least one library were further analyzed.

Based on expressional profiles of MtDIR genes' responses to abiotic stresses, they were clustered into two groups, as shown in Fig. 8. Compared to the control library, group A genes, including *MtDIR1*, *-2*, *-6*, *-7*, *-24*, *-25*, *-29*, *-32*, and *-39*, were expressed fewer under cold, freezing, salt and ABA treatments. While most MtDIR genes in group B were highly expressed in response to cold and/or freezing stress.

We used publicly available data of microarrays generated from *M. truncatula* Gene Expression Atlas (https://mtgea.noble.org/v2) to analyze expression of 19 genes under exposure to salt that were represented by at least one probe in the MtGEA (Fig. S2). These genes responded differently to 180 mM NaCl exposure. Profiling of the differentially expressed gene revealed that five genes (*MtDIR9*, *-34*, *-35*, *-37,* and *-42*) were downregulated after salt treatment for 6 h and upregulated after 24 h and 48 h. While four genes (*MtDIR4*, *-8*, *-29*, and *-33*) showed higher transcriptional expression after 6 h of salt exposure, whereas they displayed a relatively lower expression after 24 h and 48 h of salt exposure.

Discussion

DIR proteins are very important in lignan and lignin metabolisms, and plant-stress responses. The continual accumulation of data regarding whole–genome sequences and RNA-seq has created numerous opportunities for exploration of the dirigent gene family in related plants. In this study, 45 DIR genes were identified from *M. truncatula*. Domain composition of MtDIRs is simple, and all MtDIR proteins only contain one dirigent domain. As of March 2017, a total of 2163 proteins containing dirigent domain were collected from the Pfam database, of which 2016 proteins (93.2%) contained only one dirigent domain, 25 and 3 proteins contained two and

Fig. 6 Heatmap of expression profiles for MtDIR genes across different tissues. RNA-seq relative expression data from 6 tissues were used to reconstruct the expression patterns of DIR genes. The raw data were normalized and retrieved from the online database (https://mtgea.noble.org/v2/). Heatmap was drawn by ClustVis tool (https://biit.cs.ut.ee/clustvis/). Rows were centered, and unit variance scaling was applied to rows. Rows were clustered using Euclidean distance and average linkage

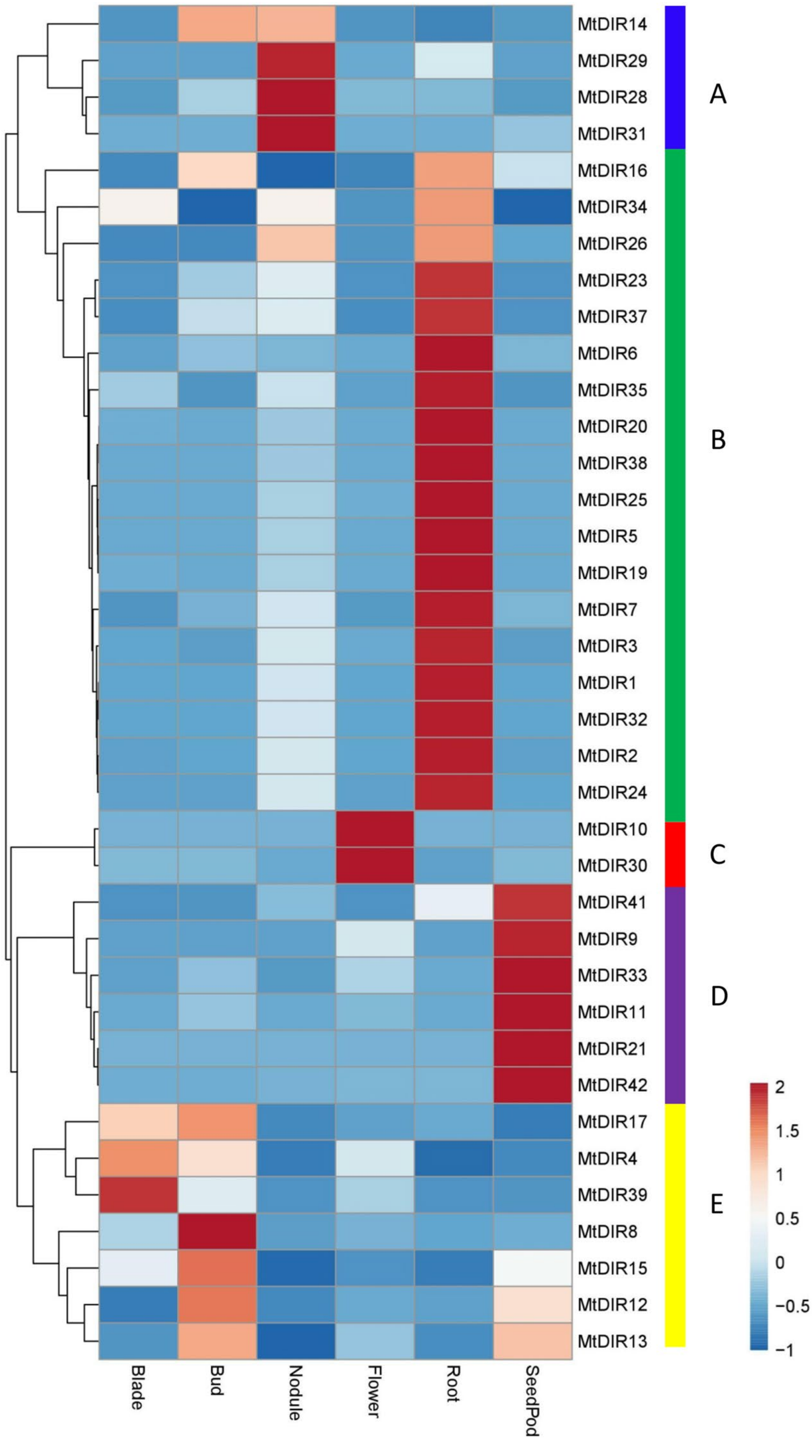
three dirigent domains in tandem, respectively. The rest of DIR proteins contained other domains. Thus, a single dirigent domain is the major domain composition of the DIR proteins.

The DIR gene structures of *M. truncatula* identified in this study and the previously reported DIR genes of *A. thaliana*, poplar, and pepper, are simple, and most of them have no introns (Khan et al. 2018). Howeve, one third of rice genes have 1–5 introns (Liao et al. 2017). This suggests that rice and dicotyledonous *M. truncatula*, poplar, and Arabidopsis may have different evolutionary patterns after differentiation.

In the expansion of a gene family, tandem duplication, segmental duplication, and transposition are the main evolutionary mechanisms (Cannon et al. 2004). There were 77.8% MtDIR genes clustered in a tandem manner (Fig. 3). These percentages are 55.1%, 40%, and 16.7% in rice, Arabidopsis, and pepper, respectively. On the other hand, 0.08% rice DIR genes, 0.04% MtDIR genes, and 62.5% pepper DIR genes belong to the fragment replication region, but no such genes were present in Arabidopsis (Ralph et al. 2007; Liao et al. 2017; Khan et al. 2018). These results suggest that the expansion of the DIR gene family differs among the different species. Tandem duplications played an important role in the DIR family expansion in *M. truncatula*, rice, and Arabidopsis, whereas, segmental duplication appeared to be a major factor contributing to DIR gene expansion in pepper.

The DIR genes that have been demonstrated by biochemical experiments to guide the formation of stereoselective conformations of lignin belong to the DIR-a subfamily, such as the cloned genes in *F. intermedia*, *Podophyllum peltatum*, and *Thuja plicata* (Davin et al. 1997; Xia et al. 2000; Kim et al. 2002b; Ralph et al. 2007). However, in recent years, it has also been found that some members of the DIR-b/d subfamily are involved in lignin or lignan metabolism, such as GmDIR22 and GhDIR1 (Shi et al. 2012; Li et al. 2017; Cheng et al. 2018). Therefore, two MtDIR genes of DIR-a subfamily and 26 MtDIR genes of DIR-b/d subfamily should be tested first for their functions involved in lignin or lignan metabolism in the future.

The DIR family genes play a role in specific tissues at certain stages of plant growth and development. The zonal DIR gene promoter was fused to the GUS reporter gene and transformed into Arabidopsis, resulting in the detection of GUS activity in vascular bundles (Kim et al. 2002a). Burlat et al. (2001) reported a Forsythia DIR gene expressed in the root, stem, and petiole. More than half of the raped DIR genes are specifically expressed in roots and/or flower buds (Thamil et al. 2013). The *GhDIR1* gene in cotton is mainly expressed in hypocotyls, whereas *GhDIR2* is mainly expressed in cotton fibers (Shi et al. 2012). Arabidopsis *ESB1* (AT2G28670) gene is involved in the formation of the Kjeldahl band in the root (Hosmani et al. 2013). This study also demonstrated that 40% MtDIR genes had a higher expression level in the root. In general, the DIR genes of *M.*


MtDIR14
MtDIR29
MtDIR28
MtDIR31
MtDIR16
MtDIR34
MtDIR26
MtDIR23
MtDIR37
MtDIR6
MtDIR35
MtDIR20
MtDIR38
MtDIR25
MtDIR5
MtDIR19
MtDIR7
MtDIR3
MtDIR1
MtDIR32
MtDIR2
MtDIR24
MtDIR10
MtDIR30
MtDIR41
MtDIR9
MtDIR33
MtDIR11
MtDIR21
MtDIR42
MtDIR17
MtDIR4
MtDIR39
MtDIR8
MtDIR15
MtDIR12
MtDIR13
A
B
C
D
E
2
1.5
1
0.5
0
−0.5
−1
Blade
Bud
Nodule
Flower
Root
SeedPod

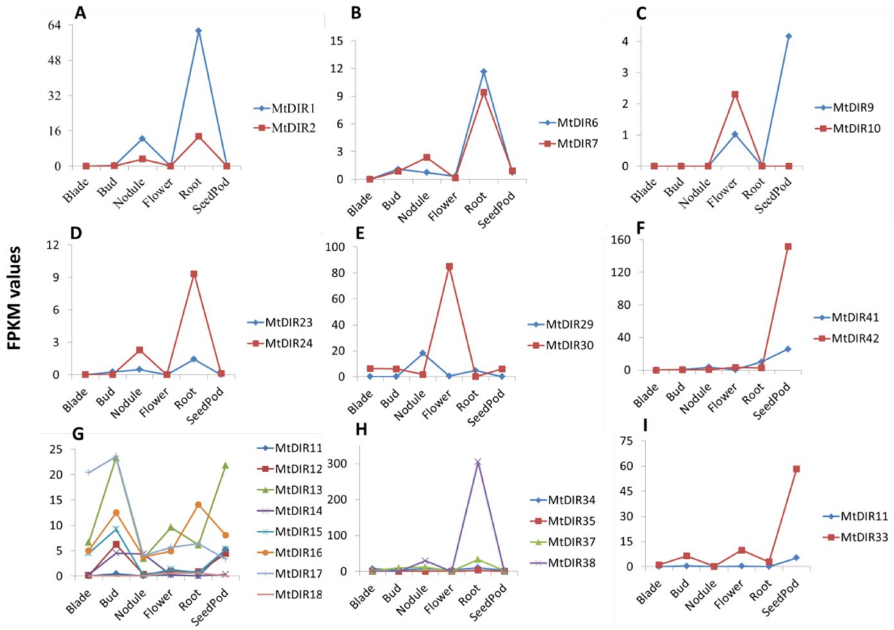


Fig. 7 The comparative analysis of expression profiles of duplicated MtDIR genes. X-axis represents different tissues of *M. truncatula*. Y-axis shows the expression values (FPKM) obtained using RNA-seq data

truncatula, same as in rice, Arabidopsis, and poplars, have low expression level in leaves.

Previous reports have shown that the DIR genes were involved in abiotic stress. *BhDIR1* of *B. hygrometrica* responded to various abiotic stresses, including drought, $CaCl_2$, ABA, H_2O_2, and EGTA (Wu et al. 2009). *T. androssowii Tadir* gene was upregulated under salt and saline stress (Gao et al. 2010). *Saccharum* spp. *ScDir* expression was increased when seedlings were exposed to hydrogen peroxide, PEG and salt stress (Guo et al. 2012). In the *B. rapa*, 11 DIR genes responded to flooding, 12 responded to ABA and 12 responded to cold stress (Arasan et al. 2013). In this study, RNA-seq data analysis showed that 30% of the MtDIR gene responded to stress. Chip data analysis showed that 38% of the MtDIR gene responded to abiotic stress.

Our results corroborate the former studies on the analysis of the cis-elements, and elements related to stress and light were found to in upstream region, indicating that stress and light may play a regulatory role in MtDIRs (Fig. 5). In addition, elements responsive to SA and MeJA have been identified in upstream region of more than 20 MtDIRs (Supplementary Table S3). Interestingly, Gibberellin responsive elements were found in 25 DIR genes. All results indicate that the patterns of MtDIR's responses to hormones are quite complex. It is speculated that different DIRs play roles in different time periods in a diversity environment.

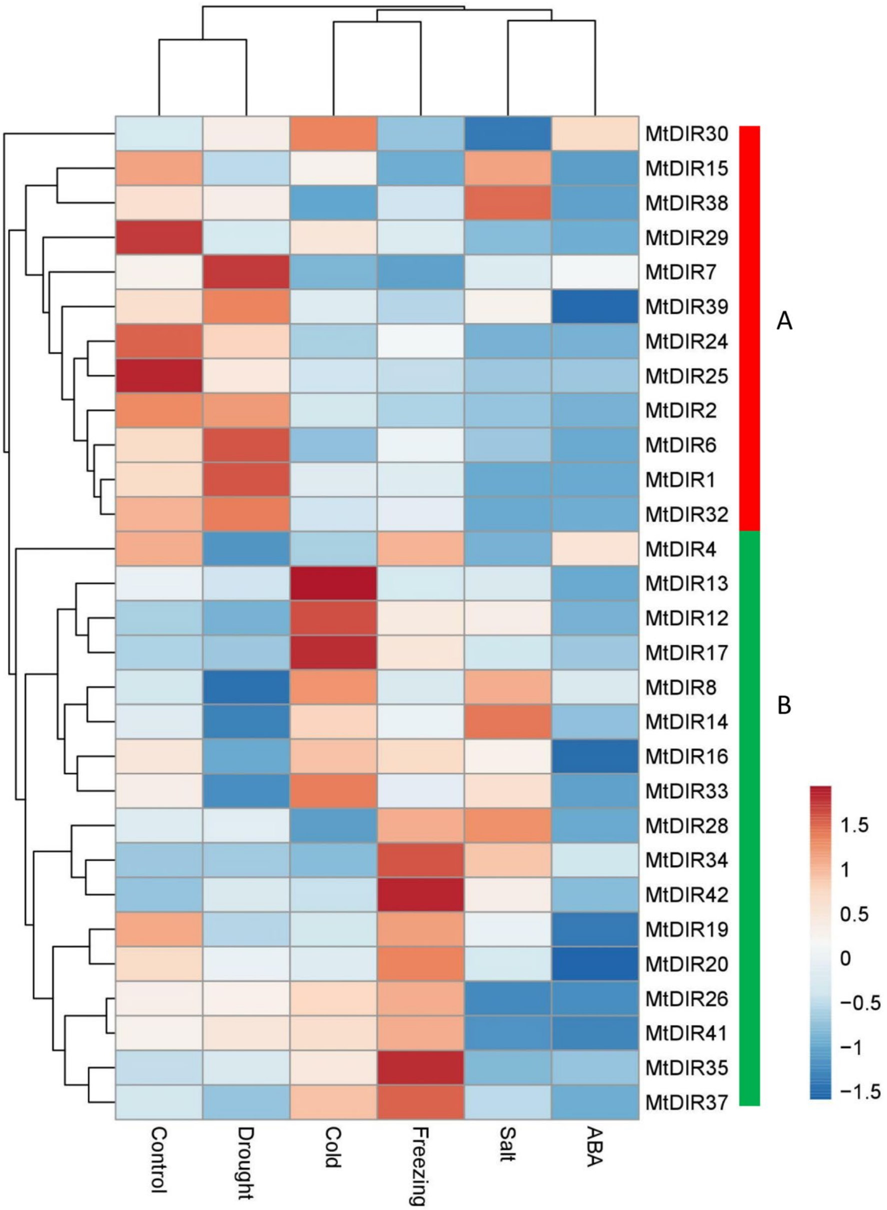


Fig. 8 Expression profile cluster analysis of MtDIR genes' responses to abiotic stresses. The genome-wide transcriptome data were from the NCBI SRA database (Accession numbers: SRX1056987–92). The 8-week-old *M. truncatula* cv. Jemalong A17 seedlings were treated. For cold-stress and freezing-stress treatments, the seedlings were transferred into at 4 ℃ or − 8 ℃, respectively. For drought-stress or salt-stress treatment, the seedlings were treated with 300 mM mannitol or 200 mM NaCl solution, respectively. For ABA treatment, the seedlings were sprayed with 100 μM ABA solution. Control (untreated) and treated seedlings were harvested at 3 h after treatment. Heatmap was drawn by ClustVis tool. Rows were centered, and unit variance scaling was applied to rows. Both rows and columns were clustered using Euclidean distance and average linkage

Author Contributions MS designed the experiments and wrote the paper. XYP analyzed the data. All authors read and approved the final manuscript.

Funding This work was supported by the Science and Technology project of Shandong Education Department to Min Song (Grant No. J15LE02), and China Postdoctoral Science Foundation funded project to Min Song (Grant No. 2018M632646).

Compliance with Ethical Standards

Conflict of interest Min Song declares that she does not have conflict of interest. Xiangyong Peng declares that he does not have conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent This article does not contain any studies with human participants.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L (2009) Meme suite: tools for motif discovery and searching. Nucl Acids Res 37:W202–W208

Boudet AM (2000) Lignins and lignification: selected issues. Plant Physiol Biochem 38:81–96

Burlat V, Kwon M, Davin LB, Lewis NG (2001) Dirigent proteins and dirigent sites in lignifying tissues. Phytochemistry 57:883–897

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4(1):10

Chen CJ, Xia R, Chen H, He YH (2018) TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. bioRxiv. doi: 10.1101/289660

Cheng X, Su X, Muhammad A, Li M, Zhang J, Sun Y et al (2018) Molecular characterization, evolution, and expression profiling of the dirigent (DIR) family genes in Chinese White Pear (*Pyrus bretschneideri*). Front Genet 9:136

Damaj MB, Kumpatla SP, Emani C, Beremand PD, Avutu S, Reddy AS et al (2010) Sugarcane DIRIGENT and O-METHYLTRANSFERASE promoters confer stem-regulated gene expression in diverse monocots. Planta 231:1439–1458

Daniels CH, Fristensky B, Wagoner W, Hadwiger LA (1987) Pea genes associated with non-host disease resistance to *Fusarium* are also active in race-specific disease resistance to *Pseudomonas*. Plant Mol Biol 8:309–316

Darzentas N (2010) Circoletto: visualizing sequence similarity with circos. Bioinformatics 26(20):2620

Davin LB, Lewis NG (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. Plant Physiol 123:453–461

Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, Sarkanen S, Lewis NG (1997) Stereoselective bimolecular phenoxy radical coupling by anauxiliary (dirigent) protein without an active center. Science 275:362–366

Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M et al (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc Natl Acad Sci USA 111(50):17797

Gao CQ, Liu GF, Wang YC, Jiang J, Yang CP (2010) Cloning and analysis of dirigent-like protein in gene from *Tamarix androssowii*. Bull Bot Res 30(1):81–86

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR., Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy Server. In: Walker JM (eds) The proteomics protocols handbook. Humana Press 571-607

Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. Yi Chuan 29:1023–1026

Guo JL, Xu LP, Su YC, Fu HY, Que YX, Xu JS (2012) A novel dirigent protein gene with highly stem-specific expression from sugarcane, response to drought, salt and oxidative stresses. Plant Cell Rep 31(10):1801–1812

Gupta, R., Jung, E. Brunak, S (2004) Prediction of N-glycosylation sites in human proteins. In preparation

Hadwiger LA, Chiang CC, Horovitz D (1992) Expression of disease resistance response genes in near-isogenic pea cultivars following challenge by *Fusarium oxysporum* race 1. Physiol Mol Plant P 40: 259-269

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucl Acids Res 35:585–587

Hosmani PS, Kamiya T, Danku J, Naseer S, Geldner N, Guerinot ML (2013) Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. Proc Natl Acad Sci USA 110:14498–14503

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, et al. (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics 2008:420747

Khan A, Li RJ, Sun JT, Ma F, Zhang HX,Jin JH, et al. (2018) Genome-wide analysis of dirigent gene family in pepper (*Capsicum annuum* L.) and characterization of CaDIR7 in biotic and abiotic stresses. Sci Rep 8:5500

Kim MK, Jeon JH, Davin LB, Lewis NG (2002a) Monolignol radical-radical coupling networks in western red cedar and Arabidopsis and their evolutionary implications. Phytochemistry 61(3):311–322

Kim MK, Jeon J-H, Fujita M, Davin LB, Lewis NG (2002b) The western red cedar (*Thuja plicata*) 8–8′ DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. Plant Mol Biol 49:199–214

Kim KW, Moinuddin SGA, Atwell KM, Costa MA, Davin LB, Lewis NG (2012) Opposite stereoselectivities of dirigent proteins in Arabidopsis and *Schizandra* species. J Biol Chem 287:33957–33972

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R et al (2012) The Arabidopsis information resource (tair): improved gene annotation and new tools. Nucl Acids Res 40:1202–1210

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41:1152–1158

Lewis NG, Davin LB (1999) Lignans: biosynthesis and function. In: Barton DHR, Nakanishi K, Meth-Cohn O (eds) Comprehensive natural products chemistry. Elsevier, London, pp 639–712

Li Q, Chen J, Xiao Y, Di P, Zhang L, Chen W (2014) The dirigent multigene family in *Isatis indigotica*: gene discovery and differential transcript abundance. BMC Genom 15:388

Li N, Zhao M, Liu T, Dong L, Cheng Q, Wu J et al (2017) A novel soybean dirigent gene GmDIR22 contributes to promotion of lignan biosynthesis and enhances resistance to *Phytophthora sojae*. Front Plant Sci 8:1185

Liao Y, Liu S, Jiang Y, Hu C, Zhang X, Cao X et al (2017) Genome-wide analysis and environmental response profiling of dirigent family genes in rice (*Oryza sativa*). Genes Genom 39(1):47–62

Liu J, Stipanovic RD, Bell AA, Puckhaber LS, Magill CW (2008) Stereoselective coupling of hemigossypol to form (+)-gossypol in moco cotton is mediated by a dirigent protein. Phytochemistry 69(18):3038–3042

Liu W, Xiang L, Zheng T, Jin J, Zhang G (2018) TranslatomeDB: a comprehensive database and cloud-based analysis platform for translatome sequencing data. Nucl Acids Res 46:D206–D212

Ma QH, Liu YC (2015) Tadir13, a dirigent protein from wheat, promotes lignan biosynthesis and enhances pathogen resistance. Plant Mol Biol Rep 33(1):143–152

Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucl Acids Res 43:W566–W570

Pickel B, Constantin MA, Pfannstiel J, Beifuss U, Schaller DA (2010) An enantiocomplementary dirigent protein for the enantioselective laccase-catalyzed oxidative coupling of phenols. Angew Chem Int Ed Engl 49(1):202–204

Pickel B, Pfannstiel J, Steudle A, Lehmann A, Gerken U, Pleiss J, Schaller A (2012) A model of dirigent proteins derived from structural and functional similarities with allene oxide cyclase and lipocalins. FEBS J 279:1980–1993

Ralph SG, Park JY, Bohlmann J, Mansfield SD (2006) Dirigent proteins in conifer defense: gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea* spp.). Plant Mol Biol 60:21–40

Ralph SG, Jancsik S, Bohlmann J (2007) Dirigent proteins in conifer defense II: extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.). Phytochemistry 68:1975–2199

Shi H, Liu Z, Zhu L, Zhang C, Chen Y, Zhou Y, Li F, Li X (2012) Overexpression of cotton (*Gossypium hirsutum*) dirigent1 gene enhances lignification that blocks the spread of *Verticillium dahliae*. Acta Biochem Biophys Sin 44:555–564

Shu Y, Liu Y, Zhang J, Song L, Guo C (2016) Genome-wide analysis of the AP2/ERF superfamily genes and their responses to abiotic stress in *Medicago truncatula*. Front Plant Sci 6(676):1247

Song M, Xu W, Xiang Y, Jia H, Zhang L, Ma Z (2013) Association of jacalin-related lectins with wheat responses to stresses revealed by transcriptional profiling. Plant Mol Biol 84:95–110

Subramanyam S, Sardesai N, Puthoff DP, Meyer JM, Nemacheck JA, Gonzalo M, Williams CE (2006) Expression of two wheat defense-response genes, *Hfr-1* and *Wci-1*, under biotic and abiotic stresses. Plant Sci 170:90–103

Subramanyam S, Smith DF, Clemens JC, Webb MA, Sardesai N, Williams CE (2008) Functional characterization of HFR1, a high-mannose N-glycan-specific wheat lectin induced by *Hessian fly* larvae. Plant Physiol 147(3):1412–1426

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731

Tang Y, Zhao PX, Murray JD, Wang M, Benedito VA, Ji H et al (2009) The *Medicago truncatula* gene expression atlas web server. BMC Bioinform 10(1):441

Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, et al. (2014) An improved genome release (version mt4.0) for the model legume *Medicago truncatula*. BMC Genom 15 (1):312–312

Thamil Arasan SK, Park JI, Ahmed NU, Jung HJ, Hur Y, Kang KK et al (2013) Characterization and expression analysis of dirigent family genes related to stresses in brassica. Plant Physiol Biochem 67(3):144–153

Voorrips RE (2002) Mapchart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One 2(8):e718

Wu R, Wang L, Wang Z, Shang H, Liu X, Zhu Y, Qi D, Deng X (2009) Cloning and expression analysis of a dirigent protein gene from the resurrection plant *Boea hygrometrica*. Prog Nat Sci 19:347–352

Xia ZQ, Costa MA, Proctor J, Davin LB, Lewis NG (2000) Dirigent-mediated podophyllotoxin biosynthesis in *Linum flavum* and *Podophyllum peltatum*. Phytochemistry 55:537–549

Zhou J, Lee C, Zhong R, Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. Plant Cell 21:48–266