Genome-wide identification, phylogenetic, and expression analysis under abiotic stress conditions of Whirly (WHY) gene family in Medicago sativa L.

Qian Ruan1,2,3, Yizhen Wang1,2,3, Haoyu Xu1,2,3, Baoqiang Wang1,2,3, Xiaolin Zhu1,2,3, Bochuang Wei1,2,3 & Xiaohong Wei1,2,3*

The WHY family is a group of plant-specific transcription factors, that can bind to single-stranded DNA molecules and play a variety of functions in plant nuclei and organelles, participating in the regulation of plant leaf senescence. It has been identified and analyzed in many species, however, the systematic identification and analysis of the WHY genes family have not yet been reported in alfalfa (Medicago sativa L.). Therefore, to explore the function of alfalfa the WHY genes, and 10 MsWHY genes were identified and further characterized their evolutionary relationship and expression patterns by analyzing the recently published genome of alfalfa. Comprehensive analysis of the chromosome location, physicochemical properties of the protein, evolutionary relationship, conserved motifs, and responses to abiotic stresses of the WHY gene family in alfalfa using bioinformatics methods. The results showed that 10 MsWHY genes were distributed on 10 chromosomes, and collinearity analysis showed that many MsWHYs might be derived from segmental duplications, and these genes are under purifying selection. Based on phylogenetic analyses, the WHY gene family of alfalfa can be divided into four subfamilies: I-IV subfamily, and approximately all the WHY genes within the same subfamily share similar gene structures. The 10 MsWHY gene family members contained 10 motifs, of which motif 2 and motif 4 are the conserved motifs shared by these genes. Furthermore, the analysis of cis-regulatory elements indicated that regulatory elements related to transcription, cell cycle, development, hormone, and stress response are abundant in the promoter sequence of the MsWHY genes. Real-time quantitative PCR demonstrated that MsWHYs gene expression is induced by drought, salt, and methyl jasmonate. The present study serves as a basic foundation for future functional studies on the alfalfa WHY family.
In recent years, as more and more studies have been conducted on the function of WHY proteins, it has been confirmed that these proteins have many important functions in plant development and stress tolerance. WHY1 acts as a transcription factor in the nucleus and is involved in regulating pathogen response pathways and the expression of downstream target genes for plant senescence, such as PR10a potatoes, WRKY53 in Arabidopsis mustard, and HvS40 in driver proteins and barley. In addition, WHY1 enrichment in cyscicloid membranes affects plant photosynthesis and redox stress. For example, a reduced abundance of WHY1 can lead to delayed chloroplast development and leaf senescence. WHY1 also has a significant effect on the synthesis and response of plant hormones related to plant growth and defense, such as ABA and SA. This is evidenced by studies of seed germination and senescence. Meanwhile, WHY1, as a co-factor of homologous recombination and DNA double-strand break repair in organelles, maintains organelle genome stability and influences telomere maintenance and microRNA synthesis. WHY1 also interacts with WHY3 to maintain organelle genomic stability and protein metabolism. WHY2 acts as a DNA/RNA-binding protein in mitochondria and activates NAD1/CCB382 gene expression, and WHY2 binds to the promoter of SWEET11/15, which encodes sucrose transporter, in the nucleus. It also increased the expression of genes involved in jasmonic acid signaling and related defense responses. These results suggest that WHY2 plays an important role in carbon redistribution between organelles and nuclei. In addition, overexpression of WHY2 led to the accumulation of starch particles in the chloroplast of pericarp cells, leading to a phenotype of wilting, yellowing, and premature aging of leaves and horny fruits. Although little has been reported about plant phenotypes overexpressing WHY1 or WHY3, transgenic tomato lines overexpressing WHY1 showed increased resistance to cryogenic stress by altering photosynthetic gene expression and modifying starch accumulation. Arabidopsis plants that overexpress WHY2 exhibit early decay. MicroRNA840 (miR840) is a PPR and WHY3 protein that occurs only in Arabidopsis and can specifically target Arabidopsis through post-transcriptional gene silencing of PPR and WHY3.

As a perennial legume forage of the genus Medicago, alfalfa has the characteristics of high yield, good palatability, and strong adaptability has a long cultivation history, and is widely planted. Alfalfa can not only be used as fodder but also has the function of water and soil conservation, soil improvement, and ecological environment protection. Therefore, cultivating tolerant alfalfa varieties is an economic and effective way to resist adversity environment. This experiment uses bioinformatics analysis to identify the alfalfa WHY gene family at the genome-wide level and further analyzes the gene structure, chromosome distribution, and promoter cis-acting elements. QRT-PCR was used to analyze WHY gene expression in alfalfa leaves at different treatment time points under salt, drought, and methyl jasmonate stress. This study can provide a theoretical basis for further research on alfalfa gene function in alfalfa.

**Materials and methods**

**Identification and data collection of alfalfa WHY family.** First, we obtained and downloaded alfalfa genome data from the Alfalfa Breeders Toolbox (https://www.alfalfatoolbox.org) of the proposed southern WHY family (AtWHY1, AtWHY2, and AtWHY3) as decoys to retrieve the alfalfa genome database at the genome-wide level. WHY members have typical Whirly domains and can further be used in Pfam tools (http://pfam.xfam.org/family) to remove homologous sequences from canonical Whirly domains. Reuse online tools NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd/) and SMART (http://smart.embl-heidelberg.de) to predict and identify all possible WHY family members in alfalfa. After deduplication, the genes left were considered alfalfa WHY genes.

**Basic physicochemical properties, secondary structure analysis, and 3D structure prediction of alfalfa WHY genes.** To determine the physical and chemical parameters of each alfalfa WHY protein, the online software ExPASy (https://web.expasy.org/protparam/) was used to calculate the molecular weight (MW), isoelectric point (PI), amino acid numbers, and the average value of hydrophilicity (GRAVY). BUSCA was used to predict protein subcellular localization (http://busca.biocomp.unibo.it) and the secondary structure of proteins was predicted using the online tool SOPMA (http://npsa.phil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) of the proposed alfalfa WHY proteins homology modeling for three-dimensional structure.

**Protein interaction network diagram construction, phylogenetic and promoter cis-acting elements analysis of the MsWHYs.** Based on the model plant Arabidopsis thaliana, the interaction of the alfalfa WHY protein network was predicted. The protein network structure diagram was constructed using STRING (http://STRINGdb.org) software (confidence limit is 0.4). We used nine species to study the evolutionary relationship between alfalfa WHY genes and other plants' WHY genes, including Arabidopsis (Arabidopsis thaliana L.), tobacco (Nicotiana tabacum L.), maize (Zea mays L.), barley medica (Medicago truncatula L.), soybean (Glycine max L.), tomato (Solanum lycopersicum L.), grape (Vitis vinifera L.), sorghum (Sorghum bicolor L.), rice (Oryza sativa L.). Multiple alignments of the above nine species and alfalfa WHY protein sequences were performed by ClustalW of MEGA X. The phylogenetic tree was constructed using the maximum-likelihood (ML) method of MEGA X, with 1000 bootstrap replications. After the phylogenetic tree was constructed, the family members were classified according to the classification criteria of Arabidopsis, tobacco, and rice. The 2000 bp sequence upstream of the WHY gene was used as the promoter of the alfalfa WHY gene. Using
MsWHY2 and MsWHY3). However, the secondary structure α-helix of MsWHY2 and MsWHY3 proteins of the secondary structure of MsWHY proteins, which is more than 55% in most MsWHY proteins (except WHY proteins play different functions in different organelles.

PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) database to predict the promoter cis-acting elements, and display them in the form of graphs.

Gene structure and conserved motif analysis. WHY family DNA and CDS sequences were selected from alfalfa whole-genome sequencing and gene annotation files, respectively. Using the web-based bioinformatic tool GDSD2.0 (http://gods.cbi.pku.edu.cn/index.php) to graphically display the exon/intron genomic structures of alfalfa WHY family. The conserved motifs of alfalfa WHY proteins were analyzed using Multiple Expectation Maximization for Motif Elicitation (MEME Suite) (http://meme-suite.org/)49, and the maximum number of patterns determined in the MEME program was adjusted to 10 and the width of the domain was set from 6 to 10046.

Analysis of chromosome localization, gene duplication, and synteny of the MsWHYs. Chromosome mapping analysis of alfalfa WHY gene family using the MapInspect software (http://mapinspect.software.informer.com/)47. On the plant genome duplication database server (http://chibba.agtec.uga.edu/duplication/index/locket), the duplicate gene pairs are detected48. The non-synonymous substitution rate (Ka), synonymous substitution rate (Ks), and the Ka/Ks ratio were calculated using TBtools49. The MsWHYs hydropathy index (GRAVY) values of all MsWHYs were located and molecular function. The subcellular localization prediction showed that MsWHY2 MsWHY1 was −0.076 to −0.388, indicating that they were hydrophilic. MsWHY hydropathy index (GRAVY) values of all MsWHYs ranged from −0.076 to −0.388, indicating that they were hydrophilic. The calculated grand average of D, and the PI variation ranged from 5.82 (MsWHY1) to 10.28 (MsWHY2) MsWHY1 being homologous (Table 1). The protein lengths varied from 6 to 10046.

RNA extraction and qRT-PCR detection. Using Shenggong's UNIQ-10 column Trizol total RNA extraction kit to extract total RNA from each sample, and use Nano-Drop 2000 UV spectrophotometer to detect RNA quality and concentration. Use M-Mu LV first-strand cDNA synthesis kit reverse transcription RNA to obtain cDNA. After detecting the concentration, uniformly dilute to 100 ng/μl as the qRT-PCR reaction template. Using NCBI Primers for qPCR were designed by primer-BLAST tools (www.ncbi.nlm.nih.gov/tools)50. The primers used in the experiment are shown in the S1 Table. Use Shenggong's 2× SG. Fast qPCR Master Mix kit, the reaction system is 20 μl, the PCR reaction program is 95 °C pre-denaturation for 10 min, and then 40 cycles including 95 °C denaturation for 15 s and 60 °C annealing for 1 min, the instrument used for Applied Biosystems 7500, the experimental results are processed by the 2−ΔΔCt method51. Each experiment was repeated three times with independent RNA samples. Analysis of variance (ANOVA) of the relative expression level of each gene at different sampling points under each abiotic stress treatment was carried out following a generalized linear model using SPSS statistical software. Significant differences in means at different sampling times were determined by Tukey's pairwise comparison tests, as indicated by different letters in the figures. The graphical representation of the experimental findings was produced by using Graphpad.

Ethical approval and consent to participate. This study does not include human or animal subjects. All experimental studies and experimental materials involved in this research are in full compliance with relevant institutional, national and international guidelines and legislation.

Results

Basic physicochemical properties and secondary structure analysis. Accurate identification and an unified nomenclature are essential for future research into the WHY gene family in alfalfa. Here, we identified a total of 10 WHY genes from the alfalfa genome and named them from MsWHY1 to MsWHY8 based on their chromosomal location, with MsWHY5.1-MsWHY5.3 being homologous (Table 1). The protein lengths varied greatly from 91 aa (MsWHY1) to 290 aa (MsWHY2). The molecular weights ranged from 10,490.38 to 32,491.09 D, and the PI variation ranged from 5.82 (MsWHY2) to 10.28 (MsWHY1). The calculated grand average of hydropathy index (GRAVY) values of all MsWHYs was −0.076 to −0.388, indicating that they were hydrophilic in nature. The determination of the subcellular localization of MsWHY proteins will help to understand the molecular function. The subcellular localization prediction showed that MsWHY1 and MsWHY2 were located in mitochondria, MsWHY4, MsWHY5.1, MsWHY5.2, and MsWHY5.3 were located in mitochondria and chloroplasts, and MsWHY3, MsWHY6, MsWHY7, and MsWHY8 were located in chloroplasts. These results indicate WHY proteins play different functions in different organelles.

To better understand the molecular properties of MsWHY proteins, we used SPOMA online software to predict the secondary structures of 10 MsWHY proteins (Table 2). We found that random coil is the main component of the secondary structure of MsWHY proteins, which is more than 55% in most MsWHY proteins (except MsWHY2 and MsWHY3). However, the secondary structure α-helix of MsWHY2 and MsWHY3 proteins...
ZmWHY1, ZmWHY2 proteins properly grouped. alfalfa subfamily III family members were associated with evolutionary process.

gene families of monocotyledons and dicotyledons are in the WHY into small branches, indicating that the WHY or similar biological functions. At the same time, it was found that the genes of monocotyledons clustered—GmWHY6, OsWHY2, GmWHY7, GmWHY8, and MsWHY4—were all distributed at the endpoints of MsChr7.1, MsChr7.2, MsChr7.3 and MsChr7.4, respectively. Alfalfa subfamily I family members were closely related to AtWHY1, AtWHY3, SlWHY2, MsWHY3, NtWHY3, MsWHY4, MsWHY5.1, MsChr5.1, MsChr5.2, and MsChr5.3 chromosomes, respectively.

For proteins that are closely related, it is speculated that they have similar GmWHY7—MsWHY4, MtWHY3, MtWHY4, MtWHY5, MtWHY6, MsWHY5.1, MS.gene63226.t1 268 29,724.73 9.3 − 0.388 Chloroplast; mitochondria

| Gene name | Gene ID | Length/aa | Molecular weight/D | PI | GRAVY | Subcellular localization |
|-----------|---------|-----------|--------------------|----|-------|--------------------------|
| MsWHY1    | MS.gene49569.t1 | 91  | 10,490.38 | 10.28 | −0.105 | Mitochondria |
| MsWHY2    | MS.gene032805.t1 | 290 | 32,491.09 | 5.82 | −0.32 | Mitochondria |
| MsWHY3    | MS.gene91941.t1 | 144 | 16,391.96 | 8.48 | −0.076 | Chloroplast |
| MsWHY4    | MS.gene025976.t1 | 190 | 20,866.89 | 9.46 | −0.085 | Chloroplast; mitochondria |
| MsWHY5.1  | MS.gene024508.t1 | 223 | 27,753.3  | 9.38 | −0.215 | Chloroplast; mitochondria |
| MsWHY5.2  | MS.gene050578.t1 | 223 | 27,753.3  | 9.38 | −0.215 | Chloroplast; mitochondria |
| MsWHY5.3  | MS.gene22868.t1 | 223 | 27,753.3  | 9.38 | −0.215 | Chloroplast; mitochondria |
| MsWHY6    | MS.gene25421.t1 | 194 | 21,709.85 | 9  | −0.267 | Chloroplast |
| MsWHY7    | MS.gene72950.t1 | 266 | 29,413.39 | 9.3  | −0.383 | Chloroplast |
| MsWHY8    | MS.gene63226.t1 | 268 | 29724.73 | 9.3  | −0.388 | Chloroplast |

Table 1. Basic information about members of the Whirly gene family in alfalfa.

counted for a large proportion (more than 35%), while the proportion of other proteins MsWHY proteins was small (MsWHY7 had the lowest 12.78%). The proportion of extended chains in MsWHY protein structure was less than 25% (except MsWHY4).

Phylogenetic analysis of MsWHYs. The molecular evolution of the WHY family is decided mainly by the evolution of increasingly sophisticated organs in plants. To investigate the phylogenetic relationships of WHY gene families in alfalfa, an unrooted phylogenetic tree was constructed from the alignment of full-length WHY proteins of maize, Arabidopsis thaliana, rice, sorghum, grape, tobacco, soybean, barrel medic, and tomato (data in S5 Table). The results showed that 46 WHY genes are grouped into four sub-families and correspondingly named categories I to IV based on a previous study (Fig. 1). The number of members in each subgroup is unevenly distributed. Among them, the subfamily II had the most members, and they were MsWHY3, MsWHY4, MsWHY5.1, MsWHY5.2, MsWHY5.3, and MsWHY6; Subfamily I and III each had two members, and the members were MsWHY7—MsWHY8 and MsWHY1—MsWHY2. Alfalfa subfamily II family members were closely related to SiWHY2. Alfalfa subfamily I family members were closely related to AtWHY1—AtWHY3, NtWHY2-NWHY3, GmWHY1-GmWHY5, VvWHY1, MtWHY3, SbWHY2, SiWHY1, and ZmWHY3-ZmWHY6 and were properly grouped. Alfalfa subfamily III family members were associated with ZmWHY1—ZmWHY2, MtWHY1, OsWHY2, and GmWHY6-GmWHY7. For proteins that are closely related, it is speculated that they have similar or similar biological functions. At the same time, it was found that the WHY genes of monocotyledons clustered into small branches, indicating that the WHY gene families of monocotyledons and dicotyledons are in the evolutionary process.

Analysis of the chromosome location, gene duplication, and synteny of the MsWHYs. According to the annotation information in the alfalfa genome, we found that 10 MsWHY genes are distributed on 10 chromosomes of alfalfa (Fig. 2). And the MsWHY1, MsWHY2, and MsWHY3 genes were distributed on MsChr5.1, MsChr5.2, and MsChr5.3 chromosomes, respectively. MsWHY4, MsWHY5.1, MsWHY5.2 and MsWHY5.3 were all distributed at the endpoints of MsChr7.1, MsChr7.2, MsChr7.3 and MsChr7.4, respectively. MsWHY6, MsWHY7 and MsWHY8 genes were distributed on MsChr8.2, MsChr8.3 and MsChr8.4 chromosomes, respectively.

In the context that alfalfa is an autotetraploid with large genomes, we further examined the duplication events in the MsWHY gene family. Repeat events include segment repetition and tandem repetition. Segmental duplications are long DNA fragments that are nearly identical and present in distant chromosome locations.
Figure 1. Phylogeny of the MsWHY gene family. The phylogenetic tree of the MsWHY gene family was constructed using all 46 WHY genes of Arabidopsis thaliana (At), Oryza sativa (Os), Nicotiana tabacum (Nt), Medicago sativa (Ms), Zea mays (Zm), Medicago truncatula (Mt), Glycine max (Gm), Solanum Lycopersicum (Sl), Vitis vinifera (Vv), and Sorghum bicolor (Sb) as outgroup. The subfamilies of the MsWHY gene family are indicated by I, II, III, and IV. The number nearby each cluster indicates the bootstrap confidence of the cluster in percentage (%).

Figure 2. The distribution of MsWHY genes on 10 chromosomes in alfalfa. The scale (Mb) bar of the left displays the length of alfalfa chromosomes. The number of the chromosome is shown at the top of the chromosome.
They occur most commonly in plants because most plants are diploidized polyploids, and retain a large number of duplicated chromosomes in their genomes54. However, tandem duplication occurred mainly in chromosome recombination region55. To better understand the expansion patterns of MsWHY genes, we performed a colinear analysis of MsWHY genes and found that nine MsWHY genes were colinear with multiple genes in the family, such as MsWHY1 and MsWHY2, MsWHY3, MsWHY6, and MsWHY7 (Fig. 3). A total of 16 pairs of genes were found to have a co-linear relationship. All nine genes were copied in fragments, suggesting that fragment repetition plays a key role in the expansion of the gene family. To elucidate the selective pressure on the duplicated MsWHY genes, we calculated the non-synonymous (Ka) and synonymous substitutions (Ks), and the Ka/Ks ratios for the 8 MsWHY gene pairs (Table 3). The value of Ka/Ks = 1 denotes that genes experienced a neutral selection; < 1 suggests a purifying or negative selection; and > 1 indicates a positive selection56.

**Figure 3.** Syntenic relationship of MsWHYs. The number on the fragments represents the positions on the corresponding chromosomes. The MsWHYs involved in segmental duplications in the MsWHY gene family are mapped to their respective locations of the alfalfa genome in the circular diagram. The red lines represent the segmental duplication pairs between the MsWHYs and the gray lines represent the segmental duplication pairs in the whole alfalfa genome.

**Table 3.** The duplication events of MsWHYs identified in alfalfa.

| No | Sequence                  | Duplication type | Ka    | Ks    | Ka/Ks | Date (Millions of years ago) |
|----|---------------------------|------------------|-------|-------|-------|-------------------------------|
| 1  | MsWHY2 & MsWHY1           | Segmental        | 0.1834| 0.3008| 0.6098| 20.051                        |
| 2  | MsWHY3 & MsWHY2           | Segmental        | 0.0518| 0.1037| 0.4996| 6.91                          |
| 3  | MsWHY5.1 & MsWHY4         | Segmental        | 0.0257| 0.0302| 0.8530| 2.012                         |
| 4  | MsWHY5.2 & MsWHY5.3       | Segmental        | 0     | 0.0066| 0     | 0.438                         |
| 5  | MsWHY5.2 & MsWHY5.1       | Segmental        | 0     | 0.0066| 0     | 0.438                         |
| 6  | MsWHY5.2 & MsWHY4         | Segmental        | 0.0257| 0.0379| 0.6789| 2.527                         |
| 7  | MsWHY5.3 & MsWHY4         | Segmental        | 0.0257| 0.0302| 0.8530| 2.012                         |
| 8  | MsWHY6 & MsWHY7           | Segmental        | 0.0414| 0.0778| 0.5325| 5.186                         |
duplication events were calculated \( T = \frac{K_s}{2\lambda} \) \( (\lambda \) represents the estimated clock-like rate of synonymous substitution, which is \( 1.65 \times 10^{-8} \) substitutions/synonymous site/year for cereals)\(^7\). Our analysis revealed 8 segmental duplication pairs in \( MsWHYs \) (Table 3), with no tandem duplicate pairs. This is because alfalfa is homotetraploid and the 10 \( MsWHYs \) are located on different chromosomes. Furthermore, their \( Ka/Ks \) values vary from 0 to 0.8530, which are all less than 1, indicating that they are subject to purification selection during the evolution process. The dates of these segmental duplication pairs were 0.438 to 20.051 million years ago. Thus, these results indicate the conserved evolution of \( MsWHY \) genes.

**Gene structure and conserved motif analysis.** To understand the structural characteristics of the \( MsWHY \) genes, the exon–intron structures, and conserved motifs of \( MsWHY \) genes were analyzed (Fig. 4). We observed that the structure of the exons and introns of the \( MsWHY \) genes of alfalfa were different among different subfamilies but relatively conserved within the same subfamily. Gene structure (Fig. 4B) analysis of \( MsWHYs \) showed that the number of introns varied from 3 to 8. Of these, \( MsWHY4, MsWHY6, MsWHY7, \) and \( MsWHY8 \) all have 6 introns, while \( MsWHY2, MsWHY5.1, MsWHY5.2, \) and \( MsWHY5.3 \) each contain 8 introns. \( MsWHY1 \) and \( MsWHY3 \) have fewer introns, 3 and 4, respectively. In addition, introns and exosomes in the \( MsWHY1, MsWHY2, \) and \( MsWHY3 \) gene structures of the same branch were significantly different, especially in the number and length of introns in the \( MsWHY2 \). Although the introns of \( MsWHY \) genes were different, the members with the highest homology had similar gene structure, intron length, and the same number of exons, such as \( MsWHY1, MsWHY2, \) and \( MsWHY3 \). In addition, we also elucidated the conserved base sequence of the \( MsWHY \) genes using the MEME (Multiple Em for Motif Element) online servers. Finally, 10 conserved motifs were identified in \( MsWHYs \) (Fig. 4A,C). Motifs owned by or shared by most members of a gene family may be indispensable components of a gene family and have important functions or structures. We found although the conserved motifs of 10 \( MsWHYs \) are different in composition, all 10 \( MsWHYs \) contain motif 2 and motif 4, indicating that motif 2 and motif 4 play an extremely important role in the \( MsWHYs \) \( MsWHY5.1, MsWHY5.2, MsWHY5.3, MsWHY7, \) and \( MsWHY8 \) contain 9 motifs, \( MsWHY4 \) and \( MsWHY6 \) contain 7 motifs, \( MsWHY1, MsWHY2, \) and \( MsWHY3 \) contain 8, 3 and 4 motifs, respectively. Furthermore, the conserved motifs of \( MsWHY5.1, MsWHY5.2, \) and \( MsWHY5.3 \) were identical, as were the conserved motifs of \( MsWHY7 \) and \( MsWHY8 \), suggesting that they may have similar molecular functions. This prediction could lead to the discovery of new members of the \( MsWHY \) gene family.
Promoter cis-acting elements analysis of MsWHYs. Promoter cis-acting elements are important transcription initiation binding regions of transcription initiation factors and play an important role in regulating gene expression. To further analyze the possible biological functions, we used the 2.0 kb sequence upstream of the MsWHY gene's promoter to predict the regulatory elements of cis action through the Plant CARE website (Fig. 5). It is speculated that there are many cis-regulatory elements related to transcription, cell cycle, light, hormone, and stress response in the WHY genes promoter region of alfalfa, some of which are related to root specificity, leaf morphology specificity, seed specificity, and meristem specificity. In addition, we also found 7 elements related to hormone signaling pathways, ABRE, AuxRR-core, CGTCA-motif, P-box, TGACG-motif, TGA-element, and TCA-element. These cis-elements are involved in methyl jasmonate (MeJA), abscisic acid (ABA), salicylic acid (SA), gibberellin (GA), and auxin (IAA) metabolism regulation. In addition to MsWHY1 and MsWHY8, the remaining 8 MsWHYs contained methyl jasmonate response elements (TGACG-motif and CGTCA-motif), and 10 MsWHYs all had ABRE response elements, which indicates that most MsWHYs can participate in JA and ABA-mediated signaling pathways. 3 cis-regulatory elements associated with response to external or environmental stresses were also present. This category includes a low-temperature responsive element (LTR), drought-inducibility element (MBS), and defense and stresses responsive element (TC-rich repeats). In the stress-related expression, the genes related to low temperature were MsWHY1, MsWHY2, MsWHY3, and MsWHY8. At the same time, MsWHY6, MsWHY7, and MsWHY8 were related to drought. In addition, 8 cis-acting elements associated with tissue-specific expression were identified, anaerobic induction elements (ARE, GC-motif), AT-rich sequence, CAT-box, circadian control element (circadian), GCN4-motif, MBSI, and O2-site. It should be noted that all MsWHYs contain components related to light response, and all 10 MsWHY genes contain G-box and GT1-motif. The expression of these genes might be regulated by phytohormones, diverse light-responsiveness cis-elements, defense signaling transduction, and abiotic stresses during alfalfa growth.

Figure 5. Putative cis-acting regulatory elements (CAREs) of the MsWHY gene family. The CAREs analysis was performed with the 2 kb upstream region using the PlantCARE online server. Hormone-responsive elements, stress-responsive elements, specific expression-related elements, and light-responsive elements are shown in different colors.

Protein interaction network diagram and three-dimensional structure prediction analysis. Using protein network interactions to connect unknown functional proteins to protein interaction networks will contribute to a further understanding of the rich biological functions of proteins and the dynamic regulatory networks among various biomolecules. Therefore, in this study, the model plant Arabidopsis thaliana was used as the background to predict the physical and chemical properties of the WHY protein and its potential function-related interacting proteins (Fig. 6).

The expected edge number of our interaction network graph is 10, the average local clustering coefficient is 0.803, and the protein–protein interaction enrichment P value is $< 0.00769$, so we consider the results to be reasonable. We identified two WHY functional molecules and five potential interacting proteins directly related to MsWHY proteins (Fig. 6). They are OSB1, RECA3, OSB3, AT3G18580, and PUM24. ATWHY2 regulates leaf premature senescence, pollen tube activity, and pod development. WHY1 plays an important role in chloroplast and nucleus. In the nucleus, WHY1 is involved in the regulation of plant disease resistance, stress resistance, and senescence. Therefore, we can infer that the functions of 10 WHYs transcription factors in alfalfa are similar to those of the above two Arabidopsis transcription factors. MsWHY1, MsWHY2, MsWHY3, MsWHY6, MsWHY7 and MsWHY8 have similar functions to WHY1, and MsWHY4, MsWHY5.1, MsWHY5.2 and MsWHY5.3 have similar functions to ATWHY2. In this study, the amino acid sequences of 10 members of the MsWHY gene family were modeled by 3D structural homology. The software Swiss-Model was used for online analysis, and the tertiary amino acid sequences of 10 members of the MsWHY gene family were highly similar (Fig. 7). Such as MsWHY4, MsWHY5.1, MsWHY5.2, and MsWHY5.3 have highly similar tertiary structures. In addition, the
three-tiered structure of MsWHY6, MsWHY7, and MsWHY8 are highly similar. The three-stage structure of MsWHY1, MsWHY2, and MsWHY3 differed significantly. However, the tertiary structures are not identical, which may be related to α-helix, β-folding, and irregular criability. These similarities or differences may account for their similar or different functions.

qRT-PCR analysis. To further determine the expression pattern of the MsWHY genes under abiotic stress (drought and salt) and hormone (MeJA) treatment, qRT-PCR was used to quantitatively detect the MsWHY gene's expression under drought, salt, and MeJA treatment (S6 Table). Compared with the control (0 h), under drought stress (Fig. 8A), the expression levels of MsWHY3, MsWHY4, MsWHY5.1, MsWHY5.2, and MsWHY5.3 were significantly up-regulated at three time points (6 h, 9 h, and 12 h). MsWHY7 and MsWHY8 were significantly up-regulated at two-time points (9 h and 12 h), indicating that the stress response of these genes was strong. However, the expression levels of MsWHY1, MsWHY2, and MsWHY6 were lower under drought induc-
Figure 8. qRT-PCR expression analysis of MsWHY genes. Treatment time: 0 h, 3 h, 6 h, 9 h, 12 h, 24 h, and 48 h. (A): Expression analysis of MsWHY genes under PEG stress; (B): Expression analysis of MsWHY genes under NaCl stress. (C): Expression analysis of MsWHY genes under MeJA treatment. Error bars represent standard errors of three biological replicates. The different lowercase letters indicate significant differences at the P<0.05 level.

Discussion
Members of the WHY protein family are found throughout the plant world, and WHY proteins play several important roles in plant development and stress tolerance. In particular, WHY1 regulates gene expression that encodes numerous housekeeping proteins and regulates plant development in response to biological and abiotic stress\textsuperscript{1,5}. First, it acts as a transcription factor in the nucleus, regulating the expression of hormones such as ABA and SA, then as an organoid in organelle chloroplasts and mitochondria\textsuperscript{5}. Barley WHY1 deficient plants showed...
RNAi-mediated loss of WHY1 function in barley, for example, affects aging and stress tolerance\(^{26,27}\). Manipulating WHY1 distribution between nucleus and chloroplast has been shown to alter senescence and cellular redox homeostasis\(^{99}\). Our study also found that ATWHY2 interacts with MsWHY4-MsWHY5.3 proteins, and we hypothesized that MsWHYs play an important role in regulating leaf aging, among others. Studies have shown that WHY2 triple locates in mitochondria, plasmids, and nuclei, and that overexpression of WHY2 in Arabidopsis leads to leaf aging and abnormal growth of longhorns, as well as changes in starch metabolism and expression of genes associated with aging\(^{68}\). Phyllogenetic tree results show that the closer the clustering relationship is, the more likely it is to have similar functions\(^{60}\). In the phylogenetic tree constructed in this study, WHY proteins could be divided into four subfamilies, among which I, II, and III subfamilies contained two, six, and two genes, respectively. By subcellular localization analysis of alfalfa WHY proteins, only MsWHY1 and MsWHY2 were located in mitochondria, MsWHY3 and MsWHY6-MsWHY8 were located in chloroplast, and the other four MsWHY proteins were located in mitochondria and chloroplast. The location of WHY proteins in different subcells may mean that their functions are also different\(^{61}\). Studies have shown that the WHY2 protein located in mitochondria is speculated to be involved in maintaining cell stability and regulating the transcription of the mitochondrial genome in mitochondria, thus playing an important role in plant growth and development, but more sufficient evidence is still lacking\(^{62}\). In chloroplast, the first WH1 protein-binding protein pTAC1, which is involved in chloroplast genome protection and damage repair\(^{83}\). Gene classification, phylogeny, and subcellular localization analysis can help to study the function of similar gene families more accurately and conveniently\(^{64}\). The analysis of gene structure showed that the structure of the MsWHY gene in alfalfa was significantly different, with a maximum of 8 introns and the minimum of only 3 introns. The individual gene structure of the alfalfa WHY family was different, and the different exon–intron structures also contributed to the diversification of gene function\(^{65}\). MsWHY proteins motif in alfalfa is highly conserved. These conserved motifs determine the relatively conserved function of the MsWHY genes. These conserved motifs determine the relatively conserved function of the MsWHY genes. In particular, some genes are missing some motifs, which may be one of the reasons for the functional diversity of the MsWHY genes\(^{66}\).

Homologous genes distributed in farther locations are usually referred to as segmental duplication events, while those located together are considered as tandem duplication events\(^{67}\). Our analysis shows that gene replication plays a major role in gene family expansion. Moreover, both Ka/Ks ratios were less than 0, suggesting that replication of the MsWHY genes occurs through purification selection and that the corresponding MsWHY proteins are considered to be relatively conservative\(^{68}\). In addition, the predicted earliest dates of duplication events in the MsWHY genes segmental duplication pairs ranged from 0.438 to 20.051 million years ago, these results suggest that this is an ancient gene family. Promoter cis-acting element analysis indicated that MsWHYs may be involved in a variety of important biological processes, such as transcription, cell cycle, development, hormone, and biological stresses\(^{71, 73}\). Similarly, the Botrytis cinera genome revealed that MsWHYs gene is primarily located in mitochondria and chloroplasts, suggesting that it plays an important role in mitochondria and chloroplasts. The division of WHY proteins are critical to their function in plant growth, development, and defense. WHY proteins are synthesized in cytosols and transmitted to mitochondria and the chloroplast domain via their targeted signaling. Although many scholars have carried out comprehensive research on WHY protein, there are still many problems to be further studied\(^{70}\).

Gene expression patterns are significant clues for clarifying gene function. In this study, we analyzed expression profiles of the 10 MsWHY genes. The expression patterns of MsWHYs under different stress treatments were significantly different, indicating that the genes responded differently to different stresses. Many studies have reported an increase in WHY transcriptome levels in plants exposed to environmental stresses such as salt and drought stress\(^{70, 71}\), heat\(^{72}\), oxidative stress, and infection with the fungus, Botrytis cinera\(^{71, 72}\). Similarly, the application of exogenous hydrogen peroxide contributes to the accumulation of WHY1 protein in the chloroplast of Arabidopsis\(^{49}\). In contrast, citral, a naturally phytotoxic aromatic compound in lemon fruits, reduced the expression of all WHY genes\(^{74}\). However, WHY2 transcripts, but not WHY1 transcripts, were increased in dehydrated chickpea seedlings\(^{75}\). WHY1 has been implicated in plant responses to biotic and abiotic stresses and has been shown to bind to the promoters of a variety of genes encoding proteins involved in stress tolerance, particularly those containing ERF elements\(^{71}\). In many reports, WHY1-dependent changes in gene expression or delayed greening and delayed photosynthesis, suggesting WHY1 is necessary for chloroplast biogenesis\(^{15, 18}\). In addition, deletion of WHY1 in maize leads to abnormal embryos and albimis\(^{48}\). This reported change in WHY-deficient plant phenotype indicates significant differences in WHY protein function across species. In this study, we found that WHY1 interacts with MsWHY1, MsWHY2, MsWHY3, MsWHY4, MsWHY6, MsWHY7, and MsWHY8 proteins, suggesting that the molecular functions of most MsWHY genes are similar to those of WHY1. Furthermore, the role of WHY1 as a transcription factor regulating leaf aging is well documented.
other WHY1 interactions in the nucleus have been associated with increased stress tolerance, such as enhanced levels of SiWHY1 and SiWHY2 transcripts during drought and salt stress\(^7\). In this paper, we conducted a series of analyses and research on the WHY genes of alfalfa and found that it can be induced expression under abiotic stress. Whether it has a certain protective effect on plants under abiotic stress and the specific mechanism of action still need to be solved gradually in the future.

**Conclusion**

In this study, the phylogeny and diversification of WHY genes in alfalfa were investigated at different levels, including gene structures, evolutionary relationships, promoter cis-acting elements, and expression patterns. All 10 MsWHY genes were divided into 4 groups, and genes in the same group shared similar evolutionary features and expression patterns, implying potentially similar functions for MsWHY genes. MsWHY genes were distributed on 10 chromosomes of alfalfa, and the tertiary structure of amino acid sequence was similar, but not identical. There are cis-regulatory elements in the MsWHY genes promoter region related to hormones, stress, specific expression, and light. Collinearity analysis showed that a high proportion of the MsWHY genes might be derived from segmental duplications with purifying selection, providing insights into possible functional divergence among members of the MsWHY gene family. The physical and chemical properties of the MsWHY protein and the potential interaction proteins related to its function were predicted. qPCR analysis showed that MsWHY genes had different degrees of response to drought and salt stress and methyl jasmonate. The results obtained in this study will provide key ideas for further research on more functions of the WHY genes in alfalfa.

**Data availability**

The genome-wide data of alfalfa is obtained from “Alfalfa Breeder’s Toolbox” (https://www.alfalfatoolbox.org/) and the Whirly protein sequences of Arabidopsis and other species data were downloaded from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html). The original contributions presented in this study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Received: 20 July 2022; Accepted: 18 October 2022
Published online: 04 November 2022

**References**

1. Krupinska, K., Diesel, C., Frank, S. & Hensel, G. WHIRLIES are multifunctional DNA-binding proteins with impact on plant development and stress resistance. *Front. Plant Sci.* **13**, 880423 (2022).

2. Cappadocia, L., Parent, J. S., Sygusch, J. & Brisson, N. A family portrait: Structural comparison of the Whirly proteins from *Arabidopsis thaliana* and *Solanium tuberosum*. *Struct. Biol. Cryst. Commun.* **69**, 1207–1211 (2013).

3. Desveaux, D., Maréchal, A. & Brisson, N. Whirly transcription factors: Defense gene regulation and beyond. *Trends Plant Sci.* **10**, 95–102 (2005).

4. Krause, K. & Krupinska, K. Nuclear regulators with a second home in organelles. *Trends Plant Sci.* **14**, 194–199 (2009).

5. Taylor, R. E., West, C. E. & Foyer, C. H. WHIRLY protein functions in plants. *Food Energy Secur.* **00**, e379 (2022).

6. Grabowski, E., Miao, Y., Mulisch, M. & Krupinska, K. Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell. *Plant Physiol.* **147**, 1800–1804 (2008).

7. Huang, D. et al. Dual-Located WHIRLY1 interacting with LHC1 alters photosynthetic activities of Photosystem I and is involved in light adaptation in Arabidopsis. *Int. J. Mol. Sci.* **18**, 2352 (2017).

8. Krause, K. et al. DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Lett.* **579**, 3707–3712 (2005).

9. Golin, S. et al. WHIRLY2 plays a key role in mitochondria morphology, dynamics, and functionality in *Arabidopsis thaliana*. *Plant Direct.* **4**, e00229 (2020).

10. Desveaux, D. et al. PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* **12**(8), 1477–1489 (2000).

11. Prikril, J., Watkins, K. P., Friso, G., Van Wijk, K. J. & Barkan, A. A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res.* **36**, 5152–5165 (2008).

12. Zhang, Y. F., Hou, M. M. & Tan, B. C. The requirement of WHIRLY1 for embryogenesis is dependent on genetic background in maize. *PLoS ONE* **8**, e67369 (2013).

13. Zhao, S. Y. et al. Overexpression of tomato WHIRLY protein enhances tolerance to drought stress and resistance to pseudomonas solanacearum in transgenic tobacco. *Biol. Plant.* **62**, 55–68 (2018).

14. Xiong, J. Y. et al. Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKPI1 to mediate transcriptional repression. *Plant Mol. Biol.* **71**, 437–449 (2009).

15. Miao, Y., Jiang, J., Ren, Y. & Zhao, Z. The single-stranded DNA-binding protein WHIRLY1 represses WRKY33 expression and delays leaf senescence in a developmental stage-dependent manner in Arabidopsis. *Plant Physiol.* **163**, 746–756 (2013).

16. Comadira, G. et al. WHY1 functions in the control of responses to nitrogen deficiency but not aphid infestation in barley. *Plant Physiol.* **168**, 1140–1151 (2015).

17. Guan, Z., Wang, W. Z., Yu, X. L., Lin, W. F. & Miao, Y. Comparative proteomic analysis of coregulation of CIPK14 and WHIRLY1/3 mediated pale yellowing of leaves in Arabidopsis. *Int. J. Mol. Sci.* **19**, 2231 (2018).

18. Krupinska, K. et al. The nucleoid-associated protein WHIRLY1 is required for the coordinate assembly of plastid- and nucleoc-encoded proteins during chloroplast development. *Planta* **249**, 1337–1347 (2019).

19. Kucharewicz, W. et al. Acceleration of leaf senescence is slowed down in transgenic barley plants deficient in the DNA/RNA-binding protein WHIRLY1. *J. Exp. Bot.* **68**, 983–996 (2017).

20. Isemer, R. et al. Plastid located WHIRLY1 enhances the responsiveness of Arabidopsis seedlings toward abscisic acid. *Front. Plant Sci.* **3**, 283 (2012).

21. Lin, W. F. et al. Dual-localized WHIRLY1 affects salicylic acid biosynthesis via coordination of ISOCHORISMATE SYNTHASE1, PHENYLALANINE AMMONIA LIAISE1, and S-ADENOSYL-L-METHIONINE-DEPENDENT METHYLTRANSFERASE1. *Plant Physiol.* **184**, 1884–1899 (2020).

22. Maréchal, A. et al. Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol.* **8**, 42 (2008).

23. Yoo, H. H., Kwon, C., Lee, M. M. & Chung, I. K. Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in Arabidopsis. *Plant J.* **49**, 442–451 (2007).
24. Swids-Barteczka, A. et al. The plastid-nucleus located DNA/RNA binding protein WHIRLY1 regulates microRNA-levels during stress in barley (Hordeum vulgare L.), RNA Biol. 15, 886–891 (2018).
25. Meng, C. et al. SWHY2 interacts with SIR2A to maintain mitochondrial function under drought stress in tomato. Plant Sci. 301, 110674 (2020).
26. Huang, C. X. et al. Triple-localized WHIRLY2 influences leaf senescence and siliqule development via carbon allocation. Plant Physiol. 184, 1348–1362 (2020).
27. Ren, Y., Li, Y., Jiang, Y., Wu, B. & Miao, Y. Phosphorylation of WHIRLY1 by CIPK14 shifts its localization and dual functions in Arabidopsis. Mol. Plant. 10, 749–763 (2017).
28. Zhuang, K. et al. WHIRLY1 maintains leaf photosynthetic capacity in tomato by regulating the expression of RbcS1 under chilling stress. J. Exp. Bot. 71, 3653–3663 (2020).
29. Ren, Y. et al. MicroRNA840 (MIR840) accelerates leaf senescence by targeting the overlapping 3'UTRs of PPR and WHIRLY3 in Arabidopsis thaliana. Plant J. Cell Mol. Biol. 109, 126–143 (2022).
30. Nian, L. et al. Genome-wide identification, phylogenetic, and expression analysis under abiotic stress conditions of LIM gene family in Medicago sativa L. PLoS ONE 16, e0252213 (2021).
31. Marshall, O. J. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR.
32. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR. Methods 25(4), 402–408 (2001).
33. Yang, Z. L., Liu, H. J., Wang, X. R. & Zeng, Q. Y. Molecular evolution and expression divergence of the populus polylacturonase supergene family shed light on the evolution of increasingly complex organs in plants. New Phytol. 197, 1353–1363 (2013).
34. Yu, J. et al. The genomes of Oryza sativa: A history of duplications. PLoS Biol. 3, 286–281 (2005).
35. Li, W. et al. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. BMC Plant Biol. 10, 282 (2010).
36. Chen, C. et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol. Plant. 13, 1194–1202 (2020).
37. Marshall, O. J. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20(15), 2471–2472 (2004).
38. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR. Method 25(4), 402–408 (2001).
39. Yang, Z. L., Liu, H. J., Wang, X. R. & Zeng, Q. Y. Molecular evolution and expression divergence of the populus polylacturonase supergene family shed light on the evolution of increasingly complex organs in plants. New Phytol. 197, 1353–1363 (2013).
40. Yu, J. et al. The genomes of Oryza sativa: A history of duplications. PLoS Biol. 3, 286–281 (2005).
41. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis. Front. Plant Sci. 5, 716 (2018).
42. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549 (2018).
43. Lescot, M. et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30(1), 325–327 (2002).
44. Hu, B., Jin, J., Guo, A., Zhang, H. & Luo, J. GSDS 2.0: An upgraded gene feature visualization server. Nucleic Acids Res. 31, 1296–1297 (2014).
45. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite.
46. Wilhelm, F., Schmitz, J., Mungall, C., Amundsen, E. & Schulten, K. Comparative genome-wide analysis of regulatory elements in the model plant Arabidopsis thaliana. Genome Biol. 11(1), 1967 (2021).
47. Wu, G. Q., Li, Z. Q., Cao, H. & Wang, J. L. Genome-wide identification and functional analysis of the WRKY genes in sugar beet (Beta vulgaris L.) under alkaline stress. PeerJ 7, e7817 (2019).
48. Wang, L. et al. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. BMC Plant Biol. 10, 282 (2010).
49. Chen, C. et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol. Plant. 13, 1194–1202 (2020).
50. Marshall, O. J. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20(15), 2471–2472 (2004).
51. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR. Method 25(4), 402–408 (2001).
52. Yang, Z. L., Liu, H. J., Wang, X. R. & Zeng, Q. Y. Molecular evolution and expression divergence of the populus polylacturonase supergene family shed light on the evolution of increasingly complex organs in plants. New Phytol. 197, 1353–1363 (2013).
53. Yu, J. et al. The genomes of Oryza sativa: A history of duplications. PLoS Biol. 3, 286–281 (2005).
54. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis. Front. Plant Sci. 5, 716 (2018).
55. Blanca, G. & Wolfe, K. H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667–1678 (2004).
56. Yang, Z. & Nielsen, R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol. Biol. Evol. 25, 568–579 (2008).
57. Bai, X. F. et al. Genome-wide analysis of plant-specific df transcription factor family in tomato. J. Integr. Plant Biol. 55, 552–566 (2013).
58. Kretzschmer, M., Croll, D. & Kronstad, J. W. Chloroplast-associated metabolic functions influence the susceptibility of maize to H2O2 as a feedback signal on dual-located WHIRLY1 associates with leaf senescence in Arabidopsis. J. Integr. Plant Biol. 55, 552–566 (2013).
59. Lin, W. et al. H2O2 as a feedback signal on dual-located WHIRLY1 associates with leaf senescence in Arabidopsis. Cells 8, 1585 (2019).
60. Christiansen, J. F. et al. Functional divergence of gene duplicates through ectopic recombination. EMBO Rep. 13(12), 1145–1151 (2012).
61. Desveaux, D., Maréchal, A. & Brisson, N. Whirly transcription factors: Defense gene regulation and beyond. Trends Plant Sci. 10(2), 95–102 (2002).
62. Huang, C. et al. Triple-localized WHIRLY2 influences leaf senescence and siliqule development via carbon allocation. Plant Physiol. 184(3), 1348–1362 (2020).
63. Ren, Y., Li, Y., Jiang, Y., Wu, B. & Miao, Y. Phosphorylation of WHIRLY1 by CIPK14 shifts its localization and dual functions in Arabidopsis. Mol. Plant 10(5), 749–763 (2017).
64. Abbdud, M. A. & Filiz, E. Whirly (Why) transcription factors in tomato (Solanum lycopersicum L.): Genome-wide identification and transcriptional profiling under drought and salt stresses. Mol. Biol. Rep. 46(4), 4139–4150 (2019).
65. Bonthala, V. S. et al. Identification of gene modules associated with low temperatures response in bambara groundnut by network-based analysis. PLoS ONE 11(2), e0148771 (2016).
66. Janack, B., Sosoi, P., Krupinska, K. & Humbeck, K. Knockdown of WHIRLY1 affects drought stress-induced leaf senescence and histone modifications of the senescence-associated gene Hv540. Plants (Basel) 5(3), 37 (2016).
67. Hsa, C. C. & McGinnis, W. Evolution of transcription factor function.Curr. Opin. Genet. Dev. 13(2), 199–206 (2013).
68. Krupinska, K., Desel, C., Frank, S. & Hensel, G. WHIRLIES are multifunctional DNA-binding proteins with impact on plant development and stress resistance. Front. Plant Sci. 13, 880423 (2022).
69. Karpinska, B. et al. WHIRLY1 functions in the nucleus to regulate barley leaf development and associated metabolite profiles. Biochem. J. 479(5), 641–659 (2022).
70. Yan, Y., Liu, W., Wei, Y. & Shi, H. McCIPK23 interacts with Whirly transcription factors to activate abscisic acid biosynthesis and regulate drought resistance in cassava. Plant Biotechnol. J. 18(7), 1504–1506 (2020).
71. Akbudak, M. & Filiz, E. Whirly (why) transcription factors in tomato (Solanum lycopersicum L.): Genome-wide identification and transcriptional profiling under drought and salt stresses. Mol. Biol. Rep. 46, 4139–4150 (2019).
72. Zhuang, K. et al. WHIRLY1 regulates HSP21.5A expression to promote thermotolerance in tomato. Plant Cell Physiol. 61, 169–177 (2020).
73. Tunc-Ozdemir, M. et al. Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. Plant Physiol. 151, 421–432 (2009).
74. Graña, E. et al. Transcriptome and binding data indicate that citral inhibits single strand DNA-binding proteins. Physiol. Plant. 169, 99–109 (2020).
75. Lande, N. V. et al. Dehydration-induced alterations in chloroplast proteome and reprogramming of cellular metabolism in developing chickpea delineate inter-related adaptive responses. Plant Physiol. Biochem. 146, 337–348 (2020).

Author contributions
X.W. and B.W. conceived and designed the research. H.X. and Y.W. conducted data analysis. B.W. and X.Z. prepared seed materials and contributed analytical tools. Q.R. wrote the manuscript. All authors read and approved the manuscript for publication.

Funding
This research was supported by the Key Laboratory of Crop Science in Arid Regions of Gansu Province, Gansu Agricultural University, National Natural Science Foundation of China (32060401).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22658-3.

Correspondence and requests for materials should be addressed to X.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022