Characterization of WRKY Gene Family in Whole-Genome and Exploration of Flowering Improvement Genes in Chrysanthemum lavandulifolium

Muhammad Ayoub Khan†, Kang Dongru†, Wu Yifei, Wang Ying, Ai Penghui and Wang Zicheng*

State Key Laboratory of Crop Stress Adaptation and Improvement, Plant Germplasm Resources and Genetic Laboratory, Kaifeng Key Laboratory of Chrysanthemum Biology, School of Life Sciences, Henan University, Kaifeng, China

WRKY is a large family of transcription factors known for a variety of functions ranging from stress resistance to plant growth and development. Due to the limited research on the WRKY family in chrysanthemums, we examined them for the first time in Chrysanthemum lavandulifolium. A total of 138 ClWRKY genes were identified, which were classified into three groups. Group III in C. lavandulifolium contains 53 members, which is larger than group III of Arabidopsis. The number of introns varied from one to nine in the ClWRKY gene family. The “WRKYGQK” motif is conserved in 118 members, while other members showed slight variations. AuR and GRE responsive cis-acting elements were located in the promoter region of WRKY members, which are important for plant development and flowering induction. In addition, the W box was present in most genes; the recognition site for the WRKY gene may play a role in autoregulation and cross-regulation. The expression of the most variable 19 genes in terms of different parameters was observed at different stages. Among them, 10 genes were selected due to the presence of CpG islands, while nine genes were selected based on their close association with important Arabidopsis genes related to floral traits. ClWRKY36 and ClWRKY45 exhibit differential expression at flowering stages in the capitulum, while methylation is detected in three genes, including ClWRKY31, ClWRKY100, and ClWRKY129. Our results provide a basis for further exploration of WRKY members to find their functions in plant growth and development, especially in flowering traits.

Keywords: WRKY transcription factors, ornamental, motifs, cis-acting elements, CpG islands, methylation, flowering traits

INTRODUCTION

Chrysanthemum lavandulifolium is one of the original species of Chrysanthemum × morifolium and uses as a model plant. It is a major ornamental plant, originated in China, an important cut flower with diversified petal color and shape having a variety of flower structures, which makes it a highly valuable ornamental plant in the floral industry worldwide (Nguyen and Lim, 2019). Chrysanthemum is among the four most important ornamental plants in the world used for...
beautification and aesthetic value (Shinozaki and Yamaguchi-Shinozaki, 2007). The flower of chrysanthemum is also known for its health benefits, and it contains important antioxidants and phenolic compounds (Lin and Harnly, 2010; Yue et al., 2018).

Due to a simple diploid (2n = 2x = 18) genetic background, C. lavandulifolium dominates over C. morifolium (Huang et al., 2012; Dong-ru et al., 2019). The wild ancestors having desired qualities against biotic and abiotic cues can be used for the genetic improvement of cultivated varieties. Identification and incorporation of such genes into commercial varieties reduces pesticide use, improves plant health, and induces photoperiod-controlling flowering mechanisms (Wu et al., 2010; Fu et al., 2013; Yang et al., 2017).

Genetic architectural study is one of the major research directions in any plant. By narrowing the path further, transcription activation and silencing is the major field of concern. Many transcription factors (TFs) have been explored to switch on and off against certain environmental conditions. WRKY TF is one of the major families first identified in plants in 1994 by the name of DNA-binding protein in sweet potato, which is labeled as SWEET POTATO FACTOR1 (SPF1) (Ishiguro and Nakamura, 1994). WRKY TFs play important roles in resistance against biotic (fungal or bacterial pathogen) and abiotic stresses, growth, and development. Several WRKY TFs have been proven to provide resistance against biotic stresses exerted by fungal or bacterial pathogens by influencing other associated genes (Turck et al., 2004; Duan et al., 2007; Zhou et al., 2008, 2011; Li et al., 2009, 2011; Chen et al., 2010; Scarpeci et al., 2013; Phukan et al., 2016; Yokotani et al., 2018; Warmerdam et al., 2020). The name WRKY has been given to this family for their distinguishing characteristic, the WRKY domain, which is composed of highly conserved 60 amino acids and a zinc finger motif. WRKY TFs are classified into three groups based on the number of (WRKYGQK) domain, and zinc-finger motif. Group I has two domains, and groups II and III have one domain. Groups I and II have C2H2 type zinc finger motifs while group III has a C2HC zinc finger motif. Members of group II were further classified into subgroups (Rushton et al., 2010).

WRKY TFs are extensively studied in Arabidopsis and most other plants. WRKY7 and OsWRKY11 expression resulted in blooming delay in Arabidopsis and rice (Cai et al., 2014; Chen et al., 2019), WRKY13 has an adverse effect on flowering time, and WRKY12, GsWRKY20 and GsWRKY50 promote flowering, possibly through the influence of gibberellic acid under short-day conditions in Arabidopsis (Luo et al., 2013; Kumar et al., 2016; Li et al., 2016), GhWRKY22 regulating pollen development (Wang et al., 2019; Ramos et al., 2021). While different WRKY family TFs play an important role in chrysanthemum, they showed resistance to salt stress (Liu et al., 2013, 2014; Li et al., 2015b; Jaffar et al., 2016; Li et al., 2017; Wang K. et al., 2017; He et al., 2018; Zhang et al., 2020), and CmWRKY1 and CmWRKY15 transgenic lines in C. morifolium interact with SA and ABA and altered plant development (Fan et al., 2015, 2016; Bi et al., 2021). CmWRKY48 showed resistance against aphid infestation (Li et al., 2015a). They also play a regulatory role in the flowering stages. The complex network of VQ20, WRKY12, and WRKY34 interaction influences the plant gametogenesis and plays a crucial part in the development of pollen and pollen tube growth, modulating flowering (Lei et al., 2017; Ma et al., 2020).

Flowering time is a complex character regulated by a network of different external and internal signals. Different pathways have been identified, including photoperiod pathway, vernalization pathway, gibberellin-regulated pathways, and autonomous floral initiation pathway (Soares et al., 2020). Therefore, flowering induction studies are more crucial in ornamental plants, especially in chrysanthemum, where all the ornamental characteristics are associated with flowering, i.e., flower color, shape, and size. Chrysanthemum is a short-day plant and needs a specific photoperiod to induce flowering, which makes it relatively expensive as well as limits the annual production of chrysanthemum flowers. We explored the WRKY gene family in C. lavandulifolium for the first time, prior, the WRKY gene family has only been studied in the C. morifolium (Song A. et al., 2014). The size of the WRKY family in C. morifolium comprised 15 members, which were named from CmWRKY1 to CmWRKY15 (GenBank: KC615355–KC615369). As WRKY is a large family, influencing floral characteristics, we need to explore them deeply. Our present study was a part of such investigation in which we tried to explore the WRKY gene family in C. lavandulifolium comprised structural gene analysis, protein motifs, phylogenetic relatedness of C. lavandulifolium and Arabidopsis thaliana, cis-acting elements, recognition of CpG rich regions, the methylation status of the genes, and expression of 19 important genes in different tissues at different stages to check their role in flowering-related traits. This investigation will lead us to identify variation in transcription regions in the WRKY family for future chrysanthemum research.

**MATERIALS AND METHODS**

**Plant Material and Treatment**

The plant materials of C. lavandulifolium were obtained from Beijing Forestry University. The plants’ G1 lines were grown under controlled environmental conditions in the growth chamber. All the plants used in this research came from the G1 line. The plants were propagated by cuttings and grown in pots. The plants were kept under long-day conditions (14 h L/10 h D) light/dark at a 25°C temperature. The photoperiod conditions were changed to short-day conditions (12 h L/12 h D) when the plants attained maturity (more than 14 leaves). The temperature was kept constant. The plants were watered every 7 days. Stem, root, leaf, buds, and capitulum samples for quantitative real-time polymerase chain reaction (qRT-PCR) were taken at different growth stages, including young, mature, and flowering. The plants were considered at a young stage when they attained eight leaves. The plants were called into the adult stage when the number of leaves was 14, and they were transferred to short-day conditions. The flower develops in short conditions, and samples were taken at different stages of flower development (Supplementary Figure 1). The little buds (1–3 mm in diameter) appeared after 7 days in short-day conditions and were taken for RNA extraction. The median buds (5–8 mm in diameter) were...
taken after a few days before flower opening. The first flower samples were taken when the flowers had just opened.

Characterization of WRKY Gene Family in Chrysanthemum lavandulifolium

The C. lavandulifolium genome has been sequenced (Wen et al., 2022). All the protein and genomic sequences of ClWRKY members were obtained from PRJNA681093, National Center for Biotechnology Information (NCBI). The genomic information of C. lavandulifolium was completed by Henan University and Beijing Forestry University, and related papers have been published (Wen et al., 2022). The redundant sequences, which lack the WRKY domain, were removed after the sequence analysis using manual inspection in the Mega X software. A total of 138 definite ClWRKY sequences were taken for further study. The ClWRKY members were named from ClWRKY1 to ClWRKY138, according to the source provider (Supplementary Table 3).

Classification of ClWRKY Proteins

Multiple protein sequence alignments of C. lavandulifolium and A. thaliana WRKY were made using ClustalX and MEGA-X version 10.2.6 (Molecular Evolutionary Genetics Analysis) (Kumar et al., 2018). The phylogenetic tree based on amino acid (aa) sequences of ClWRKY and AtWRKY conserved WRKY domains was constructed following the neighbor-joining method with 1,000 bootstraps to determine the close association of both gene families.

Gene Structure and Protein Motif Analysis

The gene structure (intron-exon) of ClWRKY members, full-length gene, and coding (CDS) sequences were analyzed in the online database, Gene Structure Display Server (GSDS)\(^1\) as described (Hu et al., 2015). Analysis of the conserved protein motifs was performed using the NCBI CDD batch\(^2\) and the TBtools software (Chen et al., 2020).

Cis-Acting Elements in the Promoter Region of ClWRKY Members

The promoter sequences of all 138 ClWRKY members were submitted to the PlantCARE software to analyze their promoter region for cis-acting elements (Lescot et al., 2002). For this purpose, a 2,000 bp sequence upstream of the gene was taken to identify cis-acting elements. The MEME suite software was used to identify de novo motifs and known enriched motifs (Bailey et al., 2015). The de novo motif analysis was performed to identify significant patterns (Bailey and Elkan, 1994)\(^3\). The known enriched motifs were identified via simple enrichment analysis using the MEME software (Bailey and Grant, 2021).\(^4\) Tomtom motif comparison tool was used to quantify the similarity between motifs (Gupta et al., 2007).

CpG Island Analysis of the Promoter Region of ClWRKY Gene Family

For CpG analysis, promoter regions from the WRKY gene family were sequenced. For the exploration of CpG islands, the “Methyl Primer express version 1.0” software was used (Methyl Primer Express Software v1.0 Quick Reference Card)\(^5\). All the sequences were submitted to the software one by one to identify CpG islands or rich regions. The minimum length of the island was kept at 200 bp, while the maximum length was 2,000 bp. The criterion of “C + G/Total bases” was 50%. The CpG observed/CpG expected ratio was 0.6.

Genomic DNA Extraction and Bisulfite Treatment

The total DNA was extracted from different tissues of C. lavandulifolium. The tissues included root, stem, leaf, and capitulum at three different growth stages, i.e., young stage, adult stage, and flowering stage. Total genomic DNA was extracted using the “Plant Genomic DNA Kit” (TOLOBIO), according to the manufacturer’s instructions. DNA was quantified using Thermo Scientific NanoDrop™ 2000/2000c Spectrophotometers, and the quality of DNA was checked by 1.5% agarose gel electrophoresis.

Total genomic DNA from each sample was treated with sodium bisulfites using the “DNA Bisulfite Conversion Kit” (TIANGEN) according to the manufacturer’s instructions. The conversion efficiency and quality were measured using Thermo Scientific NanoDrop™ 2000/2000c Spectrophotometers.

Methylation-Specific Polymerase Chain Reaction

Ten genes were selected on the basis of CpG analysis. Primers for all ten genes were designed using the “Methyl primer express version 1.0” software. For this purpose, two types of primers (methylated and unmethylated) were designed to check the possible methylated sites. The primer length ranged from 18 to 22 bp. The PCR amplicon length was between 100 and 175 bp. Methylation-Specific Polymerase Chain Reaction (MSP) was conducted to analyze the CpG island among the ten selected genes after treatment with bisulfite for methylation status. The MSP conditions were as follows: pre-denaturation for 3 min at 95°C, followed by 34 cycles of the 30s at 95°C and 30 s at 55°C, 1 min at 72°C, and a final extension for 5 min at 72°C. PCR products were resolved onto 1.5% agarose gel and observed under UV transilluminator.

Quantitative Real-Time Polymerase Chain Reaction

Nineteen (Zhou et al., 2011) variable genes on the basis of CpG island and phylogenetic analysis were selected for gene expression patterns. Primers for all the 19 ClWRKY genes were designed using the “Primer Premier 5” software. RNA was extracted from different plant tissues, i.e., root, stem, and leaves at young, adult, and flowering stages, while the capitulum samples, which

\(^1\)http://gsds.gao-lab.org/
\(^2\)https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
\(^3\)https://meme-suite.org/meme/tools/meme
\(^4\)https://meme-suite.org/meme/tools/sea
\(^5\)thermofisher.com
RESULTS
Identification and Characterization of WRKY Gene Family in Chrysanthemum lavandulifolium
WRKY genes possess a conserved “WRKYGQK” domain. The “WRKYGQK” heptapeptide is the most defining characteristic of WRKY TFs. The C. lavandulifolium WRKY sequences were obtained from PRJNA681093, NCBI. A total of 153 WRKY gene members have been selected; among them, 138 were exhibited for sequence analysis, and 15 were considered defective. The WRKY members are named from CIWRKY1 to CIWRKY138 according to their location on the chromosome. Protein motif analysis showed homology in most of the conserved regions. Among 138 CIWRKY members, 85.51% of members have homology in the “WRKYGQK” core domain, and 19 members have two conserved WRKY domains. However, 14.5% of WRKY members showed variation in the conserved heptapeptide domain and had substitutions in the core domain. Out of 14.5% variants, 13 members (CIWRKY37, 38, 39, 49, 50, 51, 52, 53, 54, 74, 76, 77, and 98) have “WRKYGKK”, substituting glutamine as a lysine aa. CIWRKY12 replaced histidine instead of lysine, two genes, CIWRKY20 and CIWRKY21, shared sixth and seventh as glutamine and asparagine substitutions. High diversity has been found in three genes, i.e., (CIWRKY27) “WKKYGEQK,” (CIWRKY97) “WKKYGEKK,” and (CIWRKY131) “WRKNGQNN” WRKY domains, respectively (Figure 1).

For the determination of important properties such as aa length, molecular weight, and isolectric point, we thoroughly analyzed the WRKY gene. The average length of WRKY proteins was 313 aa, while 74 aa was the minimum and 697 aa was the maximum length. The WRKY members with the least and highest lengths were CIWRKY58 and CIWRKY36, respectively. The molecular weight ranged from 0.02 kDa (CIWRKY55) to 76.26 kDa (CIWRKY36). The pIs (isolectric point) ranged from 4.73 (CIWRKY75) to 10.63 (CIWRKY97). Among 138 CIWRKY members, 82 were acidic (having less than 7 pl), while 56 were basic (having more than 7 pl). More details of CIWRKY features are available in Supplementary Table 1.

Phylogenetic Analysis and Classification of CIWRKY Proteins
A phylogenetic tree was constructed between C. lavandulifolium and A. thaliana to investigate the relatedness of different WRKY genes and to understand the classification of different groups among WRKY members. Arabidopsis is a model plant for genetic study and has been explored for the function of WRKY genes. The MEGA-X software was used for agglomerative construction following the neighbor-joining method to find the correlation between WRKY members of C. lavandulifolium and A. thaliana. The WRKY members are divided into three groups or seven subgroups. Group I has 27 members, and group II has 58 members with 9 members in IIa, 10 members in IIb, 10 members in IIc, 10 members in IId, and 19 members in Ile. Group III has 53 members and was found to be the largest group in this classification (Figure 2). The majority of the CIWRKY members clustered in the corresponding groups. Some of the WRKY members (CIWRKY20, 21, 55, 56, and 58) had one WRKY domain, but still, clustered in group I; by the rules, group I members should have two domains. These WRKY members may evolve from two WRKY domain genes and lose one domain in the process of evolution. Previously, such exceptional cases of WRKY members with one WRKY domain clustering in group I have been reported (Bi et al., 2016).

Some of the CIWRKY members clustered with important Arabidopsis WRKY genes that helped us identify the genes that are associated with flowering. CIWRKY36 is clustered with Arabidopsis AtWRKY2 and AtWRKY34 (Figure 2), both of which play an important role in pollen development through interaction with the VQ20 protein (Lei et al., 2017). Similarly, CIWRKY45 clustered with AtWRKY71 (Figure 2), which play a key role in the regulation of flowering, directly and indirectly, by influencing other flowering-associated genes (Yu et al., 2016, 2018).

Gene Structure of WRKY Members and Protein Motif Analysis
For the determination of intron-exon structure similarity, Gene Structure Display Server (GSDS) database has been used. Variation among CIWRKY members was observed in terms of a number of introns. The minimal number of introns in a gene was one, which occurred in seven members including CIWRKY27, CIWRKY58, CIWRKY59, CIWRKY76, CIWRKY86, CIWRKY92, and CIWRKY125 while CIWRKY90 had 9 introns, which is the maximal in the whole family. The overall average number of introns in the CIWRKY family was three (Figure 3). The number of introns in group I ranged from 1 to 5, group IIa varied from 3 to 9, group IIb from 2 to 6, group IIc from 1 to 2, group IId from 1 to 4, and group Ile from 1 to 4. Group III varied from 1 to 6 (Supplementary Table 2). The minimal number of exons was two, which appeared in CIWRKY27, CIWRKY58, CIWRKY59, CIWRKY76, CIWRKY86, CIWRKY92,

http://gsds.cbi.pku.edu.cn/
and ClWRKY125, while the maximal number of exons was 10, which occurred in ClWRKY87 and ClWRKY90 (Supplementary Table 2). The overall average number of exons in the ClWRKY family was four (Figure 3). The exon numbers varied within the groups; group I ranged from 2 to 6, group IIa from 4 to 10, group IIb from 3 to 7, group IIc from 2 to 3, group IId from 2 to 5, group Ie from 2 to 4, and group III from 2 to 7, respectively. The average number of exons within the groups is as follows: group I has 4, group IIa has 6, group IIb has 5, group IIc has 3, group IId has 3, group Ie has 3, and group III has 3, respectively (Supplementary Table 2). Protein motif analyses were performed on the ClWRKY gene family. The first two motifs represented by green and yellow colors are WRKY protein motifs (Figure 4). At least one or two WRKY protein motifs have been identified among all the gene sequences. This suggests that WRKY is a conserved domain.

CpG Island Analysis of the Promoter Region of ClWRKY Gene Family
CpG island is the region of DNA that is rich in CpG sites and is often located near the promoters of genes. Ten members were found to have CpG islands; among them, seven members contained CpG in their promoter region, two in the cross-promoter and gene region, and one member was found to have it in the gene body. All the ten members had one CpG island. ClWRKYY115 had the shortest island with 438 bp, while ClWRKY83 had the longest island with a length of 1,065 bp (Table 1). The genes with the CG-rich content are the obvious means of DNA methylation and a prominent source of epigenetic improvement.

Identification of DNA Methylation
Total genomic DNA of C. lavandulifolium was extracted from the root, stem, leaf, little buds, median buds, and first flowers (flowers at just opening time). These tissues were taken at different growth stages of the plant, which include the young stage, adult stage, and flowering stage. The methylation and unmethylation primers were designed for the ten different genes using the “Methyl Primer Express version 1.0” software. These genes were selected due to the presence of CpG islands. The MSP was carried out for the following ten genes: ClWRKY9, ClWRKY31, ClWRKY33, ClWRKY41, ClWRKY83, ClWRKY97, ClWRKY100, ClWRKY115, ClWRKY129, and ClWRKY130. The methylation status detected by MSP was analyzed through gel electrophoresis. The MSP results showed the presence of methylation in ClWRKY100 and ClWRKYY129 at different stages in different tissues. These genes showed significant methylation status in all tissue samples at all growth stages of the plant (Figures 5A–C). However, the methylation status detection through bisulfite sequencing can give us more clear results compared with the MSP method.

Cis-Acting Elements in the Promoter Region of ClWRKY
The cis-acting elements in the promoter region are really important for gene expression (Rombauts et al., 1999). The 2,000 bp sequence upstream of the genes was submitted to the PlantCARE software and searched for cis-acting elements. A range of cis-acting elements was identified in the promoter region of the ClWRKY family, which were highly diverse.

A total of 110 cis-acting elements were found in the promoter region of ClWRKY members. Most of these elements have a diverse role in the plant life cycle. According to our objectives, we screened out the elements that can play a role in plant growth and development particularly in flowering induction. Therefore, we selected 15 cis-acting elements on the basis of their possible role in flowering-related traits and the overall development of...
FIGURE 2 | A phylogenetic tree of 138 ClWRKY proteins of Chrysanthemum lavandulifolium (blue) and 74 AtWRKY proteins of Arabidopsis thaliana (red) on the basis of amino acid sequences. The tree was constructed in MEGA X. The domains were clustered into three groups, namely, group I, group II, and group III. The group II was divided into five subgroups (a–e).

the plant. The detailed features of the 15 cis-acting elements are provided in Supplementary Table 4. Most of the genes have important cis-acting elements, such as auxin responsiveness (AuR) and gibberellin responsive elements (GRE), associated with flowering-related traits (Figure 6). W-box was most widely available among the CIWRKY family. CIWRKY45 and CIWRKY36, the genes that resulted in high expression at the capitulum stage, had some important elements involved in endosperm expression, elements involved in auxin responsiveness, and ethylene-responsive elements (AuR, EE, and ERE), which are supposed to play a role in flower development. Methyl jasmonate (MeJAR), which plays important role in flowering induction, was present in most of the CIWRKY members, especially in group III members, which indicates that WRKY members may play a role in the overall development of a plant.

The de novo motif analysis in MEME identified 15 motifs (Supplementary Figure 2). Tomtom tool comparison showed homology with some important motifs, such as motifs similar to MEME12 and MEME14, which are involved in seed germination, pollen tube growth, and WRKY binding sites (Supplementary Table 5). The enriched motifs identified in MEME are mostly different from the motifs found in the PlantCare software. The simple enrichment analysis (SEA) in MEME identified 95 known enriched motifs. Among them, 49 motifs encoded ethylene, which is essential for plant growth and development. Other important motifs included are WRKY binding sites, TEOSINTE BRANCHED, cycloidea and PCF (TCP) binding sites, and so on (Supplementary Table 6).

qRT-PCR Analyses
Different CIWRKY genes having special characters were selected on the basis of previous analysis. Total RNA was extracted from different tissues (root, stem, leaves, and flowers) of C. lavandulifolium at different stages (young, adult, and flowering). The expression pattern of 19 WRKY genes was observed in all tissues at different stages of the plant’s life span (Figure 7). Among them, 10 genes were selected due to the presence of CG-rich regions, and nine genes were selected as they showed close association with important Arabidopsis WRKY genes that are associated with floral traits. Most of the genes exhibited higher expression at the flowering stage in the capitulum.
tissues. The highest expression in the capitulum tissues at the flowering stage was exhibited by CiWRKY36, CiWRKY45, CiWRKY57, CiWRKY59, and CiWRKY82 (Figure 7). The expression of these genes was higher in the flowering stage, especially in the first flowering (FF) stage and median bud stage, which suggests their role in flowering regulation. These genes were selected for qRT-PCR as they clustered with flowering-related AtWRKY genes (Figure 2). On the basis of previous analysis, CiWRKY36 clustered with CiWRKY71, so it might have a basic role in the control of flowering time and plant development, as WRKY71 homolog was found to promote flowering in woodland strawberries (Lei et al., 2020).

CIWRKY31, CIWRKY33, CIWRKY41, CIWRKY97, CIWRKY100, CIWRKY129, and CIWRKY130, which were selected on the basis of having CpG rich regions (Supplementary Table 1), showed higher expression in capitulum tissues (Figure 7). In addition, CIWRKY31, CIWRKY100, and CIWRKY129 were detected with DNA methylation (Figure 5), which is linked with flowering regulation and is one of the basic sources of epigenetics. Interestingly, CIWRKY83 remains stable across all the tissues and stages as it is constantly expressed in all samples of C. lavandulifolium. It might play a role in adaptability. Moreover, CIWRKY83 showed close association with important AtWRKY7 of Arabidopsis. The other WRKY member's expressions showed less or non-significant changes at different stages in all tissues.

We also investigated CILFY (leafy), as it not only promotes flowering time but also influences other genes that regulate flowering-related traits (Yang et al., 2016). The expression of this gene was high in the median bud and adult leaf stages, but low in stem and root (Figure 7). Besides, CILFY may promote flowering time in C. lavandulifolium by making association with CIWRKY36 and CIWRKY45.

**DISCUSSION**

WRKY TFs are one of the largest transcription families, playing an important role in plant development and resistance against different biotic and abiotic stresses. WRKY TFs are spread broadly across higher plants (Li et al., 2014). WRKY family investigation is limited to C. morifolium with 15 genes (Song A. et al., 2014). Being a model plant for chrysanthemums (Dongru et al., 2019), the characterization of the WRKY gene family in C. lavandulifolium is of great importance. For the first time, we identified 138 WRKY members in C. lavandulifolium. Out of 138 members, 118 were properly conserved in the “WRKYGQK” domain. Substitution of glutamine occurs instead of lysine in thirteen WRKY heptapeptide domains. This WRKYGKK sequence is reported to be a major variant in many studies (Song H. et al., 2014; Song et al., 2016a,b). CIWRKY12 (WRKYGHK) “Q” replaced by “H” CIWYK20 and 21 (WRKYGQN) “K” replaced by “N,” CIWRY27 (WKKYGQEK) “K” replaced by “Q,” CIWRKY97 (WKKYGEK) “Q” replaced by “E,” and CIWRY131 (WRKNQGK)”Y” replaced by “N” as their respective domains were also reported in carrot (Nan and Gao, 2019). The variation in the heptapeptide domain of WRKY genes
has been largely demonstrated (Yamasaki et al., 2005; Brand et al., 2013; Rinerson et al., 2015; Yang et al., 2020); however, variations in aa sequence were different from our findings. The variation in the WRKYGQK conserved domain mainly occurs from Q to K aa. The same kind of substitution was observed in the WRKY gene family of Camelina sativa (Song et al., 2020).
TABLE 1 | Details of each of the ten genes with a CpG island.

| NO | WRKY Gene | Island # | Start | End | Length | Region |
|----|-----------|----------|-------|-----|--------|--------|
| 1  | CIWRKY9   | 1        | –1239 | –593| 647    | Promoter |
| 2  | CIWRKY31  | 1        | –1049 | –251| 799    | Promoter |
| 3  | CIWRKY33  | 1        | –4    | 566 | 571    | Promoter + Gene Body |
| 4  | CIWRKY41  | 1        | –1296 | –688| 609    | Promoter |
| 5  | CIWRKY83  | 1        | –2000 | –935| 1,065  | Promoter |
| 6  | CIWRKY97  | 1        | –80   | 471 | 552    | Promoter + Gene Body |
| 7  | CIWRKY100 | 1        | –2000 | –1429| 571  | Promoter |
| 8  | CIWRKY115 | 1        | –2000 | –1562| 438  | Promoter |
| 9  | CIWRKY129 | 1        | 6494  | 6950| 457    | Gene Body |
| 10 | CIWRKY130 | 1        | –878  | –251| 628    | Promoter |

The 138 CIWRKY genes were analyzed for CG-rich regions using the “Methyl Primer Express version 1.0” software.

The CIWRKY gene family is composed of 138 genes, which is bigger than the ArWRKY gene family (Rushston et al., 2010). The CIWRKY family was classified into three groups. The classification of CIWRKY members was according to the classification of the WRKY family in Arabidopsis. Most of the CIWRKY members clustered in the corresponding groups. However, some of the CIWRKY members with one WRKY domain clustered in group I, which included, CIWRKY58, 95, 60, 56, 20, 21, and CIWRKY55. Among them, CIWRKY20, 21, and 55 possessed a unique pattern “WNSIVV” which was conserved in these three members. Clustering of one domain WRKY of factors in group I is consistent with the study of the WRKY TF family in Linum usitatissimum (Yuan et al., 2021). The same results are reported in Cicer arietinum (Kumar et al., 2016).

Some WRKY members with one conserved domain and clustered into group I have been reported in peanut (Arachis hypogaea L.) (Zhao et al., 2020).

The intron-exon structure seems to be diverse, starting from 1 to 9 introns. Diversification in intron-exon structure may play an important role in the evolution of genes (Xu et al., 2012). In this study, we have identified different intron constitutions within gene groups. Group I has four introns on average, group IIa has 5, group IIb has 4, and all other groups have an average of three introns. This suggests that the latter groups originated from group I and earlier subgroups of group II. The same trend of intron presence in different groups of WRKY members in eggplant was reported (Yang et al., 2020). A total of six motifs were detected; among them, two motifs were considered to be WRKY domains. The conserved domain “WRKYGQK” was almost included in all of the members; however, some of the WRKY members showed a slight variation as “WRKYGKX” domain as suggested by Liu et al. (2021) and Qu et al. (2021).

The cis-acting elements are involved in plant development and flowering induction. There are several elements, such as auxin, gibberellins, and jasmonic acid, which affect flowering-related traits. Jasmonates, cytokinins, and auxin control second bud growth in chrysanthemum (Sun et al., 2021). Auxins control plant growth and development (Singla et al., 2006). The petal is the most important factor influencing the chrysanthemum’s ornamental value. Auxin signaling is associated with petal growth in chrysanthemums (Wang J. et al., 2017). AuR is found in most CIWRKY members and may play a role in plant development and flowering induction. GRE was abundant in CIWRKY, an important hormone associated with flowering induction in chrysanthemum (Dong et al., 2017). CmBBX24 regulates flowering in chrysanthemum in conjunction with gibberellin synthesis (Zhu et al., 2020). In addition, the key recognition site for the WRKY gene, the W-box containing the TGAC sequence, was observed in most CIWRKY members. The W-box contributes under abiotic stress conditions (Dhatterwal et al., 2019) by influencing the early senescence of rice flag leaves (Liu et al., 2016). The presence of flower-associated cis-acting elements in most genes indicates the potential of CIWRKY to play a role in flowering control in C. lavandulifolium.

DNA methylation has an impact on growth and development, with CG content, or CpG, being the most abundant source of DNA methylation (Dong-ru et al., 2019). Various methods such as methylation-sensitive amplification polymorphism and MSP have been applied to methylated genes (Herman et al., 1996; Coronel et al., 2018). In this study, we have identified ten genes with strong CpG content in the promoter region (Table 1), which may cause epigenetic changes in C. lavandulifolium (Chip and Dresselhaus, 2019). Three genes having methylated sites, i.e., CIWRKY31, CIWRKY100, and CIWRKY129, were identified on the basis of MSP results. These genes can regulate gene expression and may play a role in flowering-related traits, as they are expressed significantly in capitulum tissues at the flowering stage. CmMET1 can affect methylation and regulate candidate genes associated with flowering in C. morifolium (Kang et al., 2021). DNA methylation changes in the promoter and gene body regions can influence gene expression and function of the protein.
FIGURE 6 | Cis-acting element analysis in the promoter region of the ClWRKY family. The PlantCARE software was used to search the cis-acting elements. Abbreviation of important cis-acting elements are as follows: EE, endosperm expression; ARE, auxin-responsive elements; GR, gibberellin responsiveness; LR, light responsiveness; REEAI, regulatory element essential for anaerobic induction; AuR, elements involved in auxin responsiveness; SSR, seed-specific regulation; MeJAR, methyl jasmonate responsiveness; ERE, ethylene-responsive elements; GRE, gibberellin responsive elements; ESNE, endosperm specific negative expression; LR2, elements involved in light response; PAP2L, plant-AP2 such as AuR2, part of auxin responsiveness; and W-box, recognition site of WRKY TFs.
This study focuses on the expression of CIWRKY members to find their role in plant development, especially their influence over flower induction. The expression analysis revealed variations of different genes in different organs. CIWRKY45 showed significant expression at the flowering stage in the capitulum tissue, which may be related to some morphological characters and floral organ development. In the same fashion, CIWRKY45 clustered with Arabidopsis AtWRKY71, a prominent gene that controls flowering characteristics (Yu et al., 2016, 2018).
providing strong evidence for our results. Moreover, CIWRKY36 and CIWRKY57 are significantly expressed in capitulum at the flowering stages and have close homology with Arabidopsis AtWRKY2 and AtWRKY34, indicating the role of CIWRKY36 and CIWRKY57 in flowering, as Arabidopsis homologs are the key regulators of pollen development (Lei et al., 2017). CIWRKY59 is expressed significantly in the first flower tissue and can positively regulate flowering, as CIWRKY59 homolog, AtWRKY7, expressed in the median bud stage, which indicates its potential to play a role in flowering regulation. AtWRKY7, a CIWRKY82 homolog, regulates plant growth (Chen et al., 2019). The qRT-PCR results indicate that CIWRKY83 is a stable gene as it expresses across all the tissues at different stages of plant life and may play a role in adaptability. The higher expression of three genes in capitulum tissues, including CIWRKY31, CIWRKY100, and CIWRKY129, which are methylated as well, can play an important role in flowering-related traits. These results suggest that CIWRKY33, CIWRKY36, and CIWRKY45, along with CIWRKY31, CIWRKY100, and CIWRKY129, can play a crucial role in the genetic and epigenetic improvement of the chrysanthemum. CIWRKY45, CIWRKY36, and CIWRKY129, in particular, have the potential to alter flowering-related traits. The overall investigation of the CIWRKY domain is providing theoretical insight into C. lavandulifolium transcriptional activation and inactivation mechanisms for the improvement of plant development.

CONCLUSION

This study aimed to identify the WRKY gene family in C. lavandulifolium and explore the potential of WRKY genes for genetic improvement. A total of 138 WRKY genes were analyzed in C. lavandulifolium. All the data results are unique to explore the underlying mechanism of the WRKY TF in C. lavandulifolium plant improvement. The gene structure analysis suggested the vast variation among WRKY members. The WRKY domain remains conserved across all the members of this family. Ten WRKY genes possessed CG-rich content in the promoter region where three genes exhibit methylation status. This research can provide a basis to study the role of WRKY genes in the improvement of chrysanthemum plants, especially in flowering traits.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

WZ designed the research project, coordinated, and supervised the study. MA performed all the experiments and wrote the manuscript. MA and KD did most of the experimental work and analyzed the data. WuY, AP, and WaY performed specific experiments during the research. All authors read and approved the final manuscript.

FUNDING

This research was supported by the National Key Research and Development Plan (2018YFD1000403).

ACKNOWLEDGMENTS

We thank all the colleagues who helped them a lot during this research. We specially thank WZ and KD for their support and timely response.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.861193/full#supplementary-material

REFERENCES

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36.
Bailey, T. L., and Grant, C. E. (2021). SEA: Simple Enrichment Analysis of motifs. bioRxiv [Preprint]. doi: 10.1101/2021.09.02.458722
Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. Nucleic Acids Res. 43, W39–W49.
Bi, C., Xu, Y., Ye, Q., Yin, T., and Ye, N. (2016). Genome-wide identification and characterization of WRKY gene family in Salixgoochowsensis. PeerJ 4:e2437. doi: 10.7717/peerj.2437
Bi, M., Li, X., Yan, X., Liu, D., Gao, G., Zhu, P., et al. (2021). Chrysanthemum WRKY15-1 promotes resistance to Puccinia horiana Hann. via the salicylic acid signaling pathway. Hortic Res. 8.6. doi: 10.1038/s41438-020-00436-4
Brand, L. H., Fischer, N. M., Harter, K., Kohlbacher, O., and Wanke, D. (2013). Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. Nucleic Acids Res. 41, 9764–9778. doi: 10.1093/nar/gkt732
Cai, Y., Chen, X., Xie, K., Xing, Q., Wu, Y., Li, J., et al. (2014). DFI, a WRKY transcription factor, is involved in the control of flowering time and plant height in rice. PLoS One 9:e102529. doi: 10.1371/journal.pone.0102529
Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TTools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol. Plant 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
Chen, H., Lai, Z., Shi, J., Xiao, Y., Chen, Z., and Xu, X. (2010). Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. BMC Plant Biol. 10:281. doi: 10.1186/1471-2229-10-281
Chen, W., Hao, W. J., Xu, Y. X., Zheng, C., Ni, D. J., Yao, M. Z., et al. (2019). Isolation and characterization of CsWRKY7, a subgroup IId WRKY transcription factor from Camellia sinensis, linked to development in arabidopsis. Int. J. Mol. Sci. 20:2815. doi: 10.3390/ijms20112815
Chip, L., and Dresselhaus, T. (2019). Chapter 12 and Reproductive Development in Cereals Using Chromatin. Cereal Genomics 2072, 141–156. doi: 10.1007/978-1-4939-9865-4_12

Coronel, C. J., González, A. L., Ruiz, M. L., and Polanco, C. (2018). Analysis of somaclonal variation in transgenic and regenerated plants of Arabidopsis thaliana using methylation related met AFLP and TMD markers. Plant Cell Rep. 37, 137–152. doi: 10.1007/s00299-017-2217-x

Dhatterwal, P., Basu, S., Mehrotra, S., and Mehrotra, R. (2019). Genome wide analysis of W-box element in Arabidopsis thaliana reveals TGA motif with genes down regulated by heat and salinity. Sci Rep. 9:16811. doi: 10.1038/s41598-019-38757-7

Dong, B., Deng, Y., Wang, H., Gao, R., Stephen, G. K., Chen, S., et al. (2017). Gibberelic acid signaling is required to induce flowering of chrysanthemums grown under both short and long days. Int. J. Mol. Sci. 18:12529. doi: 10.3390/ijms18061259

Dong-rui, K., Si-lan, D., Kang, G., Fan, Z., and Hong, L. (2019). Morphological variation of Chrysanthemum lavandulifolium induced by 5-azaC treatment. Sci. Hortic. 257:108645. doi: 10.1016/j.scienta.2019.108645

Duan, M. R., Nan, J., Liang, Y. H., Mao, P., Lu, L., Li, L., et al. (2007). DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein. Nucleic Acids Res. 35, 1145–1154. doi: 10.1093/nar/gkm001

Fan, Q., Song, A., Jiang, J., Zhang, T., Sun, H., Wang, Y., et al. (2016). CmWRKY1 enhances the dehydration tolerance of chrysanthemum through the regulation of ABA-associated genes. PLoS One 11:e0150572. doi: 10.1371/journal.pone.0150572

Fan, Q., Song, A., Xin, J., Chen, S., Jiang, J., Wang, Y., et al. (2015). CmWRKY15 facilitates Alternaria tenuissima infection of chrysanthemum. PLoS One 10:e0143349. doi: 10.1371/journal.pone.0143349

Fu, J., Wang, Y., Huang, H., Zhang, C., and Dai, S. (2013). Reference gene selection for RT-qPCR analysis of Chrysanthemum lavandulifolium during its flowering stages. Mol. Breed. 31, 205–215. doi: 10.1007/s11032-012-9784-x

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. Genome Biol. 8:R24. doi: 10.1186/gb-2007-8-2-r24

He, L., Wu, Y. H., Zhao, Q., Wang, B., Liu, Q. L., and Zhang, L. (2018). Chrysanthemum DgWRKY2 gene enhances tolerance to salt stress in transgenic chrysanthemum. Int. J. Mol. Sci. 19:2062. doi: 10.3390/ijms19072062

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996). DNA methylation and tumour suppressor genes. PLoS Med. 9, 9821–9826. doi: 10.1371/pnas.93.18.9821

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GDS2.0: An upgraded gene feature visualization server. Bioinformatics 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Huang, H., Cao, H., Niu, Y., and Dai, S. (2012). Expression Analysis of Nudix hydrolase genes in Chrysanthemum lavandulifolium. Plant Mol. Biol. Rep. 30, 973–982. doi: 10.11105/pnas.1308111

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5’-upstream regions of genes coding for sporamin and β-amylase from sweet potato. MGG Mol. Gen. Genet. 244, 563–571. doi: 10.1007/BF00282746

Jaffar, M. A., Song, A., Faheem, M., Chen, S., Jiang, J., Liu, C., et al. (2016). Involvement of CmWRKY10 in drought tolerance of chrysanthemum through the ABA-signaling pathway. Int. J. Mol. Sci. 17:693. doi: 10.3390/ijms17030693

Kang, D. R., Zhu, Y., Li, S. L., Ai, P. H., Khan, M. A., Ding, H. X., et al. (2021). Functional Analysis of a Novel Chrysanthemum WRKY Transcription Factor Gene Involved in Salt Tolerance. Plant Mol. Biol. Rep. 32, 282–289. doi: 10.1007/s11105-013-0639-3

Li, Q. Y., Zhong, M., Li, S., Pan, Y. Z., Jiang, B. B., Liu, G. L., Jin, J., et al. (2014). Overexpression of a chrysanthemum transcription factor gene, DgWRKY3, in tobacco enhances tolerance to salt stress. Plant Physiol. Biochem. 69, 27–33. doi: 10.1016/j.plaphy.2013.04.016

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25, 402–408. doi: 10.1016/s1046-2024(01)00090-4

Luo, X. L., Sun, X. L., Liu, B. D., Bai, X. C., Cai, H., et al. (2013). Co-expression of three WRKY genes from Chrysanthemum indicum modulates flowering time in Arabidopsis. Plant J. 74, 1455–1468. doi: 10.1111/j.1365-313X.2013.04883.x

Miguel, C., and Marum, L. (2011). An epigenetic view of plant cells cultured in vitro: Somatical variation and beyond. J. Exp. Bot. 62, 3713–3725. doi: 10.1093/jxb/err155

Nan, H., and Gao, L. Z. (2019). Genome-wide analysis of WRKY genes and their response to hormone and mechanical stresses in carrot. Front. Genet. 10.3389/fgene.2019.00363

Nguyen, T. K., and Lim, J. H. (2019). Tools for Chrysanthemum genetic research and breeding: Is genotyping-by-sequencing (GBS) the best approach? Hort. Environ. Biotechnol. 50, 625–635. doi: 10.1007/s13580-019-00660-5

Phukan, U. J., Jeena, G. S., and Shukla, R. K. (2016). WRKY transcription factors: Molecular regulation and stress responses in plants. Front. Plant Sci. 7:260. doi: 10.3389/fpls.2016.00760
Zhou, X., Jiang, Y., and Yu, D. (2011). WRKY22 transcription factor mediates dark-induced leaf senescence in Arabidopsis. Mol. Cells 31, 303–313. doi: 10.1007/s10059-011-0047-1

Zhu, L., Guan, Y., Liu, Y., Zhang, Z., Jaffar, M. A., Song, A., et al. (2020). Regulation of flowering time in chrysanthemum by the R2R3 MYB transcription factor CmMYB2 is associated with changes in gibberellin metabolism. Hortic Res. 7:96. doi: 10.1038/s41438-020-0317-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.