BMC Plant Biology

RESEARCH Open Access

Identification of *SDG* gene family members and exploration of flowering related genes in different cultivars of chrysanthemums and their wild ancestors

Ting Han[1†], Muhammad Ayoub Khan[1†], Yiming Wang[1], Wenchao Tan[1], Chenran Li[1], Penghui Ai[1], Wenqian Zhao[1], Zhongai Li[1] and Zicheng Wang[1*]

Abstract

The SET domain genes (*SDGs*) are significant contributors to various aspects of plant growth and development, mainly includes flowering, pollen development, root growth, regulation of the biological clock and branching patterns. To clarify the biological functions of the chrysanthemum *SDG* family, the *SDG* family members of four chrysanthemum cultivars and three related wild species were identified; their physical and chemical properties, protein domains and conserved motifs were predicted and analyzed. The results showed that 59, 67, 67, 102, 106, 114, and 123 *SDGs* were identified from *Chrysanthemum nankingense*, *Chrysanthemum lavandulifolium*, *Chrysanthemum seticuspe*, *Chrysanthemum* × *morifolium* cv. 'Hechengxinghuo', 'Zhongshanzigui', 'Quanxiangshuichang' and 'Jinbeidahong', respectively. The *SDGs* were divided into 5–7 subfamilies by cluster analysis; different conserved motifs were observed in particular families. The *SDGs* of *C. lavandulifolium* and *C. seticuspe* were distributed unevenly on 9 chromosomes. *SDG* promoters of different species include growth and development, photo-response, stress response and hormone responsive elements, among them, the cis-acting elements related to MeJA response had the largest proportion. The expression of chrysanthemum *SDG* genes was observed for most variable selected genes which has close association with important *Arabidopsis thaliana* genes related to flowering regulation. The qPCR results showed that the expression trend of *SDG* genes varied in different tissues at different growth stages with high expression in the flowering period. The *ClSDG29* showed higher expression in the flower and bud tissues, which indicate that *ClSDG29* might be associated with flowering regulation in chrysanthemum. In summary, the results of this study can provide a basis for subsequent research on chrysanthemum flowering time regulation.

Keywords Chrysanthemum, Allohexaploid, Diploid, *SDG*, Epigenetic, Flowering regulation

†Ting Han and Muhammad Ayoub Khan contributed equally to this work.

*Correspondence:
Zicheng Wang
wzc@henu.edu.cn

[1]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, Jinming Road, Kaifeng, Henan 475004, China

Introduction

The concept of epigenetics was first proposed to describe heritable alterations in gene activity that cannot be accounted for by modifications in DNA sequence during the processes of mitosis and meiosis [1], which mainly includes gene silencing mediated by DNA methylation, histone covalent modification, chromatin remodeling, non-coding RNA and paramutations [2]. In contemporary scientific discourse, there has been a surge of interest in the investigation of histone covalent modification. It can affect chromatin status through modification such as methylation, acetylation, phosphorylation, ubiquitination and ADP-ribosylation [3] to regulate gene transcription and expression. Methylation is one of the most important types of histone modification. Lysine residues at various sites of histone H3 (H3K4, H3K9, H3K27, H3K36, H3K79), position 20 of H4 (H3K20) and arginine residues at various sites of H3 (H3R2, H3R8, H3R17, H3R26) and position 3 of H4 (H4R3) were methylated to different degrees [4]. In contrast, histone lysine methylation is a relatively dynamic modification, whereas histone arginine methylation is more stable. The SET domain protein, named after three *Drosophila* genes (Su(var), E(z) and Trithorax), consists of about 130–150 amino acids and constitutes the specific catalytic site of histone lysine methyltransferases (HKMTs) [5]. The *SDG* family in *A. thaliana* is divided into 7 groups based on domain and function, including E(z), which catalyzes H3K27 methylation, SMYD and Ash, which catalyse H3K36 methylation, Trx and ATXR5/6, which catalyse H3K4 methylation, Suv, which catalyses H3K9 methylation, and SETD [6]. Studies have shown that *SDG* plays an important role in plant flowering [7], pollen development [8], root growth and development [9], regulation of the biological clock [10] and number of shoots [11]. In conclusion, the regulation of gene expression by histone lysine methylation plays an important role in the regulation of plant growth and development.

Chrysanthemum, an important ornamental plant possessing medicinal and food properties, originated in China and has rich flower colours and diverse flower types. As a short-day plant, it is affected by photoperiod, and the flowering period of chrysanthemum is mainly concentrated in autumn. The regulation of the flowering period is commonly achieved through the manipulation of light exposure duration, resulting in increased expenses related to labor and materials. With the publication of chrysanthemum genome, more and more gene family identification has been realized [12, 13]. Consequently, the utilization of genetic engineering techniques to control the flowering period has emerged as a prominent area of modern research, aiming to enhance economic advantages. *CDM111* and *CDM44* have been previously implicated in chrysanthemum flowering, and it has been speculated that they may regulate the opening time of flowers [14]. It was found that overexpression of *LFY* [15], *FPF1* [16], *SOC1* [17], *CO* [18], *FTL3* [19], *AP1/FUL* [20] and repression of *ERF3* [21] could lead to premature flowering in transgenic chrysanthemums. Additionally, it was discovered that flowering-related genes could be regulated by transcription factors to regulate flowering time. For example, *CmBBX8* promotes flowering by targeting *FTL1* [22]. It can be seen that the use of genetic engineering to improve the flowering time of chrysanthemum is based on the function of flowering-associated genes and related regulatory networks, which can provide genetic resources for molecular improvement of chrysanthemum flowering time and obtain new chrysanthemum with improved flowering time.

In the context of chrysanthemum flowering regulation, it has been observed that the overexpression of genes involved in flowering regulation, as well as the regulation of related genes by transcription factors, are significant considerations. Furthermore, the involvement of histone lysine methyltransferase in the regulation of flowering has been demonstrated in *Arabidopsis* and other plant species. Studies have shown that many members of the SET domain family can regulate plant flowering by methylating related genomic proteins in the flowering pathway. *AtCLF* (*AtSDG1*) not only modifies H3K27me3 at the *FT* site, thereby inhibiting *FT* expression, but also modifies H3K4 methylation at the *FLC* site to promote its expression, thereby delaying flowering and floral organ formation [23]. H3K4 methylation occurred at the *FLC* site modified by *ATXR3*(*AtSDG2*) [24], whereas H3K36 methylation occurred at *ASHH2*(*AtSDG8*) [25] and *ATXR7*(*AtSDG25*) [26], both of which promoted late flowering of *FLC* expression. The binding of *ATX1* and *FLC* loci is associated with H3K4me3, which may promote *FLC* expression and lead to late flowering, but the phenotype is not apparent after *ATX2* deletion [27]. During vernalisation, *AtSWN* (*AtSDG10*) can reduce the transcriptional efficiency of *FLC* and promote flowering [28]. Deletion of *AtSUVR5* resulted in delayed flowering in *Arabidopsis* [29]. *AtASHH1* (*AtSDG26*) may be involved in the transcriptional regulation of *SOC1* and promote flowering by binding to the *SOC1* promoter and enriching H3K36me3 in the chromatin region where *AtASHH1* resides [30], and the *sdg26* mutant showed a delayed flowering phenotype [31]. *SDG708* [32], which is highly homologous to *SDG26* in rice, also showed a late flowering phenotype after gene interference, suggesting that *SDG* can undergo different degrees of histone methylation modification at different loci of flowering-related genes, thereby regulating the flowering time. However, *SDG* has not been reported in chrysanthemum and related wild species. Given the potential biological

functions of *SDG* in the growth and development of higher plants, further analysis and identification are important.

Previously, the genomes of *Chrysanthemum nankingense*, *Chrysanthemum lavandulifolium* and *Chrysanthemum seticuspe* were obtained. At the same time, the transcriptomes of several chrysanthemum cultivars were obtained and the *SDG* family was identified based on these data. In this study, we attempted to explore the *SDG* family in chrysanthemum, including phylogenetic analysis, gene structure analysis, protein motif analysis, cis-acting element analysis, CpG island identification, and the expression of several important genes in different tissues at different growth stages. Exploring the role of *SDG* in flowering-related traits can not only lay the foundation for the research of chrysanthemum *SDG* family members in chrysanthemum flower formation mechanisms, but also provide genetic resources for the use of genetic engineering to improve chrysanthemum flowering time.

Materials and methods

Plant materials and experimental treatments

The genome data of *Chrysanthemum nankingense* [33] (http://www.amwayabrc.com),

Chrysanthemum lavandulifolium [34] (GenBank: JAHFWF000000000), *Chrysanthemum seticuspe* [35] and *C. morifolium* cv. 'Zhongshanzigui' were obtained from website. The transcriptome data of the *Chrysanthemum morifolium* cv. 'Jinbeidahong' [36], 'Hechengxinghuo', 'Quanxiangshuichang', 'Tangyujinqiu', 'Guohuadexin' and 'Mix9' were obtained from, Plant Germplasm Resources and Genetic Engineering, Henan University. *A. thaliana SDG* members were mainly downloaded from the TAIR database (https://www.arabidopsis.org/).

Identification and analysis of physicochemical properties of SDG family in different cultivars of chrysanthemum

Build a local protein database using Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/doc/blast-news/2023-BLAST-News.html), in this experiment, we choose blastp for blast [37], take *Arabidopsis* SDG amino acid sequence as the alignment sequence, and use Blast local alignment function to compare the different chrysanthemum varieties database established locally to obtain chrysanthemum SDG candidate protein, the value of E was set to 1e-5, other parameters can be defaulted. And at the same time, obtain the chrysanthemum SDG candidate protein from Pfam database [38] (http://pfam-legacy.xfam.org/). The HMM model of SDG family (PF00856) was downloaded from Pfam, and the hmmer search function in the HMMER3.0 program (http://hmmer.org/) was used to predict the chrysanthemum *SDG* members with PF00856 model as the probe. After removing the redundant sequences, all the sequences obtained were considered as the candidate chrysanthemum *SDG*. Then use the CDD program [39] (https://www.ncbi.nlm.nih.gov/cdd/), SMART [40] (http://smart.embl.de/) and InterPro (https://www.ebi.ac.uk/interpro/) to predict the screened *SDG* domain, remove the *SDG* gene with incomplete conservative domain, and finally obtain the chrysanthemum *SDG* gene families sequence. With reference to the method of Garg [41] *et* al, the ProtParam online website (https://web.expasy.org/protparam/) was used to predict the physicochemical properties of the identified *SDGs*, including the number of amino acids, molecular weight, theoretical PI, instability index, aliphatic index and grand average of hydropathicity [42]. And the WoLF PSORT online software (https://wolfpsort.hgc.jp/) was used to predict the subcellular localisation of the SDG proteins [43].

Classification of SDG proteins

Arabidopsis SDG protein sequence and chrysanthemum candidate *SDG* protein sequence were aligned by EMBL clustalw (https://www.ebi.ac.uk/Tools/msa/clustalo/), and other parameters build Neighbor-Joining (NJ) tree by default. Upload the synthesized newick format to the iTOL website. The iTOL (https://itol.embl.de/) provided visualization and beautification of the phylogenetic tree [44]. We set the required parameters according to the requirements [45].

Gene structure and Protein motif analysis

The gene structure (intron-exon) of *SDG* members were analyzed in the online database, Gene Structure Display Server(GSDS) (Gene Structure Display Server 2.0 (https://gsds.gao-lab.org/index.php) as described [46]. The analysis of protein conserved motifs was conducted with CDD program [39] (https://www.ncbi.nlm.nih.gov/cdd/), SMART [40] (http://smart.embl.de/) and InterPro (https://www.ebi.ac.uk/interpro/), the final figures were made by gene structure view (advanced) of TBtools software [47] according to our requirements.

Cis-acting elements in the promoter region of *SDG* members

The 2000 bp upstream sequence of the *SDG* genes was extracted using Gtf/Gff3 sequence extract by TBtools [47]. The promoter region and functional characteristics of cis-acting elements of different members of *SDG* were predicted online by PlantCARE [48] (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) website, and the visual mapping was performed by Excel.

CpG island analysis of the promoter region of *SDG* gene family

CpG island analysis was performed on the promoters and gene sequences of *SDG* using "Methyl Primer Express version 1.0" software (Methyl Primer Express Software v1.0 Quick Reference) [49]. The CpG island range was set to 200 bp ~ 2000 bp, the criterion of "C + G/total bases" was 50%, the ratio of CpG observed/CpG expected was 0.6, and other parameters were as defaulted.

Chromosomal mapping and gene syntenic analysis

The distribution of *SDG* gene on the chromosomes of *C. lavandulifolium* and *C. seticuspe* were analyzed using TBtools software. One step McScanX in TBtools was used to complete the syntenic analysis of *C. lavandulifolium* and *C. seticuspe* and to create the visualization [47].

Quantitative real-time polymerase chain reaction

According to the phylogenetic tree, 10 *CnSDGs* from *C. nankingense* and 13 *ClSDGs* from *C. lavandulifolium* were selected for qRT-PCR analysis, and 'Primer Premier 5' (https://www.premierbiosoft.com/products/products.html) software was designed to primers (Table S8). RNA was extracted from the hypocotyl (H), cotyledom (C), roots, stems, leaves in young (YSR, YSS, YSL), and adult (ASR, ASS, ASL) stage, steam, leaves, and buds, in budding (BSS, BSL and BSB) stage, steam, leaves and buds, in flowering (FSS, FSL and FSB) stage and first flower stage (FFS) in *C. nankingense* and *C. lavandulifolium* using the RNAprep Pure Plant Plus Kit (spin column). The sampling period of *C. lavandulifolium* is detailed in the paper [49]. The cDNA was synthesized using the Takara kit (RR037A). A Roche LightCycler® 480 PCR assay was used with chrysanthemum actin as the reference gene. PCR amplification procedure: pre-denaturation at 95℃ for 5 min, denaturation at 95℃ for 10 s, annealing at 60℃ for 10 s, 45 cycles. Relative gene expression levels were calculated by $2^{-\Delta\Delta CT}$. The qPCR analysis contains three biological replicates and three technical replicates. The data expressed as the average SEM of at least three independent experiments. Statistical significance was determined by one-way ANOVA. The GraphPad Prism software (version 9.0) was used to analyze all data with appropriate statistical analysis methods, as described in the legend. The "*" shows significance at $p < 0.05$, "**" shows significance at $0.001 < p < 0.01$, and "***" show significance at $p < 0.001$.

Results

Identification of *SDGs* in different chrysanthemum cultivars

The *SDG* of different chrysanthemum varieties was identified by local blast and HMMER3.0, and verified by Pfam, CDD and SMART which are conservative domain prediction softwares. According to the genomes of *C. nankingense*, *C. lavandulifolium* and *C. seticuspe* and the transcriptome data of *C. morifolium* cv. 'Zhongshanzigui', 'Hechengxinghuo', 'Quanxiangshuichang' and 'Jinbeidahong',59, 67, 67, 106, 102,114 and 123 candidate *SDG* genes were selected as initial genes for the study. It can be seen that the number and classification of *SDG* genes varies in different chrysanthemum species.

Analysis of physical and chemical properties and prediction of subcellular localization of chrysanthemum *SDG*

ExPASy was used to analyse the amino acid sequences of *SDG* family members from different chrysanthemum materials, such as molecular weight, isoelectric point and hydrophilicity, etc. WoLF PSORT was used to predict the subcellular localization, and the results are shown in supplement table 1-7.

The protein length of 59 *SDG* genes in *C. nankingense* ranged from 168 to 2242 amino acids, and the molecular weight ranged from 19.62 to 254.97 KDa. The theoretical isoelectric point ranged from 4.13 to 9.41, 13 SDG proteins were stable proteins and 46 SDG proteins were unstable proteins. Subcellular localization prediction results showed that the majority of *SDGs* were localized in the nucleus, while only *CnSDG33* was found in the peroxisome, plasma membrane and golgi apparatus. The protein length of 67 *SDG* genes in *C. lavandulifolium* ranged from 111 to 2356 amino acids, the molecular weight from 13.28 to 268.32 KDa, the theoretical isoelectric point from 4.63 to 10.1, 13 SDG proteins were stable proteins, 64 SDG proteins were unstable proteins. The subcellular localization prediction results showed that most of the *SDGs* located in the nucleus and only 1 *SDG* was located in the plasma membrane. The protein length of 67 *SDG* genes in *C. seticuspe* ranged from 121 to 2356 amino acids, the molecular weight ranged from 13.67 to 268.30 KDa, the theoretical isoelectric point ranged from 4.05 to 9.47, 16 SDG proteins were stable proteins, 61 SDG proteins were unstable proteins. Subcellular localization prediction results showed that the majority of *SDGs* were located in the nucleus, and only *CsSDG56* was located in the cytoskeleton. The protein length of 106 *SDG* genes in *C. morifolium* cv. 'Zhongshanzigui' ranged from 216 to 1886 amino acids, the molecular weight was 24.2 to 268.70 KDa, the theoretical isoelectric point range was 4.62 to 9.49, 24 SDG proteins were stable proteins and the rest were unstable proteins. The subcellular localization prediction results showed that the proportion of *SDGs* located in the nucleus was the highest and only Cm*SDG29* was located in the cell wall. The protein length of 114 *SDG* genes in *C. morifolium* cv. 'Quanxiangshuichang' ranged from 190 to 2355 amino acids, and their molecular weight ranged from 20.83 to 207.29 KDa. The theoretical isoelectric point ranged from 4.26

to 9.24. A total of 16 SDG proteins were stable and the rest were unstable proteins. Subcellular localization prediction results showed that the majority of *SDGs* were located in the nucleus, and only 2 *SDGs* were located in the cytoskeleton and endoplasmic reticulum. The protein length of 102 *SDG* genes in *C. morifolium* cv. 'Hechengxinghuo' ranged from 147 to 1516 amino acids, the molecular weight ranged from 16.63 to 167.23 KDa, the theoretical isoelectric point ranged from 4.28 to 9.49, 10 SDG proteins were stable proteins and the rest were unstable proteins. The prediction results of subcellular localization showed that the proportion of *SDG* located in the nucleus was the largest, and the proportion of *SDG* located in the golgi apparatus was the smallest. The protein length of 123 *SDG* genes in *C. morifolium* cv. 'Jinbeidahong' ranged from 197 to 1486 amino acids, and the molecular weight was 22.93 to 164.70 KDa. The theoretical isoelectric point ranged from 4.65 to 9.76. 14 SDG proteins were stable and the rest were unstable. Subcellular localization prediction results showed that the majority of *SDGs* were located in the nucleus, and 3 *SDGs* were located in the peroxisome. The fat coefficient of the SDG proteins identified from diploid wild species and hexaploid related species was less than 100, and they were all hydrophilic proteins.

Phylogenetic analysis of *SDG* and classification of SDG proteins

By employing the established *Arabidopsis* SDG protein as a reference and utilizing the EMBL to cluster and analyze the SDG proteins from various chrysanthemum accessions, it is possible to attain distinct classification (Fig. 1). The classification of different chrysanthemum cultivars varies. Since 'Hechengxinghuo' and 'Jinbeidahong' didn't have the fourth subgroup, their *SDG* included only six subgroups. Conversely, the other chrysanthemum varieties can be classified into seven subgroups. The specific *SDG* classification is shown in the table S1. As shown in the table, there was variation in the number of candidate genes across different chrysanthemum varieties within distinct subgroups of *SDG*. *C. lavandulifolium* exhibited varying numbers of *SDGs* across different groups; specifically groups I-VII had 4, 6, 7, 3, 16, 8, and 23 *SDGs*, respectively. In the specie *C. nankingense*, the groups I-VII exhibited SDGs 4, 5, 7, 1, 18, 7, and 17, respectively. In the case of *C. seticuspe*, groups I-VII exhibited *SDGs* genes are 5, 4, 8, 2, 18, 10, and 20, respectively. The cultivar *C. morifolium* cv. 'Zhongshanzigui' exhibited varying numbers of *SDGs* across different groups, specifically 9, 8, 14, 3, 44, 9, and 19 for groups I-VII, respectively. In the cultivar *C. morifolium* cv. 'Quanxiangshuichang', the different groups, namely I-VII, exhibited 9, 4, 14, 1, 55, 8, and 23 significant developmental genes (*SDGs*), respectively. Except IV, which is 0, I-VII is 11, 11, 14, 48, 5, 14 and 10, 20, 33, 32, 5, 23 respectively in *C. morifolium* cv. 'Hechengxinghuo' and 'Jinbeidahong'. Similar to Arabidopsis, the number of candidate genes in the fourth subgroup was the lowest across six chrysanthemum cultivars, and the seventh subgroup has the largest number of *SDG* in *C. lavandulifolium* and *C. seticuspe*, but the third subgroup in *C. morifolium* cv. 'Jinbeidahong' has the largest number, and the remaining

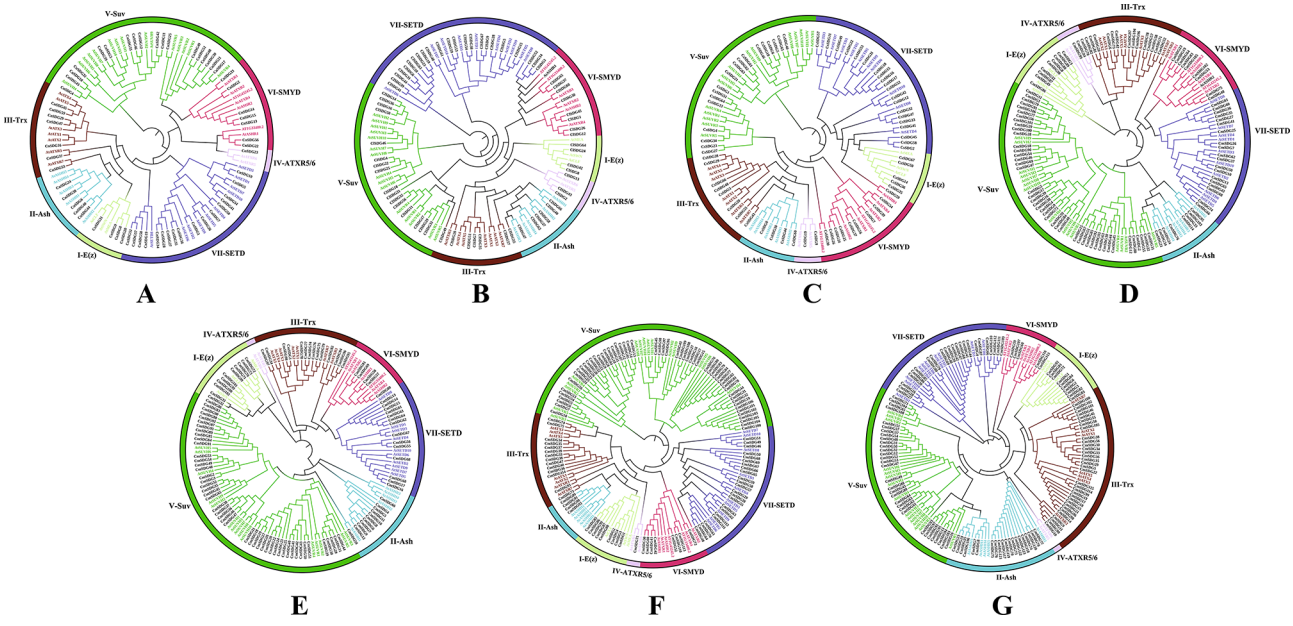


Fig. 1 Phylogenetic trees of SDG proteins of different chrysanthemum accessions and 49 AtSDG proteins of*A. thaliana* on the basis of amino acid sequences. **A**, **B**, **C**, **D**, **E**, **F**, **G** represent the evolutionary analysis of the *SDGS* of *C. nankingense*, *C. lavandulifolium* and *C. seticuspe* and the transcriptome data of *C. morifolium* cv. 'Zhongshanzigui', 'Hechengxinghuo', 'Quanxiangshuichang' and 'Jinbeidahong'

three varieties have the fifth subgroup with more SDG members.

Gene structure of *SDG* members and protein conserved motif analysis

The number and distribution of introns and exons in the *SDG* genes of diploid *C. nankingense*, *C. lavandulifolium*, *C. seticuspe* and haploid *C. morifolium* cv. 'Zhongshanzigui' were analysed using GSDS software. The number of introns and exons varied between species and families (Figs. 2, 3, 4, 5 and 6). In the seven *SDG* sub-groups of *C. nankingense*, the number of exons ranged from 1 to 24 and the number of introns from 0 to 23. In the seven *SDG* subgroups of *C. lavandulifolium*, the number of exons ranged from 1 to 35 and the number of introns ranged from 0 to 34. Among the seven *SDG* subgroups in the *C. seticuspe*, number of exons ranged from 1 to 23 and the number of introns from 0 to 22. Based on the number, *C. nankingense* and *C. seticuspe* are more similar. At the same time, we found that the number of exons and introns was not completely consistent in some sequences with close evolutionary relationship. In *C. morifolium* cv. 'Zhongshanzigui', the number of exons in 106 *SDGs* ranges from 1 to 26, and the number of introns from 0 to 25, and the number of exons and the number of introns in different genes may have different variable shapes, leading to different roles in the plant.

The detection of protein conserved motifs showed that the motifs contained in different families are not similar, but all SDG protein sequences contain SET conserved motifs, and the most conserved SET motifs of different families are in different positions in the gene, such as in the first, third, fourth and fifth families, which are near the C-terminus of the gene. In the second and seventh families, near the N-terminus of the genes, it was also found that the types and locations of motifs in the same family were similar. These motifs may be closely related to function, and different motifs in different locations may play different roles.

Analysis of the *SDG* gene promoter region in three wild chrysanthemum species

A total of 67 *ClSDGs*, 67 *CsSDGs* and 58 *CnSDGs* promoter sequences were selected to analyze cis-acting elements and CpG islands using PlantCARE database and "Methyl Primer Express version 1.0" software. However, the promoter of *CnSDG23* in *C. nankingense* was excluded because the sequence was undetermined. In terms of cis-acting element analysis, the number of cis-acting elements and the number of genes involved are shown in the Fig. 7. The analysis shows that all *SDG* promoters contain core elements, TATA-box and CAAT-box, which are mainly responsible for the accuracy and efficiency of gene transcription initiation, as well as

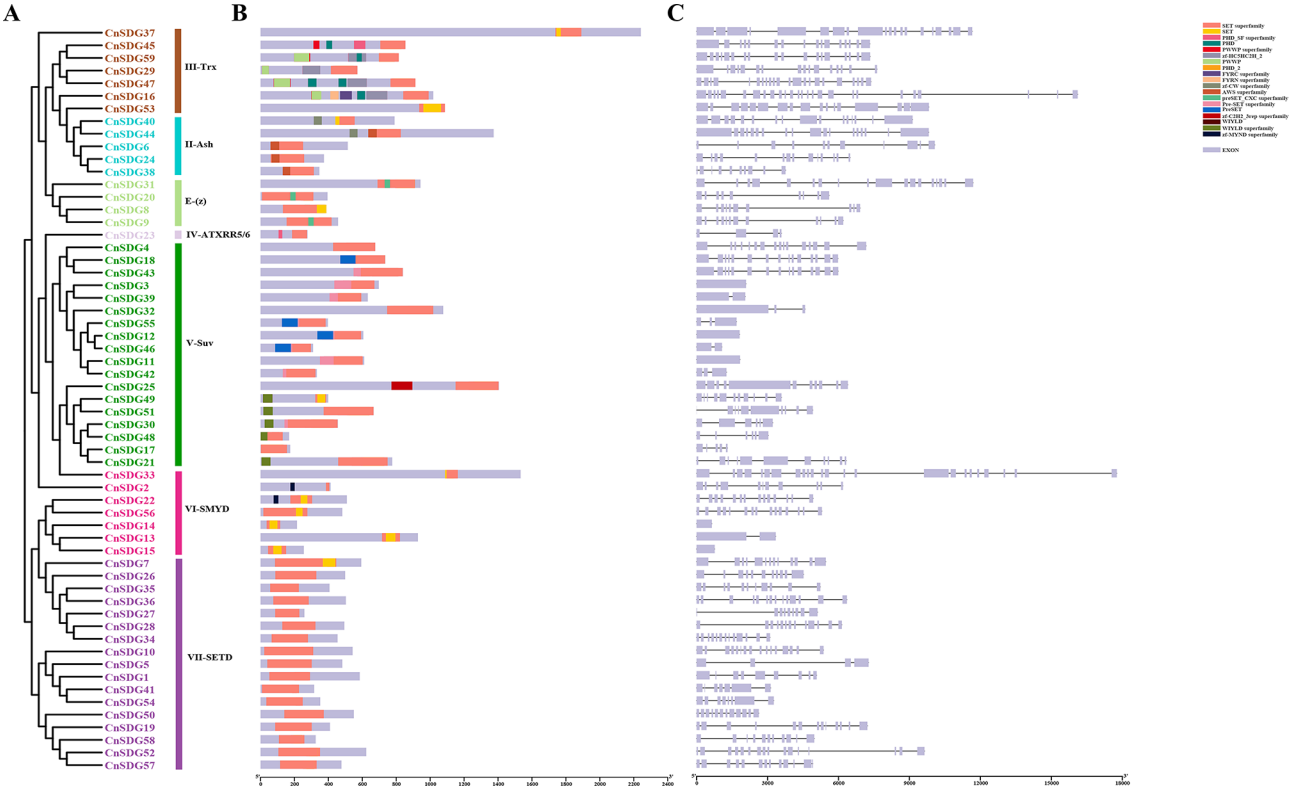


Fig. 2 Schematic diagram of conserved motif and the intron-exon structure of *SDG* members according to the phylogenetic relationship in *C. nankingense*

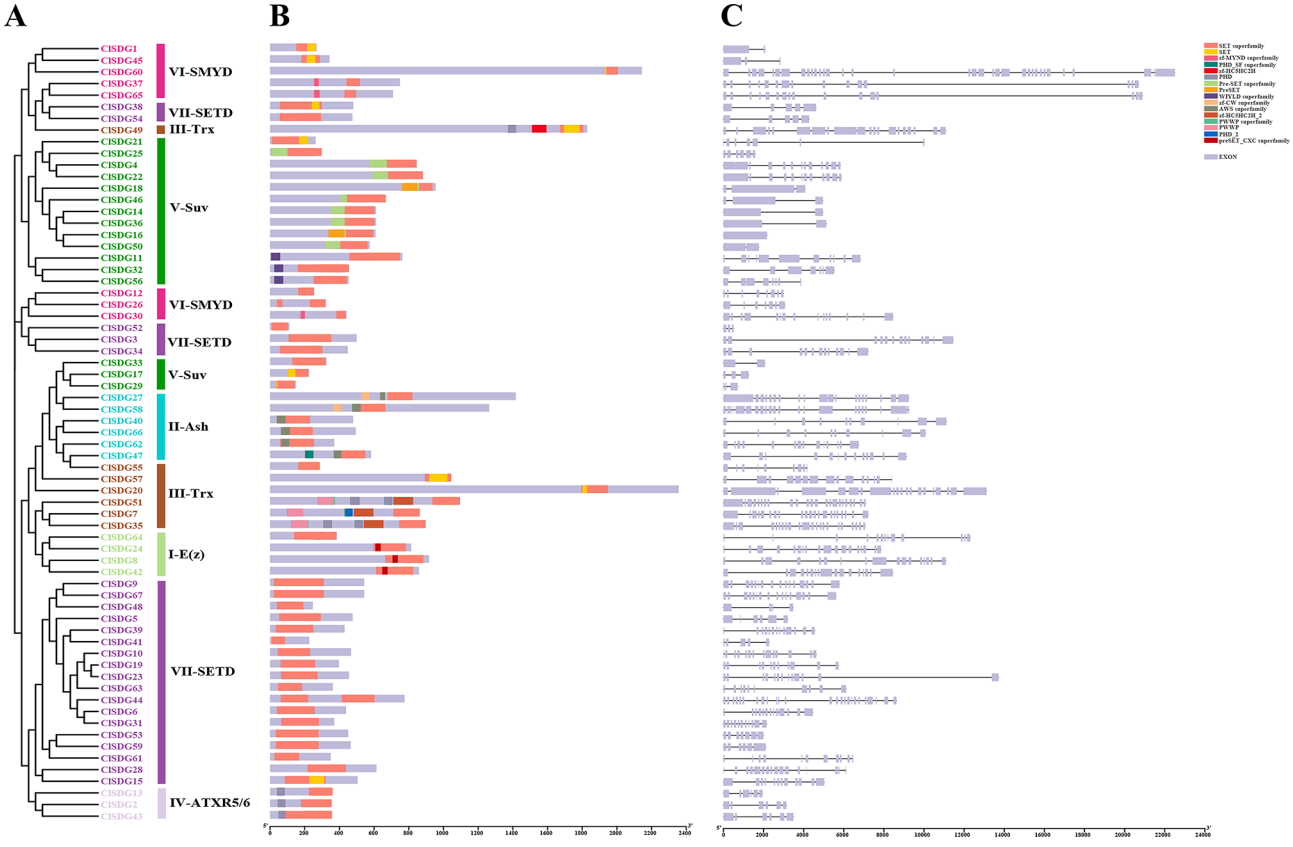


Fig. 3 Schematic diagram of conserved motif and the intron-exon structure of *SDG* members according to the phylogenetic relationship in *C. lavandulifolium*

growth and development, hormone response, stress response and light response elements. Regarding plant growth and development responsiveness, there are eight cis-acting elements that are involved in endosperm expression (AACA motif and GCN4-motif), meristem differentiation (CAT-box), palisade mesial cell differentiation (HD-zip), zein olysin metabolism (O2-site), seed-specific regulation (RY-element), circadian rhythm and germ expression (as-1), of which as-1 is the largest. Elements in the hormone response category include methyl jasmonate (AAGAA motif, CGTCA motif and TGACG motif), auxin (AuxRR and TGA element), gibberellin (GARE motif, P box and TATC box), abscisic acid (ABRE), ethylene (ERE) and salicylic acid (TCA element) response elements. Among them, most of the cis-acting elements were responsive to MeJA, followed by ABA, and may indicate that chrysanthemum *SDG* is induced by MeJA and ABA. There are 9 cis-acting elements which response to stress includes; anaerobic induction (ARE), defensive stress (STRE and TC-rich), low temperature (LTR), drought (MBS and CCAAT-box), dehydration (DRE) and trauma (WRE3 and Wun-motif). The proportion of ARE related to anaerobic induction is the highest, followed by STRE related to defence and immunity. In the photo response category, there were 20 cis-acting elements, but all of them were not present in the 3 diploid chrysanthemums, the ATC motif and the GTGGC motif was absent in the *C. nankingense*, while other 18 elements were present across the species, among them, Box4 was the most abundant, which may play a key role in photo regulation.

A1, B1, and C1 represent the number of components in *C. nankingense*, *C. lavandulifolium* and *C. seticuspe* respectively. A2, B2, and C2 represent the number of genes involved in each promoter cis-acting element in *C. nankingense*, *C. lavandulifolium* and *C. seticuspe* respectively.

In terms of CpG island analysis, CpG islands are located in the promoter region of genes and the region which can regulate genes. In diploid *C. nankingense*, *C. lavandulifolium*, and *C. seticuspe*, CpG islands were found in 12, 12, and 10 *SDG* genes, respectively. Most of the members are detected as having 1 CpG island, except *CnSDG14* and *CnSDG15*. *C. nankingense* was observed to possess CpG islands in the promoter region of three genes, while 5 genes exhibited islands in both the cross-promoter and gene regions, and six genes displayed islands in the gene body region. Interestingly, there had 3 CpG types in *CnSDG15*. The islands exhibited a range in overall length, ranging from 229 base pairs

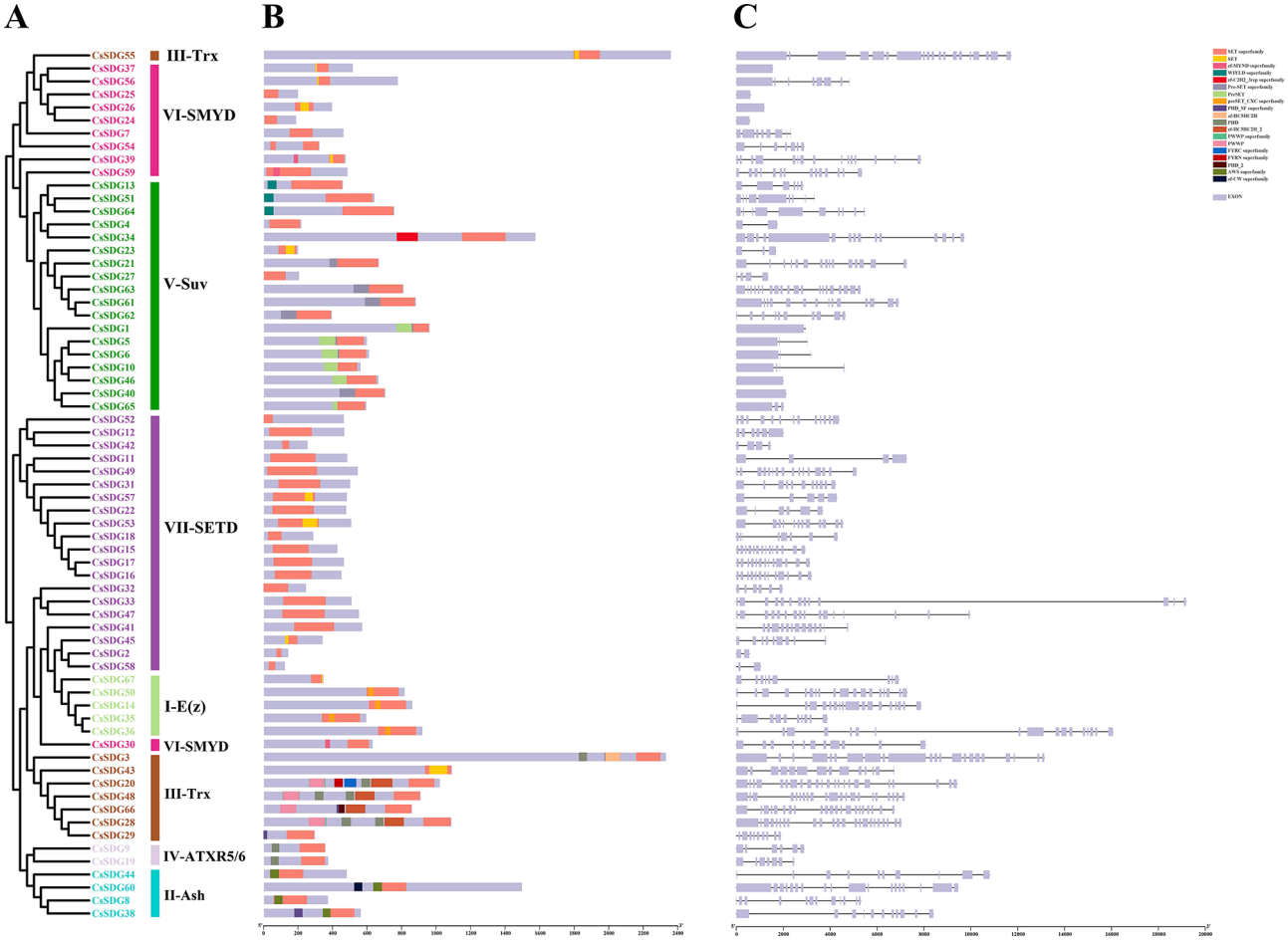


Fig. 4 Schematic diagram of conserved motif and the intron-exon structure of *SDG* members according to the phylogenetic relationship in *C. seticuspe*

(bp) to 836 bp. In *C. lavandulifolium*, four *SDG* members had islands in the promoter region; 3 *SDG* members had islands in the promoters and gene regions; five *SDG* members had islands in the gene body regions; and the CpG islands ranged in length from 274 to 608 bp. Regarding *C. seticuspe*, four *SDG* members had islands in the promoter regions, *CsSDG28* members had islands in promoters and gene regions, and five *SDG* members had in gene body regions. The length of islands within the species ranged from 199 to 817 bp (Table 1). Hence, there exists variation in the number, distribution, and size of CpG islands within *SDGs* across different species, potentially indicating a correlation with species evolution.

Chromosomal synteny analysis

To further elucidate the evolutionary relationship between *SDG* genes in different chrysanthemum varieties, we performed collinearity analysis of *SDG* gene families in *C. lavandulifolium* and *C. seticuspe* (Fig. 8). The results showed that *C. lavandulifolium* and *C. seticuspe* had homologous gene pairs. Several *SDG* genes in *C. lavandulifolium* were identified as homologs of a single *SDG* gene in *C. seticuspe*. For example, *ClSDG32* and *ClSDG56* in *C. lavandulifolium* were both derived from *CsSDG13*, while several *SDG* genes in *C. seticuspe* were derived from a single *SDG* gene in *C. lavandulifolium*. If *CsSDG66* and *CsSDG48* in *C. seticuspe* were derived from the same *ClSDG35* in *C. lavandulifolium*, it is speculated that these genes may have similar functions. This suggests that the expansion of this gene family may have happened before *C. lavandulifolium* and *C. seticuspe* differentiation.

Gene duplication is one of the main driving forces for the evolution of genomes and genetic systems, and the occurrence of replication events promotes the generation of new genes and new functions. TBtools software was used to analyse the location and gene replication events of *ClSDGs* and *CsSDGs* (Figs. 9 and 10), and the results showed that *ClSDGs* and *CsSDGs* were unevenly distributed on nine chromosomes. Among them, the 8th chromosome of *C. lavandulifolium* possess 20 *SDGs*, and had highest distribution genes. The 2nd chromosome of *C. seticuspe* was found to have the most distributed number of genes, with a total of 13 *SDGs*. The 9th chromosome

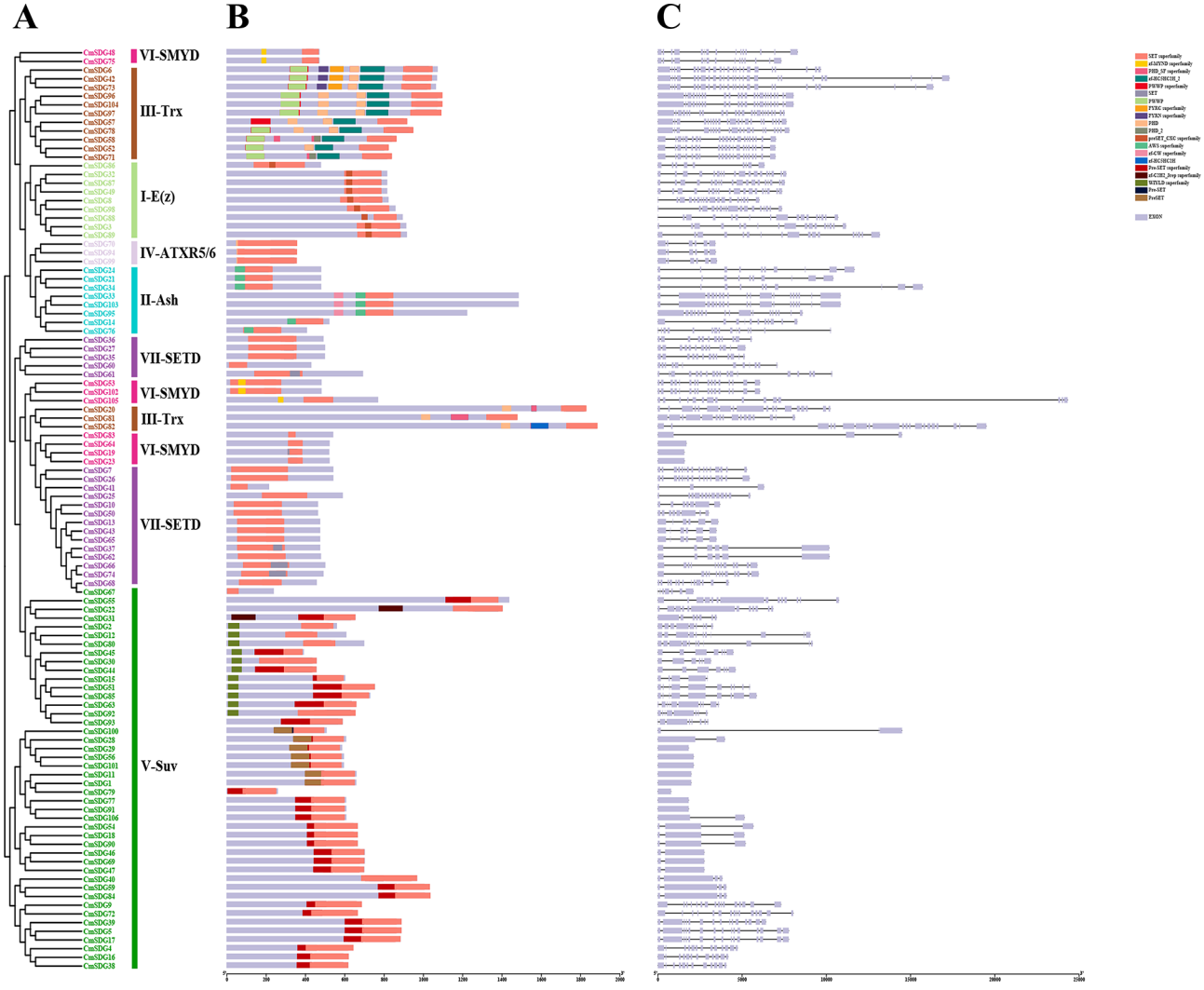


Fig. 5 Schematic diagram of conserved motif and the intron-exon structure of *SDG* members according to the phylogenetic relationship in *C. morifolium* cv.'Zhongshanzigui'

of *C. lavandulifolium* had the least distributed genes with a total of 3 *SDGs*, while in *C. seticuspe*, the 6th chromosome had 4 *SDGs*, the least in the specie. There were 1 and 3 gene replication events observed on the 3rd and 8th chromosomes of *C. lavandulifolium*, while 3 gene replication events were detected on the 2nd, 4th and 9th chromosomes of *C. seticuspe*, which provided a driving force for the evolution of chrysanthemum.

Expression pattern analysis

RNA was extracted from various plant tissues, including the hypocotyl, cotyledon, root, stem, leaves, and flowers, at different stages of growth, namely young, mature, and flowering. Subsequently, complementary DNA (cDNA) was synthesized following the guidelines provided by the manufacturer. The qRT-PCR was employed to evaluate the expression levels of 10 *SDG* genes in *C. nankingense* and 13 *SDG* genes in *C. lavandulifolium* across diverse plant tissues at different developmental stages (Figs. 11 and 12). The 10 genes which were selected for qRT-PCR in *C. nankingense* on the basis of their close association with *CnFT* includes *CnSDG25*, *CnSDG16*, *CnSDG6*, *CnSDG38*, *CnSDG40*, *CnSDG31*, *CnSDG44*, *CnSDG24*, *CnSDG53* and *CnSDG37*, respectively. The 13 *SDG* genes which were selected for expression analysis in *C. lavandulifolium* on the basis of their close association with *ClFT* comprised *ClSDG47*, *ClSDG40*, *ClSDG29*, *ClSDG66*, *ClSDG8*, *ClSDG27*, *ClSDG33*, *ClSDG57*, *ClSDG24*, *ClSDG58*, *ClSDG20*, *ClSDG62* and *ClSDG17*.

The qRT-PCR expression analysis focused on the *SDG* genes in chrysanthemum species that exhibited clustering with significant Arabidopsis genes involved in the regulation of flowering, as determined by phylogenetic tree analysis. The qPCR results showed that the expression of different *SDG* genes varied among different growth stages. The expression of *CnSDG25* was highest

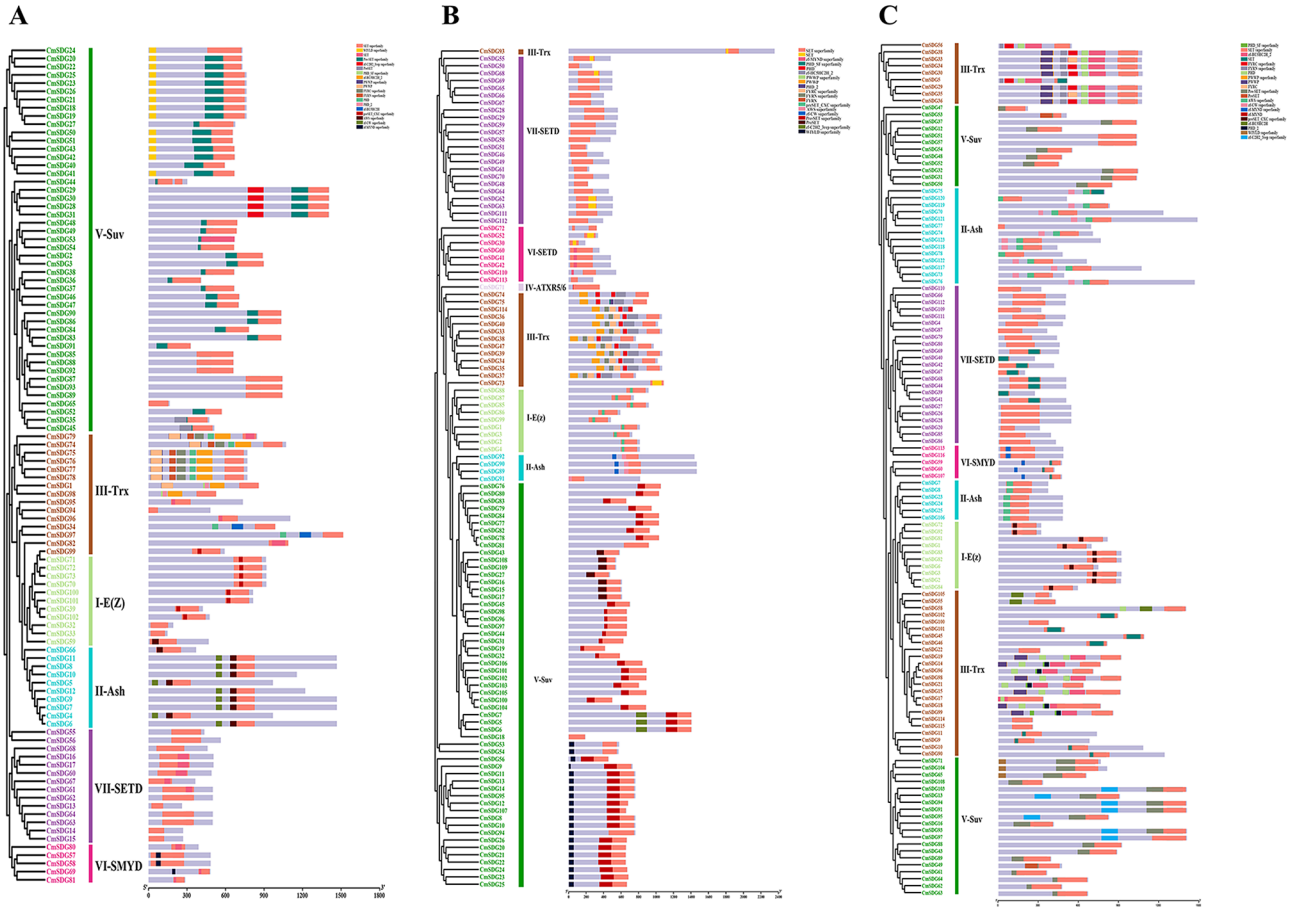
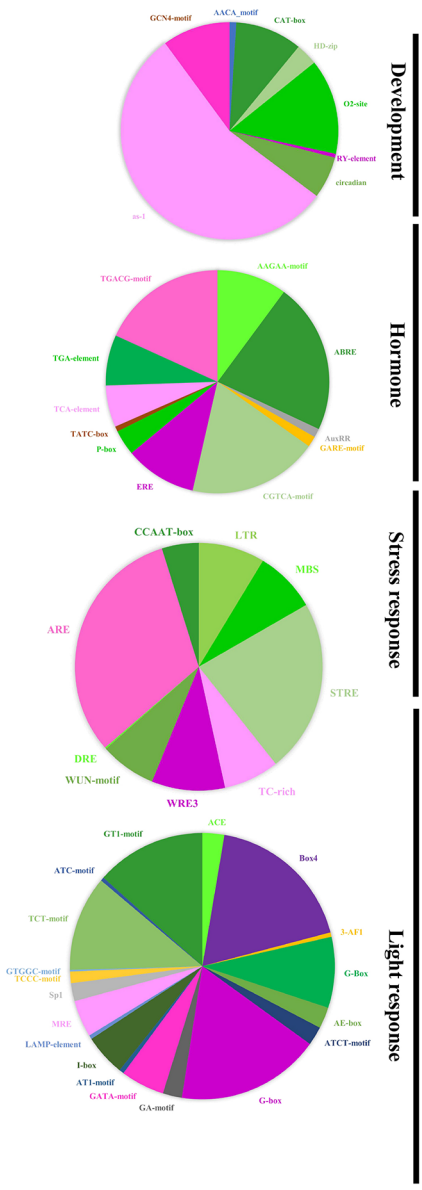


Fig. 6 Schematic diagram of conserved motif of *SDG* members according to the phylogenetic relationship in *C. morifolium* cv.'Hechengxinghuo','Quanxiangshuichang','Jinbeidahong'

in the flower buds at the flower bud stage, and most of the other *SDGs* were highest in the stems at the flowering stage of *C. nankingense*. The expression level of the *FT* gene in *C. nankingense* was higher in leaves than in stems at the flowering stage, while the expression level was lowest in buds at the flowering stage. The expression trend of *CnSDG25* in different tissues at flowering stage was the same as that of *FT*. It suggests that *CnSDG25* might be related to the regulation of flowering genes in the *C. nankingense*. Interestingly, 11 of the 13 *ClSDGs* reached their highest expression level in the stem at the flowering stage, suggesting that *SDGs* may affect the flowering process by regulating stem expression. The *FT* expression level of *C. lavandulifolium* was highest in leaves at young age, followed by stems at flowering stage. The *FT* expression level gradually increased from small bud to flowering, and the expression level of *ClSDG47* also increased with bud growth, while the expression level of *ClSDG40* showed an opposite trend. Therefore, *ClSDG40* and *ClSDG47* may have contrasting effects on *FT*. The expression levels of *CnSDG16*, *CnSDG24*, *CnSDG37*, *CnSDG40*, *CnSDG44* and *CnSDG53* in leaves at the flowering stage were second only to those in stems at the same growth stage. The expression levels of *ClSDG29* were higher in small buds and first flower, but decreased in medium buds. The *ClSDG29* gene exhibits a strong association with AtSUVR5 (Fig. 1B), which serves as a crucial regulator of the flowering process. Knockout of *SUVR5* causes a delayed flowering phenotype, while quintuple mutants lacking *SUVR1, SUVR2, SUVR3, SUVR4*, and *SUVR5* do not exhibit any other observable changes in phenotype. These findings demonstrate the importance of *ClSDG29* in the regulatory network that controls flowering [50]. Besides, *ClSDG29* also possess CpG island in the promoter region (Table 1), which increase its importance to play role in flowering regulation.

Different *SDGs* play different regulatory roles in different tissues at different stages. From the analysis of the overall qPCR results, it seems that *CnSDG25*, *ClSDG47*, *ClSDG40* and *ClSDG29* may be involved in the regulation of flowering time in chrysanthemum. Interestingly, among the selected 10 *CnSDGs* and 13 *ClSDGs*, 80% of *SDGs* were expressed at the highest level in stems during the flowering period, which may be related to



	A1	A2	B1	B2	C1	C2
AACA_motif	2	2	1	1	4	3
CAT-box	19	17	26	21	25	18
HD-zip	5	5	8	7	10	8
O2-site	28	19	36	26	36	24
RY-element			2	2	2	2
circadian	7	6	21	17	16	14
as-1	115	46	147	55	123	52
GCN4-motif	22	16	30	22	19	16
AAGAA-motif	72	39	69	44	73	42
ABRE	109	43	171	61	182	58
AuxRR	8	8	9	9	10	9
GARE-motif	13	13	8	8	12	9
CGTCA-motif	115	46	147	55	135	55
ERE	80	37	76	42	64	35
P-box	29	24	25	21	24	17
TATC-box	3	3	5	5	8	5
TCA-element	44	33	38	28	46	30
TGA-element	39	23	64	37	52	34
TGACG-motif	115	46	147	55	123	52
CCAAT-box	17	16	20	17	38	24
DRE	1	1	1	1	1	1
LTR	34	23	60	35	41	27
MBS	39	28	37	28	49	35
STRE	115	50	120	53	118	55
TC-rich	36	24	38	21	39	30
WRE3	36	28	47	35	66	39
Wun-motif	32	23	46	26	36	24
ARE	142	50	167	55	182	60
3-AF1	4	4	5	5	2	2
AAAC-motif	5	5	4	4	1	1
ACE	15	14	21	17	19	17
AE-box	17	16	19	15	17	14
AT1-motif	4	4	6	6	2	2
ATC-motif			4	4	5	5
ATCT-motif	13	12	16	14	18	17
Box4	114	50	129	54	132	58
G-Box	46	29	52	37	79	35
G-box	84	44	149	58	128	54
GA-motif	24	11	13	12	11	9
GATA-motif	36	24	37	26	37	28
I-box	27	20	42	26	35	25
LAMP-element	3	3	1	1	6	6
MRE	30	23	30	23	32	24
Sp1	11	8	25	16	9	8
TCCC-motif	9	8	11	10	9	7
TCT-motif	71	42	92	47	75	43
GT1-motif	77	41	103	50	96	47
GTGGC-motif			2	2	2	2

Fig. 7 Information of cis-acting elements in *SDG*

stem-related regulation in addition to flowering. *SDG* members are extensively involved in all stages of growth and development of *C. nankingense* and *C. lavandulifolium*, especially during flowering.

The expression of *SDG* in diploid *C. nankingense* and *C. lavandulifolium* showed dynamic changes in different tissues and developmental stages. To further explore the expression patterns of *SDG* in cultivated chrysanthemum species, *SDG* genes were screened and their expression patterns were analysed in the transcriptome data of *C. morifolium* 'Tangyujinqiu', 'Guohua dexin' and 'Mix9'. The samples of *C. morifolium* 'Tangyujinqiu' include young leaves, adult leaves and flower buds, all corresponding to different short day periods. The expression results showed that the *SDG* genes of *C. morifolium* 'Tangyujinqiu' had organisational variations (Fig. 13). Overall, the *SDG* gene expression of several members was low in leaves and high in buds, such as *CmSDG1*, *CmSDG2*, *CmSDG6*, *CmSDG8* and *CmSDG13*. The expressions of *CmSDG6* and *CmSDG67* were up-regulated in leaf tissues at different developmental stages. *CmSDG4* and *CmSDG13* showed a trend of decline, especially *CmSDG14*, *CmSDG62*, *CmSDG66*, *CmSDG68* and *CmSDG70* high at the beginning of the short day inducing expression, speculated that the five members involved in short day induced flowering regulation. In flower bud tissues, the expression levels of *CmSDG5* and *CmSDG15* decreased with flower opening,

Table 1 Details of each of the *SDG* genes with a CpG island in different chrysanthemum accessions

NO	*SDG* gene	Island	Start	End	Length	Region
C. nankingense						
1	*CnSDG4*	1	-25	527	552	Promoter + Gene Body
2	*CnSDG7*	1	337	1496	1160	Gene Body
3	*CnSDG10*	1	-1416	-1017	400	Promoter
4	*CnSDG15*	1	-6	224	230	Promoter + Gene Body
5	*CnSDG27*	1	70	359	290	Gene Body
6	*CnSDG34*	1	-1507	-1013	495	Promoter
7	*CnSDG37*	1	831	1325	495	Promoter
8	*CnSDG39*	1	33	451	419	Gene Body
9	*CnSDG42*	1	-1626	-1232	395	Promoter
10	*CnSDG45*	1	14	508	495	Gene Body
11	*CnSDG59*	1	-6	527	533	Promoter + Gene Body
C. lavandulifolium						
1	*ClSDG1*	1	-96	179	275	Promoter + Gene Body
2	*ClSDG12*	1	-21	329	350	Promoter + Gene Body
3	*ClSDG20*	1	926	1477	552	Gene Body
4	*ClSDG21*	1	-2000	-1786	215	Promoter
5	*ClSDG26*	1	-366	-137	230	Promoter
6	*ClSDG29*	1	-561	-47	515	Promoter
7	*ClSDG33*	1	-1581	-992	590	Promoter
8	*ClSDG41*	1	445	644	200	Gene Body
9	*ClSDG42*	1	-1754	-1203	552	Promoter
10	*ClSDG43*	1	-405	166	571	Promoter + Gene Body
11	*ClSDG46*	1	14	508	495	Gene Body
12	*ClSDG51*	1	33	508	476	Gene Body
13	*ClSDG64*	1	1299	-880	419	Promoter
C. seticuspe						
1	*CsSDG4*	1	-141	299	440	Promoter + Gene Body
2	*CsSDG6*	1	-1982	-1374	609	Promoter
3	*CsSDG18*	1	-1731	-1502	230	Promoter
4	*CsSDG21*	1	33	451	419	Gene Body
5	*CsSDG22*	1	-670	-138	533	Promoter
6	*CsSDG23*	1	-741	-422	320	Promoter
7	*CsSDG25*	1	235	434	200	Gene Body
8	*CsSDG28*	1	71	489	419	Gene Body
9	*CsSDG32*	1	-1144	-1502	365	Promoter
10	*CsSDG33*	1	-766	-253	514	Promoter
11	*CsSDG44*	1	-709	147	856	Promoter + Gene Body
12	*CsSDG46*	1	375	1154	780	Gene Body
13	*CsSDG55*	1	926	1477	552	Gene Body
14	*CsSDG65*	1	33	451	419	Gene Body

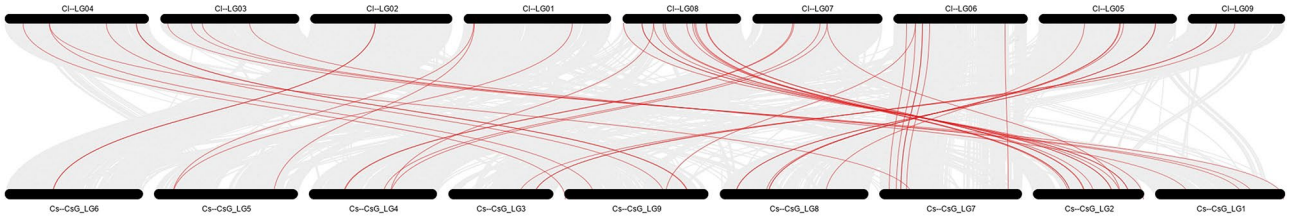


Fig. 8 Synteny of *SDG* genes between *C. lavandulifolium* and *C. seticuspe*

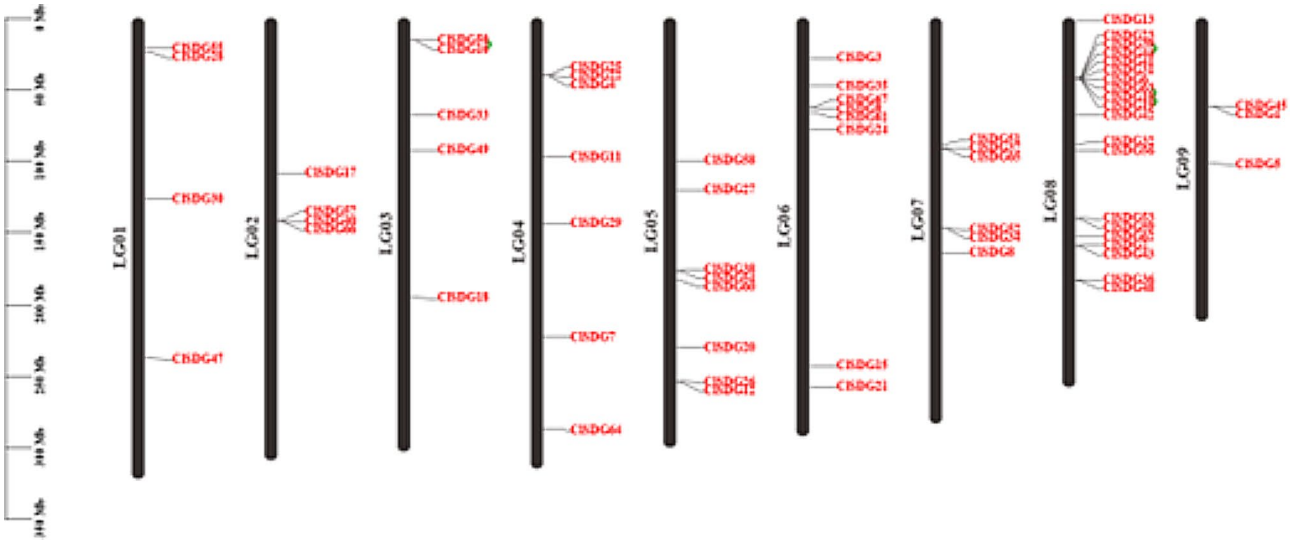


Fig. 9 The chromosome distribution of the *SDG* genes in *C. lavandulifolium*

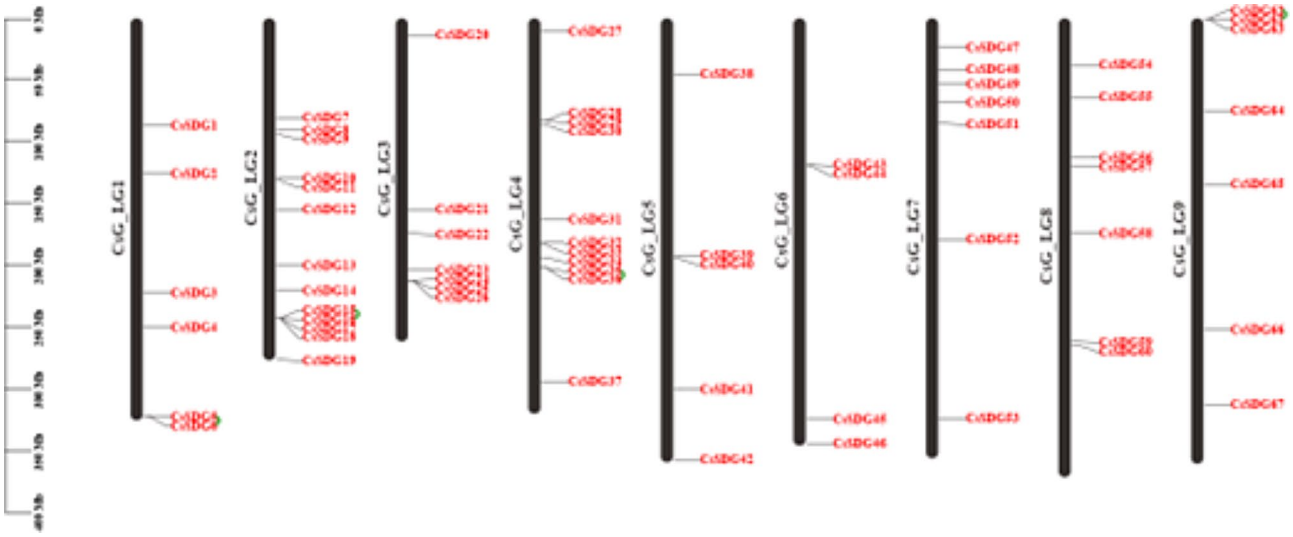


Fig. 10 The chromosome distribution of the *SDG* genes in *C. seticuspe*

while the expression levels of *CmSDG61* increased with flower opening, suggesting that these three members may be related to flower development. In addition, five *SDGs* (*CmSDG27*, *CmSDG46*, *CmSDG48*, *CmSDG57* and *CmSDG64*) were not expressed. The traditional Chinese *C. morifolium* 'Guohuadexin' cultivated species has variety degradation, after tissue culture, it can realize the recovery of fine phenotype. Analysis of leaf transcriptome data from mother plant, first generation and third generation degraded varieties showed that the expression of *SDG* in leaves showed five trends (Fig. 14), which were first increased and then decreased (Fig. 14A), first decreased and then increased (Fig. 14B), increased (Fig. 14C), decreased (Fig. 14D) and no expression (Fig. 14E) with the initial generation and several sub-generations. It is speculated that the rejuvenation of chrysanthemum after tissue culture is closely related to *SDG* activation related genes. *C. morifolium* 'Mix9' is a new strain bred by our research group that can achieve flowering under tissue culture conditions and does not require short day treatment. Analysis of the expression levels of all *SDG* members of this strain at different stages of flower development showed that *SDG* genes also showed period specificity at different stages of flower development in test tube flowering (Fig. 15), and most *SDG* members (*CmSDG1*, *CmSDG9*, *CmSDG10*, *CmSDG11* and *CmSDG12* etc.) showed a trend of first decreasing and then increasing expression levels. Considering that this stage is a key period for flower induction, some members may be closely related to flower development traits. According to the analysis of gene expression in different tissues of the three

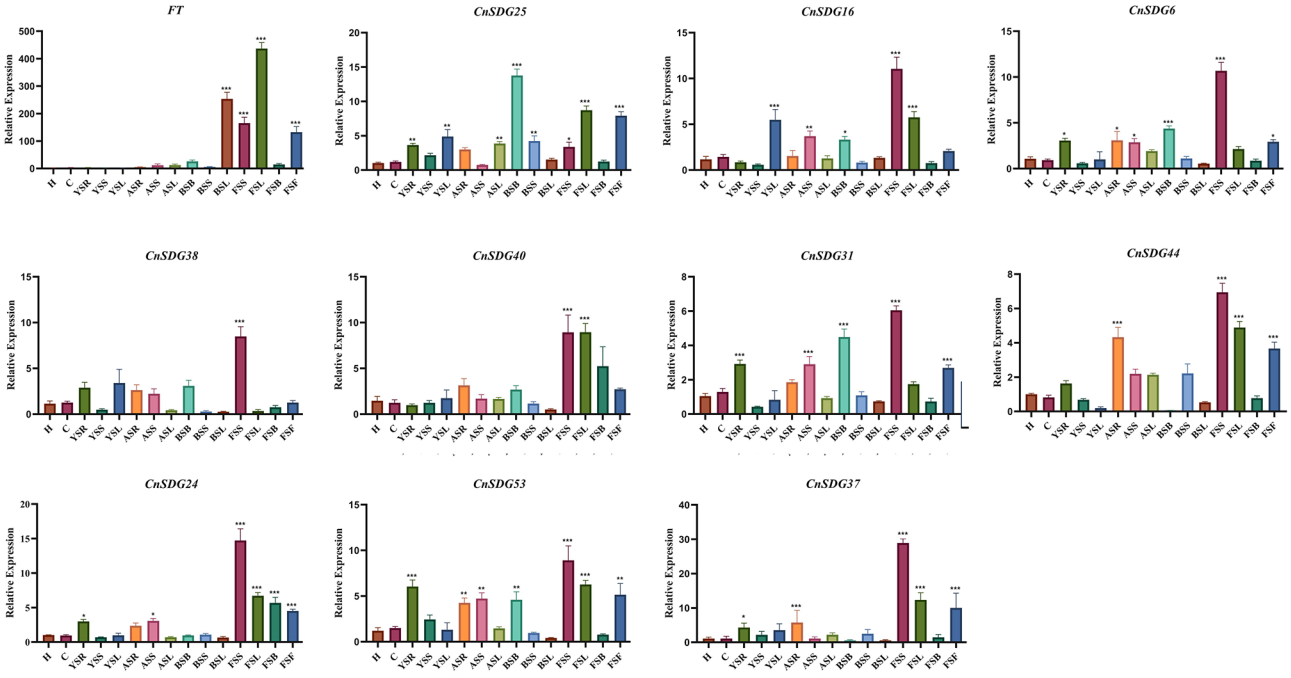


Fig. 11 Expression analysis of the selected *SDG* genes in different tissues of *C. nankingense*

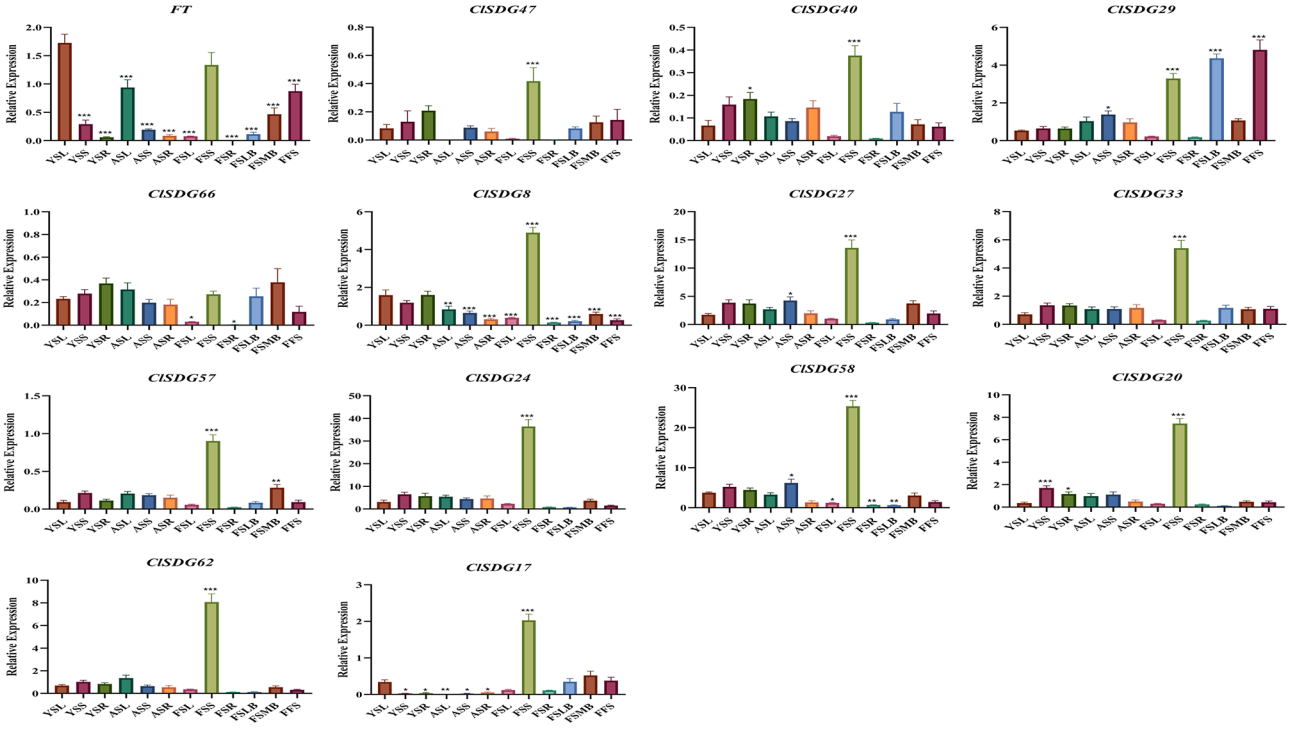


Fig. 12 Expression analysis of the selected *SDG* genes in different tissues of *C. lavandulifolium*

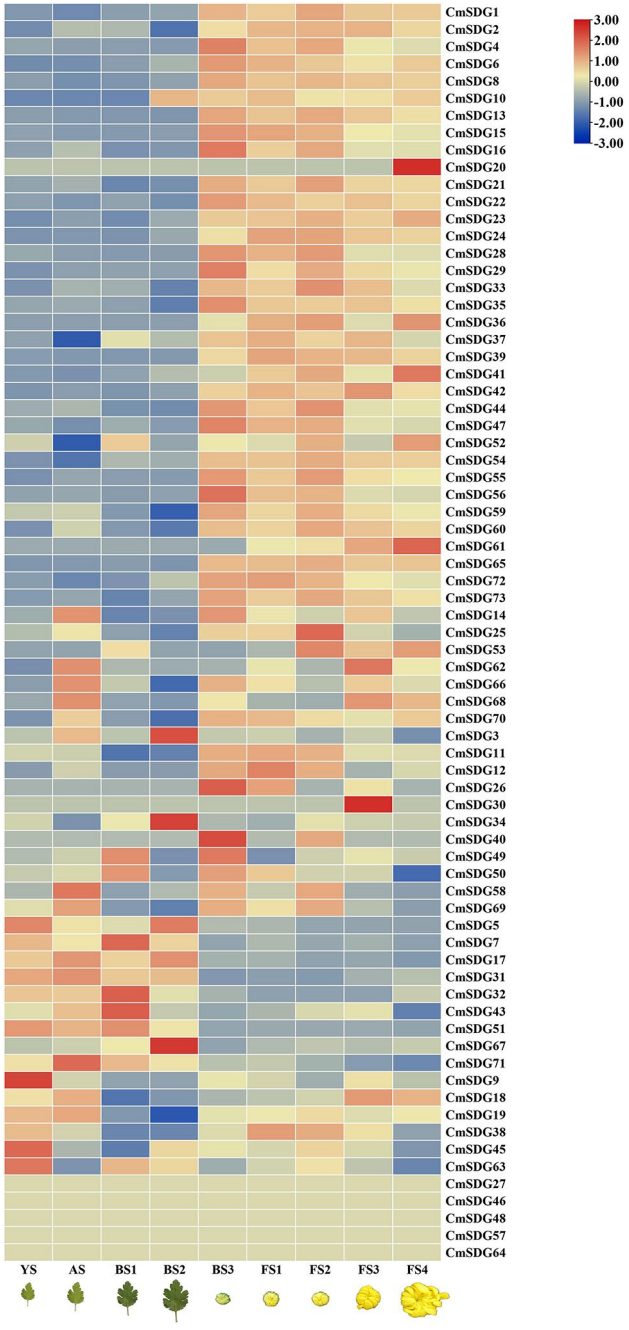


Fig. 13 Expression analysis of the *SDG* genes in subgeneration process of *C. morifolium* cv. 'Tangyujinqiu'. YS, AS, BS1, BS2 represent leaves at young age, leaves at adult age, and leaves after 25 days of short day treatment, leaves after 35 days of short day treatment respectively. BS3, FS1, FS2, FS3 and FS4 represent different stages of flower development after 50 days of short sunlight

chrysanthemum cultivars at different developmental stages, the expression of different members of *SDG* is tissue-specific and stage-specific, indicating that *SDG* plays an important role in the growth and development of chrysanthemum.

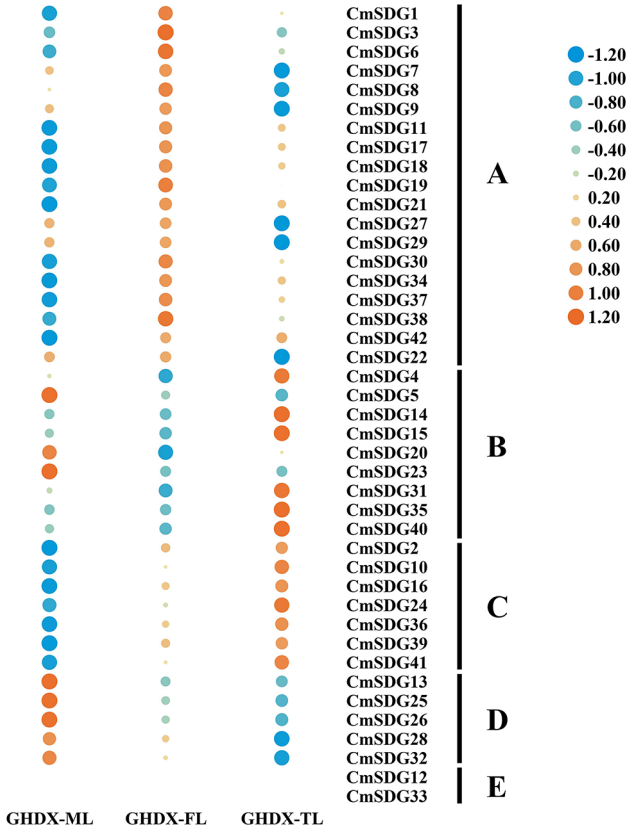


Fig. 14 Expression analysis of the *SDG* genes in sub-generation process of *C. morifolium* cv. 'Guohuadexin'. ML, FL and TL represent the leaves of the mother plant, the first generation recovery and the third generation respectively

Discussion

The *SDG* gene family in different ploidy chrysanthemum plants contains different groups and number of members

In plants, SDG proteins can dynamically regulate chromatin condensation by mediating histone methylation at lysine residues [51]. *SDG*-mediated histone methylation is involved in many biological processes, such as floral organogenesis [8]. In this study, *SDG* family members were identified from 7 materials based on genomic and transcriptomic data. Overall, the number of members in hexaploid chrysanthemum is about twice that of diploid wild species. Given that the source of hexaploid material is a haploid genome and transcriptome, there may be incomplete gene sequences or some sequences may not be transcribed. From the perspective of family classification, different families contained different conservative motifs, among which the *SDGs* of *C. morifolium* cv. 'Hechengxinghuo' and 'Jinbeidahong' were divided into 6 families, and the rest of the materials were divided into 7 families like *Arabidopsis*. Although the number distribution of *SDG* members in different plants and different families was different, the distribution trend was

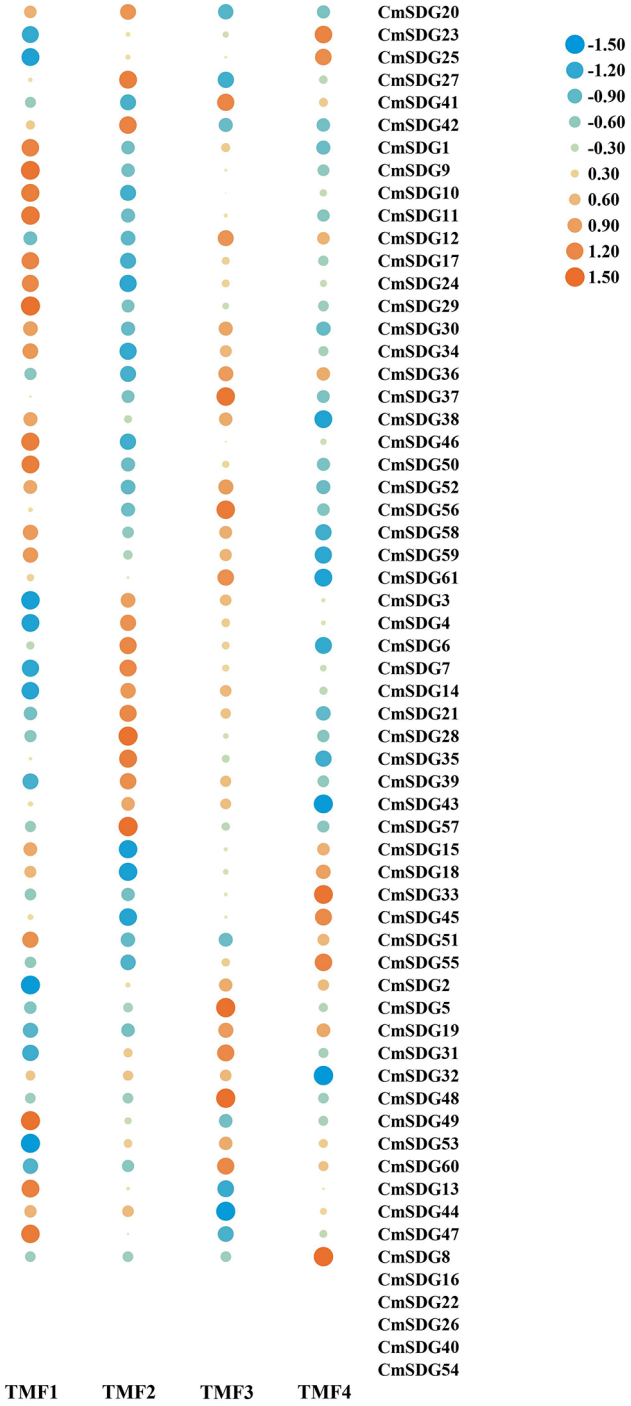


Fig. 15 Expression analysis of the *SDG* genes in sub-generation process of *C. morifolium* cv. 'Mix9'. TMF1, TMF2, TMF3 and TMF4 represent 0, 10, 20 and 30 days of tissue culture respectively

generally the same. Among the 7 families of *A. thaliana*, the number of *SDGs* in family IV accounted for the least and the number of *SDGs* in family V accounted for the most, while the number of *SDGs* in family IV of the 7 chrysanthemum materials accounted for the least among the 7 families. Among the five materials (*C. nankingense*, *C. morifolium* cv. 'Hechengxinghuo', 'Quanxiangshuichang', 'Jinbeidahong' and 'Zhongshanzigui), the number of *SDGs* in family V accounted for the largest proportion, while *C. lavandulifolium* and *C. seticuspe* had the highest number of *SDGs* in group VII, followed by the number of *SDGs* in group V. The lowest number in family IV and the highest number in family V were found in *Brassica napus* [6], *Zea mays* [52], *Citrus sinensis* [53] and *Camellia sinensis* [54], while the lowest number in family III and the highest number in family V were found in *Dimocarpus longan* Lour [55]. It can be seen that the family IV is relatively rare in most species, and the family V is relatively large. Even in the same species, such as chrysanthemum, the proportion trend of different material is different. Most studies show that families V and VII are more abundant and families IV are less abundant. In conclusion, the *SDG* quantity of 3 diploid wild chrysanthemum species was significantly lower than that of 4 hexaploid chrysanthemum cultivars due to differences in species ploidy. In the wild species, the number of members of *C. lavandulifolium* and *C. seticuspe* is relatively close, but the number distribution of different families of the three species is different. Among the four cultivars, the total number of members was similar, and the number of family members and the number of members within each family were slightly different, although they came from haploid genome and transcriptome, respectively. In conclusion, *SDG* in species with different ploidy are both similar and different, and the difference in the number of gene families is mainly due to the different ploidy. Considering that thee hexaploid chrysanthemum varieties were not sequenced, the number of *SDGs* identified by the transcriptome data may be missing, but the overall trend is that the number of genes in hexaploids is much larger than that in diploids. This is consistent with the finding that the number of *SDG* members in tetraploid *Brassica napus* is much larger than that in diploid *Brassica oelercea* and *Brassica rapa* [6]. It can be seen that there is a positive correlation between the different ploidies and the number of gene families. After the natural hybrid chromosome is doubled, the number of genes increases, providing the basis for gene variation and the generation of new genes.

The regulatory factors for chrysanthemum *SDGs* may be diverse

Nowadays, more and more studies have been conducted from the outside to the inside, from the shallow to the deep, from the external factors affecting the phenotype to the internal molecular regulatory mechanisms, and the study of upstream and downstream genes has become an important clue for research. Analysis of cis-acting elements showed that the *SDG* promoters of *C. nankingense*, *C. lavandulifolium* and *C. seticuspe* contain stress

response elements responsible for gene transcription initiation and growth regulation. Methyl jasmonate (MeJA), a naturally occurring plant growth regulator, may be involved in the production of secondary metabolites such as volatiles and may play a role in defence mechanisms against biotic and abiotic stresses [56]. The proportion of MeJA-responsive elements was largest in the 3 diploid chrysanthemum plants, and the proportion of MeJA-responsive elements was also largest in *DlSDG* [55] and *BnSDG* [6], suggesting that *SDGs* of different species may be induced by MeJA, and the content of MeJA in the *Arabidopsis* mutant *sdg26* was higher than that of the wild type. This suggests that *SDGs* may reverse regulate MeJA levels. In addition to hormone response elements, there are also stress response elements such as low temperature, drought resistance and light response elements, so it is speculated that the *SDG* identified in this study may be affected by the regulation of related genes. As mentioned above, it has been reported that many members of the *SDG* family perform histone methylation modification on genes related to the flowering pathway to regulate flowering time. In this study, the downstream genes of chrysanthemum *SDG* were not verified, but this must be the direction of future efforts.

SDGs expressed differentially in different organizations of different species at various stages

SDGs play a key role in plant growth and development. The tissue expression of *SDGs* has been extensively studied in many species and shows different expression patterns in each species, indicating their different roles in plant development. Transcriptome data can also explain the expression trend of genes. *SDGs* are expressed in buds, leaves, roots, stems and flowers of *camellia sinensis* [54] and in leaves, flowers and fruits of *Citrus sinensis* [53]. In *Brassica rapa* [57], *SDGs* are expressed in roots, stems and leaves, and the content of *SDGs* in stems of family III is higher than that in other tissues, while the expression level of *SDGs* in roots of family II is higher than that in other tissues, so the expression patterns of *SDGs* in different families may be divergent in various plants. The qPCR results of the diploid wild species of *C. nankingense* and *C. lavandulifolium* showed that 80% of the 10 *CnSDGs* and 13 *ClSDGs* were highly expressed at the stem in flowering stage, and there was also high expression at the stem in flowering stage in *Brassica napus* [6], which may indicate that the *SDGs* highly expressed in the stem are closely related to the flowering phenotype of plants. The expression trend of *CnSDG25* and *FT* in stem, leaf, bud and flower during the flowering period was consistent, suggesting that *CnSDG25* may regulate the flowering period of *C. nankingense*. The expression of *ClSDG47* and *FT* showed the same trend from bud to flower, while *ClSDG40* showed the opposite trend, which may have an effect on the flowering stage for *C. lavandulifolium*. In conclusion, the expression of *SDGs* changes with the change of species, tissues and time. Among them, *CnSDG25* in *C. nankingense* and *ClSDG47* and *ClSDG40* in *C. lavandulifolium* may regulate flowering time. The specific regulatory mechanism needs to be verified by further experiments. The transcriptome data of the hexaploid varieties 'Tangyujinqiu', 'Guohuadexin' and 'Mix9' not only showed that the expression of *SDG* varied at different developmental stages in different tissues, but also found that the expression of *SDG* was closely related to short-day flowering, tissue culture rejuvenation and test tube flowering without short day. In conclusion, the analysis of *SDG* expression in diploid and hexaploid plants indicated that *SDG* expression is tissue- and time-specific and may play an important role in the growth and development of chrysanthemum through different regulatory pathways.

Conclusion

In this study, 59, 67, 67, 102, 114, 123 and 106 *SDG* genes were identified respectively in *C. nankingense*, *C. lavandulifolium*, *C. seticuspe*, *C. morifolium* cv. 'Hechengxinghuo', 'Quanxiangshuichang', 'Jinbeidahong' and 'Zhongshanzigui'. It was found that the promoter region of *SDG* contains a variety of regulatory elements involved in growth and development, stress response, hormone response and light response. According to the types of conserved motifs, *SDGs* can be divided into up to seven subfamilies, with different materials containing different specific members. The expression patterns in different tissues of *C. nankingense* and *C. lavandulifolium* at different periods indicated that *CnSDG25*, *ClSDG40* and *ClSDG47* might regulate the flowering stage of chrysanthemum, while the expression levels of other *SDGs* were highest at the flowering stage, which might regulate the flowering.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05465-y.

Supplementary Material 1

Acknowledgements

We wish to thank the National Key Research and Development Program (2020YFD1000401) and Science and Technology Forest Promotion Program (K23040Y).

Author contributions

HT finished refining and writing the manuscript. HT performed the experiments and make all figures. HT, WYM, TWC, and LCR analyzed the experiments data. WZC, APH, ZWQ, LZA, and Muhammad Ayoub Khan refined and revised the manuscript. The author(s) read and approved the final manuscript.

Funding
This work was supported by the Key Science and technology (2202001), Science and Technology Forest Promotion Program of Henan (YLK202309), Natural Science Foundation of Henan (242300421548), National Key Research and Development Program (2020YFD1000401) and Science and Technology Forest Promotion Program (K23040Y).

Data availability
All data and material generated or analyzed during this study are included in this published article (and its supplementary materials).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 December 2023 / Accepted: 29 July 2024
Published online: 29 August 2024

References

1. Yang X, Lay F, Han H, Jones PA. Targeting DNA methylation for epigenetic therapy. Trends Pharmacol Sci. 2010;31(11):536–46.
2. Shi J, Dong A, Shen W-H. Epigenetic regulation of rice flowering and reproduction. Front Plant Sci. 2015;5.
3. Sura W, Kabza M, Karlowski WM, Bieluszewski T, Kus-Slowinska M, Pawełoszek Ł, et al. Dual role of the histone variant H2A.Z in transcriptional regulation of stress-response genes. Plant Cell. 2017;29(4):791–807.
4. Stallcup MR. Role of protein methylation in chromatin remodeling and transcriptional regulation. Oncogene. 2001;20:3014–20.
5. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol. 2005;6(8).
6. Sehrish S, Sumbal W, Xie M, Zhao C, Zuo R, Gao F et al. Genome-wide identification and characterization of SET Domain Family genes in Brassica napus L. Int J Mol Sci. 2022;23(4).
7. Liu K, Yu Y, Dong A, Shen WH. SET DOMAIN GROUP701 encodes a H3K4-methytransferase and regulates multiple key processes of rice plant development. New Phytol. 2017;215(2):609–23.
8. Cartagena JA, Matsunaga S, Seki M, Kurihara D, Yokoyama M, Shinozaki K, et al. The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Dev Biol. 2008;315(2):355–68.
9. Yao X, Feng H, Yu Y, Dong A, Shen W-H. SDG2-Mediated H3K4 methylation is required for proper Arabidopsis Root Growth and Development. PLoS ONE. 2013;8(2).
10. Malapeira J, Khaitova LC, Mas P. Ordered changes in histone modifications at the core of the Arabidopsis circadian clock. Proceedings of the National Academy of Sciences. 2012;109(52):21540-5.
11. Dong G, Ma D-P, Li J. The histone methyltransferase SDG8 regulates shoot branching in Arabidopsis. Biochem Biophys Res Commun. 2008;373(4):659–64.
12. Zhu L, Ding Y, Wang S, Wang Z, Dai L. Genome-wide identification, characterization, and expression analysis of CHS Gene Family members in Chrysanthemum nankingense. Genes. 2022;13(11).
13. Tian C, Zhai L, Zhu W, Qi X, Yu Z, Wang H et al. Characterization of the TCP Gene Family in Chrysanthemum nankingense and the Role of CnTCP4 in Cold Tolerance. Plants. 2022;11(7).
14. Shchennikova AV, Shulga OA, Immink R, Skryabin KG, Angenent GC. Identification and characterization of four Chrysanthemum MADS-Box genes, belonging to the APETALA1/FRUITFULL and SEPALLATA3 subfamilies. Plant Physiol. 2004;134(4):1632–41.
15. Shao HS, Li JH, Zheng XQ, Chen SC. Cloning of the LFY cDNA from Arabidopsis thaliana and its transformation to Chrysanthemum morifolium. J Integr Plant Biol. 1999;41:268–71.
16. Pi W, Li MY. Preliminary studies on transgenic Chrysanthemum with FPF1 gene mediated by Agrobacterium tumefaciens. J Southwest Univ. 2007;29:70–3.
17. Jin Z, Yu X, Pei Y. Ectopic expression of AtSOC1 gene driven by the inducible promoter rd29A, causes early flowering in Chrysanthemum. Sci Hort. 2020;261.
18. Tian SB, Lin GY, Zheng CS, Sun X, Ren HY, Wen LZ. Cloning and expression of CmCO and CmFT of floral development genes in Chrysanthemum. Acta Horticulturae Sinica. 2011;38:1129–38.
19. Nakano Y, Takase T, Takahashi S, Sumitomo K, Higuchi Y, Hisamatsu T. Chrysanthemum requires short-day repeats for anthesis: gradual CsFTL3 induction through a feedback loop under short-day conditions. Plant Sci. 2019;283:247–55.
20. Shulga OA, Mitiouchkina TY, Shchennikova AV, Skryabin KG, Dolgov SV. Overexpression of AP1-like genes from Asteraceae induces early-flowering in transgenic Chrysanthemum plants. In Vitro Cellular & Developmental Biology - Plant. 2011;47(5):553 – 60.
21. Cheng H, Yu Y, Zhai Y, Wang L, Wang L, Chen S, et al. An ethylene-responsive transcription factor and a B-box protein coordinate vegetative growth and photoperiodic flowering in chrysanthemum. Plant Cell Environ. 2023;46(2):440–50.
22. Wang L, Sun J, Ren L, Zhou M, Han X, Ding L, et al. CmBBX8 accelerates flowering by targeting CmFTL1 directly in summer chrysanthemum. Plant Biotechnol J. 2020;18(7):1562–72.
23. Köhler C, Villar CBR. Programming of gene expression by polycomb group proteins. Trends Cell Biol. 2008;18(5):236–43.
24. Yun J-Y, Tamada Y, Kang YE, Amasino RM. ARABIDOPSIS TRITHORAX-RELATED3/SET DOMAIN GROUP2 is required for the Winter-Annual habit of Arabidopsis thaliana. Plant Cell Physiol. 2012;53(5):834–46.
25. Zhao Z, Yu Y, Meyer D, Wu C, Shen W-H. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nat Cell Biol. 2005;7(12):1256–60.
26. Berr A, Xu L, Gao J, Cognat Vr, Steinmetz A, Dong A, et al. SET DOMAIN GROUP25Encodes a histone methyltransferase and is involved inFLOWERING LOCUS CActivation and repression of Flowering. Plant Physiol. 2009;151(3):1476–85.
27. Pien Sp, Fleury D, Mylne JS, Crevillen P, Inzé D, Avramova Z, et al. ARABIDOPSIS TRITHORAX1 dynamically RegulatesFLOWERING LOCUS CActivation via Histone 3 Lysine 4 Trimethylation. Plant Cell. 2008;20(3):580–8.
28. Wood CC, Robertson M, Tanner G, Peacock WJ, Dennis ES, Helliwell CA. The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. Proceedings of the National Academy of Sciences. 2006;103:14631–6.
29. Chen D-H, Qiu H-L, Huang Y, Zhang L, Si J-P. Genome-wide identification and expression profiling of SET DOMAIN GROUP family in Dendrobium catenatum. BMC Plant Biol. 2020;20(1).
30. Berr A, Shafiq S, Pinon V, Dong A, Shen W-H. The trxG family histone methyltransferase SET DOMAIN GROUP 26 promotes flowering via a distinctive genetic pathway. Plant J. 2015;81(2):316–28.
31. Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou J-P, Steinmetz A, et al. Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in Flowering Time Regulation and other processes in Arabidopsis thaliana. Mol Cell Biol. 2008;28(4):1348–60.
32. Liu B, Wei G, Shi J, Jin J, Shen T, Ni T, et al. SET DOMAIN GROUP 708, a histone H3 lysine 36-specific methyltransferase, controls flowering time in rice (Oryza sativa). New Phytol. 2016;210(2):577–88.
33. Song C, Liu Y, Song A, Dong G, Zhao H, Sun W, et al. The Chrysanthemum nankingense genome provides insights into the evolution and diversification of Chrysanthemum flowers and Medicinal traits. Mol Plant. 2018;11(12):1482–91.
34. Wen X, Li J, Wang L, Lu C, Gao Q, Xu P et al. TheChrysanthemum lavandulifoliumgenome and the molecular mechanism underlying diverse capitulum types. Hortic Res. 2022;9.
35. Nakano M, Hirakawa H, Fukai E, Toyoda A, Kajitani R, Minakuchi Y et al. A chromosome-level genome sequence of Chrysanthemum seticuspe, a model species for hexaploid cultivated chrysanthemum. Commun Biology. 2021;4(1).
36. Dong W, Li M, Li Z, Li S, Zhu Y, Hongxu, et al. Transcriptome analysis of the molecular mechanism of Chrysanthemum flower color change under short-day photoperiods. Plant Physiol Biochem. 2020;146:315–28.

37. Zhang Q, Li J, Wen X, Deng C, Yang X, Dai S. Genome-wide identification and characterization analysis of RWP-RK family genes reveal their role in flowering time of Chrysanthemum lavandulifolium. BMC Plant Biol. 2023;23(1).
38. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9.
39. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265–8.
40. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2020;49(D1):D458–60.
41. Garg HA, A AT, W JP. RP, M KA, MFPPI - Multi FASTA ProtParam Interface 12. 2016:74–7.
42. Wang X, Zhao S, Zhou R, Liu Y, Guo L, Hu H. Identification of Vitis vinifera MYB transcription factors and their response against grapevine berry inner necrosis virus. BMC Plant Biol. 2023;23(1).
43. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(Web Server):W585–7.
44. Ai P, Xue J, Shi Z, Liu Y, Li Z, Li T et al. Genome-wide characterization and expression analysis of MYB transcription factors in Chrysanthemum nankingense. BMC Plant Biol. 2023;23(1).
45. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024.
46. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7.
47. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative Toolkit developed for interactive analyses of big Biological Data. Mol Plant. 2020;13(8):1194–202.
48. Lescot PD, G T, K M YM. Y VdP, PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002(30):325–7.
49. Ayoub Khan M, Dongru K, Yifei W, Ying W, Penghui A, Zicheng W. Characterization of WRKY Gene Family in whole-genome and exploration of flowering improvement genes in Chrysanthemum lavandulifolium. Front Plant Sci. 2022;13.
50. Grewal S, Caro E, Stroud H, Greenberg MVC, Bernatavichute YV, Feng S et al. The SET-Domain protein SUVR5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation–independent manner. PLoS Genet. 2012;8(10).
51. Berr A, Shafiq S, Shen W-H. Histone modifications in transcriptional activation during plant development. Biochimica et Biophysica Acta (BBA) -. Gene Regul Mech. 2011;1809(10):567–76.
52. Qian Y, Xi Y, Cheng B, Zhu S, Kan X. Identification and characterization of the SET domain gene family in maize. Mol Biol Rep. 2014;41(3):1341–54.
53. Xu J, Xu H, Liu Y, Wang X, Xu Q, Deng X. Genome-wide identification of sweet orange (Citrus sinensis) histone modification gene families and their expression analysis during the fruit development and fruit-blue mold infection process. Front Plant Sci. 2015;6.
54. Chen Q, Hu S, Guo F, Zhao H, Wang M, Ni D et al. Characterization of the SET DOMAIN GROUP gene family members in Camellia sinensis and functional analysis of the SDG43 gene in abiotic stresses. Environ Exp Bot. 2021;182.
55. Shen HCX, F PXXWH, Q LX. Genome-wide identification and expression analysis of SDG Gene Family during early somatic embryogenesis in Dimocarpus longan lour. Chin J Trop Crops. 2019;40:1889–901.
56. Zhou TWT. Effects of exogenous methyl jasmonate treatment on the volatile composition and content of Rosa rugosa. Mol Plant Breed. 2019;17:4776–84.
57. Dong H, Liu D, Han T, Zhao Y, Sun J, Lin S et al. Diversification and evolution of the SDG gene family in Brassica rapa after the whole genome triplication. Sci Rep. 2015;5(1).

Publisher's Note