The origin, evolution, and functional divergence of the *Dicer-like* (DCL) and *Argonaute* (AGO) gene families in plants

Li-Yao Su[1#], Shan-Shan Li[2#], Hui Liu[1], Zong-Ming (Max) Cheng[1] and Ai-Sheng Xiong[1*] 🆔

[1] *State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Ministry of Agriculture and Rural Affairs Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China*
[2] *Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, Zhejiang, China*
[#] Authors contributed equally: Li-Yao Su, Shan-Shan Li
[*] Corresponding author, E-mail: xiongaisheng@njau.edu.cn

Abstract

Dicer-like (DCL) and Argonaute (AGO) proteins play a crucial role in plant epigenetics. However, the evolutionary origins and roles of these gene families in plant adaptation, stress response, and development remain unclear. This study explores the origin and functional evolution of DCLs and AGOs across 36 plant species spanning diverse taxonomic groups. Member identification, phylogenetic analysis, evolutionary trajectory analysis, and functional divergence analysis were conducted. The results show that the DCL and AGO originated in Rhodophytes and underwent two major expansions: during algal terrestrialization and the transition from lower to higher plants. In seed plants, DCLs diversified into four classes following two whole-genome duplication (WGD) events, whereas AGOs diversified into seven classes through two WGD events and one tandem duplication event. Expression analyses in *Physcomitrium patens*, *Zea mays*, *Arabidopsis thaliana*, and *Fragaria vesca* revealed high expression of these gene families in reproductive tissues, with notably lower expression in pollen. Additionally, the expression of these genes exhibits different responses to various environmental stresses in *A. thaliana* and *Z. mays*, highlighting their important roles in adaptation to environmental fluctuations. The present research reveals the functional diversification of DCLs and AGOs and their crucial roles in facilitating terrestrial adaptation and rapid land colonization.

Introduction

Epigenetics refers to heritable changes in gene expression that do not alter the DNA sequence but affect gene activity, such as DNA methylation, histone modifications, and non-coding RNAs. In plants, the dicer-like (DCL) and argonaute (AGO) gene families not only participate in non-coding RNA production and function but also in RNA-directed DNA methylation (RdDM)[1–3]. RNA interference (RNAi) is a critical biological process that involves both post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS) mediated by small RNAs. This process begins with the generation of small RNAs, which are then incorporated into the RNA-induced silencing complex (RISC). The generation of mature small RNAs (sRNAs) is primarily facilitated by DCL proteins, whereas the AGO proteins play a pivotal role as carriers, guiding sRNAs to recognize and base-pair with target mRNA sequences, ultimately regulating gene expression[2]. Beyond transcriptional and post-transcriptional silencing, DNA methylation represents another critical regulatory mechanism in various plant growth and developmental processes[1,3]. RdDM, often referred to as the canonical RdDM pathway, is a widespread epigenetic regulatory mechanism in plants. Both canonical and non-canonical RdDM pathways heavily rely on the functions of DCLs and AGOs[4].

DCLs function as molecular factories for processing plants small RNAs (sRNAs), serving highly conserved roles across plant biology. These proteins typically contain several domains, including DExD, Helicase-C, DUF283, PAZ, RNase III, and dsRNA-binding, all belonging to the ribonuclease III family[5]. The *Arabidopsis thaliana* (*A. thaliana*) genome contains four *DCL* genes, designated as *DCL1* through *DCL4*, each playing a unique role in RNA silencing and plant physiological responses. For example, DCL1 is primarily responsible for the biogenesis of microRNAs, indirectly affecting normal plant development and environmental adaptation[6]. DCL2 mainly produces small interfering RNAs (siRNAs), which are crucial for plant defense mechanisms and developmental processes[7–9]. DCL3 is predominantly involved in the synthesis of 24-nucleotide siRNAs and is essential for the RdDM pathway, maintenance of genomic stability, regulation of gene expression, and responses to environmental stimuli[10,11]. DCL4 produces 21-nucleotide siRNAs, which play key roles in post-transcriptional gene silencing, especially in antiviral defense mechanisms[12]. Additionally, *DCL5* (previously known as *DCL3b*) is found in monocots; it enhances the activity of DCL3 and plays a specialized role in reproductive processes[13].

In plants, the *AGO* family genes interact with sRNAs to form RISC, which act as specific regulators of gene expression across various biological processes. AGO proteins modulate gene expression through several mechanisms including transcript cleavage, suppression of PTGS, and influencing DNA methylation through RdDM, along with other specialized functions[14,15]. The AGO family exhibits significant evolutionary diversity and can be categorized into three main phylogenetic groups: *AGO1/5/10*, *AGO2/3/7*, and *AGO4/6/8/9*[16,17]. *AGO1* is a widely expressed member that plays a central role in multiple sRNA-mediated silencing pathways, especially those associated with PTGS[18,19]. The function of *AGO5* is less well understood but it is thought to be involved in gene silencing during viral infections[20,21]. AGO10 selectively binds to 21-nt siRNAs and is involved in transcriptional gene silencing pathways[22,23]. *AGO2* is recognized for its role in defense against viruses, it also binds to 21-nt siRNAs to participate in PTGS[24–26]. *AGO3* and *AGO7*, despite being phylogenetically close to *AGO2*, display functional divergence—AGO3 binds to 24-nt siRNAs and primarily participates in RdDM to maintain genomic and transposon stability[27], whereas *AGO7* interacts with miR390 to trigger the production of trans-acting siRNAs from TAS3 transcripts[28]. *AGO4*, a core component of

the RdDM pathway, guides 24-nt siRNAs to DNA sites to promote DNA methylation, thereby silencing their target genes[1]. AGO6 shares functional similarities with *AGO4*, often acting as its functional complement. Additionally, *AGO9* has been shown to also participate in RdDM[29–31]. The specific roles of *AGO8* and its associated siRNAs remain unclear. However, *AGO8*, along with its paralogs *AGO4*, *AGO6*, and *AGO9*, is crucial for early megaspore formation[30,32].

Although extensive research has detailed the evolution and function of *DCLs* and *AGOs* in plants, most studies have focused on their classification and diversity, with less attention given to their potential functional divergence during evolution[33,34]. In the present study, the distribution, evolution, and expansion of *DCLs* and *AGOs* were examined across a wide range of species. By constructing phylogenetic trees, their possible evolutionary trajectories within angiosperms were inferred. Their expression profiles were further analyzed in various tissues and under different stress conditions to explore the potential functions of these two gene families. The findings significantly advance the understanding of the functional evolution of *DCLs* and *AGOs* in angiosperms, and offer valuable insights that could inform future breeding strategies aimed at developing improved plant varieties.

Results

Identification of DCL and AGO family members in plants

To investigate the origins and evolutionary histories of *DCL* and *AGO* genes in plants, this study used Arabidopsis *DCL* and *AGO* genes as seed sequences. BLAST software was employed to identify homologous sequences in 36 plant species, spanning groups such as rhodophytes, chlorophytes, charophytes, bryophytes, ferns, gymnosperms, basal angiosperms, monocots, and eudicots. Additional validation with InterProScan confirmed the presence of requisite domains in the identified sequences. This screening process resulted in the identification of 113 *DCLs* and 334 *AGOs* across the 36 species.

To delineate the evolutionary relationships among the *DCL* and *AGO* genes, phylogenetic trees were constructed using the maximum likelihood method. The phylogenetic analysis of DCLs revealed two main branches, which can be further divided into four clades (Fig. 1a). This tree suggests that the *DCL* genes originated in rhodophytes and remained relatively stable in chlorophytes and charophytes. Notably, a significant expansion of *DCLs* occurred in bryophytes, marked by the emergence of the *DCL1*, *DCL3*, and *DCL4* clades, with seed plants exhibiting widespread representation across all four clades. The *AGO* gene family tree consists of three main branches and seven clades (Fig. 1b), with phylogenetic evidence indicating that the ancestors of the AGO4/6/8/9 and AGO2/3/7 groups were present in algae. These genes underwent further expansion in bryophytes and ferns and fully evolved in seed plants. Taken together, these results highlight a largely consistent evolutionary history for *DCLs* and *AGOs*, suggesting synchronous evolution among these gene families. Additionally, the identification of *DCL* and *AGO* members were expanded using transcriptome-based gene annotations from the 1KP database, which covers over 1,000 plant species (Supplementary Figs S1 & S2). The results from this broader analysis are consistent with those derived from the initial 36 species. Based on sequence homology and phylogenetic insights, the origins and evolutionary trajectories of the *DCL* and *AGO* gene families across various plant lineages have been inferred, providing a comprehensive overview of their development through evolutionary history.

Multiple sequence alignments of sRNA-related functional domains in DCLs and AGOs were performed. Overall, the RNase III domains of DCLs in all four branches are highly conserved at both the C- and N-termini. Additionally, there are variations among the RNase III domains between different branches, with those in higher plants being more conserved (Supplementary Fig. S3). This conservation may reflect adaptations to diverse environmental pressures through more sophisticated RNA regulatory mechanisms, leading to the synthesis of a wider variety of sRNAs. Furthermore, we aligned the MID and PAZ domains of AGOs across different branches. Members of the AGO2/3/7 and AGO4/6/8/9 branches have largely lost the MID domain, and those that retain it show less conservation.
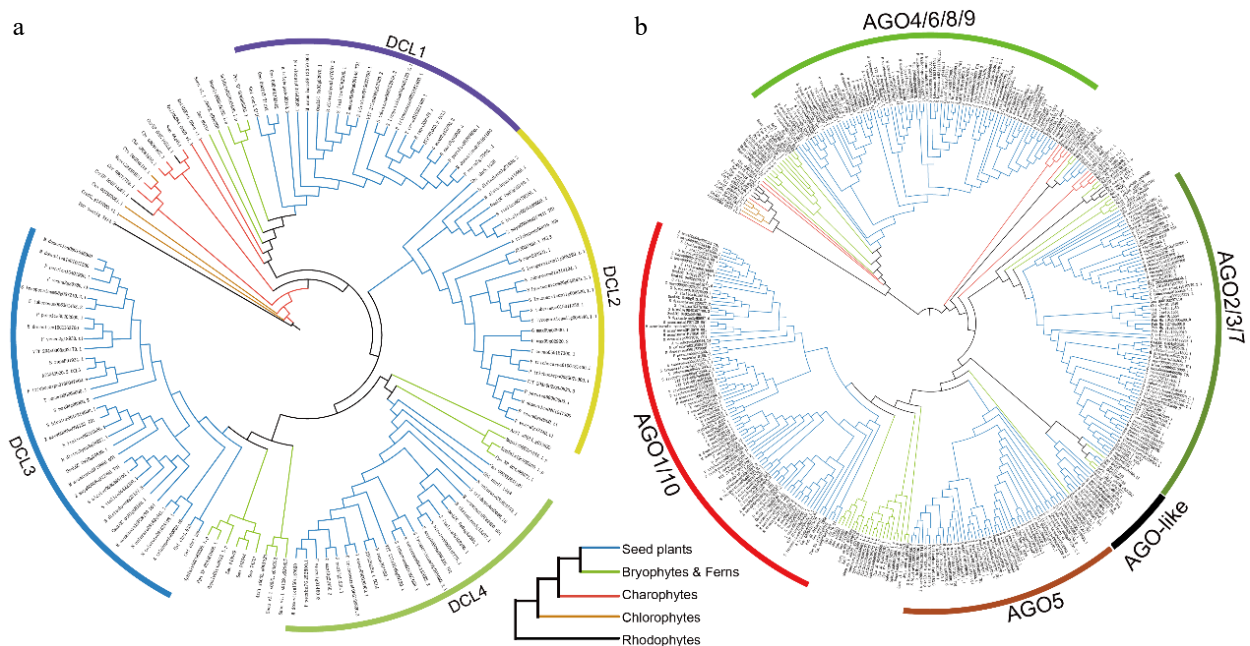

Fig. 1 Phylogenetic trees of the (a) *DCL,* and (b) *AGO* gene families across 36 plant species. Branches are color-coded to denote different plant groups: black for rhodophytes, yellow for chlorophytes, red for charophytes, green for ferns and bryophytes, and blue for seed plants.

In contrast, the AGO1/5/10 branch retains a highly conserved MID domain (Supplementary Fig. S4). Similarly, the PAZ domain shows significant differences among branches, with high conservation in the AGO1/5/10 branch, including in lower plants, while the other two branches exhibit lower conservation (Supplementary Fig. S5). Highly conserved MID and PAZ domains are typically associated with fundamental miRNA processing, whereas less conserved domains may relate to specific functional requirements and adaptive changes.

In plants, the copy number of *DCL* genes does not appreciably vary across different evolutionary branches (Fig. 2), ranging from 25 to 40 across the four groups and predominantly existing as single-copy genes throughout plant evolution. The primary factor influencing this gene copy number variation among different plants is attributed to whole-genome duplication (WGD) events during specific evolutionary processes. By contrast, the copy number of *AGO* genes exhibit considerable variation across branches. Specifically, the major clades of *AGO4/6/8/9*, *AGO2/3/7*, and *AGO1/5/10* contain 95, 93, and 173 genes, respectively (Fig. 2). The number of *AGO* genes notably exceeds that of *DCL* genes and shows diversification into more clades, suggesting that the *AGO* genes are more frequently retained during duplication events. Additionally, *AGO4* and *AGO6*, which encode key enzymes in plant methylation through

siRNA processing were analyzed separately (Fig. 2). Results show that *AGO4* is prevalent in basal angiosperms, whereas *AGO6* is restricted to monocots and dicots. Based on these findings, it is hypothesized that the evolution of *AGO4* and *AGO6* may be linked to significant shifts in reproductive strategies and the development of floral organs during the transition from gymnosperms to angiosperms. Given the unique and critical roles of *AGO4* and *AGO6* in methylation processes, the expansion and loss of these genes was investigated within the *AGO4/6/8/9* clade across 36 species (Fig. 3). Phylogenetic analysis reveals that these genes exist as single copies in algae, ferns, gymnosperms, and basal angiosperms. Following the ε duplication event, both *AGO4* and *AGO6* were retained in embryophytes. Unlike *AGO6*, which did not undergo significant expansion after its formation, *AGO4* experienced multiple duplication events. In the Brassicaceae, the α and β duplication events led to the emergence of *AGO8* and *AGO9*. Similarly, duplication events in crops such as potatoes, tomatoes, and monocots also contributed to the expansion of *AGO4*.

Phylogenetic and collinearity network analysis of the *DCLs* and *AGOs*

To elucidate the evolutionary differences between *DCLs* and *AGOs* in plants, a collinearity network analysis was conducted on 18
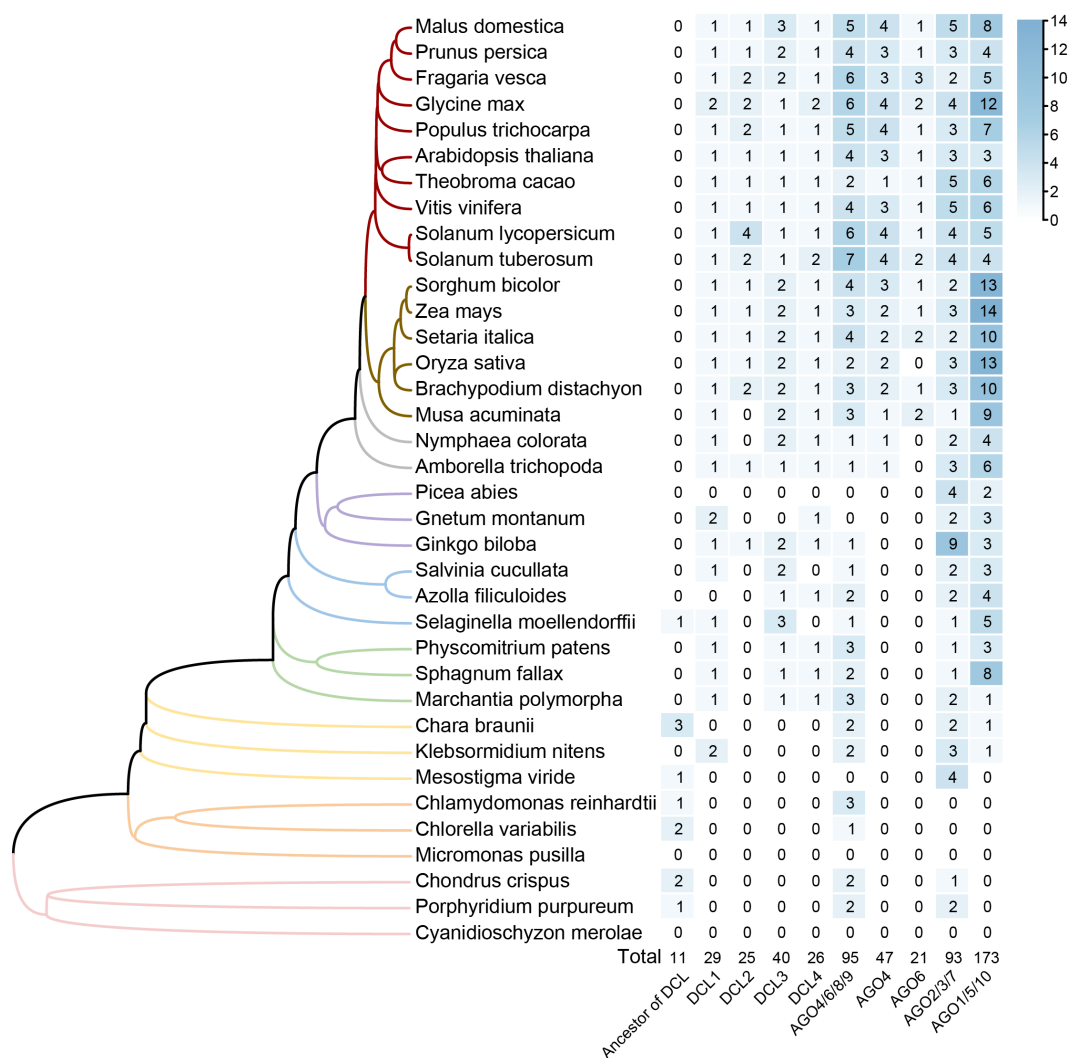

Species	Ancestor of DCL	DCL1	DCL2	DCL3	DCL4	AGO4/6/8/9	AGO4	AGO6	AGO2/3/7	AGO1/5/10
Malus domestica	0	1	1	3	1	5	4	1	5	8
Prunus persica	0	1	1	2	1	4	3	1	3	4
Fragaria vesca	0	1	2	2	1	6	3	3	2	5
Glycine max	0	2	2	1	2	6	4	2	4	12
Populus trichocarpa	0	1	2	1	1	5	4	1	3	7
Arabidopsis thaliana	0	1	1	1	1	4	3	1	3	3
Theobroma cacao	0	1	1	1	1	2	1	1	5	6
Vitis vinifera	0	1	1	1	1	4	3	1	5	6
Solanum lycopersicum	0	1	4	1	1	6	4	1	4	5
Solanum tuberosum	0	1	2	1	2	7	4	2	4	4
Sorghum bicolor	0	1	1	2	1	4	3	1	2	13
Zea mays	0	1	1	2	1	3	2	1	3	14
Setaria italica	0	1	1	2	1	4	2	2	2	10
Oryza sativa	0	1	1	2	1	2	2	0	3	13
Brachypodium distachyon	0	1	2	2	1	3	2	1	3	10
Musa acuminata	0	1	0	2	1	3	1	2	1	9
Nymphaea colorata	0	1	0	2	1	1	1	0	2	4
Amborella trichopoda	0	1	1	1	1	1	1	0	3	6
Picea abies	0	0	0	0	0	0	0	0	4	2
Gnetum montanum	0	2	0	0	1	0	0	0	2	3
Ginkgo biloba	0	1	1	2	1	1	0	0	9	3
Salvinia cucullata	0	1	0	2	0	1	0	0	2	3
Azolla filiculoides	0	0	0	1	1	2	0	0	2	4
Selaginella moellendorffii	1	1	0	3	0	1	0	0	1	5
Physcomitrium patens	0	1	0	1	1	3	0	0	1	3
Sphagnum fallax	0	1	0	1	1	2	0	0	1	8
Marchantia polymorpha	0	1	0	1	1	3	0	0	2	1
Chara braunii	3	0	0	0	0	2	0	0	2	1
Klebsormidium nitens	0	2	0	0	0	2	0	0	3	1
Mesostigma viride	1	0	0	0	0	0	0	0	4	0
Chlamydomonas reinhardtii	1	0	0	0	0	3	0	0	0	0
Chlorella variabilis	2	0	0	0	0	1	0	0	0	0
Micromonas pusilla	0	0	0	0	0	0	0	0	0	0
Chondrus crispus	2	0	0	0	0	2	0	0	1	0
Porphyridium purpureum	1	0	0	0	0	2	0	0	2	0
Cyanidioschyzon merolae	0	0	0	0	0	0	0	0	0	0
Total	11	29	25	40	26	95	47	21	93	173

Fig. 2 Phylogenetic distribution and gene copy number analysis of *DCL* and *AGO* gene families across 36 plant species. The phylogenetic tree on the left represents the evolutionary relationships of species investigated, with branches colored to represent different groups. The heatmap on the right displays gene copy numbers for each clade of gene family across the species, with higher numbers represented by darker shades. The total counts for each clade across all species are provided at the bottom of the heatmap.
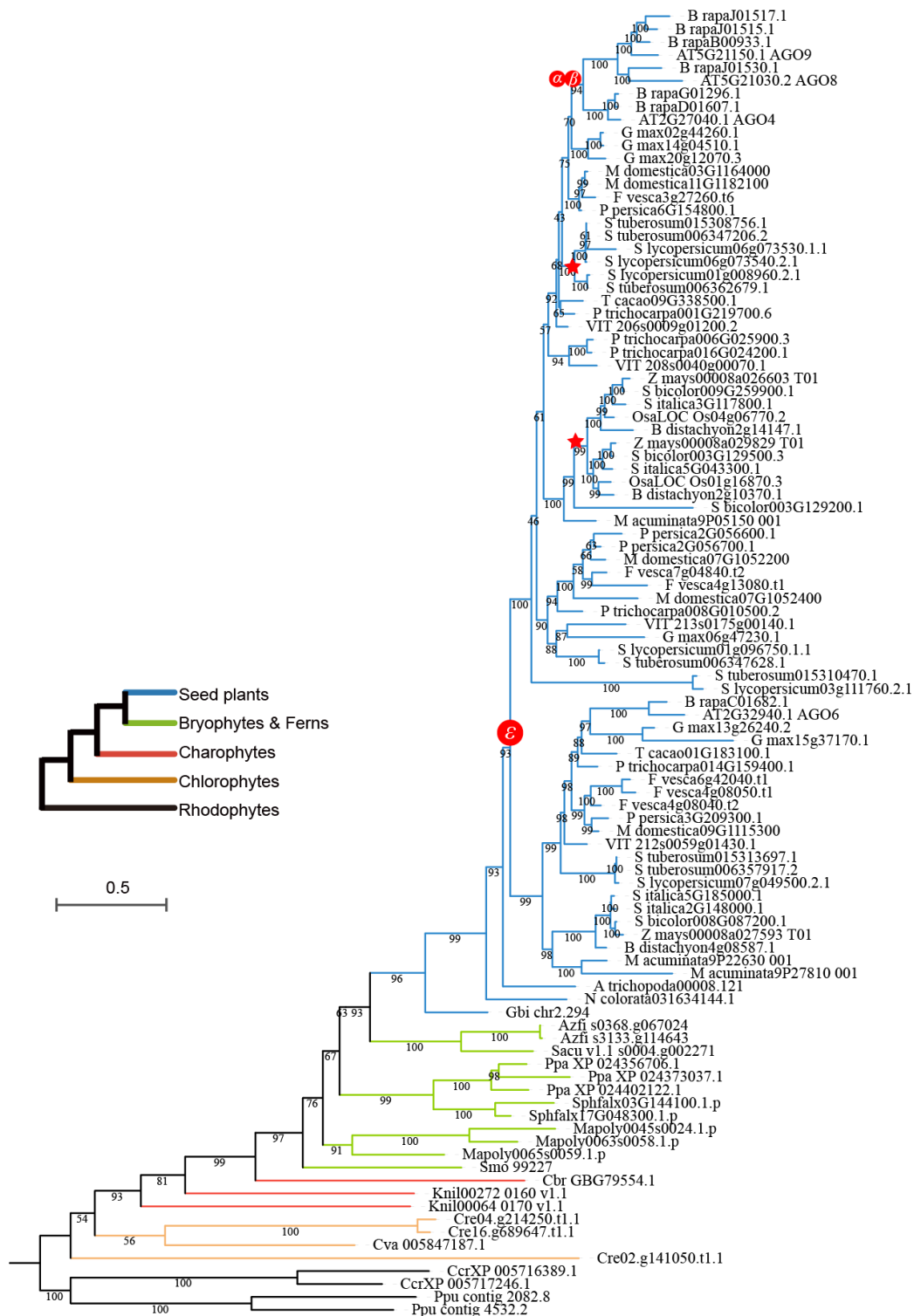
Fig. 3 Phylogeny of the *AGO4/6/8/9* clade within the *AGO* gene family. Different colored branches represent distinct plant groups. The symbols ε, α, and β represent the epsilon angiosperm-wide WGD event, the alpha duplication event, and the beta duplication event, respectively. Red stars along the branches indicate specific whole-genome duplication events.

plants, including both monocots and dicots, based on their phylogenetic relationships. The analysis identified 542 syntenic gene pairs, grouping the *DCL* genes into four clusters that represent four distinct evolutionary trajectories. Additionally, four WGD events and three tandem duplication pairs were detected in apple, soybean, and tomato, suggesting lineage-specific expansions of the *DCLs* during evolution (Fig. 4a). Therefore, we integrated both phylogenetic and collinearity data were integrated (Fig. 4b) and the

evolutionary history of *DCL* genes was reconstructed. It is proposed that two ancestral *DCL* genes existed before the emergence of seed plants, which subsequently underwent two WGD events. This process resulted in the loss of three branches, leaving four extant *DCL* groups.

In the *AGO* gene family, 1,090 syntenic gene pairs were identified and subsequently clustered into seven groups (Fig. 4c). Notably, *AGO4*, *AGO8*, and *AGO9* formed a cluster, demonstrating their
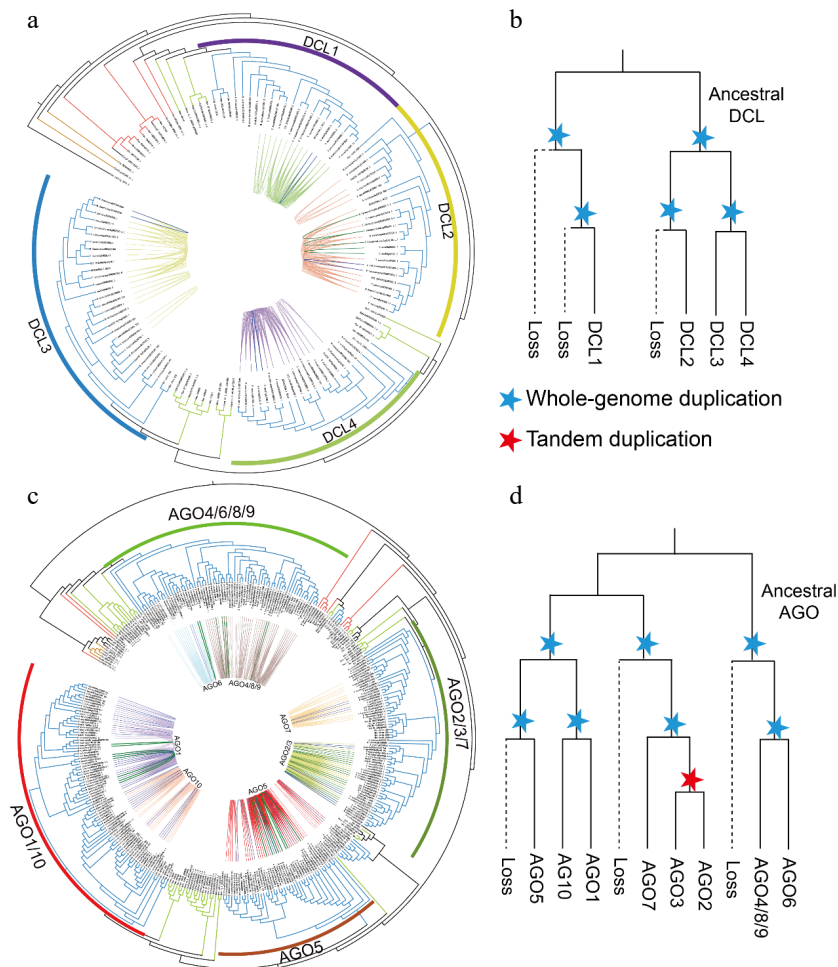
Fig. 4 Phylogenetic analysis and synteny identification of the *DCL* and *AGO* genes. (a), (c) Phylogenetic and syntenic relationships of the *DCLs* and *AGOs*. The blue and green lines indicate gene pairs resulting from WGD and tandem duplication in the *DCLs* and *AGOs*, respectively. (b, d) Schematic representation of the proposed evolutionary histories of the *DCL* and *AGO* gene families. The dashed lines indicate gene loss. Blue stars mark either the ancient seed plant-wide or angiosperm-wide genome duplication events. Red stars represent tandem duplication events of genes.

evolutionary homology. A similar homologous relationship is observed between *AGO2* and *AGO3*. Additionally, 43 syntenic gene pairs resulting from intraspecific duplications across various *AGO* groups were found. Moreover, 23 tandem duplication pairs were identified, they distributed across the syntenic gene clusters of *AGO1*, *AGO4/8/9*, *AGO5*, and *AGO6*, with most tandem duplications occurring within *AGO2/3*. Based on these findings, it is inferred that the *AGO* family originated from three ancestral genes before the emergence of seed plants, with clusters retained through two WGD events. Furthermore, *AGO2* and *AGO3* appear to have arisen from tandem duplications (Fig. 4d).

Expression patterns of *DCLs* and *AGOs* in key green plants

To explore the functional differences between *DCLs* and *AGOs* throughout plant evolution, expression patterns were analyzed using publicly available data from various tissues of *P. patens*, *Z. may*, *A. thaliana*, and *F. vesca*. Among these species, *F. vesca* exhibits the highest number of *DCL* members, totaling six. The expression profiles of various tissues were largely consistent across the four species, with *DCLs* showing high expression in reproductive tissues (Fig. 5a, Supplementary Fig. S6a). In *F. vesca*, the *DCL2* and *DCL3* branches each contain two gene members, however, in each branch, only one gene exhibits high expression. This contrasts with *Z. mays*, where the expression patterns of the two *DCL3* members are similar. In *P. patens*, *DCL3* is more prominently expressed in

vegetative tissues, whereas in *A. thaliana*, *Z. may*, and *F. vesca*, it shows high expression in reproductive tissues. For the *AGOs*, the overall expression profiles are similar to those of the *DCLs*, with high expression in reproductive tissues (Fig. 5b, Supplementary Fig. S6b). The *AGO4/6/8/9* groups demonstrate functional complementation in reproductive tissues in ancestral species, whereas in *A. thaliana*, *Z. mays*, and *F. vesca*, *AGOs* are ubiquitously expressed in all reproductive tissues. In *P. patens*, *AGO1* group members are highly expressed only in reproductive tissues in *P. patens*, however, they maintain high expression levels across all tissues in *A. thaliana* and *F. vesca*. Interestingly, despite their overall low expression levels in pollen both *DCLs* and *AGOs* from various groups remain active in reproductive organs in *A. thaliana* and *F. vesca*.

The expression of *DCLs* and *AGOs* in response to various stresses in *A. thaliana* and *Z. may* were further analyzed. The results indicate that *A. thaliana DCLs* respond to all stressors except irradiation, whereas *Z. may DCLs* primarily respond to heat, salt, drought, and nutrient deficiency (Fig. 6a). The stress responses of *AGOs* vary across different groups (Fig. 6b). For example, in *A. thaliana*, *AGO6*, *AGO7*, and *AGO10* are involved in most stress pathways. In *Z. mays*, *AGO6*, *AGO4/8/9*, *AGO2/3*, and *AGO10* participate in multiple stress responses. In both *A. thaliana* and *Z. mays*, *AGO5* exhibits minimal responsiveness to stress, whereas *AGO4* responds to similar stresses, including heat, salt, drought, cold, shade, and nutrient deficiency. This comprehensive analysis underscores the specificity and
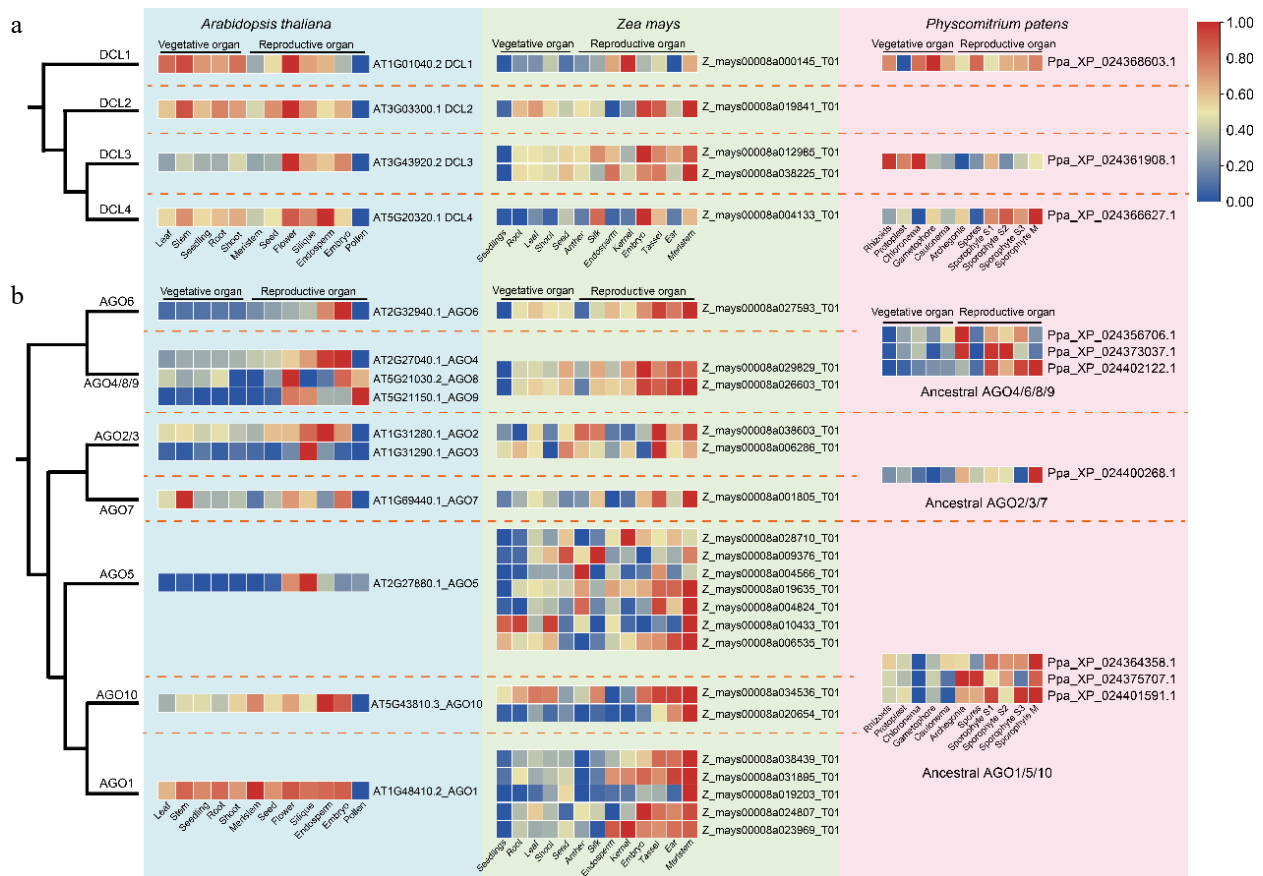
Fig. 5 Expression profiles of *DCLs* and *AGOs* in different tissues of *A. thaliana*, *Z. may*, and *P. patens*. (a) Comparative expression profiles of *DCL* gene family members. (b) Comparative expression profiles of *AGO* gene family members. Dashed lines demarcate distinct clades, with the heatmap displaying relative expression levels from low (blue) to high (red).

variability of *DCL* and *AGO* responses to environmental stresses, highlighting their essential adaptive functions in plant stress physiology.

Discussion

As plants transitioned from aquatic to terrestrial environments, they encountered more variable habitats and increased exposure to air. This shift prompted the expansion of numerous gene families to adapt to these diverse environmental challenges[35–37]. The present findings align with previous studies, revealing that *DCLs* are predominantly classified into four groups, with *DCL2* being exclusive to seed plants (Fig. 1a). The origin of *DCLs* was traced back to rhodophytes, detecting homologs in *Chondrus crispus* and *Porphyridium purpureum*, a finding supported by data from the 1KP database (Supplementary Fig. S1). Similarly, the present analysis suggests that *AGOs* also originated from rhodophytes, demonstrating the conservation of the RNAi pathway across plant species. Contrary to Li et al., who reported a single ancestral lineage for *AGO*, the present study identifies ancestral positions for *AGO4/6/8/9* and *AGO2/3/7* in *Porphyridium purpureum*[16]. Furthermore, a clear differentiation of *AGOs* into two distinct groups in rhodophytes, chlorophytes, and charophytes was observed (Fig. 1b). The expansion of the *AGO* family from charophytes to bryophytes likely represents an evolutionary adaptation crucial for terrestrial colonization[38,39]. Furthermore, these findings underscore the significant role of epigenetics in the terrestrial adaptation of plants.

DCLs exhibit a single ancestral branch in all algae, however, their expansion during the transition from aquatic to terrestrial environments coincides with that of the *AGOs*. This expansion

continuous as lower plants evolved into higher plants, leading to the present diversification of these gene families. As key components of the RNAi mechanism, *DCLs*, and *AGOs* are crucial for the generation and function of miRNA. Specific miRNAs in algae that are conserved in seed plants, such as miR167, miR172, miR395, miR414, miR418, and miR419, are missing in mosses and ferns[40–43]. These miRNAs are essential for flower development, stress resistance, and root development in higher plants[44,45]. Their absence in mosses and ferns highlights differences in the miRNA-mediated gene silencing pathways between lower and higher plants and underscores the adaptive changes during plant evolution, reflecting species-specific survival strategies and developmental needs in diverse environments. Further analysis of the evolutionary trajectories of *DCLs* and *AGOs* in seed plants revealed that *DCLs* underwent two rounds of WGDs in land plants without a significant increase in their numbers (Figs 2 & 4). By contrast, *AGOs* experienced two WGD events and one tandem duplication (Fig. 4), and maintained specific expansions within different species (Fig. 3). These findings reveal distinct evolutionary paths of these gene families and their crucial roles in adapting RNAi mechanisms for plant survival and development across diverse ecological settings.

WGD and various forms of gene duplication are the primary mechanisms that drive the expansion of gene families. The retention of duplicated genes throughout evolution has facilitated better adaptation in plant growth and development[46–48]. Extensive research has established the functions of *DCLs* and *AGOs*. This study discovers that most *DCL* and *AGO* genes in *P. patens* are highly expressed in meristematic tissues, a pattern that is conserved in *A. thaliana*, *Z. mays*, and *F. vesca* (Fig. 5 & Supplementary Fig. S3). This expression profile is likely attributed to active DNA methylation
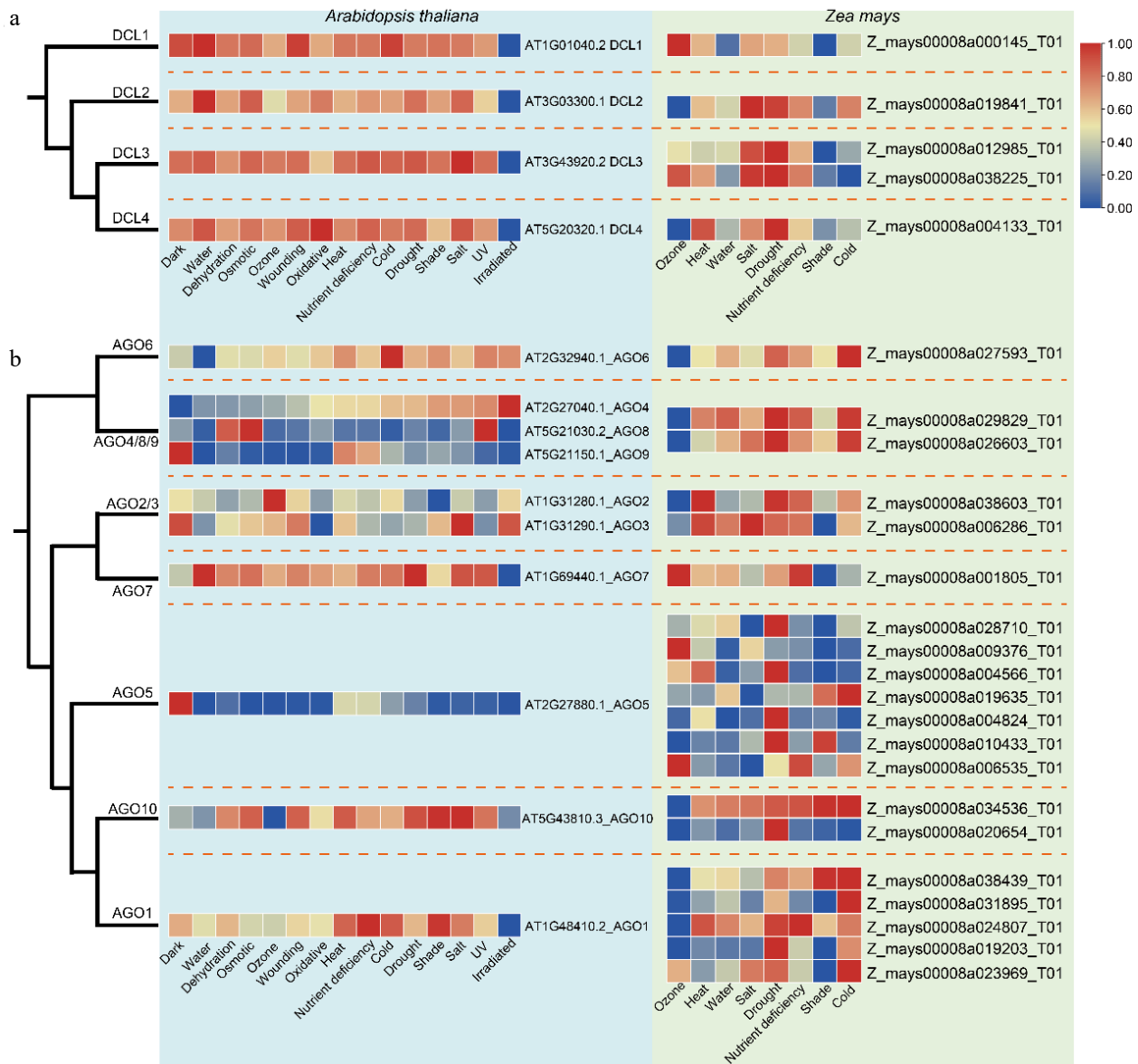
Fig. 6 Expression of the *DCLs* and *AGOs* under different stress conditions in *A. thaliana* and *Z. mays*. (a) Comparative expression profiles of *DCL* gene family members. (b) Comparative expression profiles of *AGO* gene family members. Dashed lines denote distinct clades, with the heatmap displaying relative expression levels from low (blue) to high (red).

and RNAi regulation within these tissues[10,49,50], underscoring the intricate genetic regulation essential for plant development. Additionally, *DCL3* and *DCL4* in *P. patens* show high expression in vegetative and meristematic tissues, respectively, suggesting the specialization of their functions (Figs 5 & Supplementary Fig. S3). In *Arabidopsis*, *F. vesca*, and *Z. mays*, the *DCL2/3/4* genes are highly expressed in various tissues and response to stress (Fig. 6). The sRNAs they produce vary, indicating that diverse types of sRNAs are extensively involved in the life cycles of plants. This diversity enables plants to adapt to environmental fluctuations and supports their growth and development[51–53]. Additionally, DCL serves as the factory for manufacturing and producing miRNAs. Its tissue-specific expression, along with upregulated expression under stress conditions, is closely associated with the miRNAs it produces. For example, the miRNA156/SPL module can participate in root development and vegetative growth while also enhancing the plant's tolerance to abiotic stress. miR169 targets different members of the NF-YA gene family, which is involved in multiple developmental processes and stress responses. Moreover, miR159, miR397, and miR393 possess diverse functions in plant growth, development, and stress

tolerance[44,45]. By contrast, the functions of the *AGO4/6/8/9* clade have remained largely unchanged throughout evolution, with no special functions emerging from their expansion in *Z. may* and *A. thaliana* (Figs 5, 6, & Supplementary Fig. S3). Previous studies have demonstrated functional complementarity among AGO4/6/8/9 proteins[10,54,55], which is essential for maintaining critical biological processes under varying conditions. *AGO2/3/7* appear to be predominantly involved in stress responses (Fig. 6), consistent with findings of previous studies[56,57]. In *A. thaliana*, *F. vesca*, and *Z. mays*, *AGO1/5/10* exhibit significant functional divergence. Specifically, in *A. thaliana*, *AGO1* and *AGO5* exist as single-copy genes. However, *AGO5* has largely lost its regulatory functions in tissue development and stress response, whereas *AGO1* retains all these functions (Figs 5 & 6). In *Z. mays*, although there is a significant expansion of members within the *AGO1/5/10* group, their functions remain largely similar to those observed in *P. patens*. These results suggest that subfunctionalization and neofunctionalization are two potential evolutionary outcomes of gene duplication[58], and also demonstrate the role of epigenetic regulation in directing species-specific evolutionary trajectories in plants.

Materials and methods

Identification of *DCLs* and *AGOs*

Genomic data for 36 plant species used in this study were obtained from databases such as Phytozome (Supplementary Table S1). Gene screening and alignment of the 1KP transcriptome data were conducted using the ONEKP online platform (https://db.cngb.org/onekp). Protein sequences of the four *DCLs* and ten *AGOs* from *A. thaliana* served as query sequences for BLASTP analysis against the proteomic data of the remaining 35 plant genomes, using an e-value threshold of 1e-20. The resulting sequences were then analyzed using InterProScan to identify and annotate conserved domains using the Pfam, PANTHER, and SMART databases[59]. Only sequences that contained domains consistent with those found in *A. thaliana* were selected for constructing a phylogenetic tree using FastTree. Branches exhibiting abnormal lengths were manually removed to ensure the accuracy of the inferred phylogenetic relationships[60].

Phylogenetic tree construction and collinearity analysis

First, multiple sequence alignment of all DCL and AGO protein sequences was performed using MAFFT. Gap positions were then removed from the aligned sequences using Phyutility with a cutoff parameter of 0.5. Next, ProtTest was used to predict the best-fit substitution models for constructing the DCL and AGO phylogenetic trees. The DCL and AGO trees were constructed using IQ-TREE with the JTT + F + R6 and LG + I + G + F models, respectively, with 1000 Bootstrap replicates[61]. For genes indexed in the 1KP database, phylogenetic trees were constructed using FastTree. Additionally, collinearity blocks were identified by comparing coding sequences across species using the Python version of MCScan[62].

Expression analysis of *DCL* and *AGO* genes

Expression data for different tissues of *P. patens* and *F. vesca* were obtained from the Physcomitrium eFP Browser and the Strawberry eFP Browser, respectively (http://bar.utoronto.ca). For *A. thaliana* and *Z. mays*, tissue-specific and stress-induced expression profiles were downloaded from https://plantrnadb.com. Detailed data sources are provided in Supplementary Table S2. The expression of *DCL* and *AGO* genes across these species was visualized using TBtools[63].

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Xiong AS, Cheng ZM, Su LY; data analysis, draft manuscript preparation: Su LY, Li SS, Liu H. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Liu P, Liu R, Xu Y, Zhang C, Niu Q, et al. 2023. DNA cytosine methylation dynamics and functional roles in horticultural crops. *Horticulture Research* 10:d170

2. Paudel L, Kerr S, Prentis P, Tanurdžić M, Papanicolaou A, et al. 2022. Horticultural innovation by viral-induced gene regulation of carotenogenesis. *Horticulture Research* 9:uhab008

3. Shi M, Wang C, Wang P, Yun Y, Liu Z, et al. 2023. Role of methylation in vernalization and photoperiod pathway: a potential flowering regulator? *Horticulture Research* 10:uhad17

4. Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. *Nature Plants* 2:16163

5. Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, et al. 2006. The evolution and diversification of Dicers in plants. *FEBS Letters* 580:2442−50

6. Kurihara Y, Watanabe Y. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America* 101:12753−58

7. Jia J, Ji R, Li Z, Yu Y, Nakano M, et al. 2020. Soybean DICER-LIKE2 regulates seed coat color via production of primary 22-nucleotide small interfering RNAs from long inverted repeats. *The Plant Cell* 32:3662−73

8. Taochy C, Gursanscky NR, Cao J, Fletcher SJ, Dressel U, et al. 2017. A genetic screen for impaired systemic rnai highlights the crucial role of DICER-LIKE 2. *Plant Physiology* 175:1424−1437

9. Wu YY, Hou BH, Lee WC, Lu SH, Yang CJ, et al. 2017. DCL2- and RDR6-dependent transitive silencing of *SMXL4* and *SMXL5* in Arabidopsis *dcl4* mutants causes defective phloem transport and carbohydrate over-accumulation. *The Plant Journal* 90:1064−78

10. Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* 15:394−408

11. Wang Q, Xue Y, Zhang L, Zhong Z, Feng S, et al. 2021. Mechanism of siRNA production by a plant Dicer-RNA complex in dicing-competent conformation. *Science* 374:1152−57

12. Liu Y, Teng C, Xia R, Meyers BC. 2020. PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *The Plant Cell* 32:3059−80

13. Teng C, Zhang H, Hammond R, Huang K, Meyers BC, Walbot V. 2020. *Dicer-like 5* deficiency confers temperature-sensitive male sterility in maize. *Nature Communications* 11:2912

14. Carbonell A, Carrington JC. 2015. Antiviral roles of plant ARGONAUTES. *Current Opinion in Plant Biology* 27:111−17

15. Fang X, Qi Y. 2016. RNAi in plants: an argonaute-centered view. *The Plant Cell* 28:272−85

16. Li Z, Li W, Guo M, Liu S, Liu L, et al. 2022. Origin, evolution and diversification of plant ARGONAUTE proteins. *The Plant Journal* 109:1086−97

17. Zhang H, Xia R, Meyers BC, Walbot V. 2015. Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Current Opinion in Plant Biology* 27:84−90

18. Garcia-Ruiz H, Carbonell A, Hoyer JS, Fahlgren N, Gilbert KB, et al. 2015. Roles and programming of arabidopsis argonaute proteins during *Turnip mosaic virus* infection. *PLOS Pathogens* 11:e1004755

19. Wang XB, Jovel J, Udomporn P, Wang Y, Wu Q, et al. 2011. The 21-Nucleotide, but not 22-nucleotide, viral secondary small interfering rnas direct potent antiviral defense by two cooperative argonautes in *Arabidopsis thaliana*. *The Plant Cell* 23:1625−38

20. Brosseau C, Moffett P. 2015. Functional and genetic analysis identify a role for arabidopsis argonaute5 in antiviral RNA silencing. *The Plant Cell* 27:1742–54

21. Tucker MR, Okada T, Hu Y, Scholefield A, Taylor JM, et al. 2012. Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in *Arabidopsis*. *Development* 139:1399–404

22. Yu Y, Ji L, Le BH, Zhai J, Chen J, et al. 2021. Correction: ARGONAUTE10 promotes the degradation of miR165/6 through the SDN1 and SDN2 exonucleases in *Arabidopsis*. *PLOS Biology* 19:e3001120

23. Zhu H, Hu F, Wang R, Zhou X, Sze SH, et al. 2011. Arabidopsis argonaute10 specifically sequesters MIR166/165 to regulate shoot apical meristem development. *Cell* 145:242–56

24. Gao M, Wei W, Li MM, Wu YS, Ba Z, et al. 2014. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Research* 24:532–41

25. Schuck J, Gursinsky T, Pantaleo V, Burgyán J, Behrens SE. 2013. AGO/RISC-mediated antiviral RNA silencing in a plant *in vitro* system. *Nucleic Acids Research* 41:5090–103

26. Zhang X, Zhao H, Gao S, Wang WC, Katiyar-Agarwal S, et al. 2011. *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393-mediated silencing of a golgi-localized SNARE gene, *MEMB12*. *Molecular Cell* 42:356–66

27. Zhang Z, Liu X, Guo X, Wang XJ, Zhang X. 2016. *Arabidopsis* AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing. *Nature Plants* 2:16049

28. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, et al. 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *The Plant Cell* 19:926–42

29. Duan CG, Zhang H, Tang K, Zhu X, Qian W, et al. 2014. Specific but interdependent functions for *Arabidopsis* AGO4 and AGO6 in RNA-directed DNA methylation. *The EMBO Journal* 34:581–92

30. Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, et al. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* 464:628–32

31. Zheng X, Zhu J, Kapoor A, Zhu JK. 2007. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *The EMBO Journal* 26:1691–701

32. Hernández-Lagana E, Rodríguez-Leal D, Lúa J, Vielle-Calzada J. 2016. A multigenic network of argonaute4 clade members controls early megaspore formation in *Arabidopsis*. *Genetics* 204:1045–56

33. Bélanger S, Zhan J, Meyers BC. 2023. Phylogenetic analyses of seven protein families refine the evolution of small RNA pathways in green plants. *Plant Physiology* 192:1183–203

34. Wang S, Liang H, Xu Y, Li L, Wang H, et al. 2021. Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. *Communications Biology* 4:412

35. Li S, Wei L, Gao Q, Xu M, Wang Y, et al. 2024. Molecular and phylogenetic evidence of parallel expansion of anion channels in plants. *Plant Physiology* 194:2533–48

36. Su L, Zhang T, Yang B, Dong T, Liu X, et al. 2023. Different evolutionary patterns of *TIR1/AFBs* and *AUX/IAAs* and their implications for the morphogenesis of land plants. *BMC Plant Biology* 23:265

37. Wu Y, Wen J, Xia Y, Zhang L, Du H. 2022. Evolution and functional diversification of R2R3-MYB transcription factors in plants. *Horticulture Research* 9:uhac058

38. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, et al. 2018. The *chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* 174:448–64

39. Rensing SA. 2018. Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology* 42:49–54

40. Alaba S, Piszczalka P, Pietrykowska H, Pacak AM, Sierocka I, et al. 2015. The liverwort *Pellia endiviifolia* shares microtranscriptomic traits that are common to green algae and land plants. *New Phytologist* 206:352–67

41. Axtell MJ, Snyder JA, Bartel DP. 2007. Common functions for diverse small RNAs of land plants. *The Plant Cell* 19:1750–69

42. Fattash I, Voß B, Reski R, Hess WR, Frank W. 2007. Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biology* 7:13

43. Lin PC, Lu CW, Shen BN, Lee GZ, Bowman JL, et al. 2016. Identification of miRNAs and their targets in the liverwort *Marchantia polymorpha* by integrating RNA-Seq and degradome analyses. *Plant and Cell Physiology* 57:339–58

44. Dong Q, Hu B, Zhang C. 2022. MicroRNAs and their roles in plant development. *Frontiers in Plant Science* 13:824240

45. Zhan J, Meyers BC. 2023. Plant Small RNAs: Their biogenesis, regulatory roles, and functions. *Annual Review of Plant Biology* 74:21–51

46. Conant GC, Birchler JA, Pires JC. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology* 19:91–98

47. Liu S, Liu Y, Yang X, Tong C, Edwards D, et al. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* 5:3930

48. Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant Physiology* 171:2294–316

49. He S, Feng X. 2022. DNA methylation dynamics during germline development. *Journal of Integrative Plant Biology* 64:2240–51

50. Melnyk CW, Molnar A, Bassett A, Baulcombe DC. 2011. Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of *Arabidopsis thaliana*. *Current Biology* 21:1678–83

51. Nielsen CPS, Arribas-Hernández L, Han L, Reichel M, Woessmann J, et al. 2024. Evidence for an RNAi-independent role of Arabidopsis DICER-LIKE2 in growth inhibition and basal antiviral resistance. *The Plant Cell* 36:2289–309

52. Parent JS, Bouteiller N, Elmayan T, Vaucheret H. 2015. Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *The Plant Journal* 81:223–32

53. Wu H, Li B, Iwakawa HO, Pan Y, Tang X, et al. 2020. Plant 22-nt siRNAs mediate translational repression and stress adaptation. *Nature* 581:89–93

54. Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, et al. 2010. The *Arabidopsis* RNA-directed dna methylation argonautes functionally diverge based on their expression and interaction with target loci. *The Plant Cell* 22:321–34

55. Ortiz-Vasquez Q, León-Martínez G, Barragán-Rosillo C, González-Orozco E, Deans S, et al. 2023. Genomic methylation patterns in pre-meiotic gynoecia of wild-type and RdDM mutants of Arabidopsis. *Frontiers in Plant Science* 14:1123211

56. He F, Xu C, Fu X, Shen Y, Guo L, et al. 2018. The MicroRNA390/TRANS-ACTING SHORT INTERFERING RNA3 module mediates lateral root growth under salt stress via the auxin pathway. *Plant Physiology* 177:775–91

57. Yin W, Xiao Y, Niu M, Meng W, Li L, et al. 2020. ARGONAUTE2 enhances grain length and salt tolerance by activating *BIG GRAIN3* to modulate cytokinin distribution in rice. *The Plant Cell* 32:2292–306

58. Birchler JA, Yang H. 2022. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell* 34:2466–74

59. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, et al. 2023. InterPro in 2022. *Nucleic Acids Research* 51:D418–D427

60. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490

61. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–34

62. Tang H, Bowers JE, Wang X, Ming R, Alam M, et al. 2008. Synteny and collinearity in plant genomes. *Science* 320:486–88

63. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–202