

RESEARCH ARTICLE

Open Access



Expression divergence of expansin genes drive the heteroblasty in *Ceratopteris chingii*

Yue Zhang¹, Yves Van de Peer^{2,3,4,5}, Bei Lu^{1,6}, Sisi Zhang⁷, Jingru Che^{1,6}, Jinming Chen^{1*} , Kathleen Marchal^{2,8*} and Xingyu Yang^{7,9*}

Abstract

Background Sterile-fertile heteroblasty is a common phenomenon observed in ferns, where the leaf shape of a fern sporophyll, responsible for sporangium production, differs from that of a regular trophophyll. However, due to the large size and complexity of most fern genomes, the molecular mechanisms that regulate the formation of these functionally different heteroblasty have remained elusive. To shed light on these mechanisms, we generated a full-length transcriptome of *Ceratopteris chingii* with PacBio Iso-Seq from five tissue samples. By integrating Illumina-based sequencing short reads, we identified the genes exhibiting the most significant differential expression between sporophylls and trophophylls.

Results The long reads were assembled, resulting in a total of 24,024 gene models. The differential expressed genes between heteroblasty primarily involved reproduction and cell wall composition, with a particular focus on expansin genes. Reconstructing the phylogeny of expansin genes across 19 plant species, ranging from green algae to seed plants, we identified four ortholog groups for expansins. The observed high expression of expansin genes in the young sporophylls of *C. chingii* emphasizes their role in the development of heteroblastic leaves. Through gene coexpression analysis, we identified highly divergent expressions of expansin genes both within and between species.

Conclusions The specific regulatory interactions and accompanying expression patterns of expansin genes are associated with variations in leaf shapes between sporophylls and trophophylls.

Keywords Expansin, Ferns, *Ceratopteris chingii*, Phylogeny, Coexpression network, Heteroblasty

*Correspondence:

Jinming Chen
jmchen@wbcas.cn
Kathleen Marchal
Kathleen.Marchal@UGent.be
Xingyu Yang
sanmu109@163.com

¹ Aquatic Plant Research Center, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

² Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

³ VIB Center for Plant Systems Biology, 9052 Ghent, Belgium

⁴ Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa

⁵ College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, China

⁶ University of Chinese Academy of Sciences, Beijing 100049, China

⁷ Wuhan Institute of Landscape Architecture, Wuhan 430081, China

⁸ Department of Information Technology, IDLab, IMEC, Ghent University, 9052 Ghent, Belgium

⁹ Hubei Ecology Polytechnic College, Wuhan 430200, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Heteroblasty is a developmental trajectory event where plants have rapid ontogenetic changes across multiple traits, as exemplified by plants switching from distinct juvenile to adult leaves [1]. Heteroblastic leaves, which exhibit different morphologies and functions, enable plants to adapt to environmental heterogeneity, serving as a prime example of adaptive evolution. For thriving in undulating air-water environments, amphibious and aquatic plants tend to form heteroblastic leaves at different developmental stages [2, 3]. While previous studies uncovered some of the underlying differences of heteroblasty at the morphological, physiological, and molecular levels in aquatic seed plants [4–7], the mechanism behind sterile-fertile leaf dimorphy, i.e., heteroblasty with divergent functionalities, remains relatively unexplored in aquatic ferns.

Ferns are an ancient lineage of vascular plants and occupy a key phylogenetic position as a sister to the seed plants [8–10]. Given their wide distribution and diverse habitats, ranging from submerged environments to alpine environments, ferns exhibit a remarkable diversity in leaf shape, and sterile-fertile heteroblasty is common [11, 12]. The genus *Ceratopteris*, a model aquatic fern with heteroblastic leaves, displays distinctive patterns of genomic evolution compared to terrestrial ferns and seed plants [13, 14]. Among the species in this genus, *C. chingii* ($2n = 78$) is a leptosporangiate fern species closely related to the model system *C. richardii* ($2n = 78$), and it has a wide geographic distribution in Asia [15–18]. In *C. chingii*, the abaxial leaf surface of the sporophyll is curved to enclose the simple-structured sporangium originating from a single superficial initial cell in the leaf through transverse division [19, 20]. On the other hand, the trophophyll, which is responsible for primary buoyancy and plant floating, grows with an unfolded leaf surface (Fig. 1A) [15, 21]. The dramatic morphological differences between sporo- and trophophylls make *C. chingii* an ideal material for studying the molecular mechanism underlying heteroblastic leaf formation in aquatic ferns. Additionally, expansin (EXP), a plant cell-wall loosening protein, played a crucial role in cell shape plasticity and leaf morphology by facilitating the disassembly, remodeling, and adjustment of the cell wall throughout leaf development [22–24]. However, it remains largely unknown whether the expression patterns of expansins are distinct in heteroblastic leaves, particularly in ferns.

Because the heteroblastic leaves were produced by the same plant, differences in the gene expression and regulation play important roles [25, 26]. However, due to the large genome size of most ferns, reference assembly by whole genome sequencing is not straightforward and limits genomic research [27]. Recently, only a few fern

genomes have been sequenced [15, 28–30], including five homosporous ferns: *C. richardii* [15], *Adiantum capillus-veneris* [30], *A. nelumboides* [31], *Marsilea vestita* [32], and *Alsophila spinulosa* [28], as well as two heterosporous ferns: *Azolla filliculoides* and *Salvinia cucullata* [29]. *C. chingii* belongs to the homosporous ferns, which generally have relatively large genomes compared to their heterosporous counterparts [18, 27, 33]. Full-length transcriptome sequencing has provided an attractive alternative for gathering information on gene transcripts and predicting gene models without a reference genome [34, 35]. The strategy of combining long- and short-read transcriptome sequencing has been widely applied to study species-specific traits in seed plants [36, 37].

In this study, we applied the same strategy of combining long- and short-read RNA sequencing to construct reference gene models and study gene expression relevant to sporophyll, trophophyll, and root tissues in *C. chingii*. Given the key roles of expansin genes in the leaf development of seed plants [22, 38], we aimed to investigate the expression patterns of *C. chingii* expansin (*CcEXP* genes) and how they regulated the formation of heteroblastic leaves in *C. chingii*. We found that *CcEXP* genes were highly expressed in young sporophylls and trophophylls. Building on the previous study of expansin in several ferns [39, 40], we further constructed phylogenies of the expansin gene family and identified, by means of coexpression analysis, regulatory factors that positively or negatively affect *CcEXP* gene expression. In addition, we compared the expression conservation of orthologs between ferns and lycophytes, as well as the co-expressed orthologous relationships between *C. chingii* and the model seed plant *Arabidopsis thaliana*.

Results

Gene annotation by full-length transcriptome sequencing

To obtain a representative transcriptome for *C. chingii*, we performed a PacBio full-length RNA sequencing (Fig. 1A) and obtained a total of 915,702 reads with an average length of 89,338 bp (Additional file 1: Fig. S1, Table 1). After the removal of low-quality reads, circular consensus sequencing reads were classified into 513,861 (73.49%) full-length non-chimeric reads (FLNCs) and 155,285 (26.51%) non-full-length reads (Table 1). By error-correcting these FLNCs using non-full-length reads and Illumina Seq short reads, 37,428 non-redundant high-quality full-length transcripts with an average length of 1373 bp and an N50 of 1563 bp were obtained (Additional file 1: Fig. S1, Table 1). Using these non-redundant high-quality full-length transcripts, we constructed de novo gene models. In total, we obtained 24,024 transcript families, each of which corresponded to one gene model. Among these

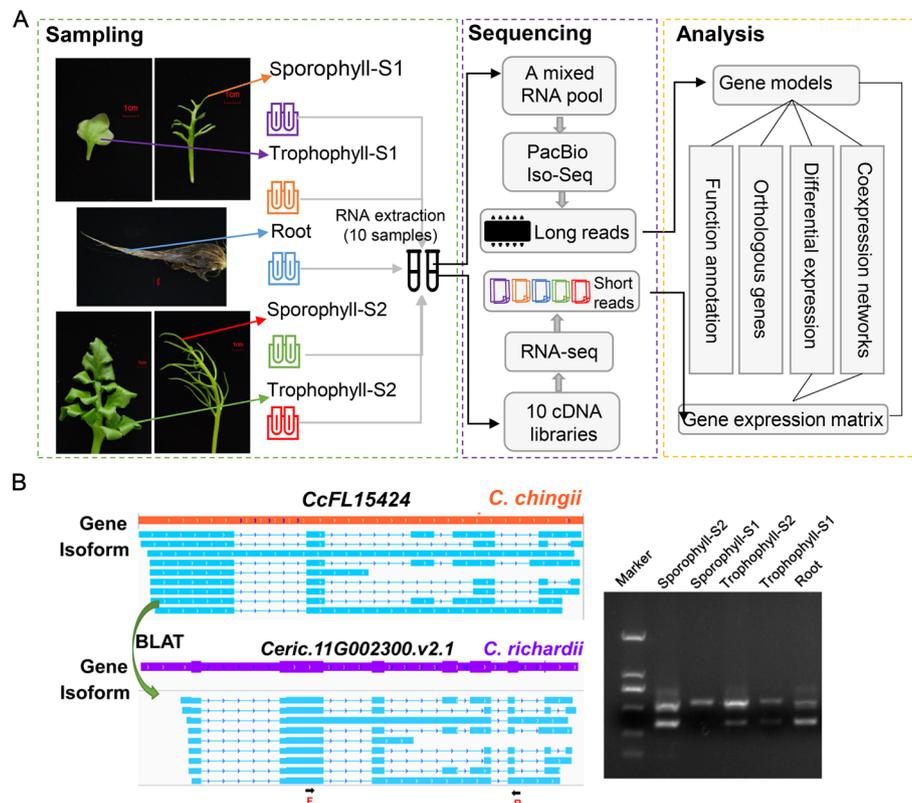


Fig. 1 Generation of the dataset and gene models for *C. chingii*. **A** Flowchart for gene annotation by PacBio full-length sequencing and Illumina RNA-seq for *C. chingii*. Sampling: photos show 5 tissue samples from different developmental stages (two biological repeats per developmental stage), of which the root system was shown but the single root sample was collected, and the total RNA of 5 tissue samples was extracted. Sequencing: RNA of each sample was used to build a cDNA library. Each library was subjected to Illumina short-read sequencing. In parallel, an RNA pool was made by mixing equal quantities of RNAs taken from each 10 samples. This pool was subjected to PacBio full-length sequencing. Analysis: reliable gene models of *C. chingii* were assembled by combining long with short read sequence data. In addition, mapping the short reads on the inferred gene models allowed quantifying the expression of each gene per developmental stage (expression matrix). **B** RT-PCR validation of a high-confident gene model. The isoforms of *CpFL15424* were mapped to *C. richardii* using BLAT and further clustered in *Ceric.11G002300.v2.1* gene regions. The arrows show the loci of the PCR primers (F, forward, and R, reverse) on the last isoform. The RT-PCR amplifications in *C. chingii* for the different tissues are shown in the right panel

Table 1 Summary of the full-length transcript sequencing of *C. chingii* on the PacBio Sequel platform

Category	Nanopore full-length sequencing
Raw read number	915,702
Base number of raw reads (Gbp)	81.81
Mean length of raw reads (bp)	89,338
Circular consensus sequencing reads number	669,146
FLNC number	513,861
FLNC percentage (%)	73.49%
N50 of FLNC (bp)	1551
Mean length of FLNC (bp)	1368
Max length of FLNC (bp)	3181
Non-full-length number	155,285
Non-full-length percentage (%)	26.51%

transcript families, 9267 were UniTransModels and 14,757 transcripts belonged to singleton genes with one isoform only. BUSCO analysis of the gene models showed that our gene models achieved 72.2% completeness according to the eukaryote (odb10) core gene dataset, indicating a comprehensive transcriptome dataset.

Full-length transcriptome sequencing only captures information on exonic regions. To predict the intronic regions of the gene models as well, the high-quality full-length transcripts were mapped to the chromosome-level reference assembly of *C. richardii*, assuming a high similarity in gene content and structure between *C. richardii* and *C. chingii*. If all transcripts were assigned to a *C. chingii* gene model mapped to a unique gene locus on the *C. richardii* genome, the gene model was regarded as a high-confident homologous match.

This was the case for 18,079 gene models (75.25%), and the sequence intron region was transferred from the *C. richardii* to *C. chingii* gene. *C. chingii* gene models with ambiguous mapping (4262 (17.74%)) or those without mapping (1692 (7.04%)) were classified as low-confident gene models. To validate the accuracy of the inferred gene structure of the *C. chingii* genes, 10 randomly selected splicing events identified in the inferred gene models were subjected to RT-PCR (Fig. 1B, Additional file 1: Fig. S2). The size of all of the amplified products corresponded to the size of the exons based on the predicted gene structure. This indicates a high similarity between the gene coding regions of both fern genomes exists (i.e., high-confident genes in *C. chingii*), which further supports the validity in extrapolation of the gene structure information.

To functionally annotate our gene models in *C. chingii*, we conducted a comprehensive local alignment against various sequence databases, including GO, KEGG, KOG, NCBI Nr, and SwissProt. A total of 19,565 (81.43%) gene models showed matches to sequences with known annotation in at least one of the five databases. Furthermore, 5079 (21.14%) gene models exhibited matches with annotation in all five databases (Additional file 1: Fig. S3). To annotate transcription factors [41, 42], the *C. chingii* gene models were mapped to the PlantTFDB database. This analysis led to the identification of 603 transcription factors belonging to 72 families, with the C3H and the bHLH families being the largest (Additional file 1: Fig. S4). Next to protein-coding genes, we also discovered 303 long non-coding RNAs (LncRNAs), potentially involved in the post-transcriptional regulation.

Expansin (EXP) genes are differentially expressed between trophophylls and sporophylls

To quantify the expression level of genes in different tissues of *C. chingii*, we conducted short-read RNA-seq on five tissue samples, with two biological repeats per tissue sample (non-pooled). The resulting short reads were mapped to the identified gene models to determine the read counts and gene expression levels, measured as FPKM. To identify tissue-dependent expression patterns that contribute to tissue morphology [43], we performed differential expression analysis between any two pairs of tissues (Fig. 2A). The largest number of differentially expressed genes (DEGs) was identified between trophophyll-S2 and root (7741), with 3846 upregulated genes and 3895 downregulated genes (Fig. 2A). Conversely, the smallest number of DEGs were observed between sporophyll-S2 and trophophyll-S2, with 2461 genes, of which 925 upregulated and 1536 downregulated in the sporophyll compared to the trophophyll (Fig. 2A).

We found three significantly enriched GO terms “reproduction,” “flower development,” and “post-embryonic development,” in the comparison of sporophyll-S1 and trophophyll-S1, suggesting their potential involvement in sexual reproduction. Additionally, three significantly enriched GO terms in the cellular component class were found overrepresented between trophophyll and sporophyll, namely “extracellular region,” “external encapsulating structure,” and “cell wall” (Fig. 2B, Additional file 1: Fig. S5), all of which are associated with cell wall morphology. Furthermore, based on their functional annotation of the Swissprot database, 20 genes within these three significantly enriched GO terms were predicted to be expansin (EXP) proteins and were the major component. These EXP genes exhibited differential expression in at least one of the four comparisons between sporophyll and trophophyll (Additional file 1: Fig. S6), i.e., trophophyll-S1 vs sporophyll-S1, trophophyll-S1 vs sporophyll-S2, trophophyll-S2 vs sporophyll-S1, and trophophyll-S2 vs sporophyll-S2. We visualized their expression pattern in a heatmap (Fig. 2C), which revealed that 12 (60%) of the EXP genes show higher expression levels in sporophyll-S1 compared to other tissue samples (Fig. 2C). This expression pattern suggests that the increased expression of EXP gene in the sporophyll-S1 may influence cell morphology and potentially contribute to sporophyll formation.

To validate the upregulation of EXP genes at the protein levels, we quantified the EXP content in the five same tissue samples used for RNA-seq. Notably, the EXP content was found to be the highest in sporophyll-S1 and showed significant differences when compared to trophophyll-S1 and trophophyll-S2 (t -test, $p < 0.01$, Fig. 2D). Similarly, significantly more expansin protein was identified in sporophyll-S2 than in trophophyll-S1 and trophophyll-S2 (t -test, $p < 0.05$, Fig. 2D). These findings suggest that the increased abundance of expansin proteins may contribute to the more intricate cellular organization and distinct tissue phenotype observed in sporophyll compared to trophophyll. Overall, this highlights the significant association between differentially expressed EXP genes and the divergent phenotypes of heteroblastic leaves in *C. chingii*.

Identification and phylogenetic analysis of expansin gene family in *C. chingii*

Previous phylogenetic analyses have classified expansins into at least four subfamilies [44, 45]. However, these analyses were restricted to seed plants and a few ferns [39, 40]. To expand the phylogeny to include more fern lineages, we first identified expansin genes in *C. chingii*. We selected all expansin-like genes that contained the conserved DPBB-1 and the Expansin_C domain, known

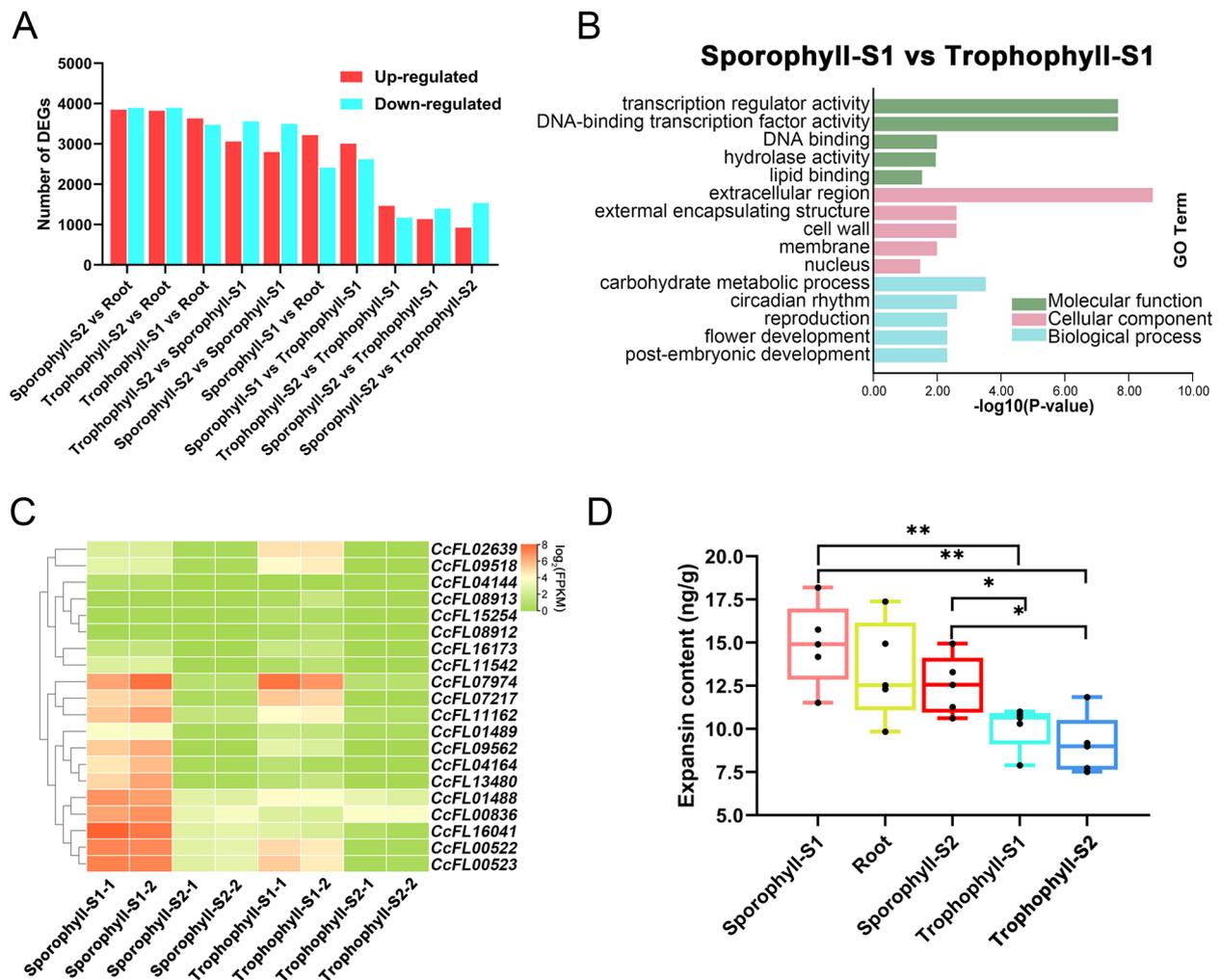


Fig. 2 Differential gene expression patterns in *C. chingii*. **A** Bar chart showing the number of differentially expressed genes between five tissues. Red bars indicate upregulated genes, and blue bars are downregulated genes. **B** GO enrichment of the genes that are differentially expressed between young sporophylls and young trophophylls. **C** Heatmap showing the expression pattern of 20 expansin genes in eight tissue samples. **D** Box graph showing the expansin content in five tissues. The significance of the differences in the expansin content between tissues was tested by the t-test. * p -value < 0.05, ** p -value < 0.01

to be a characteristic of the expansin protein function. This resulted in 26 *CcEXPs* from the 28 candidate genes, with an average of 271 amino acids (Additional file 2: Table S1). We applied similar criteria to identify candidate expansin genes in 19 additional plant genomes, including an outgroup (*Chlamydomonas reinhardtii*), chlorophytes (*Klebsormidium flaccidum*, *Mesotaenium endlicherianum*, *Spirogloea muscicola*), bryophytes (*Marchantia polymorpha*, *Physcomitrium patens*, *Sphagnum fallax*), lycophytes (*Isoetes taiwanensis*, *Selaginella moellendorffii*), ferns (*Azolla filiculoides*, *C. chingii*, *C. richardii*, *Salvinia cucullata*), and spermatophytes (*Amborella trichopoda*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Ceratophyllum demersum*, *Cycas*

panzhihuaensis, *Picea abies*). This resulted in 538 EXP candidate genes (Fig. 3A). By including more genomes, the number of EXP genes largely increased (Fig. 3A). Interestingly, the basal species within each cluster have relatively fewer EXP genes compared to other species in the same cluster (Fig. 3A). Using the maximum likelihood (ML) method, we generated a phylogeny of all EXP genes, with Chlorophyta as an early divergent lineage (Fig. 3B). In seed plants, the expansins are subdivided into 4 subfamilies: EXPA, EXPB, EXLA, and EXLB. *C. chingii* only has representatives in the EXPA and EXPB subfamilies (Additional file 1: Fig. S7, Additional file 2: Table S1). We also observed the absence of EXLA and EXLB genes also for other ferns, suggesting that ferns lack these subfamily

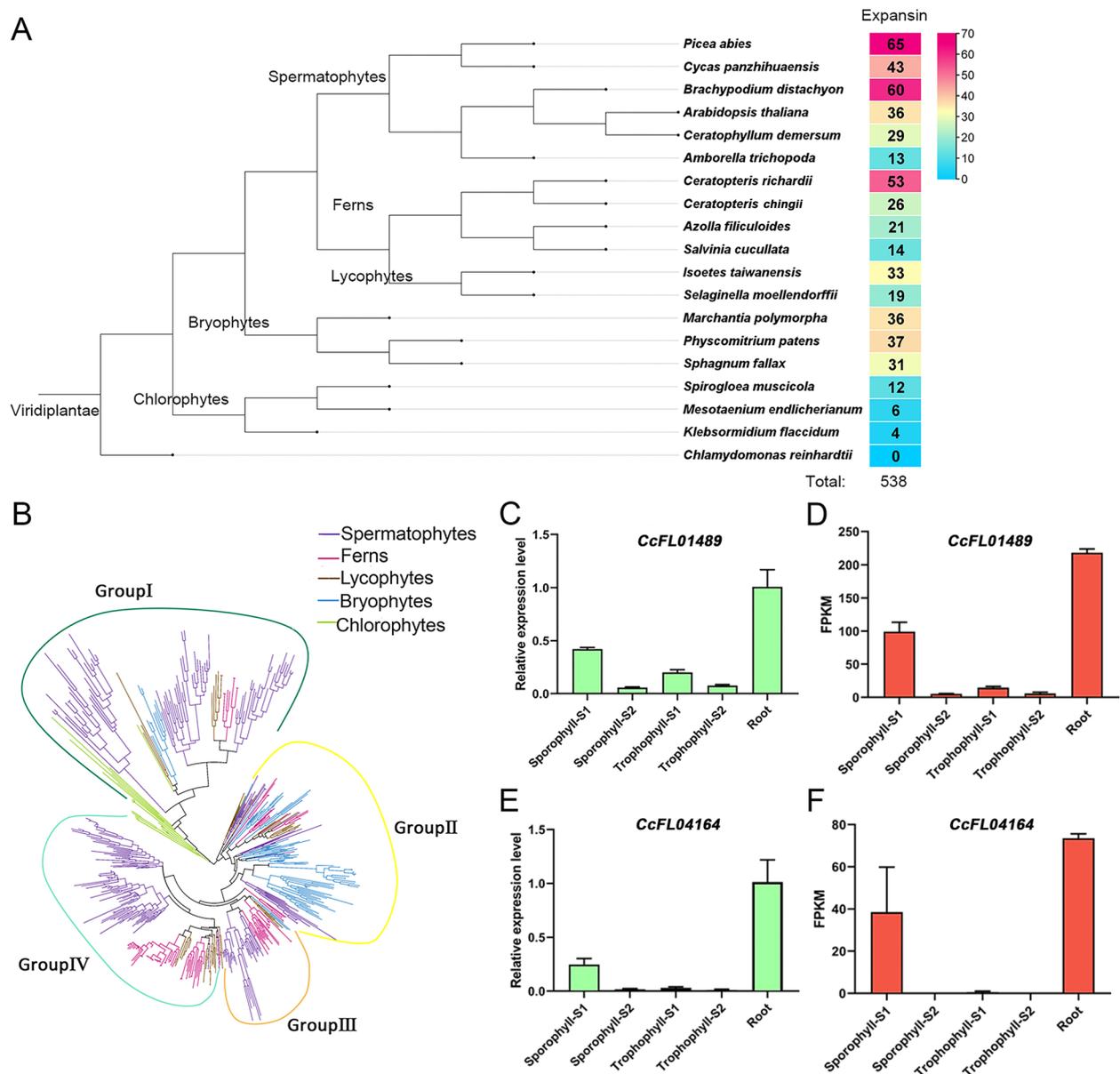


Fig. 3 Phylogeny of the expansin gene family. **A** Evolutionary tree topology representing the species from which EXP genes were retrieved. Numbers indicate the number of identified EXP sequences in these 19 species. **B** Phylogenetic tree of expansins, built with maximum likelihood. Branches are colored, depending on the species of origin from which the extant sequence was derived (spermatophyte, ferns, lycophytes, bryophyte, and chlorophyte). Four different phylogenetic groups of EXP genes can be distinguished. **C–F** Bar plots showing the qRT-PCR results and RNA-seq derived expression levels of *CcFL01489* and *CcFL04164* in five tissues

expansin genes. Using MEME, we identified ten conserved motifs in the *C. chingii* EXP proteins (Additional file 1: Fig. S8).

The phylogenetic tree of EXP genes is composed of four distinct phylogenetic groups, labeled as group I to group IV (Fig. 3B, Additional file 2: Table S2). Group I consisted of EXPB, EXLA, and EXLB genes, while groups II~IV

only contained EXPA genes (Fig. 3B). Group I EXP genes are found in all 18 plant genomes, whereas group II and group III consist of EXP genes from 15 plant genomes, excluding chlorophytes (Additional file 1: Fig. S9). Interestingly, group IV exclusively contains EXP genes from vascular plants, including lycophytes, ferns, and spermatophytes. These findings suggest that vascular plants

possess a class of specific EXP genes, and their diversification likely contributes to the development of the vascular system.

In *C. chingii*, 69.23% of *CcEXP* genes (18 genes) were present in group III and group IV, with 14 *CcEXP* genes showing differentially expressed in at least one of the four comparisons mentioned above (Additional file 1: Fig. S6). Our results indicate that seven and five *CcEXP* genes are relatively more highly expressed in sporophyll-S1 and trophophyll-S1, respectively, compared to other leaf tissue. These results highlight the distinct expression patterns of *CcEXP* genes in these two groups. To validate the accuracy of the expression pattern of these *CcEXP* genes, we performed qRT-PCR experiments in all five tissues (Additional file 1: Fig. S10 and Fig. 3C–F for the tissue-specific *CcEXP* genes). Due to the high sequence similarity between gene pairs *CcFL01488-CcFL01489* and *CcFL08912-CcFL08913*, we tested only one gene from each pair, namely *CcFL01489* and *CcFL08913* (Fig. 3C and Additional file 1: Fig. S10I). The relative expression level of the tested *CcEXP* genes in the five tissues showed an average correlation of 0.84 with their FPKM expression level (Additional file 2: Table S3), confirming the expression patterns derived from RNA-seq data.

Gene co-expression networks of *CcEXP* genes representative of different phylogenetic groups

In line with previous studies [42], we employed coexpression analysis using the *CcEXP* genes as query genes to identify additional genes in *C. chingii* that are involved in the same process as those query genes and to infer transcription factors that regulate the *CcEXP* query genes. A total of 23 filtered *CcEXP* genes were utilized to construct a co-expression network, comprising two *CcEXP* genes from group I, five from group II, seven from group III, and nine from group IV. The degree of coexpression between a *CcEXP* query gene and other genes was evaluated using the Spearman correlation coefficient (r). Only a few strong coexpression relations ($|r| > 0.95$) were detected between the *CcEXP* query genes themselves, particularly within genes of the same group, indicating distinct expression behavior among different *CcEXP* genes. By applying a stringent threshold for correlation ($|r| > 0.95$ and p -value $< 1e-5$), 1501 genes were identified to be significantly positively ($r > 0.95$) or negatively ($r < -0.95$) coexpressed with *CcEXP* genes, including 74 transcription factors and 26 lncRNAs. The majority of these genes (907, 60.42%) showed significantly coexpressed with one *CcEXP* gene only, while 594 (39.57%) genes showed a coexpression behavior with multiple *CcEXP* genes. Most of the significant coexpression relations (2511, 86.82%) identified using the Spearman

correlation were ranked in the top 1%, confirming the robustness of the results. The limited overlap in coexpression relationships among *CcEXP* genes suggests their involvement in distinct regulatory networks and their different roles in tissue morphogenesis. Furthermore, no DEGs related to reproduction were identified as significant coexpressed genes with any *CcEXP* genes, indicating a rare interaction between spore development and heteroblastic leaf formation in *C. chingii*.

Figure 4A–D depicts the specific coexpression network for each group of *CcEXP* genes. Each network was constructed using only the *CcEXP* genes from a particular group to calculate coexpression. In the coexpression networks of the group I and group II *CcEXP* genes, with the exception of one pair (*CcFL00522* and *CcFL00523*), most query *CcEXP* genes do not share any coexpressed genes (Fig. 4A, B). This suggests that *CcEXP* genes, which have a common ancestor in spermatophytes and bryophytes, have likely diverged in function throughout the species' evolution. In contrast, the coexpression network of the group III and group IV *CcEXP* genes show more shared coexpressed genes, indicating a more similar level of expression regulation (Fig. 4C, D). GO enrichment analysis of the coexpressed genes reveals their involvement in the extracellular region and cell wall for group III and group IV *CcEXP* genes (Additional file 1: Fig. S11).

Given the significance of transcriptional regulation, we focused on identifying transcription factors and lncRNAs in each of the coexpression networks corresponding to the different phylogenetic groups. In these coexpression networks, members of the bHLH family were most abundant, followed by members of the WRKY family (Additional file 1: Fig. S12). Additionally, 37 lncRNA were predicted to interact with 17 *CcEXP* genes (Additional file 2: Table S4). Among *CcEXP* genes, 17 (73.91%) were predicted to be regulated by at least one transcription factor or one lncRNA. However, these inferred regulatory relationships were rarely shared between different *CcEXP* genes. Only the *CcEXP* genes of group III were regulated by at least one transcription factor or lncRNA. This suggests that for most *CcEXP* genes the regulatory interactions are quite specific.

Predicting regulatory relations based on coexpression is prone to false positives. To further validate the inferred regulatory interactions between *CcEXP* genes and transcription factors, the *CcFL01489* gene of group III along with four transcription factors, inferred to regulate this gene were subjected to a yeast two-hybrid assay (Fig. 4E, Additional file 1: Fig. S13). Since the *CcFL01489* gene exhibited a high expression level in sporophyll-S1 and the highest correlation coefficient between RNA-seq and RT-PCR (Additional file 2:

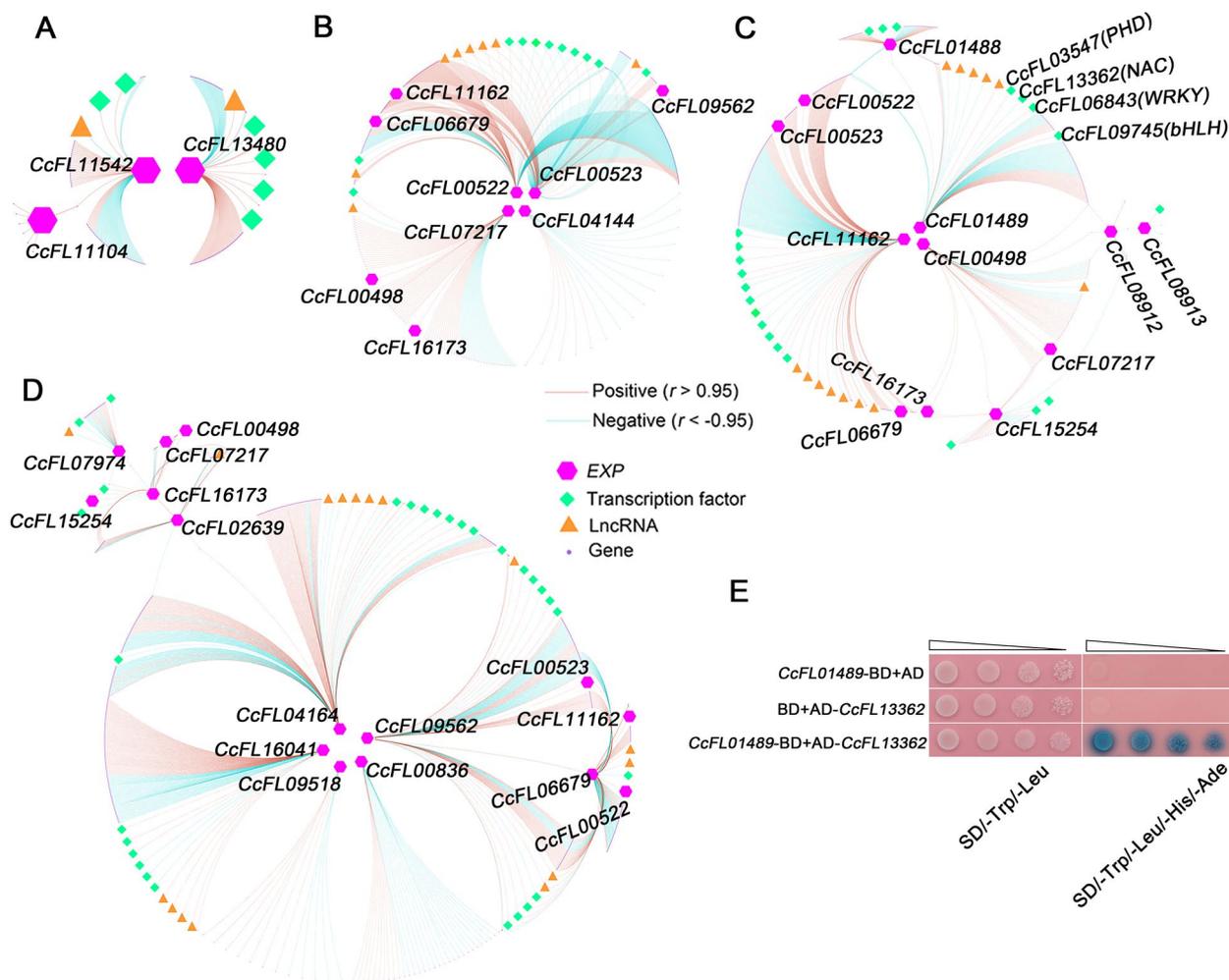


Fig. 4 Coexpression interactions of *C. chingii* expansin genes. **A–D** Gene coexpression networks of the *CcEXP* genes belonging to the different phylogenetic groups: group I (**A**), group II (**B**), group III (**C**), group IV (**D**). The hexagons represent the *CcEXP* genes of a group, rhombus represent the transcription factors, triangles represent the lncRNA, and circles represent the genes coexpressed with the *CcEXP* genes. Red links represent significant ($r > 0.95$) positive coexpression interactions and blue links represent significant ($r < -0.95$) negative coexpression interactions. **E** Yeast two-hybrid assay showing how *CcFL01489* interacts with NAC transcription factor (*CcFL13362*). Empty vector pGKBT7 (BD) and pGADT7 (AD) as the negative control. Recombinant plasmids were transformed into Y2H gold yeast strain and grown on selection medium at 30 °C for 3 days

Table S3), it might be associated with the formation of heteroblasty in *C. chingii*. We intentionally measured direct protein-protein interactions, rather than protein DNA interactions, as we did not have access to the non-coding sequence of the *CcEXP* genes in *C. chingii*. As shown in Fig. 4E, the positive transformants of the yeast strain (white) grow normally on the SD/-Trp/-Leu dropout medium. However, only one yeast strain (blue) harboring G1-BD+CcFL13362 survived on the SD/-Trp/-Leu/-His/-Ade + X- α -gal dropout medium. Therefore, our results indicate that *CcFL01489* interacts with one transcription factor *CcFL13362* and that this interaction likely results in the observed high expression of *CcFL01489* in *C. chingii* sporophyll-S1.

Expression patterns and coexpression interactions of EXP genes are largely species-specific

As differences in the expression patterns of orthologs often relate to phenotypic interspecies differences [46], we investigated how EXP orthologs from different fern species differ in their expression behavior. To do this, we built the gene expression matrices of leaf and root samples (Additional file 2: Table S5) from three fern species (*C. chingii*, *C. richardii*, *S. cucullate*) and two lycophyte species (*I. taiwanensis*, *S. moellendorffii*) (see the “Methods” section). To account for the differences in the number and developmental stage of the tissue samples available for each species, we used the method described by Tirosh et al. [47] to determine the degree to which the

expression behavior is conserved between pairs of highly similar orthologs in *C. chingii* and respectively each of the abovementioned species (EC expression conservation, see the “Materials” section). In our genome-wide analysis, considering all identified highly similar orthologous gene pairs in each species combination, we found the highest level of expression conservation (EC) for orthologous pairs of the closely related lineages *C. chingii* and *C. richardii* (Fig. 5A). However, when comparing *C. chingii* with other species (*S. cucullata*, *I. taiwanensis*, and *S. moellendorffii*), orthologous pairs showed, on average, a low level of EC ($-0.05 \sim 0.1$), indicating a high divergence in expression behavior in leaf and root tissues (Fig. 5A) between ferns and between fern and lycophyte.

When we specifically analyzed the EC for orthologous pairs of EXP genes, we found that, consistent with genome-wide analysis, the EC level of the close orthologous EXP gene pairs was significantly higher in the comparison of *C. chingii* vs *C. richardii* than in comparisons of *C. chingii* with other ferns (Mann-Whitney *U* test, $p < 0.05$, Fig. 5B). Rarely, we observed the EC of close orthologous EXP genes in the top 5% EC values observed in the genome-wide comparison. This was true for each of the different species comparisons, further confirming that orthologous EXP genes evolved largely distinct expression patterns.

To investigate how the coexpression interactions of EXP genes belonging to a specific group evolved during species differentiation, we compared the gene

coexpression network constructed for *C. chingii* involving the EXP genes in each phylogenetic group with that of the corresponding orthologous EXP genes in *C. richardii* and the outgroup *A. thaliana* (see the “Methods” section) (Additional file 1: Fig. S14). Similar to what we observed for *C. chingii*, coexpression relationships between EXP in *C. richardii* or *A. thaliana* were rarely significant (Additional file 1: Fig. S14), indicating the dramatically distinct expression patterns for this gene family, even within species. Moreover, the coexpression networks of EXP genes belonging to the same group in the comparisons between *C. richardii* and *C. chingii* and between *A. thaliana* and *C. chingii* contain transcription factors of the same family. For example, the MYB transcription factors were coexpressed with group III EXP genes in respective *C. chingii* and *C. richardii* coexpression networks; bHLH and bZIP transcription factors were coexpressed with group II EXP genes in respective *C. chingii* and *A. thaliana* coexpression networks. This indicates a conserved regulation of EXP genes and transcription factors between plant species. Genes coexpressed with the *AtEXP* genes of group IV (Additional file 1: Fig. S15A) were enriched for pathways involved in response to light and abiotic stimuli, highlighting the involvement of environmental stress in the formation of spermatophyte-specific EXP genes. After comparing the GO-enriched results of coexpressed genes of group IV EXP between *A. thaliana* and *C. chingii*, the divergent functions were identified, which could be related to the

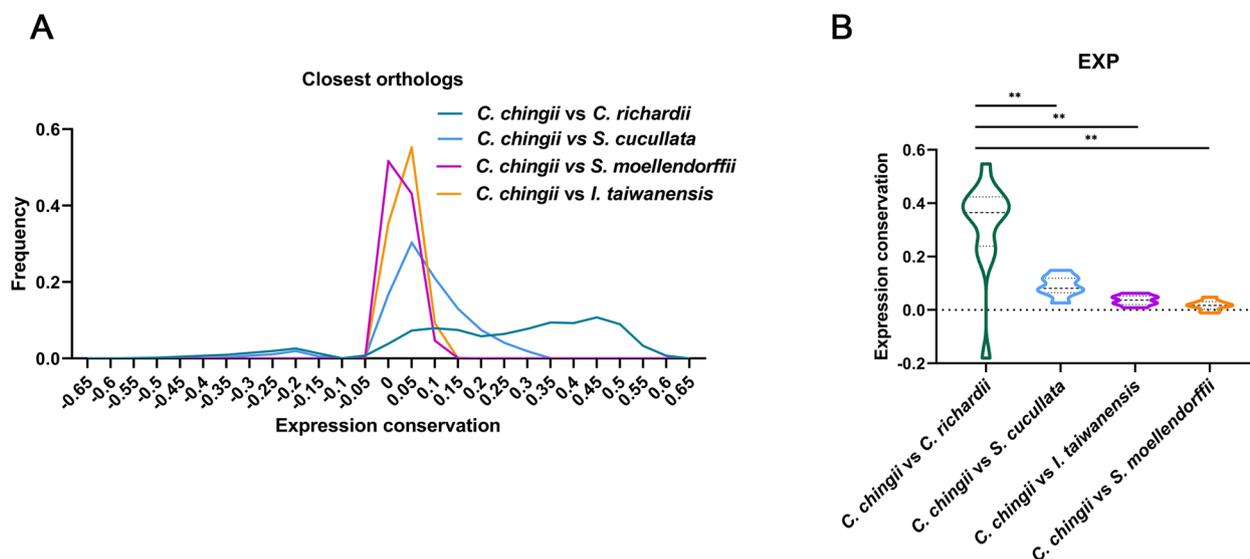


Fig. 5 Expression conservation of orthologous gene pairs. These plots show the degree to which highly similar orthologs tend to have the same expression behavior in two compared species (expression conservation). Expression conservation (EC) and highly similar orthologs are defined as described in materials and methods. **A** The frequency plot shows the distribution of the EC values for the comparison of all highly similar orthologs between the two indicated species. **B** Violin plot showing the distribution of the EC values for the comparison of orthologous EXP gene pairs between the two indicated species. Significance in EC was assessed by the Mann-Whitney *U* test. * p -value < 0.05 , ** p -value < 0.01

phenotypic differences between ferns and seed plants (Additional file 1: Fig. S15A-B).

Discussion

Comparing the gene expression between trophophylls and sporophylls provides insight into the developmental mechanisms underlying the diversity of fern leaves and the regulatory processes driving the formation of these phenotypically differentiated leaves [11]. In this work, we focused on *C. chingii*, a species closely related to the model fern *C. richardii* [15], which exhibits phenotypically differentiated trophophylls and sporophylls. To analyze the gene expression without a reference genome, we first generated gene models based on long-read transcriptome data of a mixed RNA pool. By combining these gene models with short-read RNA data of the separate leaf tissues (trophophylls and sporophylls sampled at two developmental stages and roots (as reference)), we were able to study the gene expression patterns. The unique data and gene models generated in this work will provide a valuable resource for future phylogenetic and molecular studies in ferns.

We identified 24,024 gene models, of which 75.25% showed a high level of similarity with the genes annotated in the *C. richardii* genome. This high similarity allowed extrapolating the intronic/exonic gene structure from *C. richardii* to *C. chingii*, which was verified for 10 randomly sampled genes by RT-PCR (Fig. 1B, Additional file 1: Fig. S2). Differential expression analysis between trophophylls and sporophylls of two developmental stages revealed a high degree of gene expression divergence between the two types of heteroblasty. Genes related to sexual reproduction were differentially expressed but only identified in the comparison between trophophyll-S1 and sporophyll-S1 (Fig. 2B). This suggests that spore development of *C. chingii* primarily occurs during the juvenile stage (i.e., S1). Notably, 20 differentially expressed genes belonged to the expansin gene family in all comparisons of heteroblastic leaves at different developmental stages. Expansin (EXP) is a plant cell wall-loosening protein. Recent studies of leaf development in seed plants revealed that overexpressing expansin activity increases leaf size and affects leaf shape by the expansion and enlargement of plant cells [48–50]. As a key regulator of cellulose/xyloglucan networks, expansin contributes to cell wall expansion [22, 51]. The dynamic expression of expansins is key to plant growth and development [22, 38], and it contributes to cell shape plasticity and organ morphology. The differential expression of EXP genes between trophophylls and sporophylls in *C. chingii* indicates that this gene family is also involved in shaping the differences in the cell wall between these two leaf types. RT-PCR confirmed that 12 of the EXP genes were most

highly expressed in young sporophylls (Fig. 2C). Consistent with these results, we observed higher EXP content in young sporophyll compared to other tissues. This suggests that increased EXP gene expression contributes to curving sporophyll in *C. chingii*.

Having the specific gene models for ferns allowed us to reconstruct an EXP phylogeny including both ferns and seed plants. This phylogenetic tree suggests that EXP members have progressively expanded in the order of chlorophytes, bryophytes, lycophytes, ferns, and spermatophytes (Fig. 3A). Consistent with previous studies in bryophytes [52, 53], only members of the EXPA and EXPB subfamilies were identified in *C. chingii* other ferns, and lycophytes, but no members of the EXLA and EXLB subfamilies. Our phylogenetic tree also shows the presence of a vascular plant-specific group, indicating that the sequence evolution of orthologs of this gene family reflects the divergence of plant species [54, 55]. We speculate that these vascular plant-specific EXP genes may contribute to the formation of vascular systems and the divergent functions of coexpressed genes of EXP may be associated with the different phenotypes between ferns and seed plants (Additional file 1: Fig. S15). By combining the results of RNA-seq and qRT-PCR experiments, we were able to demonstrate that EXP genes in *C. chingii* exhibit a unique expression pattern that suggests an important role in the formation of young sporophylls and young trophophylls. To understand the regulatory network of heteromorphic leaf development in *C. chingii*, we performed gene coexpression network analysis [42, 56]. Intriguingly, shared transcription factors, lncRNA, and coexpression behavior were rarely observed between EXP genes of *C. chingii*, indicating that each of these *CcEXP* genes is involved in distinct regulatory networks and functions. This diversity in gene expression behavior and regulation, coupled with the expansion of a gene family, is often associated with tissue-specificity and suggests neofunctionalization [57].

Analysis of the degree to which the expression behavior of highly similar orthologs is conserved between *C. chingii* and respectively other fern species shows that, except for the high degree of expression conservation between the two related *Ceratopteris* species, the expression patterns of *C. chingii* genes including the EXP genes are relatively different between fern species (low EC). Given the dramatic differences in fern leaf shapes, our result suggests that expression behavior has largely diverged, particularly within fern species. This low degree of coexpression conservation between EXP genes of different species was further confirmed by analyzing the degree to which interactions in the coexpression networks of the EXP genes were conserved across species. Coexpression interactions were mainly conserved

between *CcEXP* and *CrEXP* genes, as well as between *CcEXP* and *AtEXP* genes. Conclusively, the significant degree of sub and neofunctionalization observed suggests that EXP genes are associated with the formation of heteroblasty, which exhibits phenotypic differences between species.

Methods

Plant sample collection and RNA isolation

The wild *Ceratopteris chingii* plants were collected from the Zhangdu Lake at Wuhan (114° 31' E, 30° 52' N) and cultivated in the greenhouse in the Wuhan botanical garden. A single root sample was collected. We harvested whole leaves of both trophophylls and sporophylls at two developmental stages. Stage 1 (S1, young) had a blade length of 2~3 cm, while stage 2 (S2, old) had a blade length of 5~6 cm. The samples were washed with distilled water and then frozen in liquid nitrogen. Total RNA was extracted from each sample using the RNeasy Plant Kit (TIANGEN). RNA concentration and integrity were examined using the Qubit RNA Assay Kit in a Qubit 2.0 Fluorometer (Life Technologies) and the Agilent 2100 Bioanalyzer (Agilent Technologies).

PacBio Iso-Seq of a mixed RNA sample

Equal quantities of RNA were extracted from five tissue samples, including root, trophophyll-S1, sporophyll-S1, trophophyll-S2, and sporophyll-S2, for the construction of a PacBio cDNA library. Briefly, 1 µg of RNA from the pooled RNA sample was reverse transcribed using the Clontech SMARTer polymerase chain reaction (PCR) cDNA Synthesis Kit and oligo (dT) to generate the first-strand cDNA. The second strand was then synthesized and amplified with the 5' PCR primer to obtain full-length cDNA with barcodes. Size selection was carried out using the BluePippin Size Selection System protocol, with two size bins for the mixed sample: 1–4 kb and > 4 kb. Finally, the cDNA library was sequenced on the PacBio Sequel platform using a SMRT cell.

Iso-Seq data analysis

The standard Iso-Seq protocol (SMRTlink 4.0) was used to process raw sequencing data (81.81Gb) with min-length > 200 and min-read-score > 0.75. A total of 669,146 circular consensus sequencing (CCS) reads were generated from subread sequences with accuracy > 0.8 and further classified into full-length non-chimeric reads or non-full-length reads depending on whether the 5' primer, 3' primer, and a poly-A tail signal preceding the 3' prime were present. Full-length reads were clustered and assembled into consensus sequences with iterative clustering for error correction (ICE). Subsequently,

polished consensus reads were obtained by removing sequence errors using the non-full-length reads and Illumina RNA-seq data with Arrow and LoRDEC. To obtain the final non-redundant high-quality full-length transcripts, the error-corrected full-length polished consensus transcripts were merged and redundancy removed using CD-HIT with the following parameters: -c 0.99 -aS 0.99 -AS 30.

Gene structural annotation

Due to the lack of a reference genome for *C. chingii*, we processed non-redundant high-quality full-length transcripts were processed using the Coding GENome Reconstruction Tool (Cogent, <https://github.com/Magdoll/Cogent>). In summary, Cogent clusters transcripts into families based on their *k*-mer similarity and subsequently converts each transcript family into one or several unique transcript models (UniTransModels, further regarded as gene models) using a De Bruijn graph-based method. To improve the accuracy of the UniTransModels, the non-redundant high-quality full-length transcripts were mapped to the chromosome-level genome assembly of *C. richardii* using GMAP with -min-trimmed-coverage=0.85 --min-identity=0.9. If the transcripts from a *C. chingii* gene model were mapped to a unique gene locus on the *C. richardii* genome, the *C. chingii* gene model was considered a high-confident gene. A *C. chingii* gene model was considered a low-confident gene if a gene model's specific transcript (1) maps ambiguously to multiple gene loci in *C. richardii*, (2) partially maps to a unique gene locus in *C. richardii*, or (3) partially maps to multiple gene loci in *C. richardii*. Ten randomly selected splice events derived from 10 high-confident gene models were examined by RT-PCR, using the primers listed in Additional file 2: Table S6. RNA of each sample (1 µg) was reverse transcribed into cDNA and used for PCR amplification. PCR fragments were tested in 1% agarose gel.

Gene functional annotation and long non-coding RNA identification

For the functional annotation of the gene models, we used the following databases: Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Port, Pfam, and NCBI Non-redundant Protein (Nr). To analyze the enrichment of gene functions, we employed TBtools. To identify which of the gene models corresponded to transcription factors (TF), we used PlantTFDB/Itak. The coding potential of the gene models was evaluated by the following four tools: Coding Potential Calculator (CPC2), Coding-Non-Coding Index (CNCI), Pfam-scan, and Coding Potential Assessment

Tool (CPAT). For reliable identification of non-coding RNAs, we considered those with sequence lengths longer than 200 nt and an average FPKM over 0.1. The genes targeted by these lncRNAs were predicted using the LncTar (<http://www.cuilab.cn/lncTar>) and RPISeq (<http://pridb.gdcb.iastate.edu/RPISeq/>).

RNA-seq and transcriptome analysis

For each sample, the cDNA sequencing library was constructed and sequenced on the Illumina platform following the standard protocol. The reads were then trimmed and filtered by removing adapters and low-quality reads. Afterward, the trimmed and filtered reads were error-corrected to eventually obtain the polished consensus full-length sequences mentioned earlier. To estimate the gene expression, the trimmed and filtered reads were mapped to the obtained gene models using Hisat2. The expression level of each gene model (measured in fragments per kilobase of exon model per million mapped fragments (FPKM)) in each of the different samples was estimated using StringTie. Differentially expressed genes (DEGs) were identified using DESeq2, with a fold change (FC) greater than 2 and a false discovery rate (FDR) less than 0.05 as the criteria. Furthermore, all 12 *CcEXP* genes showing tissue-specific expression in sporophyll-S1 and trophophyll-S1 were selected for qRT-PCR experiments. However, due to the high sequence similarity of two pairs of genes (i.e., respectively *CcFL01488-CcFL01489* and *CcFL08912-CcFL08913*), we selected one representative of each pair only. As a result, qRT-PCR experiments were conducted on ten uniquely differentially expressed *CcEXP* genes. The experiments for the five RNA-seq samples were performed under the following conditions: 95 °C for 30 s; 40 cycles of 95 °C for 5 s, 60 °C for 30 s, and 72 °C for 15s; and 95 °C for 10 s. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with the *GAPDH* gene used as the internal standard. The primers for these 10 DEGs can be found in Additional file 2: Table S7.

Expansin gene identification and phylogenetic analysis

Homologous genes were identified with Orthofinder [58] with genomic data from a range of taxa sampled across the plant tree of life, including (a) four green algae: *Chlamydomonas reinhardtii* [59], *Klebsormidium flaccidum* [60], *Mesotaenium endlicherianum* [61], and *Spiroglaea muscicola* [61]; (b) three bryophytes: *Marchantia polymorpha* [62], *Physcomitrium patens* [63], and *Sphagnum fallax* [64]; (c) two lycophytes: *Isoetes taiwanensis* [65] and *Selaginella moellendorffii* [66]; (d) four ferns: *Azolla filiculoides* [29], *C. chingii* (this study), *C. richardii* [15], and *Salvinia cucullate* [29]; and (e) six seed plants: *Amborella trichopoda* [67], *Arabidopsis thaliana* ([\[www.arabidopsis.org\]\(http://www.arabidopsis.org\)\), *Brachypodium distachyon* \[68\], *Ceratophyllum demersum* \[69\], *Cycas panzhihuaensis* \[70\], and *Picea abies* \[71\]. All reported expansin protein sequences of *A. thaliana* were downloaded from TAIR11 \(<http://www.arabidopsis.org>\). The *A. thaliana* expansin \(*AtEXPs*\) protein sequences were mapped to each of the genomic datasets listed above using BLAST with an *e*-value < 1e−5. The mapped genes, along with the genes in each of the investigated genomes that belong to the same orthologous group as the *AtEXPs*, were identified as candidate orthologs. Candidate genes were further filtered by only retaining those that had both conserved domains: DPBB-1 and Expansin_C, characteristic for expansin proteins. This resulted in 538 expansin genes extracted from all previously mentioned genomes, except from *Chlamydomonas reinhardtii*. As for the latter species, the candidates did not contain an Expansin_C domain, and they were not retained. Filtered candidate proteins were aligned using MAFFT v7.453, and positions with more than 95% gaps were removed using Phyutility v.2.7.1 \(-clean 0.05\). A maximum likelihood \(ML\) tree of selected genes was constructed using IQTree with 1000 bootstrap replicates and an MFP model. Gene subfamilies were delineated, based on the position of outgroups and previously identified expansins of *A. thaliana*.](http://</p>
</div>
<div data-bbox=)

Coexpression network analysis of EXP genes in *C. chingii*, *C. richardii*, and *A. thaliana*

The expression level of each gene model was estimated for each of the different tissues (see gene expression matrix). The gene coexpression network was built as follows: (1) non-expressed genes with an average FPKM < 0.1 and constitutively expressed genes with a coefficient of variation (CV) < 0.5 were removed, while the expression level of the remaining genes was log-transformed with log₂; (2) the correlation between an expansin gene and any other gene was evaluated using Spearman correlation; and (3) only pairwise correlations > |0.95| and with a *p*-value < 1e−5 were retained to construct a coexpression network.

To generate a list of the most reliable coexpression interactions, we calculated the mutual rank (MR) of pairwise correlations:

$$MR(ab) = \sqrt{(\text{Rank}(a \rightarrow b) \times \text{Rank}(b \rightarrow a))}$$

Only coexpression interactions identified by the Spearman correlation that were also in the top 1% of the interactions identified by the MR were considered reliable.

To perform a comparative analysis of gene coexpression networks across species, we used *C. richardii* as a species closely related to *C. chingii* and *A. thaliana* as a representative of seed plants (outgroup). A total of 12 RNA-seq *C. richardii* samples (4 trophophylls,

4 sporophylls, and 4 roots) and 21 RNA-seq *A. thaliana* samples (8 mature leaves, 8 young leaves, and 5 roots) were downloaded from previous studies (Additional file 2: Table S5). For each sample, the filtered and trimmed reads were mapped to their respective reference genomes using Hisat2. StringTie was used to estimate the expression levels in the different tissues. The EXP gene coexpression networks in *C. richardii* and *A. thaliana* were constructed in the same way as the one of *C. chingii* (see above).

Comparative analysis of gene expression behavior between orthologs

To compare the expression divergence of orthologous genes between *C. chingii* and related fern and lycophyte species, we downloaded, in addition to the RNA of *C. richardii* (see above), also the RNA-seq samples of leaf and root in *S. cucullata*, *I. taiwanensis*, and *S. moellendorffii* from the NCBI SRA database (Additional file 2: Table S5). For each species, the clean reads of the different tissue samples were mapped to their respective reference genome using Hisat2 [72], and gene expression matrices were constructed. For each *C. chingii* gene, we identified its closest relative in other species, i.e., the ortholog that is most similar to the *C. chingii* gene using the following criteria: (1) orthologous genes should belong to the same orthologous group as the considered *C. chingii* genes (identified by the Orthofinder) and (2) should be the best hit of BLASTP with e-value < 1e-5 in bidirectional comparisons of two species. We used these pairs of “highly similar orthologs” to perform the comparative expression analysis.

To cope with the difference in the number and type of tissue samples that were available for each of the compared species, we used the iterative algorithm from Tirosh et al. [47] to assess the expression conservation of the identified orthologous gene pairs. For each species, a gene expression matrix was built that lists, in the rows, the genes in the n orthologous gene pairs ranked in the same order and, in the columns, the profiled conditions in each of the species ($E_{g,x}^{\text{SpeciesA}}, E_{g,y}^{\text{SpeciesB}}$, where $g = 1..n$). Each of these expression matrices is converted to a corresponding correlation matrix ($R_{g,g}^{\text{SpeciesA}}, R_{g,g}^{\text{SpeciesB}}$, where $g = 1..n$) by calculating the Pearson correlation coefficient (PCC) between the gene expression profiles ($n \times n$ matrix). The initial estimation of expression conservation (EC_0) is obtained by comparing the equivalent rows of the two correlation matrices (i.e., comparing the degree to which a pair of orthologs shares similar correlations with other orthologs pairs, i.e., a correlation of the vector in each matrix that contains the correlation coefficient of the gene i with any other gene g in the matrix).

$$EC_0(i) = \text{PPC}(R_{i,g}^{\text{SpeciesA}}, R_{i,g}^{\text{SpeciesB}})$$

However, when comparing a pair of orthologs, one would like to focus on their correlations with other orthologous pairs of which the expression has been conserved as this allows compensating for the differences in the profiled conditions. The iterative algorithm of expression conservation (EC) therefore calculates a weighted correlation, where the weight assigned for correlation with each gene is provided by the EC of that gene from the previous iteration. Genes with negative weights are excluded from the calculation

$$EC_k(i) = \text{PPCw}(R_{i,g'}^{\text{SpeciesA}}, R_{i,g'}^{\text{SpeciesB}})$$

where $\text{PPCw}(x,y) = \frac{\sum w_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum w_i(x_i - \bar{x})^2 \sum w_i(y_i - \bar{y})^2}}$, $w_i = EC_{k-1}(i)$, $g' = \{l \in g | EC_{k-1}(l) > 0\}$

The criteria for the end of the iteration are as follows:

$$\sum_{i \in g} [EC_k(i) - EC_{k-1}(i)]^2 < 0.1$$

Expansin content measurements

The amount of expansin was measured in trophophyll-S1, sporophyll-S1, trophophyll-S2, sporophyll-S2, and roots using five repeats for each tissue type. Approximately 0.5 g fresh weight tissues were cleaned, blotted dry, and snapfrozen in liquid nitrogen. Tissue samples were ground with 2-mL ice-cold extraction buffer with 50 mM Tris, 0.1 mM EDTA, and 15 mM MgCl₂ at PH 8 plus 10% glycerol. The homogenate of tissue samples was centrifuged at 4 °C for 15 min, and the supernatant was obtained and stored at -80 °C. The expansin content in the supernatant was measured using the Plant EXPANSIN Elisa Kit (Zhen Ke Biological Technology, Shanghai) following the operating instructions. .

Yeast two-hybrid (Y2H)

Y2H assays were performed using the Matchmaker Gold Two-Hybrid system (Clontech). The full-length CDS fragments of *CcFL01489*, *CcFL03547*, *CcFL06843*, *CcFL09745*, and *CcFL13362* were amplified from *C. chingii* cDNA. These fragments were then fused as BamHI/EcoRI fragments into pGKBT7 (BD) and pGADT7 (AD) using the Basic Seamless Cloning and Assembly Kit (TransGen, Beijing, China) to generate *CcFL01489*-BD (bait vector), *CcFL03547*-AD (prey vector), *CcFL06843*-AD (prey vector), *CcFL09745*-AD (prey vector), and *CcFL13362*-AD (prey vector), respectively. The bait vector and prey vector were co-transformed into

the Y2H Gold strain performed as previously described [73], with the empty vector as a negative control. Positive transformants were selected on SD/-Trp/-Leu dropout medium and further screened on SD/-Leu/-Trp/-His/-Ade + X- α -gal (40 $\mu\text{g mL}^{-1}$) dropout medium. The cultures were then incubated at 30 °C for 3 days. The primers used in the Y2H assay are listed in Additional file 2: Table S8.

Abbreviations

BUSCO	Benchmarking universal single-copy orthologs
DEG	Differentially expressed gene
EC	Expression conservation
EXP	Expansin
FLNCs	Full-length non-chimeric reads
FPKM	Fragments per kilobase per million mapped reads
GO	Gene Ontology
ICE	Iterative clustering for error correction
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOG	Eukaryotic Orthologous Groups
MEME	Multiple expectation maximization for motif elicitation
ML	Maximum likelihood
RT-PCR	Reverse transcription-polymerase chain reaction
qRT-PCR	Real-time quantitative reverse transcription-polymerase chain reaction

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-023-01743-7>.

Additional file 1. Fig. S1. Density distribution of raw reads and full-length non-chimeric reads obtained by PacBio Iso-seq. **Fig. S2.** RT-PCR validation of nine high-confidence gene models. **Fig. S3.** The UpSet plot summarizes the presence of genes in five databases. **Fig. S4.** The pie diagrams showing the number of transcription factors in different families. **Fig. S5.** GO enrichment analysis of differentially expressed genes in the comparisons of sporophyll-S1 vs trophophyll-S2, sporophyll-S2 vs trophophyll-S1, and sporophyll-S2 vs trophophyll-S2. **Fig. S6.** Heatmap showing the differential and non-differential expression of the expansin genes in different pairwise comparisons as obtained from transcriptome analysis. **Fig. S7.** Phylogenetic tree of expansins in *C. chingii* and *A. thaliana*, built with Maximum likelihood. **Fig. S8.** Analysis of conserved domains of the *CcEXP* proteins. **Fig. S9.** Phylogenetic profiling of four groups of expansin genes in 19 species. **Fig. S10.** Bar plots showing the differentially expressed expansin genes in pteridophyte-specific groups. **Fig. S11.** GO enrichment analysis of genes that were identified to be significantly coexpressed with *CcEXP* genes. **Fig. S12.** Heatmap showing the member number of transcription factors in different families that are coexpressed with at least one *CcEXP* gene in the four groups. **Fig. S13.** Yeast two-hybrid assay showing how *CcFL01489* interacts with three transcription factor (*CcFL09745*, *CcFL03547*, and *CcFL06843*). **Fig. S14.** Gene coexpression networks of *CrEXP* (A) and *AtEXP* (B) genes of the different groups. **Fig. S15.** GO enrichment analysis of genes that were identified to be significantly coexpressed with *AtEXP* (A) and *CcEXP* (B) genes in group IV.

Additional file 2. Table S1. EXP genes in *C. chingii*. **Table S2.** Topology of EXP gene tree of 19 species. **Table S3.** Correlation coefficient of FPKM and relative expression in *CcEXP* genes. **Table S4.** Summer of lncRNA targeted genes. **Table S5.** Summary of the RNA-Seq samples in this study. **Table S6.** Summary of the RT-PCR primer for validating the splicing events. **Table S7.** Summary of the qRT-PCR primer for expansin genes. **Table S8.** Summer of Yeast two-hybrid primer for *CcFL01489* genes and its coexpressed transcription factors.

Acknowledgements

We thank all individuals who participated in this study.

Authors' contributions

KM and JC2 conceived the idea. BL, SSZ, and XY collected the tissues. YZ and JC1 analyzed the data and finished the confirmatory experiment. YVdP, YZ, and MK wrote the manuscript. YVdP, XY, MK, and JC2 revised the manuscript.

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB31000000) and the National Natural Science Foundation of China (Nos. 31570220, 31870208, 32300198). The work was supported by grants of the Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO) [3G046318, G.0371.06] and UGent BOF.

Availability of data and materials

The PacBio full-length sequencing dataset generated for this work is accessible through NCBI Sequence Read Archive (SRA) under accession number PRJNA924672 [74]. The Illumina RNA-seq dataset is accessible through NCBI SRA accession number PRJNA924581 [75].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 30 July 2023 Accepted: 23 October 2023

Published online: 06 November 2023

References

- Zotz G, Wilhelm K, Becker A. Heteroblasty—a review. *Bot Rev.* 2011;77(2):109–51. <https://doi.org/10.1007/s12229-010-9062-8>.
- Li G, Hu S, Hou H, Kimura S. Heterophylly: phenotypic plasticity of leaf shape in aquatic and amphibious plants. *Plants.* 2019;8(10):420. <https://doi.org/10.3390/plants8100420>.
- Forster MA, Bonser SP. Heteroblastic development and shade-avoidance in response to blue and red light signals in *Acacia implexa*. *Photochem Photobiol.* 2009;85(6):1375–83. <https://doi.org/10.1111/j.1751-1097.2009.00605.x>.
- Nakayama H, Nakayama N, Seiki S, Kojima M, Sakakibara H, Sinha N, et al. Regulation of the KNOX-GA gene module induces heterophyllic alteration in North American lake cress. *Plant Cell.* 2014;26(12):4733–48. <https://doi.org/10.1105/tpc.114.130229>.
- Kim J, Joo Y, Kyung J, Jeon M, Park JY, Lee HG, et al. A molecular basis behind heterophylly in an amphibious plant, *Ranunculus trichophyllus*. *PLoS Genet.* 2018;14(2):e1007208. <https://doi.org/10.1371/journal.pgen.1007208>.
- Li G, Yang J, Chen Y, Zhao X, Chen Y, Kimura S, et al. SHOOT MERISTEMLESS participates in the heterophylly of *Hydrophila difformis* (Acanthaceae). *Plant Physiol.* 2022;190(3):1777–91. <https://doi.org/10.1093/plphys/kiac382>.
- Farquharson KL. Examining the molecular basis of heterophylly in North American lake cress. *Plant Cell.* 2014;26(12):4567. <https://doi.org/10.1105/tpc.114.135434>.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, et al. The timescale of early land plant evolution. *Proc Natl Acad Sci U S A.* 2018;115(10):E2274–83. <https://doi.org/10.1073/pnas.1719588115>.
- Schuettpelz E, Schneider H, Smith AR, Hovenkamp P, Prado J, Rouhan G, et al. A community-derived classification for extant lycophytes and ferns. *J Syst Evol.* 2016;54(6):563–603. <https://doi.org/10.1111/jse.12229>.
- Moran RC. Sterile-fertile leaf dimorphism and evolution of soral types in *Polybotrya* (Dryopteridaceae). *Syst Bot.* 1987;12(4):617–28. <https://doi.org/10.2307/2418896>.

11. Vasco A, Moran RC, Ambrose BA. The evolution, morphology, and development of fern leaves. *Front Plant Sci.* 2013;4:345. <https://doi.org/10.3389/fpls.2013.00345>.
12. Nagalingum NS, Schneider H, Pryer KM. Molecular phylogenetic relationships and morphological evolution in the heterosporous fern genus *Marsilea*. *Syst Bot.* 2007;32(1):16–25. <https://doi.org/10.1600/036364407780360256>.
13. Sun J, Li GS. Leaf dorsoventrality candidate gene *CpARF4* has conserved expression pattern but divergent tasiR-ARF regulation in the water fern *Ceratopteris pteridoides*. *Am J Bot.* 2020;107(11):1470–80. <https://doi.org/10.1002/ajb2.1570>.
14. Zheng B, Xing K, Zhang J, Liu H, Ali K, Li W, et al. Evolutionary analysis and functional identification of ancient brassinosteroid receptors in *Ceratopteris richardii*. *Int J Mol Sci.* 2022;23(12):6795. <https://doi.org/10.3390/ijms23126795>.
15. Marchant DB, Chen G, Cai S, Chen F, Schafran P, Jenkins J, et al. Dynamic genome evolution in a model fern. *Nat Plants.* 2021;8(9):1038–51. <https://doi.org/10.1038/s41477-022-01226-7>.
16. Dong YH, Wang QF, Gituru RW. Effect of habitat modification on the distribution of the endangered aquatic fern *Ceratopteris pteridoides* (Parkeriaceae) in China. *Am Fern J.* 2012;102(2):136–46. <https://doi.org/10.1640/0002-8444-102.2.136>.
17. Alonso DP, Alvarez MVN, Amorim JA, de Sá ILR, de Carvalho DP, Ribeiro KAN, et al. *Mansonia* spp. population genetics based on mitochondrial whole-genome sequencing alongside the Madeira River near Porto Velho, Rondonia, Brazil. *Brazil Infect Genet Evol.* 2022;103:105341.
18. Yu JH, Zhang R, Liu QL, Wang FG, Yu XL, Dai XL, et al. *Ceratopteris chunii* and *Ceratopteris chingii* (Pteridaceae), two new diploid species from China, based on morphological, cytological, and molecular data. *Plant Diversity.* 2022;44(3):300–7. <https://doi.org/10.1016/j.pld.2021.10.002>.
19. Zumajo-Cardona C, Pabon-Mora N, Ambrose BA. The evolution of *euA-PETALA2* genes in vascular plants: from plesiomorphic roles in sporangia to acquired functions in ovules and fruits. *Mol Biol Evol.* 2021;38(6):2319–36. <https://doi.org/10.1093/molbev/msab027>.
20. Stern WL. Comparative morphology of vascular plants. *AIBS Bulletin.* 1960;10(5):39–39. <https://doi.org/10.2307/1292723>.
21. Conway SJ, Di Stilio VS. An ontogenetic framework for functional studies in the model fern *Ceratopteris richardii*. *Dev Biol.* 2020;457(1):20–9. <https://doi.org/10.1016/j.ydbio.2019.08.017>.
22. Bashline L, Lei L, Li S, Gu Y. Cell wall, cytoskeleton, and cell expansion in higher plants. *Mol Plant.* 2014;7(4):586–600. <https://doi.org/10.1093/mp/ssu018>.
23. Goh HH, Sloan J, Dorca-Fornell C, Fleming A. Inducible repression of multiple expansin genes leads to growth suppression during leaf development. *Plant Physiol.* 2012;159(4):1759–70. <https://doi.org/10.1104/pp.112.200881>.
24. Pien S, Wyrzykowska J, McQueen-Mason S, Smart C, Fleming A. Local expression of expansin induces the entire process of leaf development and modifies leaf shape. *Proc Natl Acad Sci U S A.* 2001;98(20):11812–7. <https://doi.org/10.1073/pnas.191380498>.
25. Ostría-Gallardo E, Ranjan A, Chitwood DH, Kumar R, Townsley BT, Ichihashi Y, et al. Transcriptomic analysis suggests a key role for *SQUAMOSA PROMOTER BINDING PROTEIN LIKE*, *NAC* and *YUCCA* genes in the heteroblastic development of the temperate rainforest tree *Gevuina avellana* (Proteaceae). *New Phytol.* 2016;210(2):694–708. <https://doi.org/10.1111/nph.13776>.
26. Silva PO, Batista DS, Cavalcanti JHF, Koehler AD, Vieira LM, Fernandes AM, et al. Leaf heteroblasty in *Passiflora edulis* as revealed by metabolic profiling and expression analyses of the microRNAs miR156 and miR172. *Ann Bot.* 2019;123(7):1191–203. <https://doi.org/10.1093/aob/mcz025>.
27. Kuo LY, Li FW. A roadmap for fern genome sequencing. *Am Fern J.* 2019;109(3):212–23. <https://doi.org/10.1640/0002-8444-109.3.212>.
28. Huang X, Wang W, Gong T, Wickell D, Kuo LY, Zhang X, et al. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants.* 2022;8(5):500–12. <https://doi.org/10.1038/s41477-022-01146-6>.
29. Li FW, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux PM, et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants.* 2018;4(7):460–72. <https://doi.org/10.1038/s41477-018-0188-8>.
30. Fang Y, Qin X, Liao Q, Du R, Luo X, Zhou Q, et al. The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nat Plants.* 2022;8(9):1024–37. <https://doi.org/10.1038/s41477-022-01222-x>.
31. Zhong Y, Liu Y, Wu W, Chen J, Sun C, Liu H, et al. Genomic insights into genetic diploidization in the homosporous fern *Adiantum nelumboides*. *Genome Biol Evol.* 2022;14(8):evac127. <https://doi.org/10.1093/gbe/evac127>.
32. Rahmatpour N, Kuo LY, Kang J, Herman E, Lei L, Li M, et al. Analyses of *Marsilea vestita* genome and transcriptomes do not support widespread intron retention during spermatogenesis. *New Phytol.* 2023;237(5):1490–4. <https://doi.org/10.1111/nph.18652>.
33. Szovenyi P, Gunadi A, Li FW. Charting the genomic landscape of seed-free plants. *Nat Plants.* 2021;7(5):554–65. <https://doi.org/10.1038/s41477-021-00888-z>.
34. Li H, Teng K, Yue Y, Teng W, Zhang H, Wen H, et al. Seed germination mechanism of *Carex rigescens* under variable temperature determined using integrated single-molecule long-read and Illumina sequence analysis. *Front Plant Sci.* 2022;13: 818458. <https://doi.org/10.3389/fpls.2022.818458>.
35. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. *Trends Genet.* 2018;34(9):666–81.
36. Xu Q, Zhu J, Zhao S, Hou Y, Li F, Tai Y, et al. Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front Plant Sci.* 2017;8:1205. <https://doi.org/10.3389/fpls.2017.01205>.
37. Li J, Sun M, Li H, Ling Z, Wang D, Zhang J, Shi L. Full-length transcriptome-referenced analysis reveals crucial roles of hormone and wounding during induction of aerial bulbils in lily. *BMC Plant Biol.* 2022;22(1):415. <https://doi.org/10.3389/fpls.2017.01205>.
38. Marowa P, Ding AM, Kong YZ. Expansins: roles in plant growth and potential applications in crop improvement. *Plant Cell Rep.* 2016;35(5):949–65. <https://doi.org/10.1007/s00299-016-1948-4>.
39. Kim JH, Cho HT, Kende H. α -Expansins in the semiaquatic ferns *Marsilea quadrifolia* and *Regnellidium diphyllum*: evolutionary aspects and physiological role in rachis elongation. *Planta.* 2000;212(1):85–92. <https://doi.org/10.1007/s004250000367>.
40. de Vries J, Fischer AM, Roettger M, Rommel S, Schlupepmann H, Bräutigam A, et al. Cytokinin-induced promotion of root meristem size in the fern *Azolla* supports a shoot-like origin of euphyllophyte roots. *New Phytol.* 2016;209(2):705–20. <https://doi.org/10.1111/nph.13630>.
41. Liu X, Wu R, Bulley SM, Zhong C, Li D. Kiwifruit MYB51-like and GBF3 transcription factors influence l-ascorbic acid biosynthesis by activating transcription of GDP-L-galactose phosphorylase 3. *New Phytol.* 2022;234(5):1782–800. <https://doi.org/10.1111/nph.18097>.
42. Zhao Y, Liu XJ, Wang MK, Bi QX, Cui YF, Wang LB. Transcriptome and physiological analyses provide insights into the leaf epicuticular wax accumulation mechanism in yellowhorn. *Hortic Res.* 2021;8(1):134. <https://doi.org/10.1038/s41438-021-00564-5>.
43. Yang R, Wang X. Organ evolution in angiosperms driven by correlated divergences of gene sequences and expression patterns. *Plant Cell.* 2013;25(1):71–82. <https://doi.org/10.1105/tpc.112.106716.43>.
44. Sampedro J, Cosgrove DJ. The expansin superfamily. *Genome Biol.* 2005;6(12):242. <https://doi.org/10.1186/gb-2005-6-12-242>.
45. Hepler NK, Bowman A, Carey RE, Cosgrove DJ. Expansin gene loss is a common occurrence during adaptation to an aquatic environment. *Plant J.* 2020;101(3):666–80. <https://doi.org/10.1111/tpj.14572>.
46. Healy TM, Schulte PM. Phenotypic plasticity and divergence in gene expression. *Mol Ecol.* 2015;24(13):3220–2. <https://doi.org/10.1111/mec.13246>.
47. Tirosch I, Barkai N. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol.* 2007;8(4):R50. <https://doi.org/10.1186/gb-2007-8-4-r50>.
48. Marowa P, Ding AM, Xu ZC, Kong YZ. Overexpression of *NtEXPA11* modulates plant growth and development and enhances stress tolerance in tobacco. *Plant Physiol Biochem.* 2020;151:477–85. <https://doi.org/10.1016/j.plaphy.2020.03.033>.
49. Ramakrishna P, Duarte PR, Rance GA, Schubert M, Vordermaier V, Vu LD, et al. EXPANSIN A1-mediated radial swelling of pericycle cells positions anticlinal cell divisions during lateral root initiation. *Proc Natl Acad Sci U S A.* 2019;116(17):8597–602. <https://doi.org/10.1073/pnas.1820882116>.

50. Calderini DF, Castillo FM, Arenas-M A, Molero G, Reynolds MP, Craze M, et al. Overcoming the trade-off between grain weight and number in wheat by the ectopic expression of expansin in developing seeds leads to increased yield potential. *New Phytol.* 2021;230(2):629–40. <https://doi.org/10.1111/nph.17048>.
51. Cosgrove DJ. Loosening of plant cell walls by expansins. *Nature.* 2000;407(6802):321–6. <https://doi.org/10.1038/35030000>.
52. Carey RE, Cosgrove DJ. Portrait of the expansin superfamily in *Phycomitrella patens*: comparisons with angiosperm expansins. *Ann Bot.* 2007;99(6):1131–41. <https://doi.org/10.1093/aob/mcm044>.
53. Carey RE, Hepler NK, Cosgrove DJ. *Selaginella moellendorffii* has a reduced and highly conserved expansin superfamily with genes more closely related to angiosperms than to bryophytes. *BMC Plant Biol.* 2013;13:4. <https://doi.org/10.1186/1471-2229-13-4>.
54. Chen YC, Li Z, Zhao YX, Gao M, Wang JY, Liu KW, et al. The *Litsea* genome and the evolution of the laurel family. *Nature Com.* 2020;11(1):1675.
55. Villaverde T, Pokorny L, Olsson S, Rincon-Barrado M, Johnson MG, Gardner EM, et al. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* 2018;220(2):636–50. <https://doi.org/10.1111/nph.15312>.
56. Liu WY, Yu CP, Chang CK, Chen HJ, Li MY, Chen YH, et al. Regulators of early maize leaf development inferred from transcriptomes of laser capture microdissection (LCM)-isolated embryonic leaf cells. *Proc Natl Acad Sci U S A.* 2022;119(35): e2208795119. <https://doi.org/10.1073/pnas.2208795119>.
57. Zhang XX, Li XX, Zhao R, Zhou Y, Jiao YN. Evolutionary strategies drive a balance of the interacting gene products for the CBL and CIPK gene families. *New Phytol.* 2020;226(5):1506–16. <https://doi.org/10.1111/nph.16445>.
58. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
59. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science.* 2007;318(5848):245–50. <https://doi.org/10.1126/science.1143609>.
60. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, et al. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat Commun.* 2014;5:3978. <https://doi.org/10.1038/ncomms4978>.
61. Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, et al. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell.* 2019;179(5):1057–1067.e1014. <https://doi.org/10.1016/j.cell.2019.10.019>.
62. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* Genome. *Cell.* 2017;171(2):287–304.e215. <https://doi.org/10.1016/j.cell.2017.09.030>.
63. Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, et al. The *Phycomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 2017;93(3):515–33. <https://doi.org/10.1111/tpj.13801>.
64. Weston DJ, Turetsky MR, Johnson MG, Granath G, Lindo Z, Belyea LR, et al. The Sphagnum Project: enabling ecological and evolutionary insights through a genus-level sequencing project. *New Phytol.* 2018;217(1):16–25. <https://doi.org/10.1111/nph.14860>.
65. Wickell D, Kuo LY, Yang HP, Dhabalia Ashok A, Irisarri I, Dardas A, et al. Underwater CAM photosynthesis elucidated by Isoetes genome. *Nat Commun.* 2021;12(1):6348. <https://doi.org/10.1038/s41467-021-26644-7>.
66. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science.* 2011;332(6032):960–3. <https://doi.org/10.1126/science.1203810>.
67. Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Palmer JD, et al. The *Amborella* genome and the evolution of flowering plants. *Science.* 2013;342(6165):1241089. <https://doi.org/10.1126/science.1241089>.
68. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, et al. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature.* 2010;463(7282):763–8. <https://doi.org/10.1038/nature08747>.
69. Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, et al. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat Plants.* 2020;6(3):215–22. <https://doi.org/10.1038/s41477-020-0594-6>.
70. Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, et al. The *Cycas* genome and the early evolution of seed plants. *Nat Plants.* 2022;8(4):389–401. <https://doi.org/10.1038/s41477-022-01129-7>.
71. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The *Norway spruce* genome sequence and conifer genome evolution. *Nature.* 2013;497(7451):579–84. <https://doi.org/10.1038/nature12211>.
72. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnol.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
73. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.* 2007;2(1):31–4. <https://doi.org/10.1038/nprot.2007.13>.
74. The full-length transcriptome sequencing of a mixed RNA pool in *Ceratopteris chingii*. NCBI BioProject accession: PRJNA924672. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA924672>. (2023)
75. Five *Ceratopteris chingii* Illumina RNA-seq samples. NCBI BioProject accession: PRJNA924581. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA924581>. (2023)

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

