Supplementary information

Title: Resolving the stasis-dynamism paradox: Genome evolution in tree ferns

Authors: Zuoying Wei[1,2,3,4,5], Hengchi Chen[4,5], Chao Feng[1,3,8], Zengqiang Xia[1,2,3], Yves Van de Peer[4,5,6,7*], Ming Kang[1,3,8*], Jing Wang[1,3,8*]

Supplementary Note

Supplementary Note 1. Genome sequencing and assembly

Fresh leaves of *S. brunoniana* and *G. denticulata* were collected from South China National Botanical Garden and Nankunshan National Forest Park (Guangdong, China), respectively. To further extend the contiguity of the *S. lepifera* genome assembly, we collected a fresh sample of *S. lepifera* from Shanghai Chenshan Botanical Garden (Shanghai, China) for Hi-C sequencing. High-quality genomic DNA was extracted using a cetyltrimethylammonium bromide (CTAB) protocol (Winnepenninckx et al. 1993). For Illumina sequencing, short read sequencing libraries with ~350 bp insert size were constructed and sequenced on the Illumina NovaSeq 6000 platform to generate 150 bp read pairs. For PacBio SMRT sequencing, DNA libraries were constructed based on the PacBio library preparation protocol and further sequenced on the PacBio Sequel II platform. HiFi data was then obtained through the CCS software (https://github.com/PacificBiosciences/ccs). Hi-C libraries were constructed following a published protocol (Lieberman-Aiden et al. 2009). Generally, nuclear DNA was cross-linked, digested using the restriction enzyme MboI, biotinylated and sheared. The prepared DNA fragment libraries were sequenced on Illumina PE150. For transcriptome sequencing, RNA was extracted from fronds, roots and stems with the RNAprep Pure Plant Kit (TIANGEN). RNA-seq libraries were prepared and then subjected to PE150 sequencing on the Illumina NovaSeq 6000 platform.

To obtain prior knowledge of the newly assembled genomes, genome size, heterozygosity and repeat content of *S. brunoniana* and *G. denticulata* were estimated

by *k*-mer frequency analysis (*k*-mer=17 as the default) with GCE (Liu et al. 2013) based on Illumina short reads. For genome assembly, the PacBio HiFi reads of *S. brunoniana* and *G. denticulata* were initially assembled into contigs using hifiasm (v0.16.1-r37526) (Cheng et al. 2021) with default parameters. The Hi-C sequencing reads were aligned to the final contigs by Juicer (v1.6) (Durand et al. 2016) with default parameters. Contigs were mapped to pseudochromosomes using the 3D-DNA (v201008) pipeline (Dudchenko 2017). Juicebox (v1.11.08) (Durand et al. 2016) was further used to visually correct the assembly errors, including the orientation, order and misjoins of contigs. The completeness of the genome assembly was finally assessed using BUSCO (v5.4.3) (Simão FA et al. 2015) with the Eukaryota_odb10 database.

For Estimation of insertion time of the LTR-RTs, all the LTRs sequences identified with complete 5′-LTR and 3′-LTR were used. The insertion time for intact LTR-RTs was calculated using LTR_retriever (v.2.9.5) (Ou and Jiang 2018) with the following formula: $T = K/2r$ (divergence between LTRs/substitution per site per year). The nucleotide substitution rate r was set to 1.77×10^{-9} substitutions per site per year (Huang et al. 2022).

Supplementary Figures

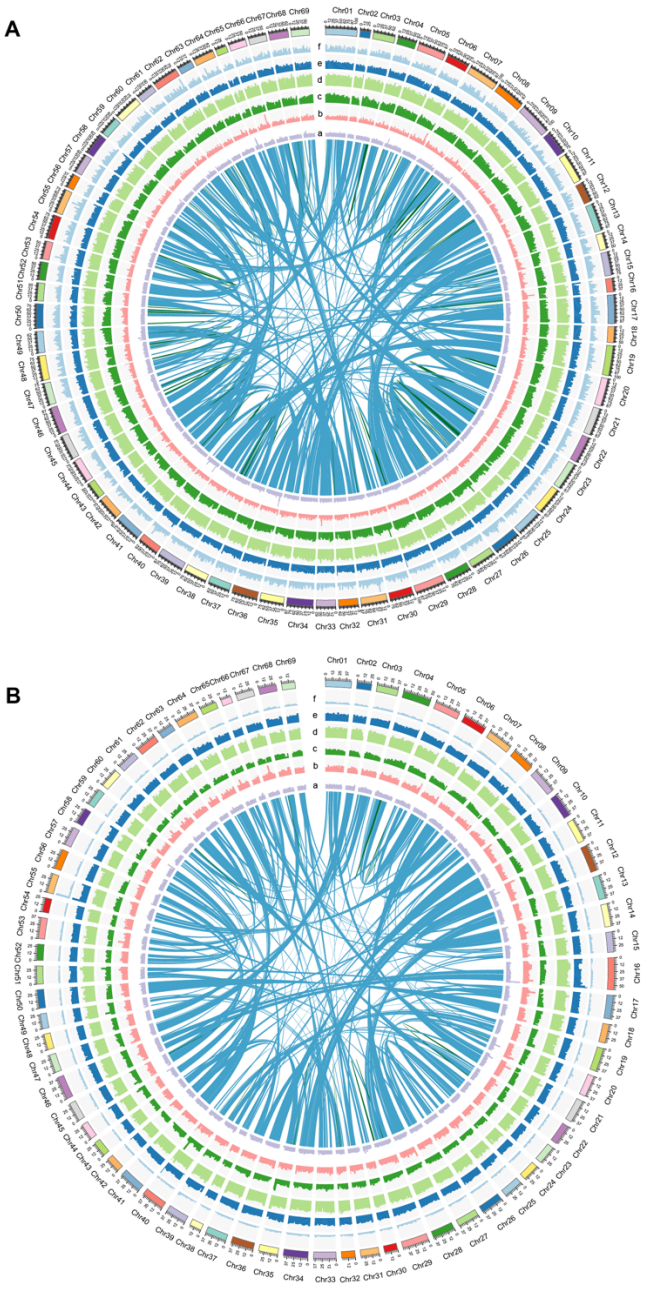

Fig. S1 Genome features of *S. brunoniana* and *S. lepifera*. (A, B) The genomic landscape for the *S. lepifera* and *S. brunoniana*, respectivelly. Tracks 'a-f' represent tandem repeat density, LTR/*Gypsy* density, LTR/*Copia* density, TE density, GC content, and gene density, respectively. The inner lines represent both inter-chromosomal (blue) and intra-chromosomal (green) synteny.

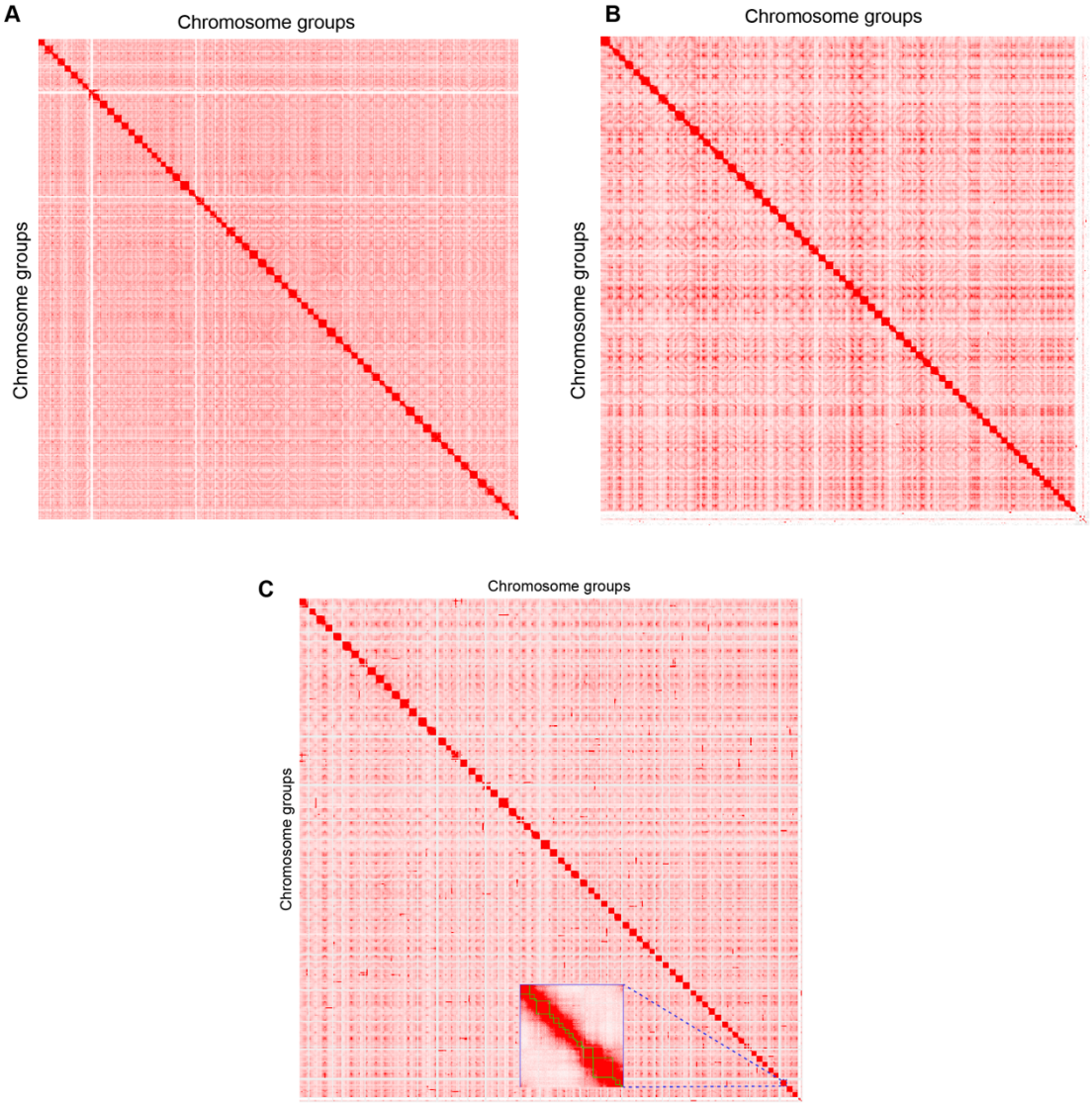

Fig. S2 Hi-C interaction maps of the corrected three assemblies ordered by the *Alsophila spinulosa* reference. (A, B, C) Hi-C interaction map of *S. brunoniana*, *S. lepifera*, and *G. denticulata*, respectivelly. The inset highlights a zoomed-in view of chromosome 66.

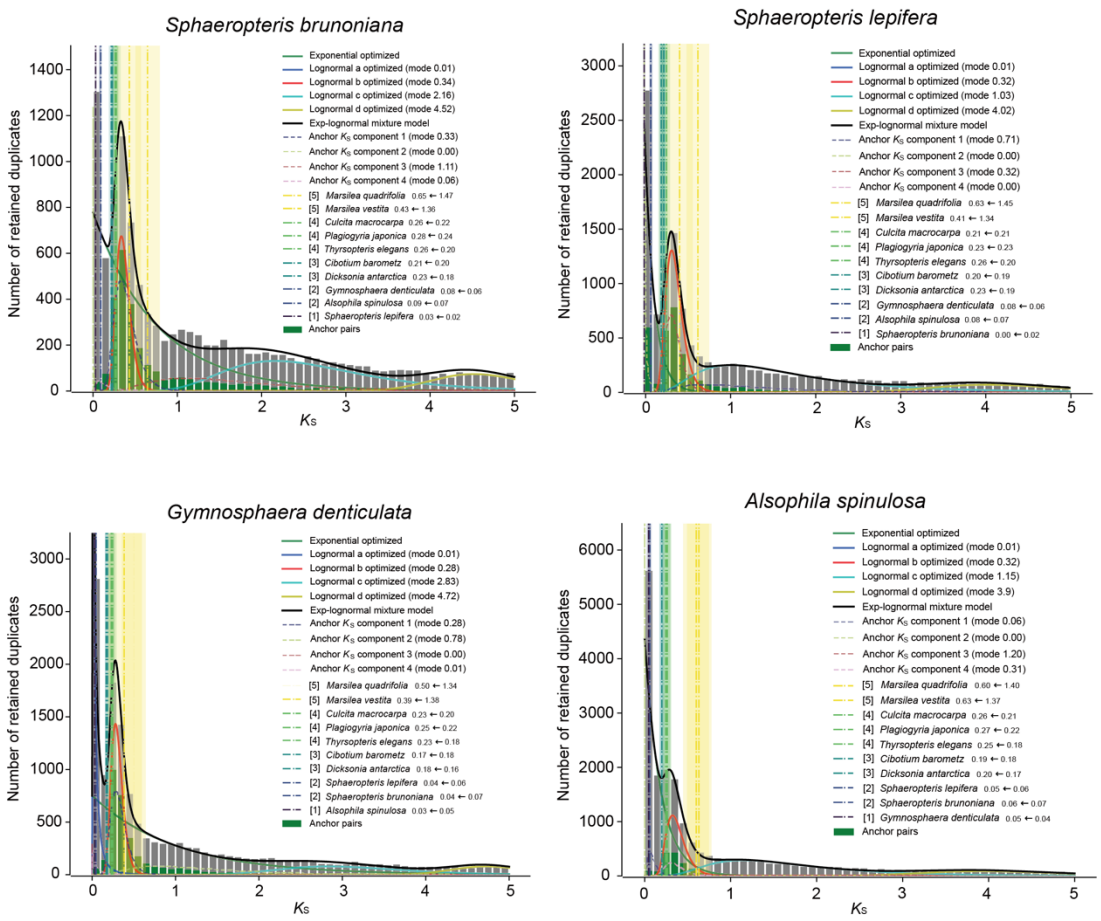

Fig. S3 K_S age distribution with substitution rate corrections for the paranome of four Cyatheaceae species. The solid and dashed lines showing the inferred components of mixture modeling analysis of the whole-paranome and anchor pair K_S distribution, respectively. The corrected divergence times were marked by the dash-dotted lines with the associated standard deviation.

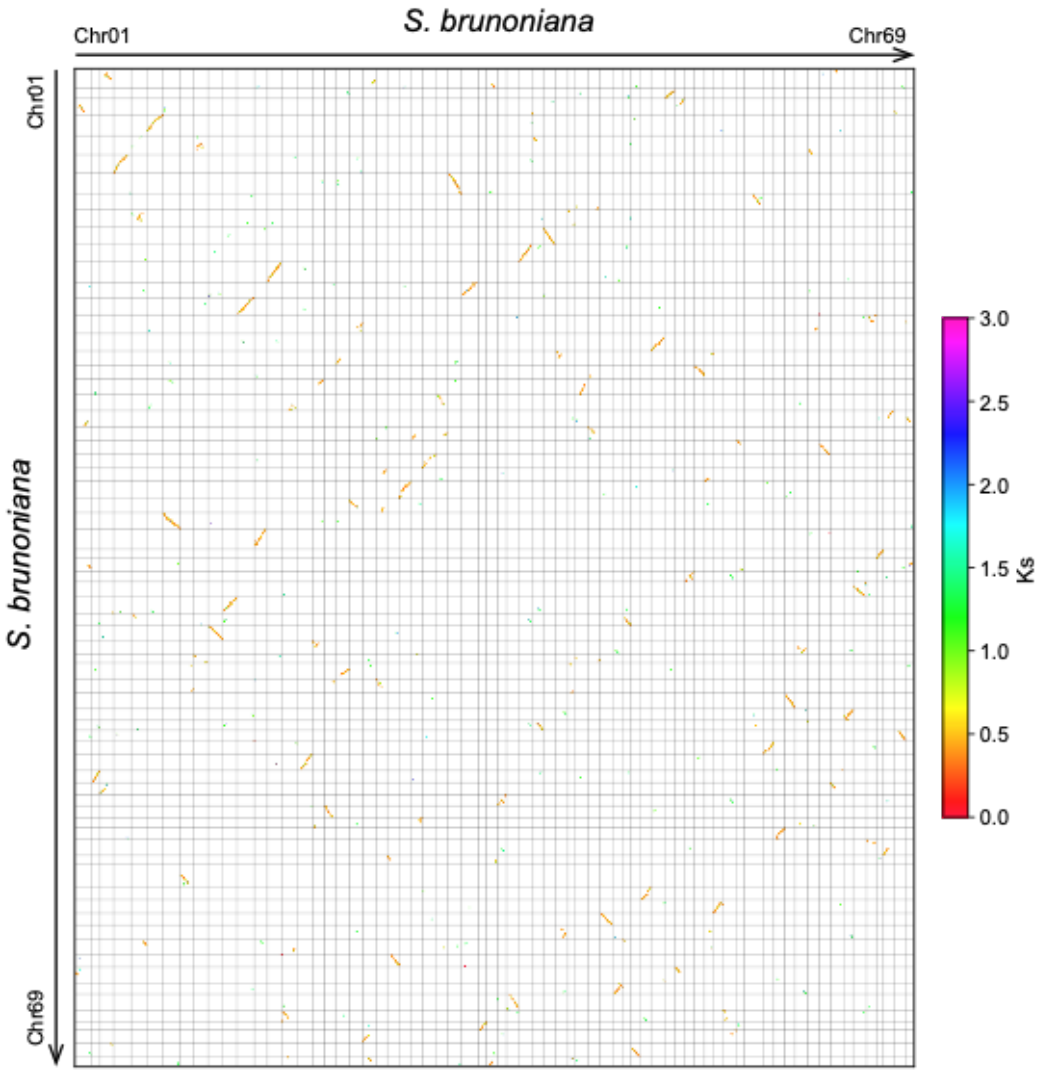

Fig. S4 Intragenomic comparison of *S. brunoniana*. Syntenic dotplot and K_S dotplot displaying the paralogs in the *S. brunoniana* genome.

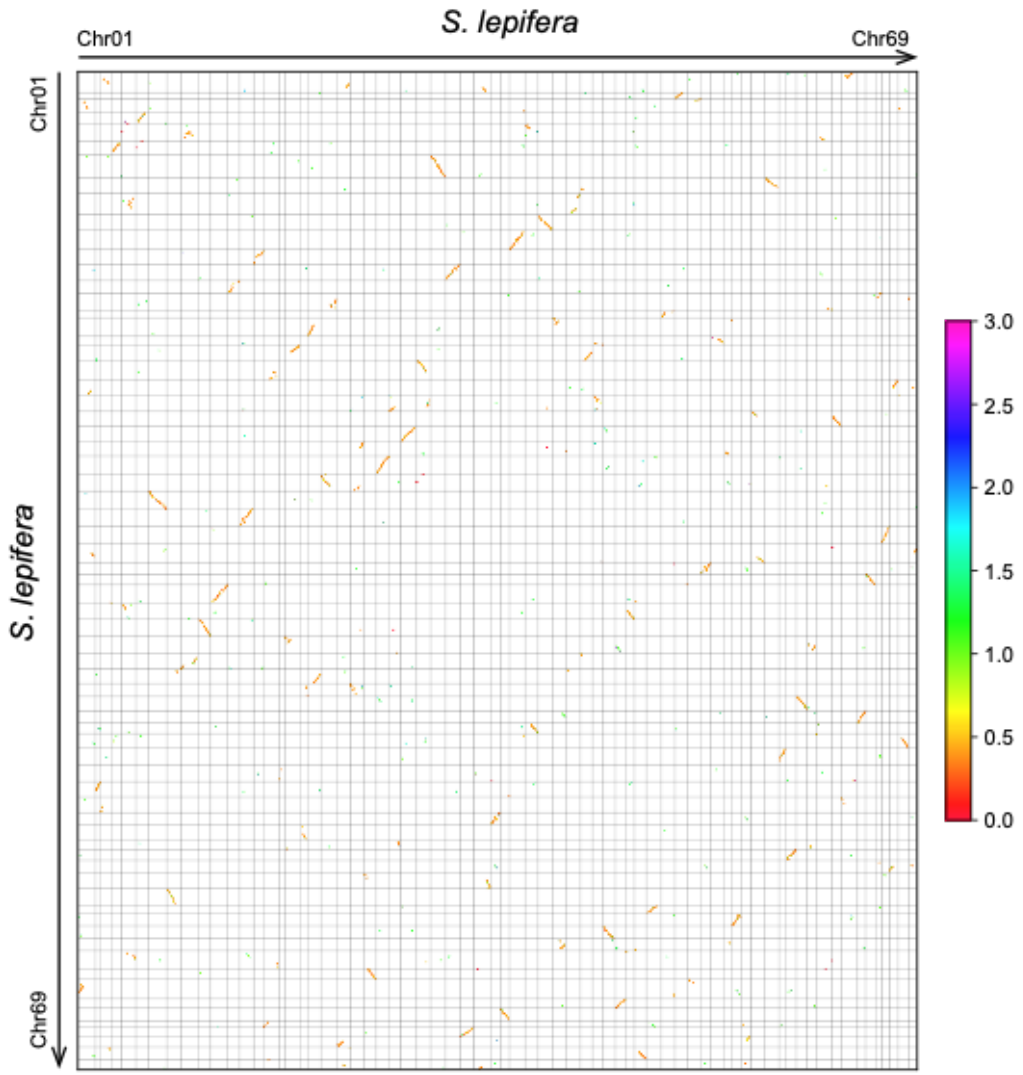

Fig. S5 Intragenomic comparison of *S. lepifera*. Syntenic dotplot and K_S dotplot displaying the paralogs in the *S. lepifera* genome.

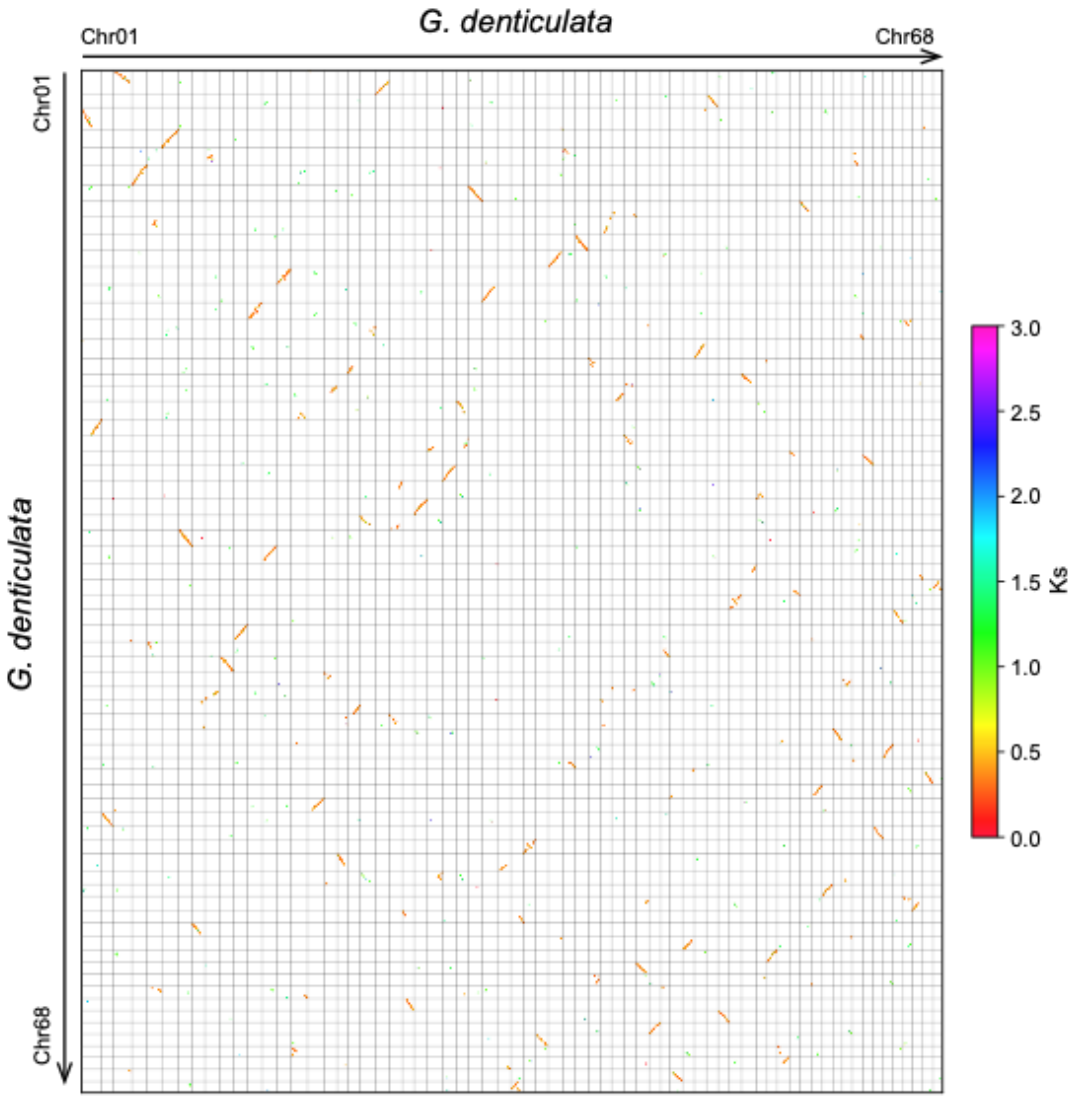

Fig. S6 Intragenomic comparison of *G. denticulata*. Syntenic dotplot and K_S dotplot displaying the paralogs in the *G. denticulata* genome.

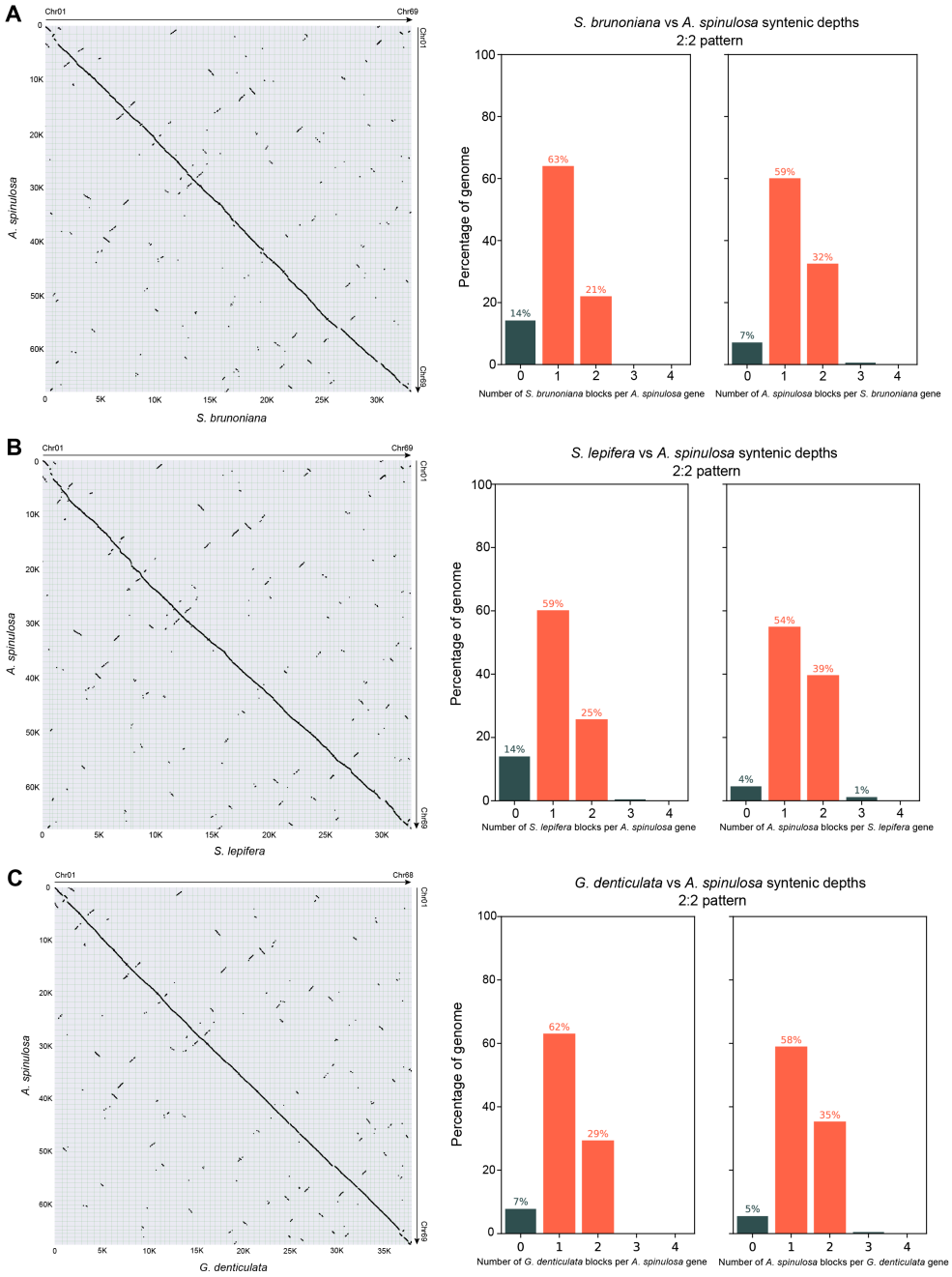

Fig. S7 Intergenomic comparisons. (A) Syntenic dotplot (left) and syntenic depths (right) between the *A. spinulosa* and *S. brunoniana* genomes. **(B)** Syntenic dotplot (left) and syntenic depths (right) between the *A. spinulosa* and *S. lepifera* genomes. **(C)** Syntenic dotplot (left) and syntenic depths (right) between the *A. spinulosa* and *G. denticulata* genomes.

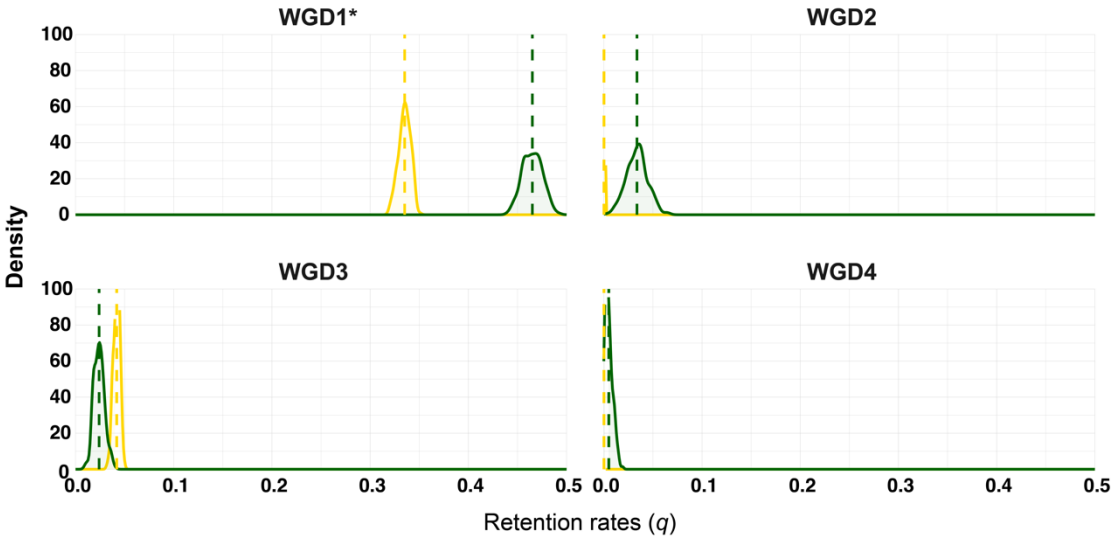

Fig. S8 The posterior distributions of the WGD retention rates (*q*) for the four putative WGDs under the relaxed branch-specific model (green) and the critical branch-specific model (yellow) in WHALE. The dotted lines show the posterior mean of each posterior distribution. Hypothetical WGD models which gained significant support and 'accepted' in this study are marked with asterisks.

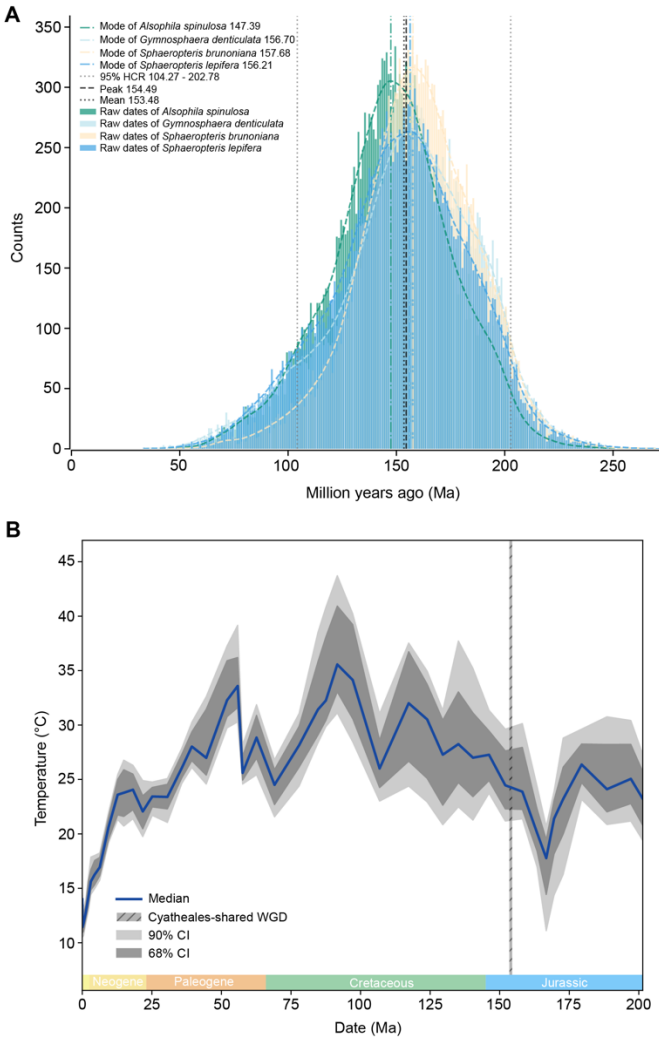

Fig. S9 Absolute dating of the ancient Cyathealean WGD event and dynamic of paleoclimates. (A) Posterior distributions, KDE curves, modes, overall means, peaks and 95% HCR of the date estimation of the Cyathealean WGD event. **(B)** The temporal dynamics of the global mean surface temperature in which the blue solid curve represents the median while the dark and light gray bands represent the 68% and 90% confidence intervals, respectively (Judd et al. 2024). The vertical line represents the inferred date for the Cyathealean WGD event.

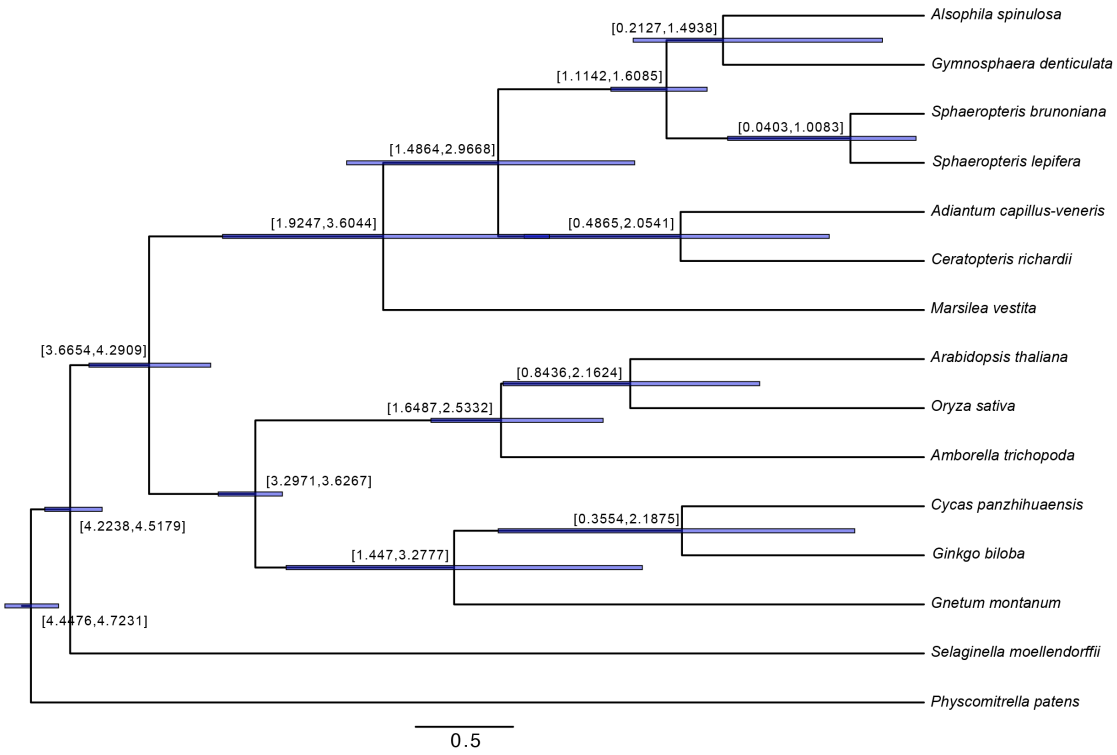

Fig. S10 Time-calibrated phylogeny. The phylogenetic tree shows the topology and divergence times for 15 plant species. Blue bars and numbers at nodes represent 95% credibility interval of the divergence time.

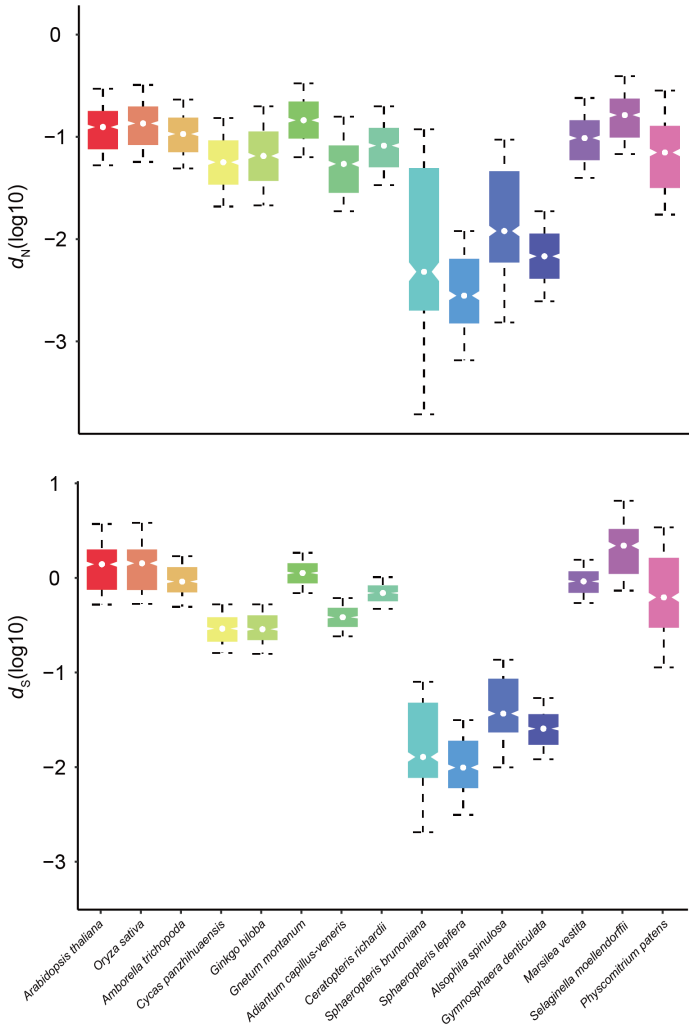

Fig. S11 Comparative analysis of evolutionary rates across 15 species. The box plots of d_S and d_N across the phylogeny of 15 species with the concatenated 583 low-copy nuclear (LCN) ortholog dataset.

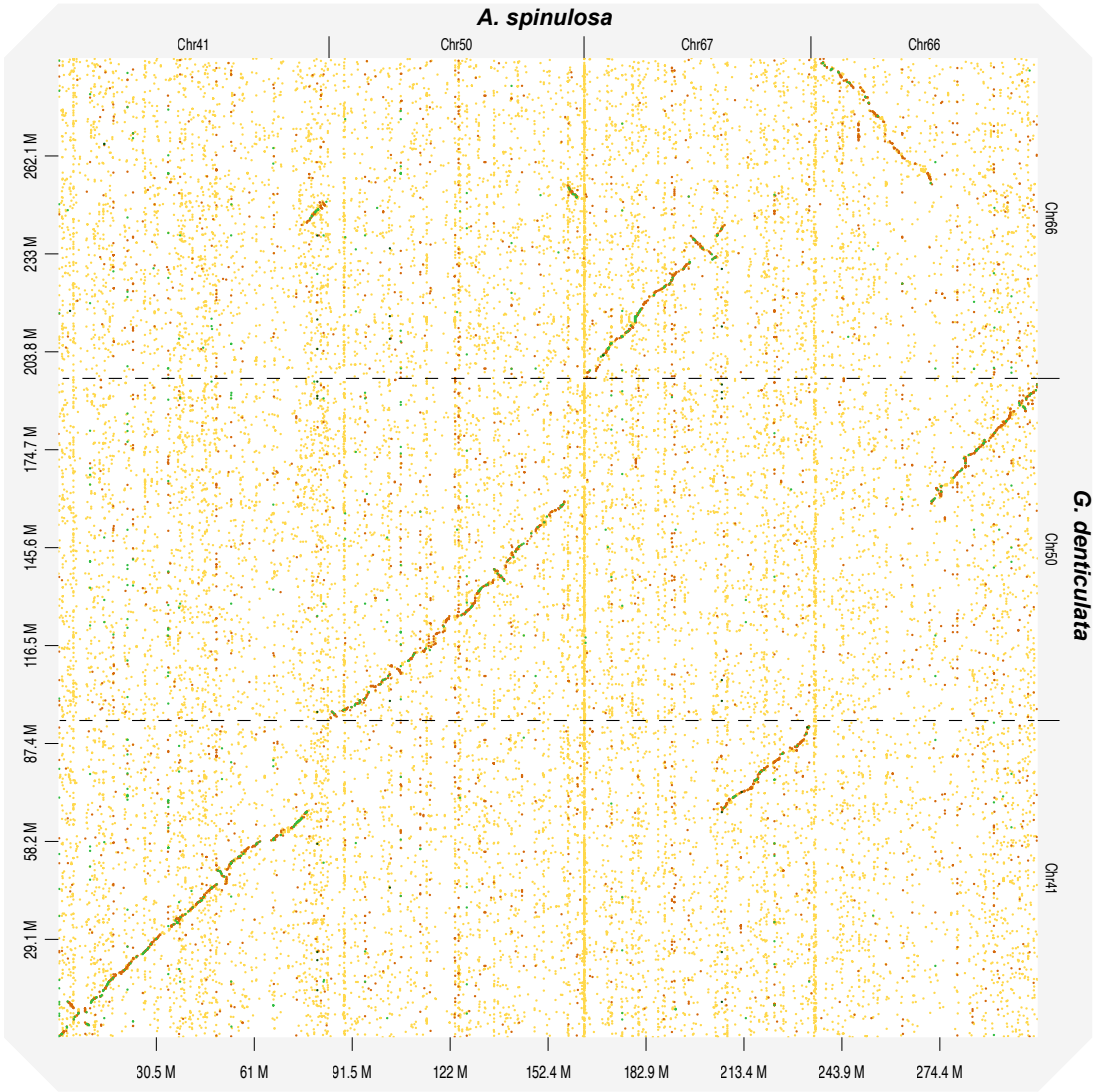

Fig. S12 Synteny map. Synteny map of *G. denticulata* chromosomes 41, 50 and 66 compared with *A. spinulosa* chromosomes 41, 50, 66 and 67.

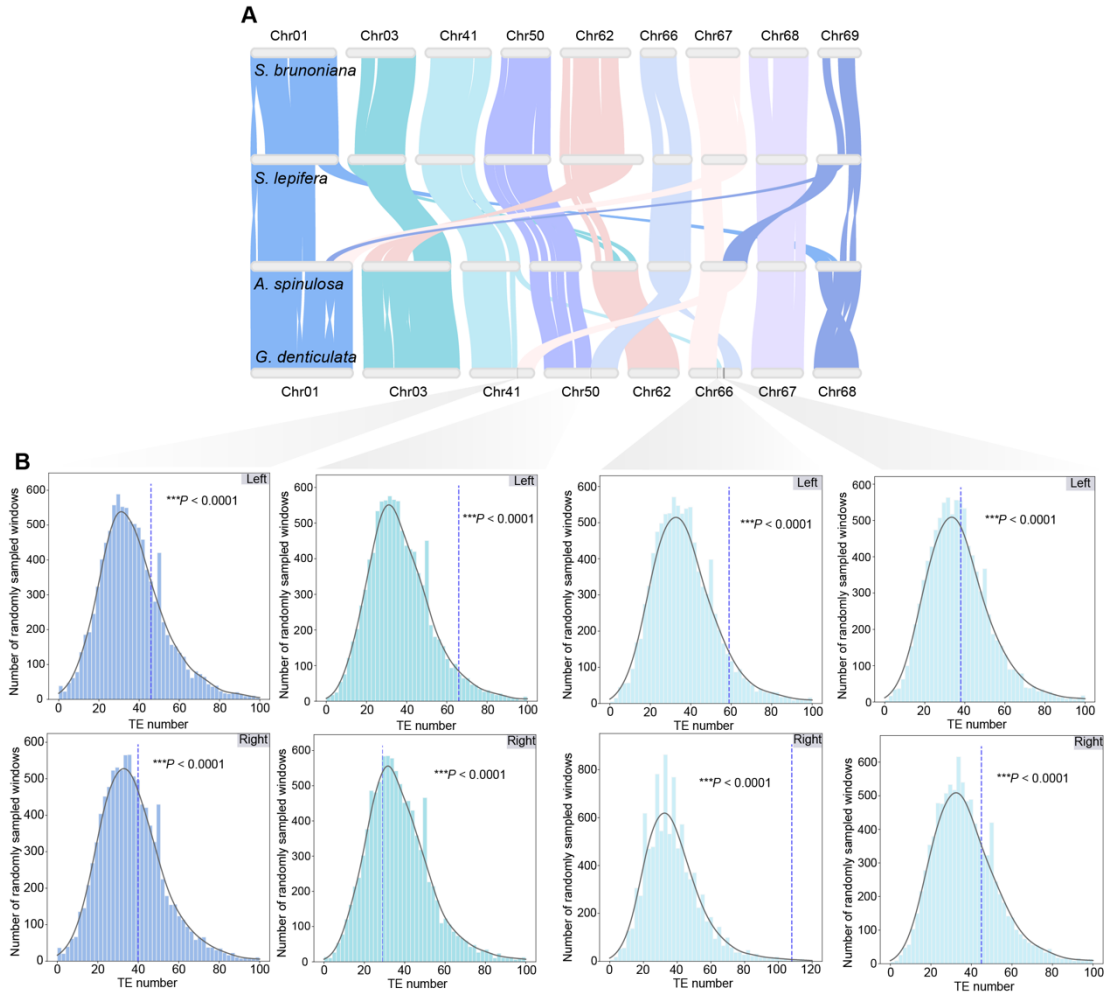

Fig. S13 Several chromosomal rearrangement events across four Cyatheaceae species and the vicinal TE density in all fusion events in *G. denticulata*. **(A)** Synteny of several chromosomes with potential macro-rearrangements among the four Cyatheaceae species. Each horizontal bar represents one chromosome, and chromosome IDs of *S. brunoniana*, *S. lepifera* and *A. spinulosa* were labeled at the top and chromosome IDs of *G. denticulata* were labeled at the bottom. Syntenic blocks of each chromosome are highlighted in color based on the *S. brunoniana*. **(B)** Compared the vicinal TE density of the breakpoints to the background TE density of the chromosome represented by the TE densities of 10,000 randomly sampled chromosomal regions in the same length. Wilcoxon test shows that the vicinal TE density of the breakpoints is significantly higher than the background TE density of the chromosome ($P < 0.0001$) in all fusion events.

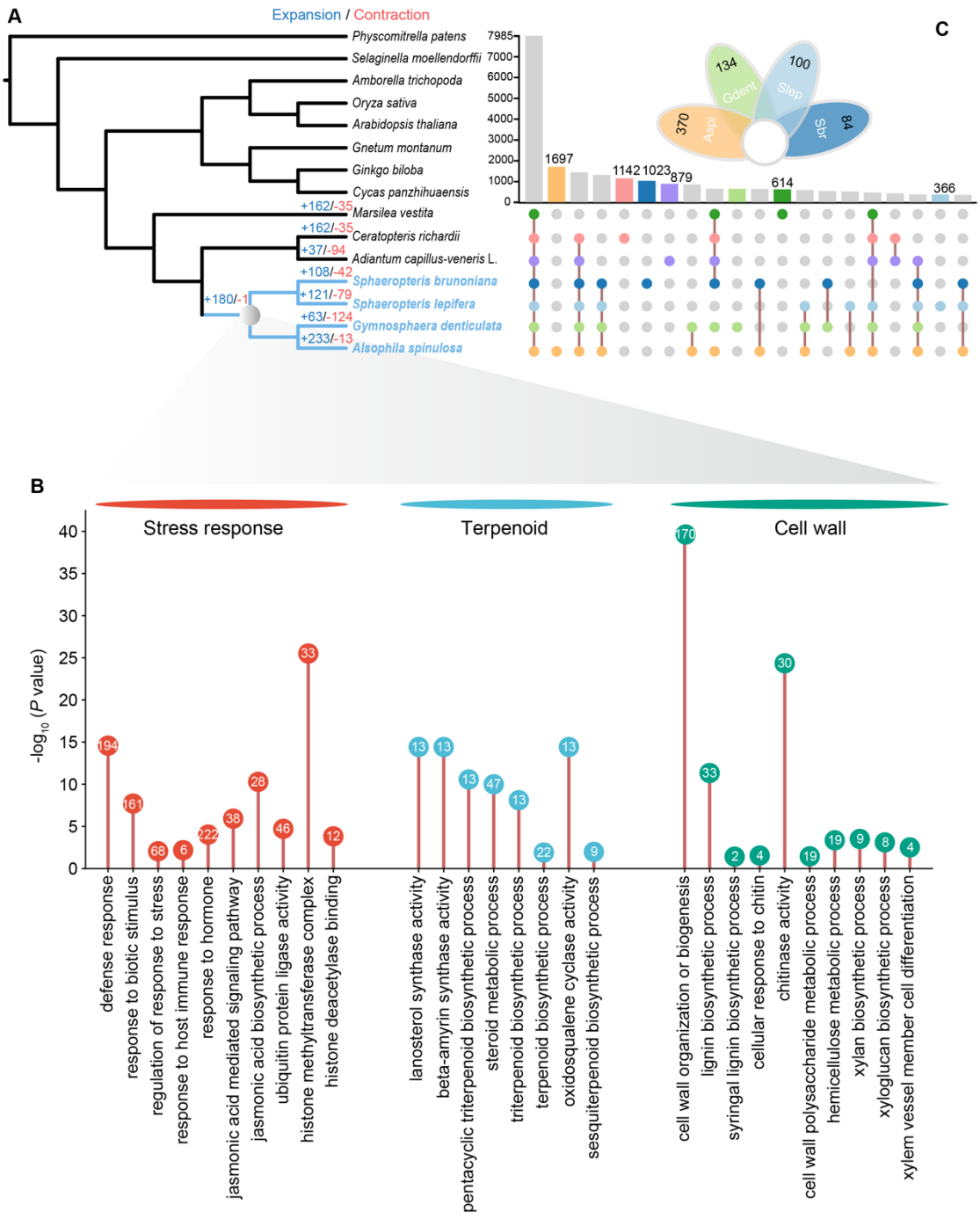

Fig. S14 Phylogenetic relationship and comparative genomics analyses. (A) Phylogenetic tree of 15 plant species and evolution of gene families. Numbers at each branch represent significantly expanded gene families (blue) and significantly contracted gene families (red). **(B)** GO enrichment of significantly expanded gene families at the crown node of the Cyatheaceae species. GO terms in stress response (red), terpenoid (blue), and cell wall (green) categories are indicated in different colors. The numbers in the circles indicate the number of enriched genes. Source data are

provided as a Source Data file. **(C)** Evolutionary analysis of orthologous clusters across the seven ferns. The values at the top of each bar shows the number of specific gene families for each fern species. The petal diagram illustrates the number of species-specific expanded genes, with the numbers labeled within the corresponding petals. Aspi, *A. spinulosa*; Gdent, *G. denticulata*; Sbr, *S. brunoniana*; Slep, *S. lepifera*.

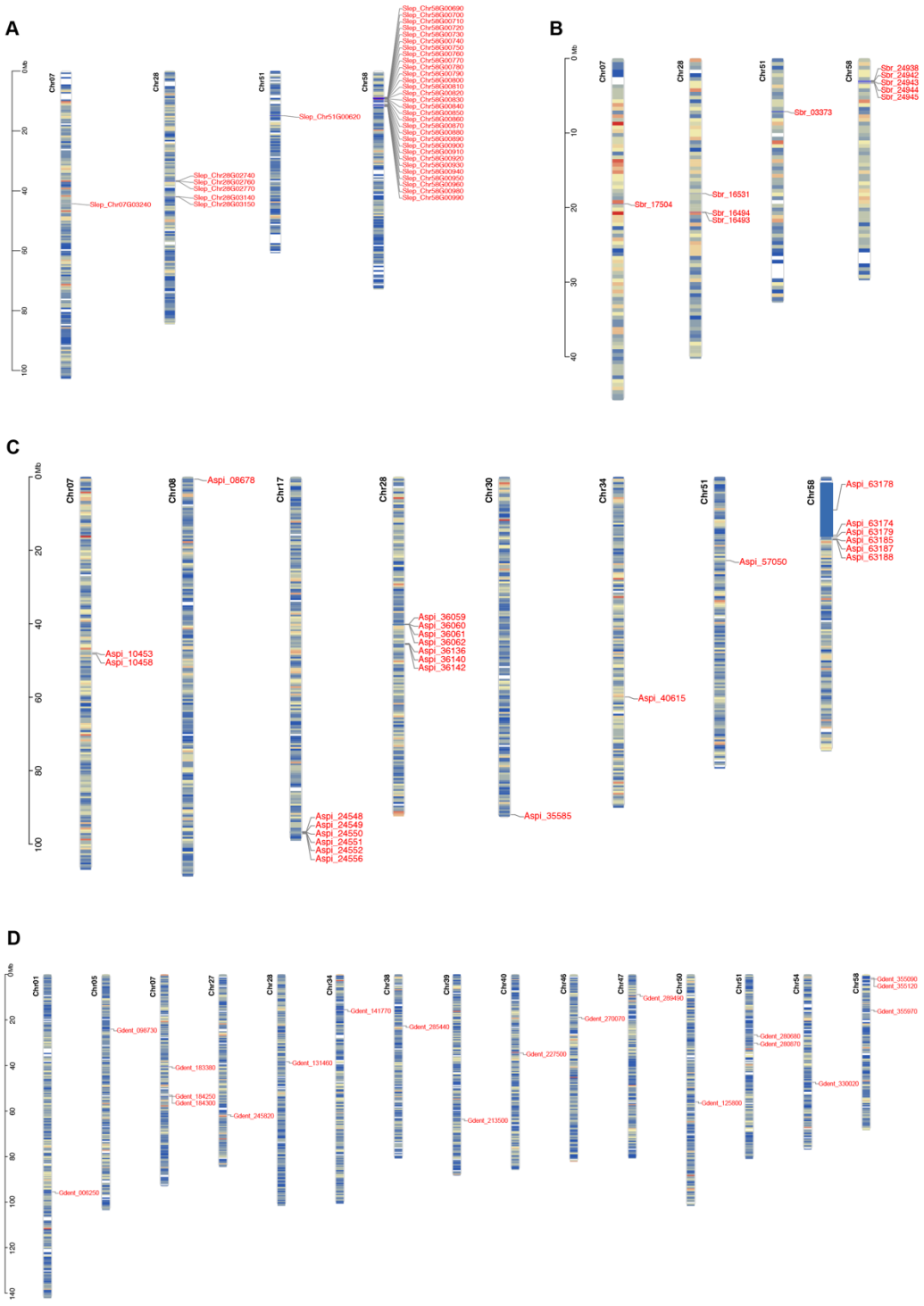

Fig. S15 Distributions of 2,3-oxidosqualene cyclase (OSC) genes in four Cyatheaceae species drawn by TBtools (Chen et al. 2023). (A) OSCs located on the four chromosomes in the *S. lepifera* genome. **(B)** OSCs located on the four chromosomes in the *S. brunoniana* genome. **(C)** OSCs located on the eight chromosomes in the *A. spinulosa* genome. **(D)** OSCs located on the 15 chromosomes in the *G. denticulata* genome.

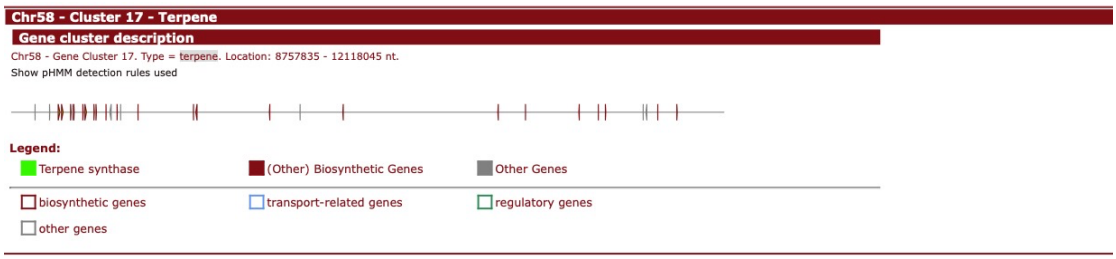

Fig. S16 Visualization of terpene-related biosynthetic gene clusters (BGCs) in *S. lepifera* identified by plantiSMASH (Kautsar et al. 2017). Various core enzymes were annotated, and the related genes were colored according to the enzymatic classes.

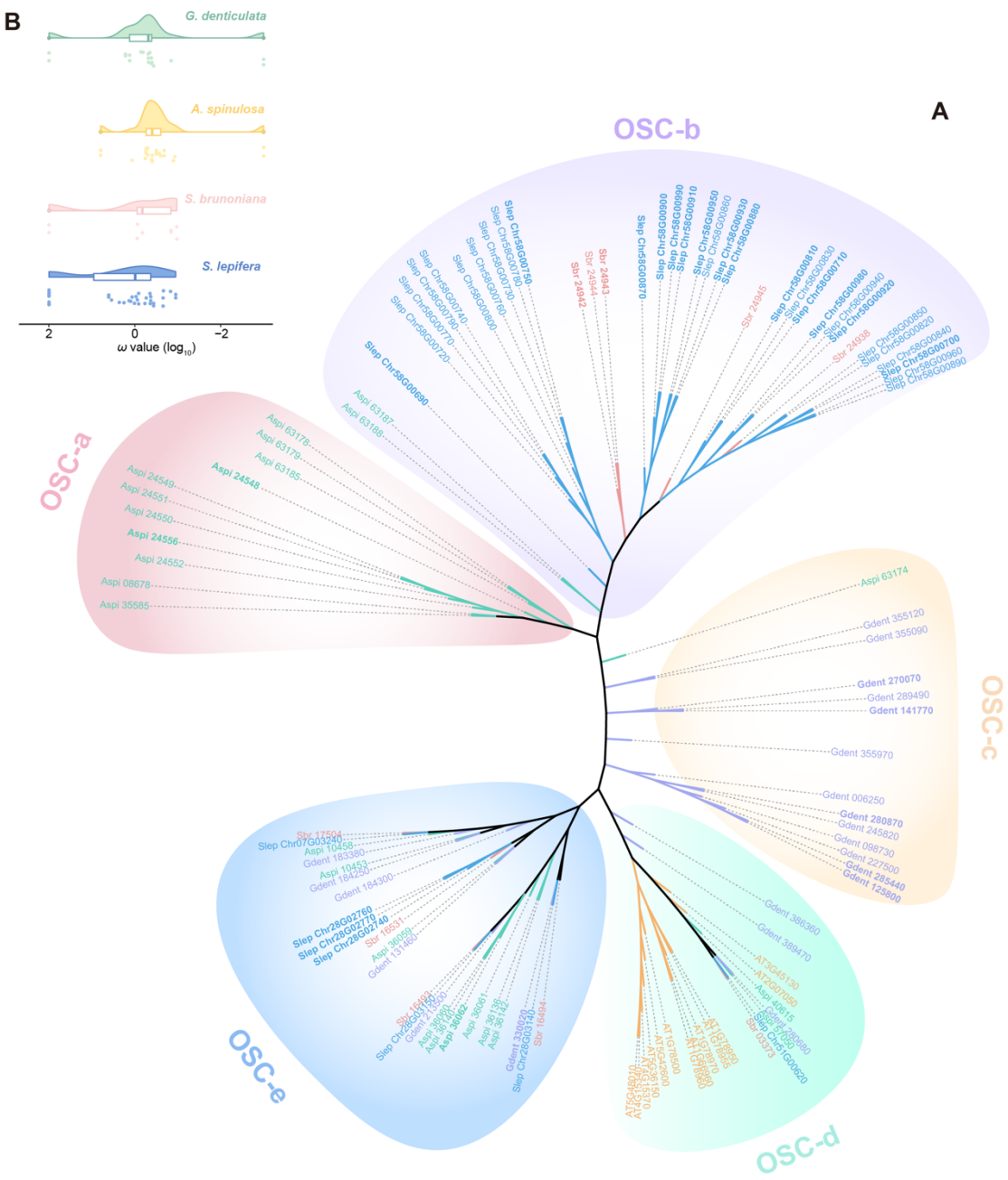

Fig. S17 The phylogenetic inference and selection pressure of OSC enzymes. (A) The phylogenetic inference of OSC enzymes for the five species. Genes from different species are indicated in different colors. Aspi, *A. spinulosa*; Gdent, *G. denticulata*; Sbr, *S. brunoniana*; Slep, *S. lepifera*; Atha, *Ar. thaliana*. **(B)** The ω values of OSC genes in the four Cyatheaceae species. The colored points below the box plot represent the distribution of ω values for each gene pair in each species. The box-plot elements are defined as: center line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range; points, outliers.

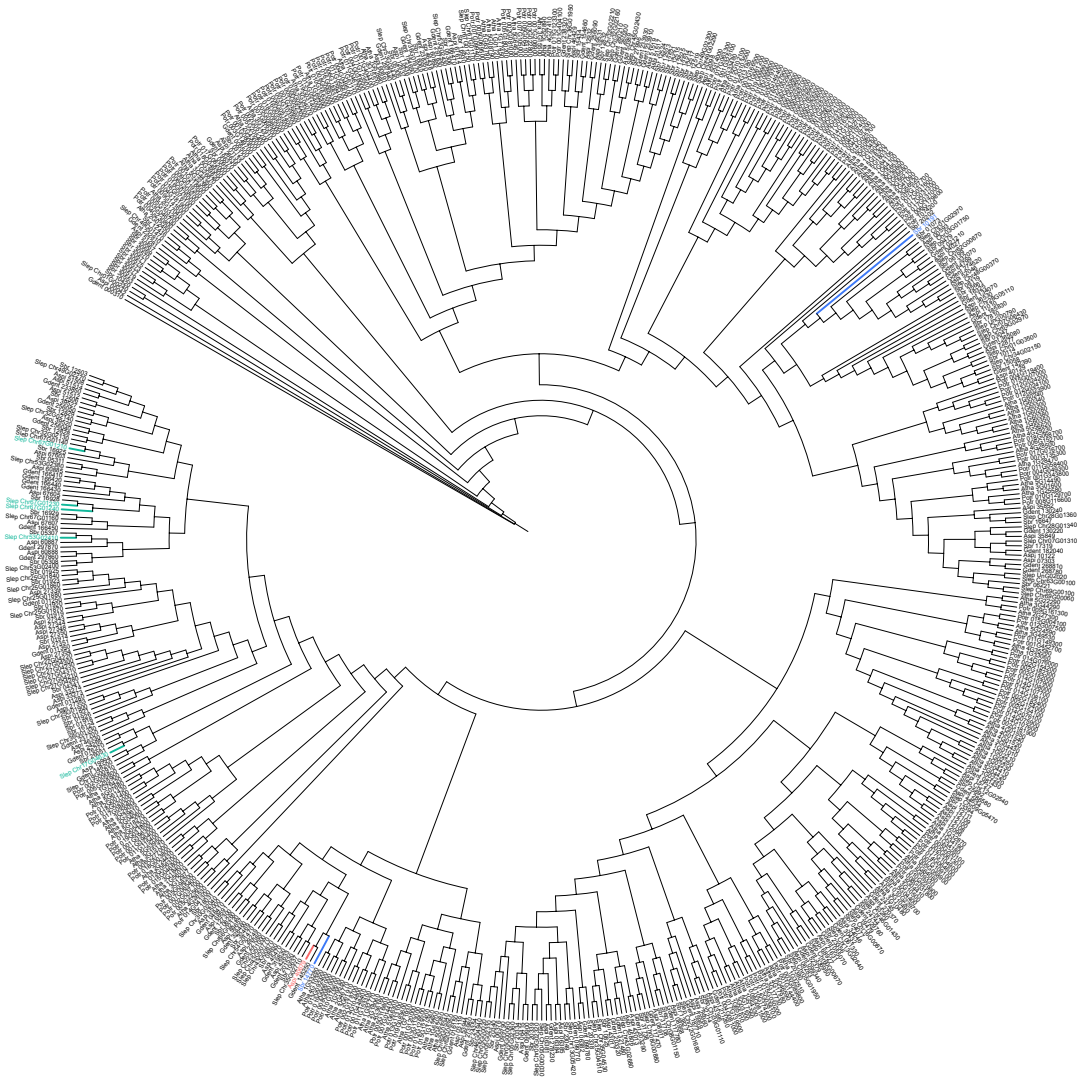

Fig. S18 Phylogenetic inference of NAC for the six species using a maximum likelihood tree. Genes are from are from *Ar. thaliana* (Ath), *P. trichocarpa* (Potr), *A. spinulosa* (Aspi), *S. brunoniana* (Sbr), *S. lepifera* (Slep) and *G. denticulata* (Gdent). Colored branches and labels indicate genes that are preferentially retained following the WGD.

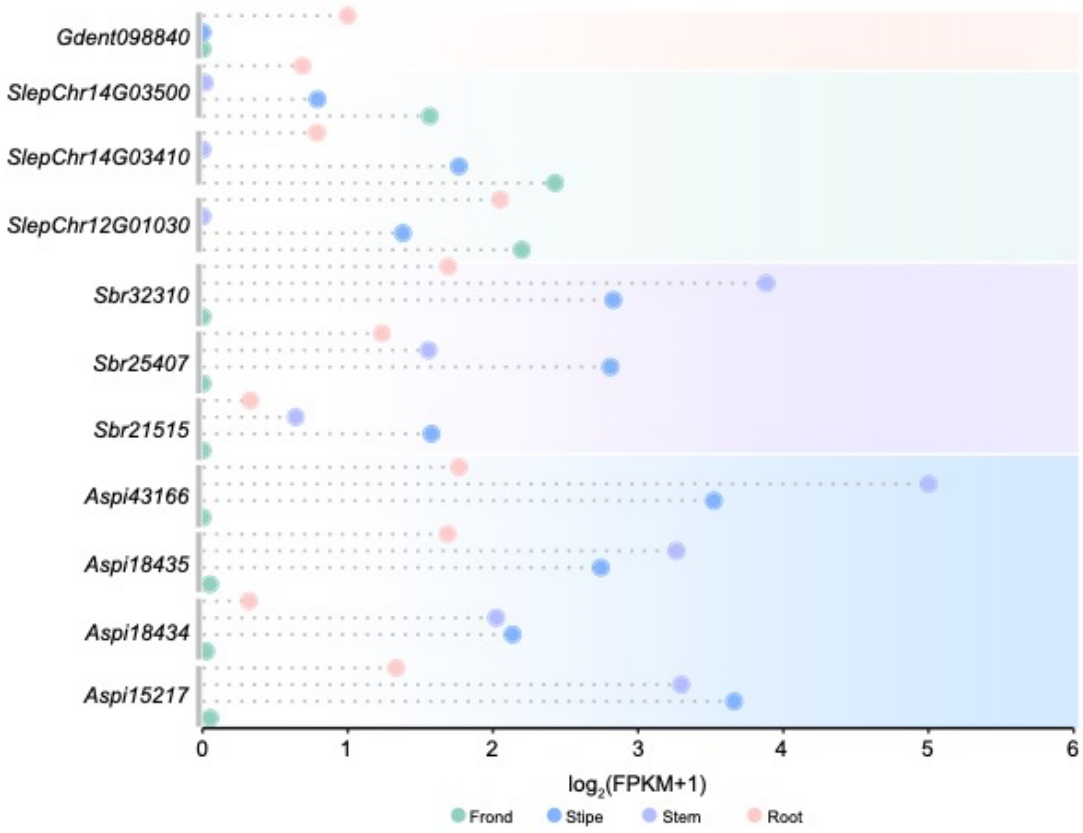

Fig. S19 Expression patterns of *POX*/ *LAC* (peroxidase/laccases) in different tissues in the four species. Only genes considered to be expressed (FPKM > 1) are shown. FPKM, fragments per kilobase per million mapped reads.

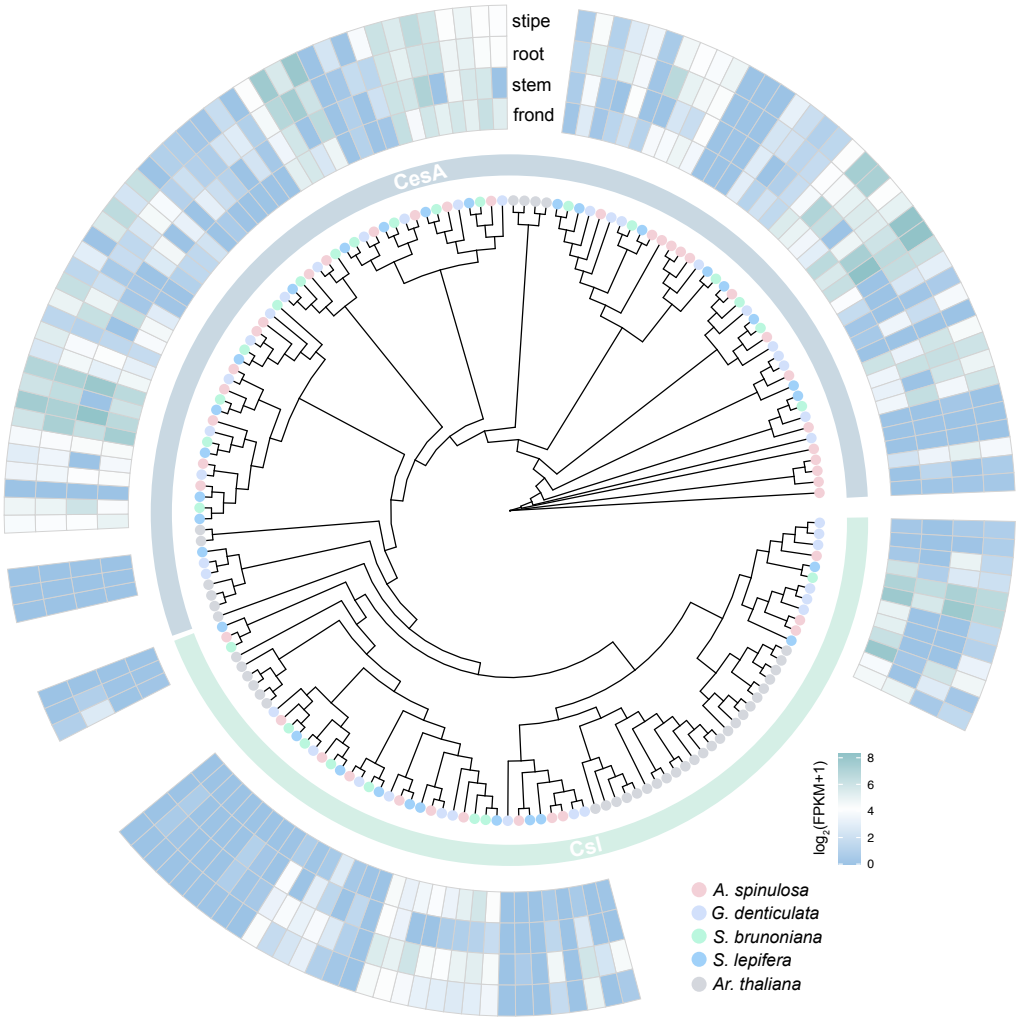

Fig. S20 Expression patterns of *CesA*/*Csl* (peroxidase/laccases) in different tissues in the four Cyatheaceae species. The expression values were scaled by $\log_2$ (FPKM + 1).

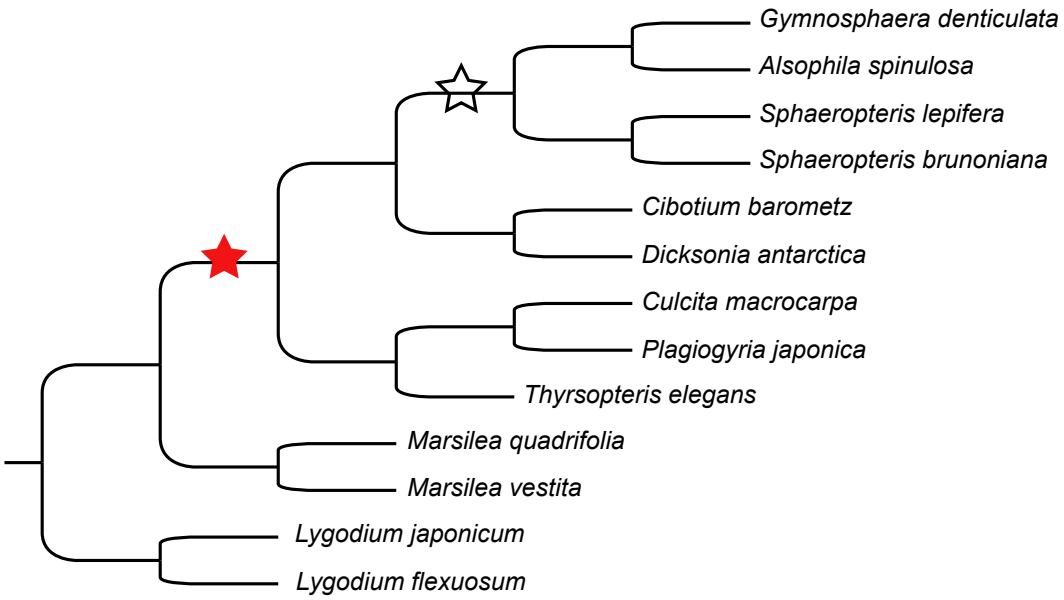

Fig. S21 Correction of substitution rate and phylogenetic placement of WGDs. The phylogenetic tree for substitution rate correction and the proposed scenario of the phylogenetic location of the Cyathealean WGD event based on the results of mixture modeling and rate correction analysis (see text for details). Red stars indicate evidence supporting WGD event, while hollow stars indicate the absence of support for WGD.

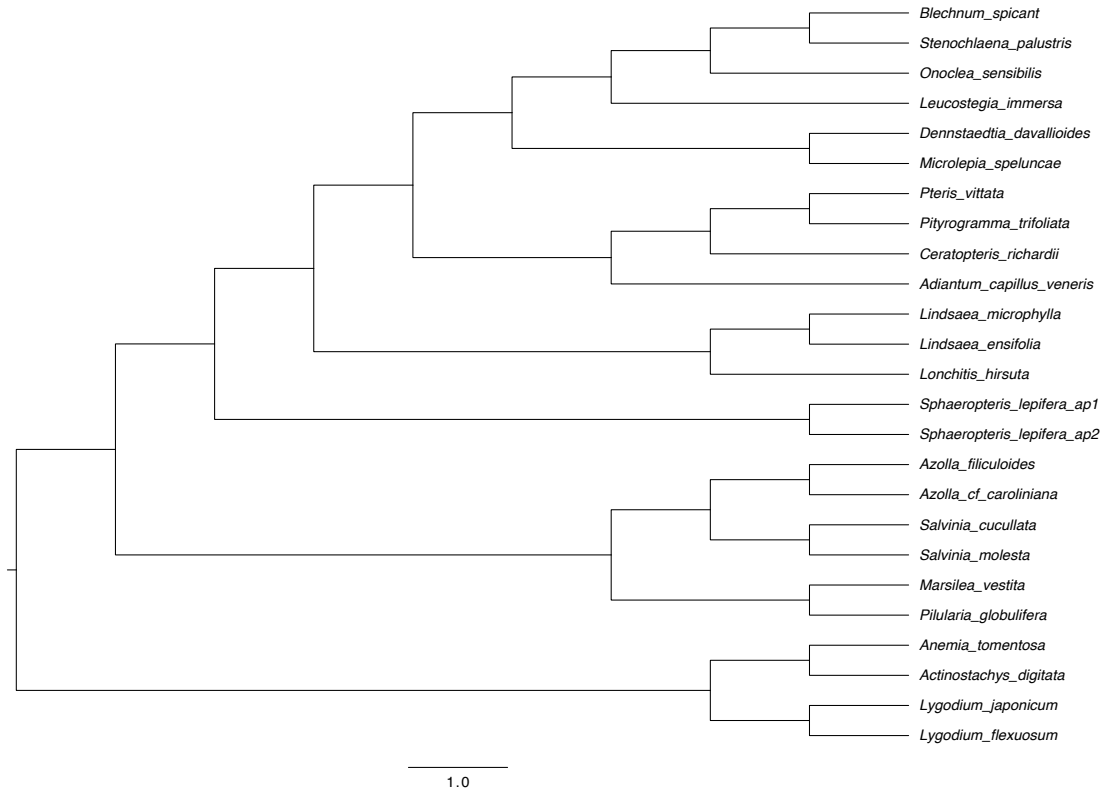

Fig. S22 Phylogenetic tree for the absolute dating of the WGD event. The maximum likelihood tree of the 25 species using the *S. lepifera* as an example. Details of resources for species used are provided in supplementary table S18.

Supplementary Tables

Table S1 Statistics of genome features of the *Sphaeropteris brunoniana*, *Sphaeropteris lepifera* and *Gymnosphaera denticulata*.

	S. brunoniana (2n=2x=138)	*S. lepifera* (2n=2x=138)	*G. denticulata* (2n=2x=136)
Sequencing			
Raw bases of PacBio Sequel (Gb)	79.40	177.47	195.95
Raw bases of Illumina (Gb)	333.24	285.97	343.41
Raw bases of Hi-C (Gb)	425.97	868.59	763.90
Assembly			
Total assembly size (Gb)	2.54	5.50	6.25
Number of scaffolds	69	69	68
N50 of scaffolds (Mb)	38.74	82.91	93.33
Number of contigs	1,722	3,005	2,284
N50 of contigs (Mb)	6.09	3.73	7.57
Anchor rates (%)	99.22	97.25	99.65
Complete BUSCOs (%)	97.2	87.5	98
GC ratio (%)	40.70	41.20	41.59
Annotation			

Number of protein-coding genes	39,097	33,993	40,140
Average gene length (bp)	13,224.92	17,098.33	19602.84
Average CDS length (bp)	1339.10	1306.70	1267.84
Average exon length (bp)	301.50	286.10	294.36
Average intron length (bp)	3,453.73	4,427.84	5,545.23
Complete BUSCOs (%)	95.30	86.60	98.00
Functional annotation			
InterPro	28,489 (72.87%)	32,081 (94.35%)	36,534 (91.02%)
Nr	35,732 (91.39%)	31,354 (92.21%)	35,301 (87.94%)
GO	29,345 (75.06%)	22,545 (66.30%)	24,803 (61.79%)
KEGG	34,225 (87.54%)	8,430 (24.79%)	9,453 (23.55%)
KOG	14,658 (37.49%)	2,850 (8.38%)	3,208 (7.99%)
Total	36,176 (92.53%)	32,862 (96.64%)	37,663 (93.83%)

Table S2 Statistics of repeat sequences in *Sphaeropteris brunoniana*, *Sphaeropteris lepifera* and *Gymnosphaera denticulata* genomes.

	Sphaeropteris brunoniana		*Sphaeropteris lepifera*		*Gymnosphaera denticulata*	
Type	Repeat size (bp)	% of genome	Repeat size (bp)	% of genome	Repeat size (bp)	% of genome
Trf	213,224,494	8.40	505,259,320	9.18	658,211,458	10.53
De novo	1,975,805,549	77.83	4,654,283,549	84.57	4,949,416,198	79.19
Proteinmask	611,871,213	24.10	1,331,897,803	24.20	1,024,113,894	16.39
Total	2,076,949,849	81.82	4,792,634,720	87.08	5,118,292,657	81.89

Table S3 WHALE (gene tree–species tree reconciliation) analysis for four hypothetical WGDs under the DL + WGD model. Hypothetical WGDs, posterior mean of retention rate (q) and the Bayes Factor (K) to compare the likelihood of $q = 0$ (H_0) to the likelihood of $q > 0$ (H_1) using the Savage-Dickey density ratio.

	Relaxed branch-specific model		Critical branch-specific model	
Hypotheses	$\hat{q}$	K	$\hat{q}$	K
WGD1	0. 4652	0. 0076	0. 3351	0. 0104
WGD2	0. 0337	3. 5020	0. 0002	6570. 0420
WGD3	0. 0240	1. 5620	0. 0419	0. 1374
WGD4	0. 0051	118. 6104	0.0001	7224.7676

Table S4 (Microsoft Excel format) The significantly enriched GO terms of the retained gene duplicates from the Cyathealean polyploidy event for the four Cyatheaceae species.

Table S5 (Microsoft Excel format) The significantly enriched GO terms of the biased retained genes following WGD for *S. brunoniana*.

Table S6 (Microsoft Excel format) The significantly enriched GO terms of the biased retained genes following WGD for *S. lepifera*.

Table S7 (Microsoft Excel format) The significantly enriched GO terms of the biased retained genes following WGD for *A. spinulosa*.

Table S8 (Microsoft Excel format) The significantly enriched GO terms of the biased retained genes following WGD for *G. denticulata*.

Table S9 Intergenomic synteny comparison between *A. spinulosa* and the other three species.

Intergenomic synteny comparisons	Block size > 5		Block size > 10		Block size > 20		Block size > 50	
	The number of anchor gene pairs	The number of syntenic blocks	The number of anchor gene pairs	The number of syntenic blocks	The number of anchor gene pairs	The number of syntenic blocks	The number of anchor gene pairs	The number of syntenic blocks
S. brunoniana vs. *A. spinulosa*	23,516	1,419	17,659	655	12,151	263	5,943	67
S. lepifera vs. A. spinulosa	16,543	1,396	10,043	544	4,070	118	1,127	16
G. denticulata vs. A. spinulosa	14,955	1,238	9,100	471	4,210	126	1,081	17

Table S10 The estimated absolute substitution rates (substitutions/site/year) of 15 species based on the 583 low-copy nuclear (LCN) genes.

Species	Rates
Physcomitrella patens	6.07×10^{-10}
Selaginella moellendorffii	1.67×10^{-9}
Marsilea vestita	1.39×10^{-9}
Gymnosphaera denticulata	$\mathbf{4.16 \times 10^{-10}}$
Alsophila spinulosa	$\mathbf{9.73 \times 10^{-10}}$
Sphaeropteris brunoniana	$\mathbf{1.08 \times 10^{-9}}$
Sphaeropteris lepifera	$\mathbf{4.50 \times 10^{-10}}$
Ceratopteris richardii	2.20×10^{-9}
Adiantum capillus-veneris	1.41×10^{-9}
Gnetum montanum	2.22×10^{-9}
Ginkgo biloba	1.45×10^{-9}
Cycas panzhihuaensis	1.24×10^{-9}
Amborella trichopoda	1.67×10^{-9}
Oryza sativa	2.71×10^{-9}
Arabidopsis thaliana	2.63×10^{-9}

Table S11 The d_S and d_N of 15 species based on the 583 low-copy nuclear (LCN) genes.

Species	d_S				d_N			
	Mean	Median	SD	Multiple comparisons	Mean	Median	SD	Multiple comparisons
Physcomitrella patens	1.084	0.623	1.088	d	0.103	0.070	0.116	f
Selaginella moellendorffii	2.193	2.192	1.404	a	0.178	0.163	0.120	a
Marsilea vestita	0.944	0.919	0.547	e	0.109	0.097	0.074	ef
Gymnosphaera denticulata	**0.045**	**0.026**	**0.113**	**h**	**0.024**	**0.007**	**0.077**	**j**
Alsophila spinulosa	**0.086**	**0.037**	**0.153**	**h**	**0.049**	**0.012**	**0.100**	**i**
Sphaeropteris brunoniana	**0.069**	**0.013**	**0.221**	**h**	**0.063**	**0.005**	**0.140**	**hi**
Sphaeropteris lepifera	**0.028**	**0.010**	**0.076**	**h**	**0.014**	**0.003**	**0.039**	**j**
Ceratopteris richardii	0.718	0.693	0.381	f	0.093	0.082	0.068	fg
Adiantum capillus-veneris	0.396	0.384	0.221	g	0.063	0.054	0.052	hi
Gnetum montanum	1.220	1.128	0.733	c	0.164	0.145	0.099	ab
Ginkgo biloba	0.356	0.287	0.273	g	0.092	0.065	0.090	fg
Cycas panzhihuaensis	0.333	0.291	0.267	g	0.078	0.057	0.089	gh
Amborella trichopoda	0.396	0.916	0.652	d	0.122	0.101	0.080	de
Oryza sativa	1.520	1.428	1.051	b	0.150	0.135	0.101	bc
Arabidopsis thaliana	1.513	1.395	1.000	b	0.138	0.125	0.089	cd

Notes: Means with the same lowercase letter do not significantly differ by the least significant difference (LSD) test at P-value < 0.05.

Table S12 Statistics on transposable elements (TE).

	Sphaeropteris brunoniana		*Sphaeropteris lepifera*		*Gymnosphaera denticulata*	
Type	Length (bp)	% of genome	Length (bp)	% of genome	Length (bp)	% of genome
DNA	282,618,916	11.13	367,024,204	6.67	793,776,402	12.70
LINE	133,739,867	5.27	322,354,251	5.86	536,292,732	8.58
SINE	323,171	0.01	6,692,860	0.12	19,962,460	0.32
Total LTR	1,614,483,374	63.60	3,866,876,871	70.26	3,358,102,951	53.72
LTR/Gypsy	824,716,658	32.49	1,514,507,630	27.52	1,591,494,890	25.46
LTR/Copia	767,267,065	30.22	1,715,728,095	31.17	995,035,404	15.92
Unclassified	51,600,929	2.03	202,205,039	3.67	391,031,747	6.26
Total TE	2,019,943,401	79.57	4,742,720,580	86.17	5,042,013,026	80.66

Table S13 Solo / Intact LTR ratios in LTR retrotransposons in ferns

Species	Number of Solo LTR	Number of Intact LTR	Ratio
S. brunoniana	297,651	49,002	6.07
S. lepifera	584,219	336,694	1.74
G. denticulata	828,616	209,269	3.96
A. spinulosa	769,400	130,786	5.88
M. vestita	70,800	55,581	1.27
Ad. capillus-veneris	654,597	54,782	11.95
C. richardii	928,075	39,249	23.65

Table S14 (Microsoft Excel format) The significantly enriched GO terms of the significantly expanded gene families at the branch leading to Cyatheaceae.

Table S15 The significantly enriched GO terms for species-specific family expansions of three tree ferns.

Species	Term	Description	*P*value
A. spinulosa	GO:0002213	defense response to insect	0.00051206
A. spinulosa	GO:0006952	defense response	0.00065941
A. spinulosa	GO:0009625	response to insect	0.00074057
A. spinulosa	GO:2000068	regulation of defense response to insect	8.449E-05
A. spinulosa	GO:0042742	defense response to bacterium	0.00119356
A. spinulosa	GO:0010039	response to iron ion	3.3008E-06
A. spinulosa	GO:0001666	response to hypoxia	4.4231E-06
A. spinulosa	GO:0009607	response to biotic stimulus	0.00035945
A. spinulosa	GO:0002831	regulation of response to biotic stimulus	0.00030508
A. spinulosa	GO:0051707	response to other organism	0.00031997
A. spinulosa	GO:0043207	response to external biotic stimulus	0.00033269
A. spinulosa	GO:0009873	ethylene-activated signaling pathway	0.00011852
A. spinulosa	GO:0009723	response to ethylene	0.00312996
A. spinulosa	GO:0071369	cellular response to ethylene stimulus	0.00018778
A. spinulosa	GO:0031347	regulation of defense response	0.04103273
A. spinulosa	GO:0045492	xylan biosynthetic process	1.953E-05
A. spinulosa	GO:0045491	xylan metabolic process	6.0285E-05
A. spinulosa	GO:0010410	hemicellulose metabolic process	0.00261142

A. spinulosa	GO:0010051	xylem and phloem pattern formation	0.0076538
A. spinulosa	GO:0009834	plant-type secondary cell wall biogenesis	0.00580197
A. spinulosa	GO:0071554	cell wall organization or biogenesis	0.00015441
A. spinulosa	GO:0010383	cell wall polysaccharide metabolic process	0.01072421
A. spinulosa	GO:0042545	cell wall modification	0.01600489
A. spinulosa	GO:0044036	cell wall macromolecule metabolic process	0.02466624
A. spinulosa	GO:0070592	cell wall polysaccharide biosynthetic process	0.00080737
A. spinulosa	GO:0042546	cell wall biogenesis	0.00088415
A. spinulosa	GO:0008202	steroid metabolic process	0.03045441
S. lepifera	GO:0009627	systemic acquired resistance	0.0000122
S. lepifera	GO:0045087	innate immune response	0.00016911
S. lepifera	GO:0006955	immune response	0.00324069
S. lepifera	GO:0002376	immune system process	0.00607529
S. lepifera	GO:0071369	cellular response to ethylene stimulus	0.03160462
S. lepifera	GO:0009723	response to ethylene	0.0371521
S. lepifera	GO:0009873	ethylene-activated signaling pathway	0.03160462
S. lepifera	GO:0000988	transcription factor activity, protein binding	0.02498735
S. lepifera	GO:0000989	transcription factor activity, transcription factor binding	0.01254812
S. lepifera	GO:0001071	nucleic acid binding transcription factor activity	0.01531632
S. lepifera	GO:0003700	transcription factor activity, sequence-specific DNA binding	0.01531632
S. lepifera	GO:0019722	calcium-mediated signaling	0.00617759

S. lepifera	GO:0061631	ubiquitin conjugating enzyme activity	0.00028858
S. lepifera	GO:0061650	ubiquitin-like protein conjugating enzyme activity	0.00028858
S. lepifera	GO:0006722	triterpenoid metabolic process	1.03E-23
S. lepifera	GO:0016104	triterpenoid biosynthetic process	1.03E-23
S. lepifera	GO:0000250	lanosterol synthase activity	1.16E-23
S. lepifera	GO:0031559	oxidosqualene cyclase activity	1.16E-23
S. lepifera	GO:0042300	beta-amyrin synthase activity	1.16E-23
S. lepifera	GO:0016114	terpenoid biosynthetic process	2.49E-08
S. lepifera	GO:0006721	terpenoid metabolic process	3.79E-08
S. lepifera	GO:0006720	isoprenoid metabolic process	4.35E-13
S. brunoniana	GO:0080134	regulation of response to stress	1.88E-18
S. brunoniana	GO:0009725	response to hormone	0.00826395
S. brunoniana	GO:0009696	salicylic acid metabolic process	0.00881823
S. brunoniana	GO:0010279	indole-3-acetic acid amido synthetase activity	3.98E-18
S. brunoniana	GO:0080123	jasmonate-amino synthetase activity	1.79E-39
S. brunoniana	GO:0009867	jasmonic acid mediated signaling pathway	5.92E-27
S. brunoniana	GO:0031349	positive regulation of defense response	9.79E-33
S. brunoniana	GO:0009694	jasmonic acid metabolic process	3.35E-26
S. brunoniana	GO:0009864	induced systemic resistance, jasmonic acid mediated signaling pathway	1.14E-38
S. brunoniana	GO:0009682	induced systemic resistance	1.04E-35
S. brunoniana	GO:0045088	regulation of innate immune response	5.54E-31

S. brunoniana	GO:0050776	regulation of immune response	4.49E-30
S. brunoniana	GO:0002682	regulation of immune system process	1.97E-29
S. brunoniana	GO:0002218	activation of innate immune response	2.21E-35
S. brunoniana	GO:0002253	activation of immune response	2.21E-35
S. brunoniana	GO:0002684	positive regulation of immune system process	1.46E-33
S. brunoniana	GO:0045089	positive regulation of innate immune response	1.46E-33
S. brunoniana	GO:0050778	positive regulation of immune response	1.46E-33
S. brunoniana	GO:0002252	immune effector process	9.79E-33
S. brunoniana	GO:0045087	innate immune response	3.77E-24
S. brunoniana	GO:0006955	immune response	1.26E-23
S. brunoniana	GO:0002376	immune system process	2.73E-22
S. brunoniana	GO:0042742	defense response to bacterium	7.19E-17
S. brunoniana	GO:0010333	terpene synthase activity	0.0000286
S. brunoniana	GO:0010411	xyloglucan metabolic process	1.23E-31
S. brunoniana	GO:0009969	xyloglucan biosynthetic process	1.62E-09
S. brunoniana	GO:0016762	xyloglucan:xyloglucosyl transferase activity	5.56E-27
S. brunoniana	GO:0071395	cellular response to jasmonic acid stimulus	8.82E-27
S. brunoniana	GO:0030243	cellulose metabolic process	0.0000018
S. brunoniana	GO:0010410	hemicellulose metabolic process	2.15E-26
S. brunoniana	GO:0009753	response to jasmonic acid	5.97E-26
S. brunoniana	GO:0010383	cell wall polysaccharide metabolic process	1.44E-24

S. brunoniana	GO:0071555	cell wall organization	5.09E-24
S. brunoniana	GO:0044036	cell wall macromolecule metabolic process	4.03E-21
S. brunoniana	GO:0042546	cell wall biogenesis	1.47E-17
S. brunoniana	GO:0042545	cell wall modification	1.75E-14
S. brunoniana	GO:0005618	cell wall	1.79E-17

Table S16 Comparisons of the number of genes involved in light signaling in 15 plant species.

Species	Cryptochromes (CRY1/2)	ETHYLENE INSENSITIVE3/EIN3-LIKE1 (EIN3/EIL1)	YUCs
Al. spinulosa	8	9	11
Sp. lepifera	7	6	11
Sp. brunoniana	7	5	10
Gy. denticulata	8	6	10
Ad. capillus-veneris	5	1	5
Ce. thalictroides	6	2	6
M. vestita	4	2	8
Cy. panzhihuaensis	2	2	6
Gi. Biloba	2	3	6
Gn. montanum	2	1	4
Am. trichopoda	2	2	7
Ar. thaliana	3	2	9
O. sativa	2	3	8
Se. moellendorffii	2	1	2
P. patens	2	2	5

Table S17 Fossil calibrations adopted for the absolute dating of the WGD event in this study.

Clade	Minimum age constraint/mya	Maximum age constraint/mya
Crown Schizaeales	174.1	345
Stem Lygodiaceae	168.3	299
Stem *Anemia*	163.5	251
Stem Salviniales	174.1	Unconstrained
Stem Marsileaceae	145	201
Stem *Azolla*	83.6	Unconstrained
Stem Cyatheaceae	168.3	251
Stem Lindsaeaceae	99.6	201.6
Stem *Ceratopteris*	37.2	145.5
Crown Pteridaceae	70.6	201.6
Stem Dennstaedtia-Microlepia clade	70.6	201.6
Stem *Ongelea*	55.8	15.5

Table S18 Information on species used for substitution rate correction and absolute dating of WGD in this study.

Species	Data type	Source
Gymnosphaera denticulata	Genome	This study
Sphaeropteris lepifera	Genome	This study
Sphaeropteris brunoniana	Genome	This study
Alsophila spinulosa	Genome	(1)
Cibotium barometz	Transcriptome	(2)
Dicksonia antarctica	Transcriptome	(2)
Culcita macrocarpa	Transcriptome	(3)
Plagiogyria japonica	Transcriptome	(3)
Thyrsopteris elegans	Transcriptome	(3)
Marsilea quadrifolia	Transcriptome	(2)
Marsilea vestita	Genome	(4)
Lygodium japonicum	Transcriptome	(3)
Lygodium flexuosum	Transcriptome	(2)
Actinostachys digitata	Transcriptome	(3)
Anemia tomentosa	Transcriptome	(3)
Pilularia globulifera	Transcriptome	(3)
Salvinia molesta	Transcriptome	(3)
Salvinia cucullata	Genome	(5)
Azolla cf. caroliniana	Transcriptome	(3)
Azolla filiculoides	Genome	(5)
Lonchitis hirsuta	Transcriptome	(3)
Lindsaea ensifolia	Transcriptome	(3)
Lindsaea microphylla	Transcriptome	(3)
Adiantum capillus-veneris	Genome	(6)
Ceratopteris richardii	Genome	(7)
Pityrogramma trifoliata	Transcriptome	(3)
Pteris vittata	Transcriptome	(3)
Microlepia speluncae	Transcriptome	(3)
Dennstaedtia davallioides	Transcriptome	(3)
Leucostegia immersa	Transcriptome	(3)
Onoclea sensibilis	Transcriptome	(3)
Stenochlaena palustris	Transcriptome	(3)

Blechnum spicant	Transcriptome	(3)

Notes:

1. Huang, X. et al. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat. Plants* **8**, 500–512 (2022).

2. Shen, H. et al. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* **7**, 1–11 (2018).

3. Ali, Z. et al. Comparative transcriptomics in ferns reveals key innovations and divergent evolution of the secondary cell walls. *Nat. Plants* 11, 1028–1048 (2025).

4. Rahmatpour, N. et al. Analyses of *Marsilea vestita* genome and transcriptomes do not support widespread intron retention during spermatogenesis. *New Phytol*. **237**, 1490–1494 (2023).

5. Li, F. W. et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **4**, 460–472 (2018).

6. Fang, Y. H. et al. The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nat. Plants* **8**, 1024–1037 (2022).

7. Marchant, D. B. et al. Dynamic genome evolution in a model fern. *Nat. Plants* **8**, 1038–1051 (2022).

Table S19 Information on genomes of 15 plant species used for phylogenetic construction in this study.

Species	Source
Physcomitrella patens	(1)
Selaginella moellendorffii	(1)
Marsilea vestita	(2)
Gymnosphaera denticulata	This study
Sphaeropteris lepifera	This study
Sphaeropteris brunoniana	This study
Alsophila spinulosa	(3)
Adiantum capillus-veneris	(4)
Ceratopteris richardii	(5)
Gnetum montanum	(6)
Ginkgo biloba	(7)
Cycas panzhihuaensis	(8)
Amborella trichopoda	(1)
Oryza sativa	(1)
Arabidopsis thaliana	(1)

Notes:

1. Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-1186 (2012).

2. Rahmatpour, N. et al. Analyses of *Marsilea vestita* genome and transcriptomes do not support widespread intron retention during spermatogenesis. *New Phytol*. **237**, 1490-1494 (2023).

3. Huang, X. et al. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. Nat. Plants 8, 500-512 (2022).

4. Fang, Y. H. et al. The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nat. Plants* **8**, 1024-1037 (2022).

5. Marchant, D. B. et al. Dynamic genome evolution in a model fern. *Nat. Plants* **8**, 1038-1051 (2022).

6. Wan, T. et al. The *Welwitschia* genome reveals unique biology underpinning extreme longevity in deserts. *Nat. Commun* **12**, 4247 (2021).

7. Liu, H. L. et al. The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. *Nat. Plants* **7**, 748–756 (2021).

8. Liu, Y. et al. The *Cycas* genome and the early evolution of seed plants. *Nat. Plants* **8**, 389–401 (2022).

Table S20 The pfam numbers and blastp queries for gene family identification in this study.

Gene family	Pfam number	Query
CesA/*Csl*	PF00535 PF03552 PF13632	All known *Arabidopsis CesA*/*Csl* genes in Supplementary Fig. 20
NAC	PF01849 PF02365	All known *Arabidopsis* and *Populus trichocarpa* NAC in Supplementary Fig. 18
OSC	PF13243 PF13249	All known Arabidopsis OSC genes in Supplementary Fig. 17
CRYs	-	AT4G08920
	-	AT1G04400
EIN3/EIL1	-	AT2G25490
	-	AT5G25350
YUC	-	AT4G32540
	-	AT1G04180
	-	AT2G33230
	-	AT1G04610
	-	AT5G25620
	-	AT4G28720
	-	AT5G43890
		AT1G48910

References

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. 2023. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol Plant*. 16:1733-1742.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 18:170-175.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, *et al*. 2017. *De novo* assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*. 356:92–95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 3:99-101.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 3:95-98.

Judd EJ, Tierney JE, Lunt DJ, Montañez IP, Huber BT, Wing SL, Valdes PJ. 2024. A 485-million-year history of Earth's surface temperature. *Science*. 385:eadk3705.

Huang X, Wang W, Gong T, Wickell D, Kuo LY, Zhang X, Wen J, Kim H, Lu F, Zhao H, *et al*. 2022. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants*. 8:500-512.

Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. 2017. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*. 45:W55-W63.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al*. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 326:289-293.

Liu BH, Shi YJ, Yuan JY, Hu XS, Zhang H, Li N, Li ZY, Chen YX, Mu DS, Wei F. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. arXiv.org arXiv:1308.2012.

Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol*. 176:1410-1422.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov, EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210-3212.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet*. 9:407.