

# **Comparative transcriptomics in ferns reveals key innovations and divergent evolution of the secondary cell walls**

---

In the format provided by the authors and unedited

---

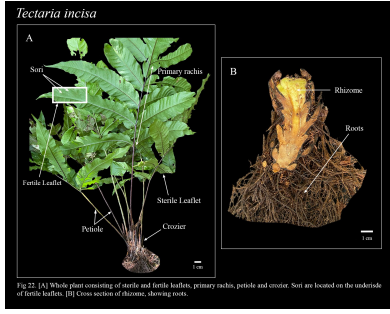
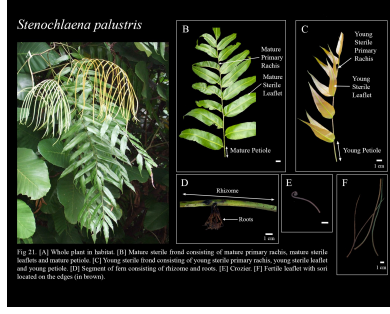
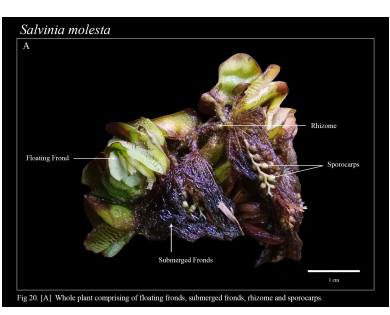
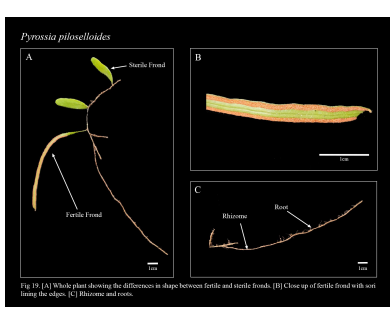
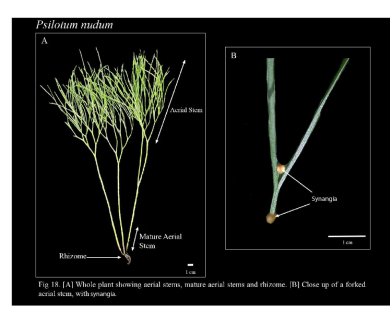
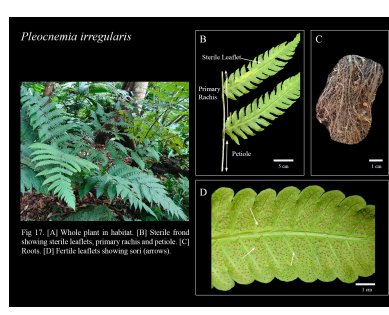
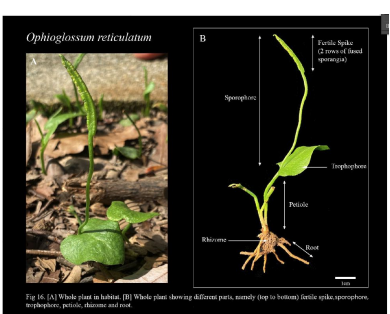
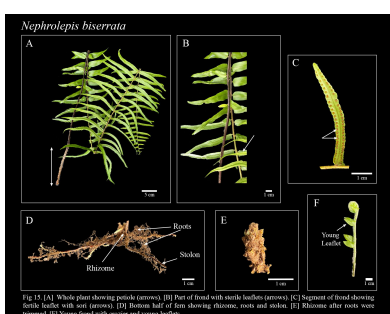
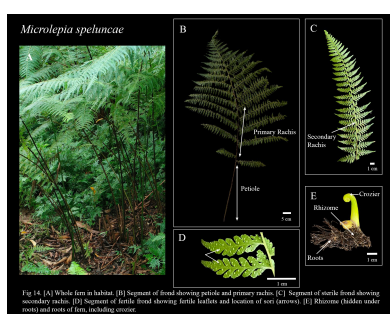
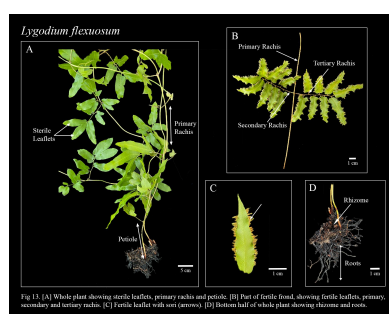
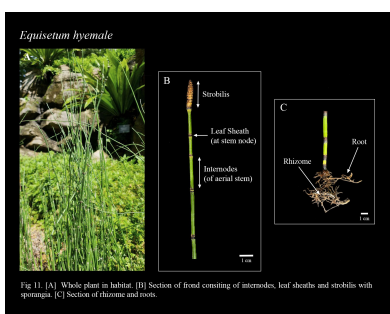
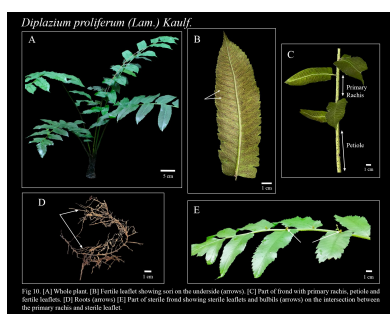
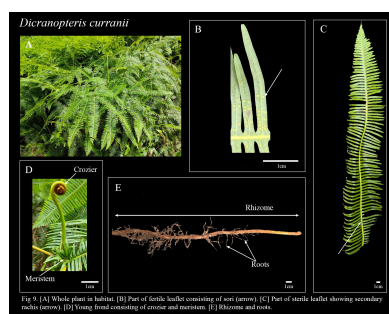
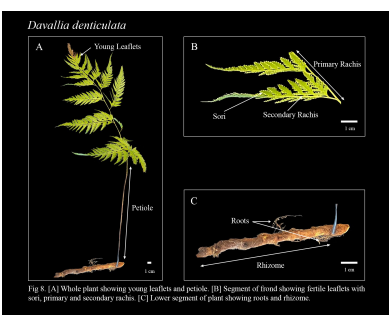
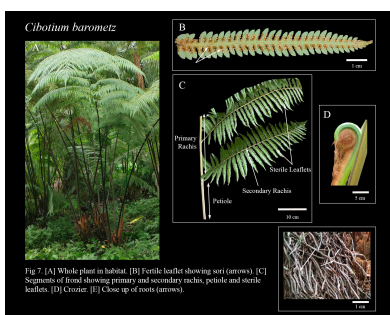
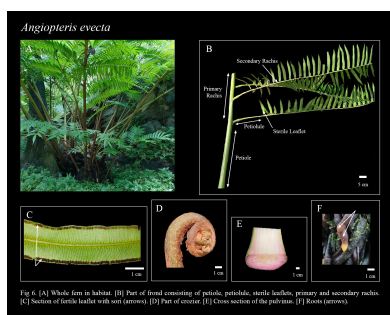
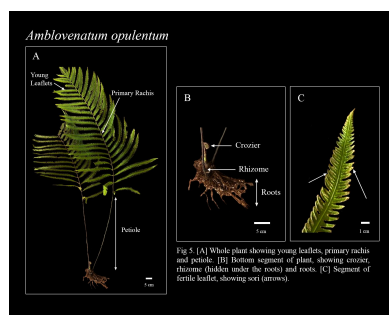
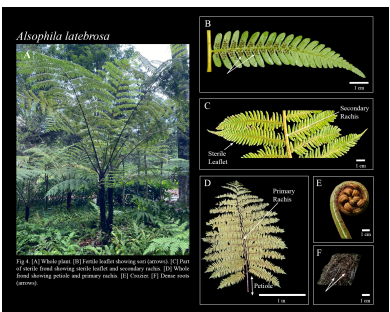
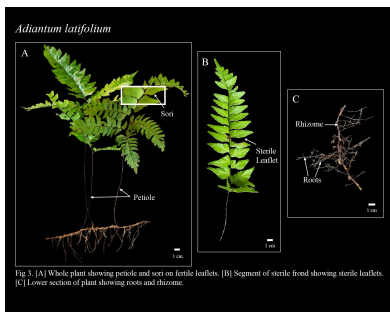
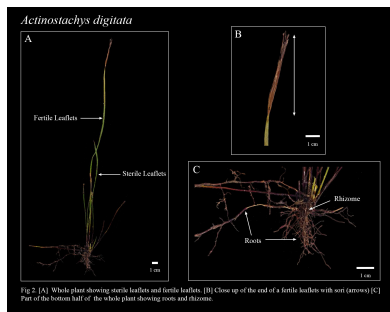
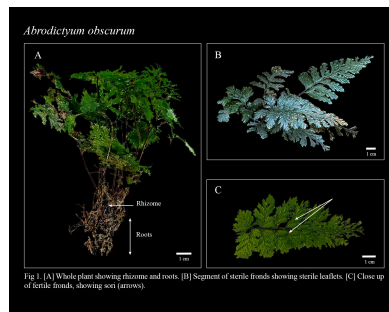
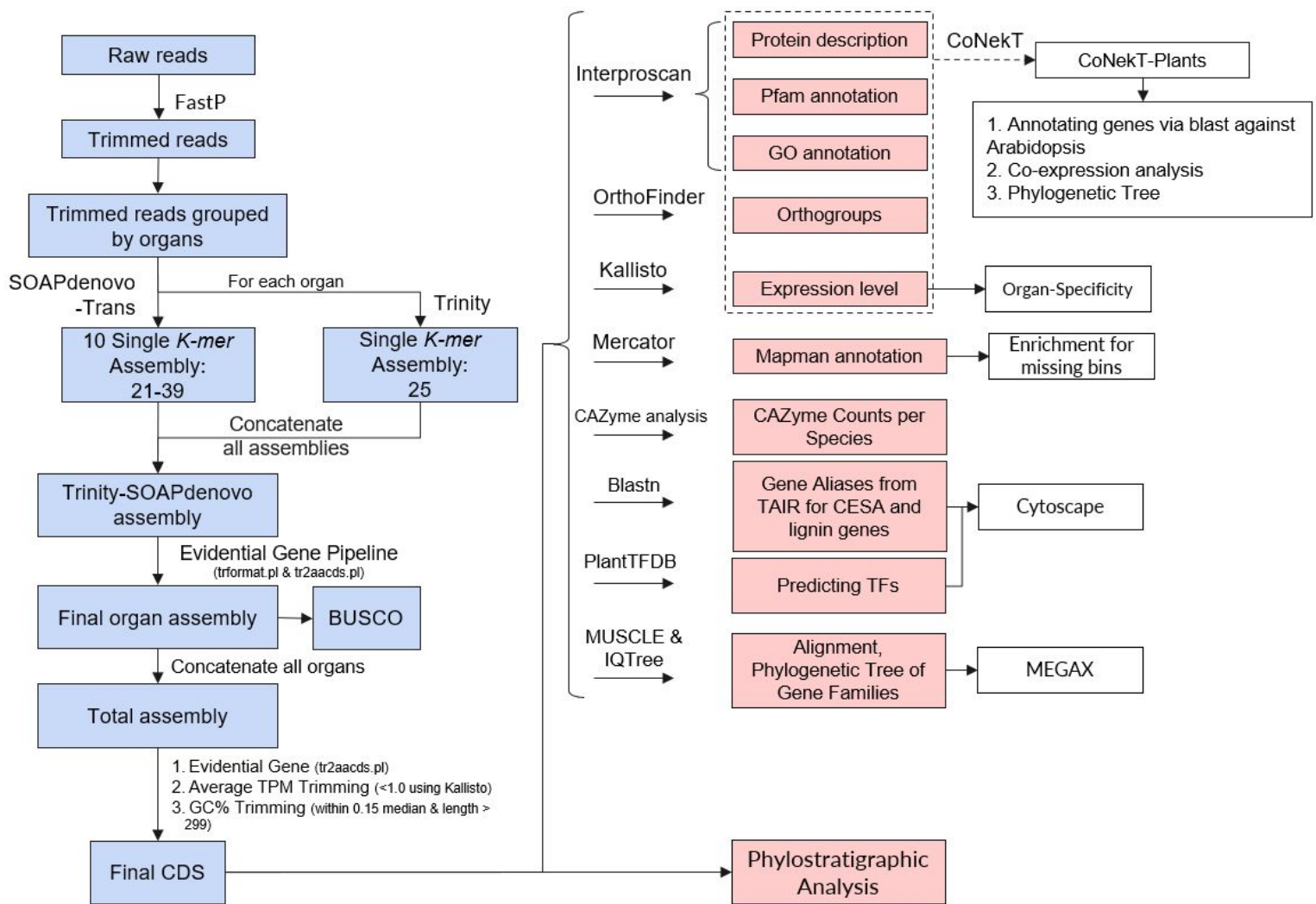
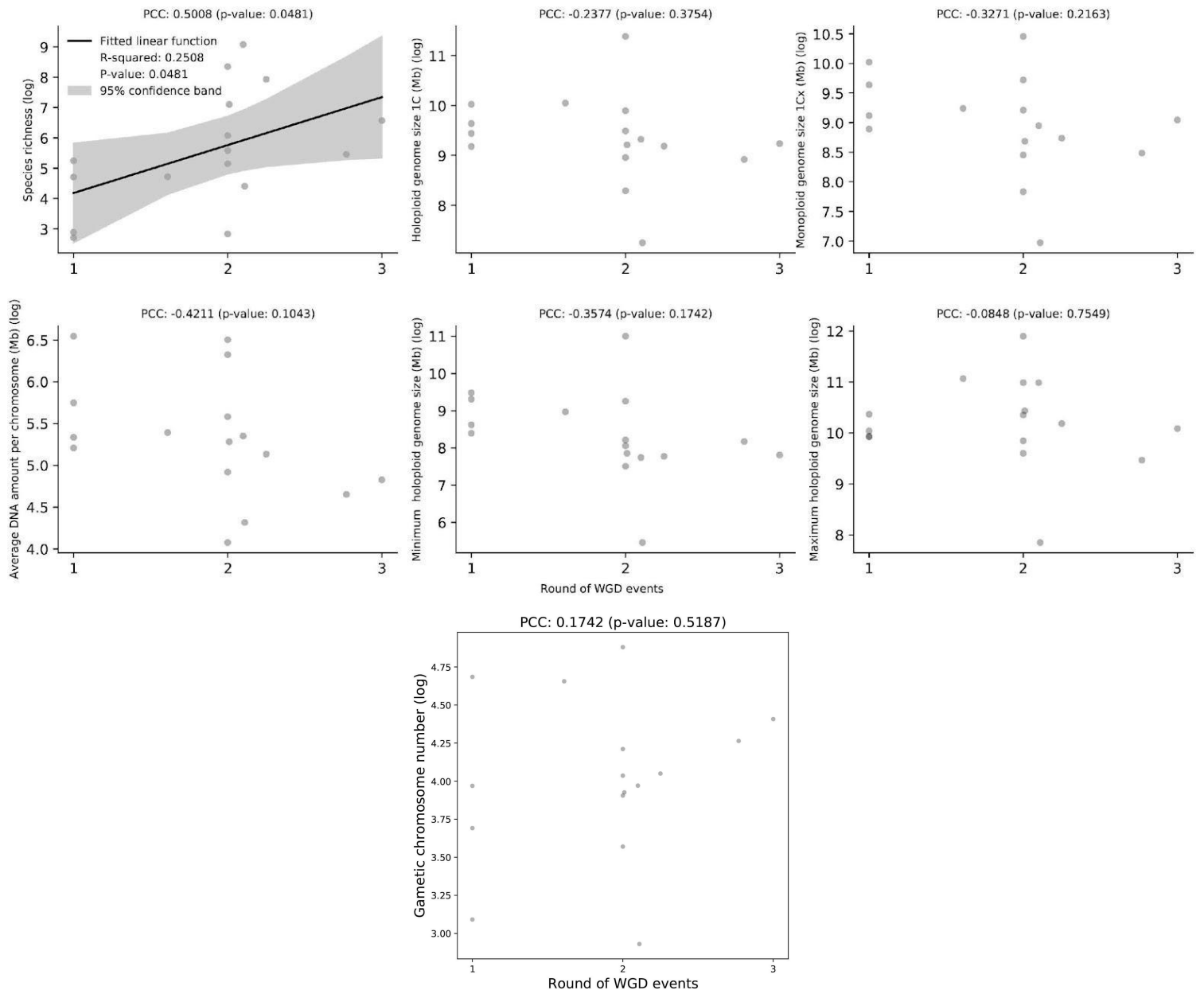


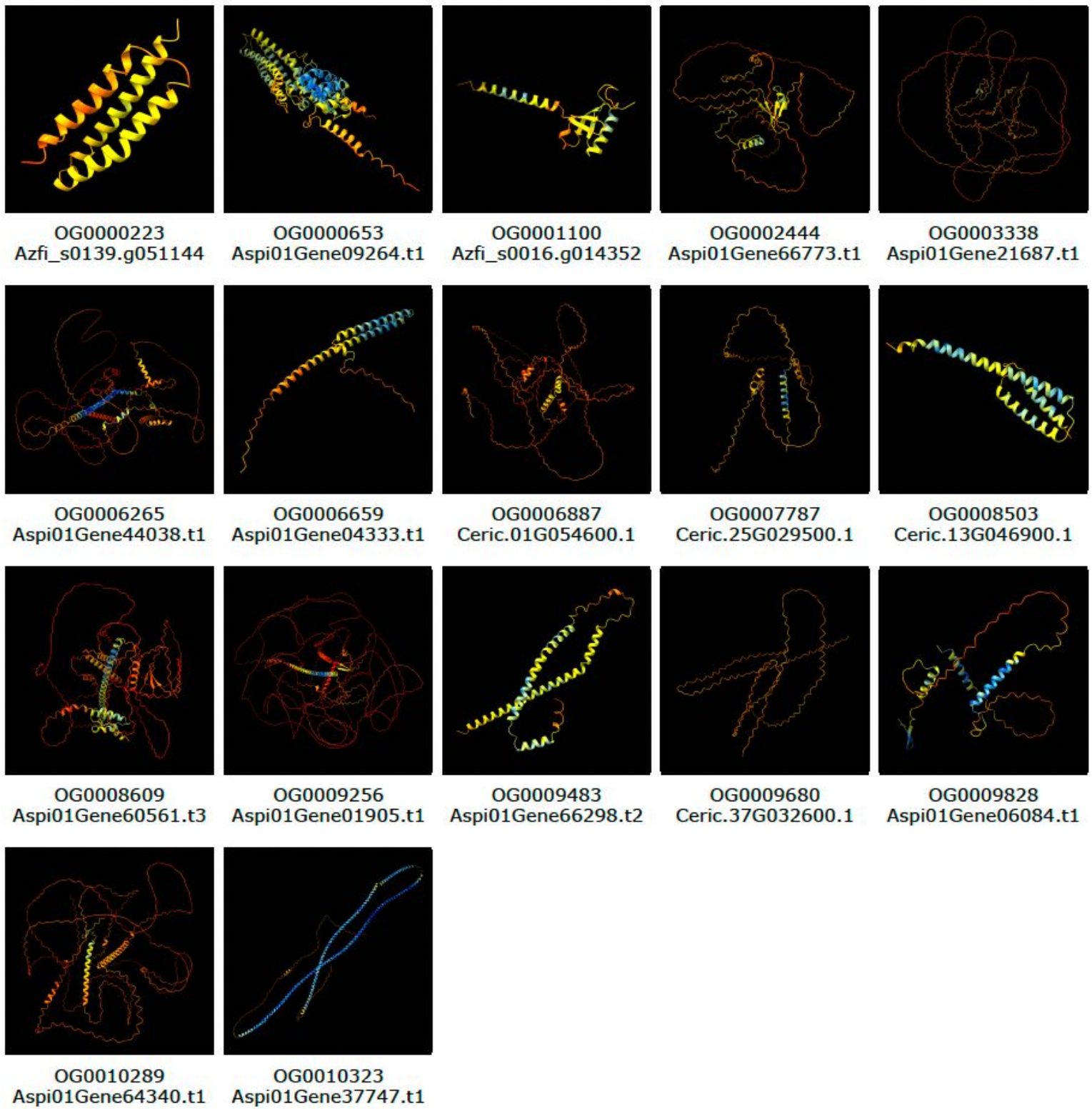
Figure S1. Pictures of the 22 ferns and their sampled organs.



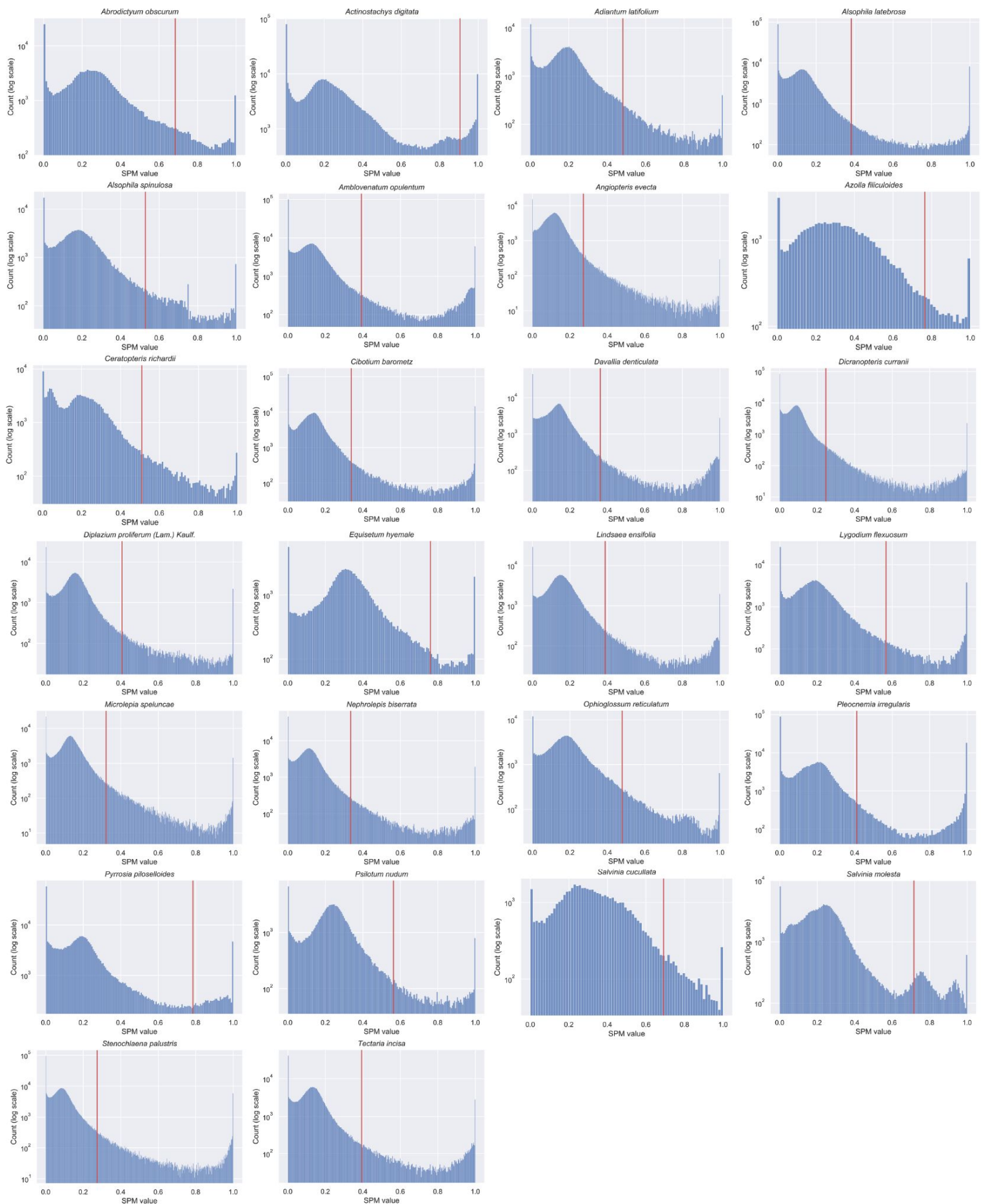
**Figure S2. Transcriptome assembly (blue boxes) and subsequent analyses (red boxes).**



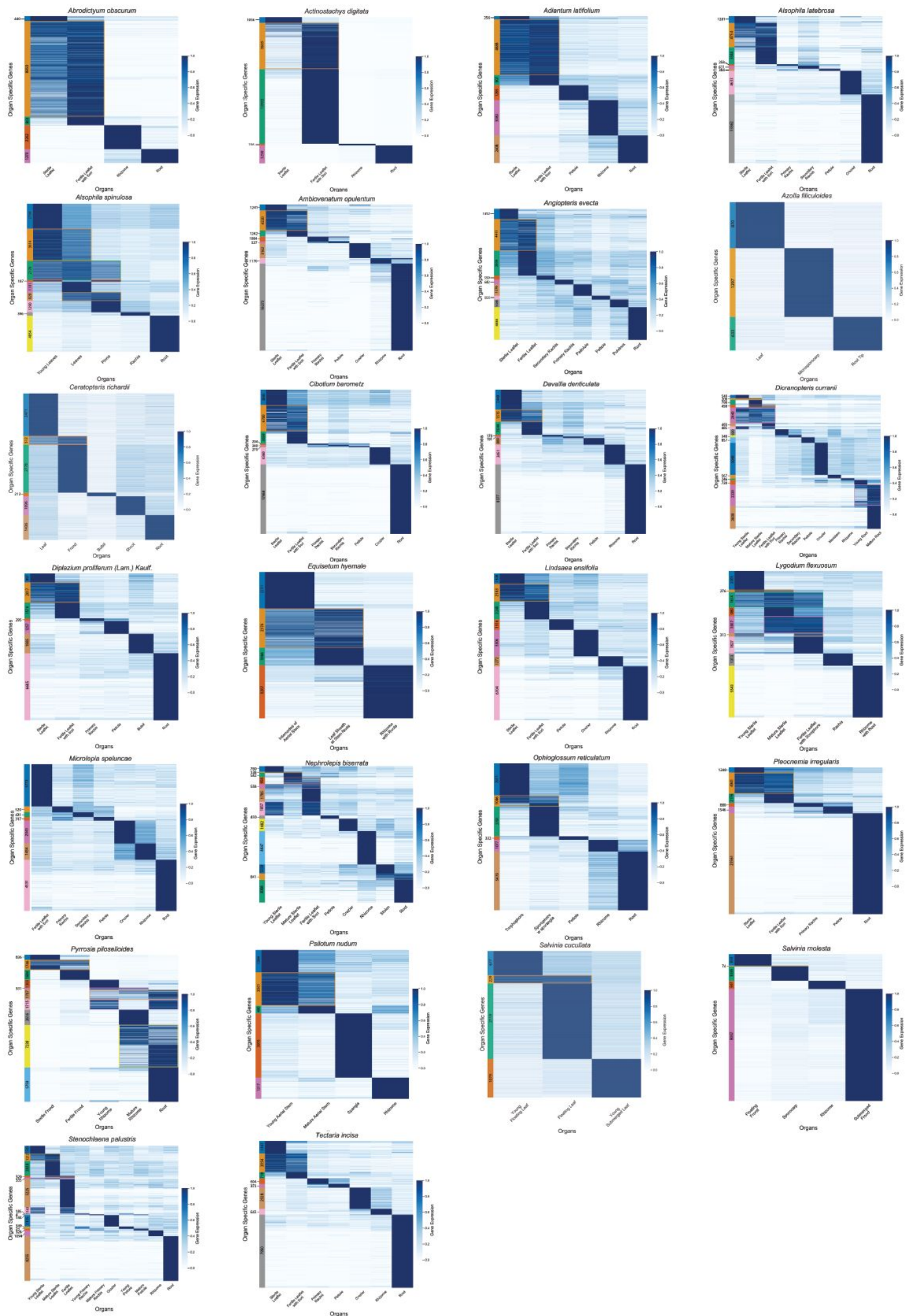
**Figure S3. Correlation analysis between the round of WGD events and the species richness and genomic properties across different fern clades.** The x-axis represents the round of WGD events in different fern clades while the y-axis represents the corresponding species richness or different genomic properties (natural log-transformed to reduce skewness). The title of each subplot denotes the calculated Pearson correlation coefficient (PCC) and the associated p-value. The round of WGD events only shows significant correlation with the species richness (p-value < 0.05), of which the fitted linear function, R-squared, p-value and 95% confidence band derived from the additional linear regression analysis are plotted.



**Figure S4. AlphaFold3-derived structures of the 17 fern-specific proteins.** The colors indicate confidence scores of the structures.



**Figure S5. Number of genes (y-axis) with a given SPM value (x-axis). The SPM value cutoff is indicated by the red line.**



**Figure S6** Gene expression profiles of organ-specific genes. Each gene's expression has been scaled to range from 0 to 1.

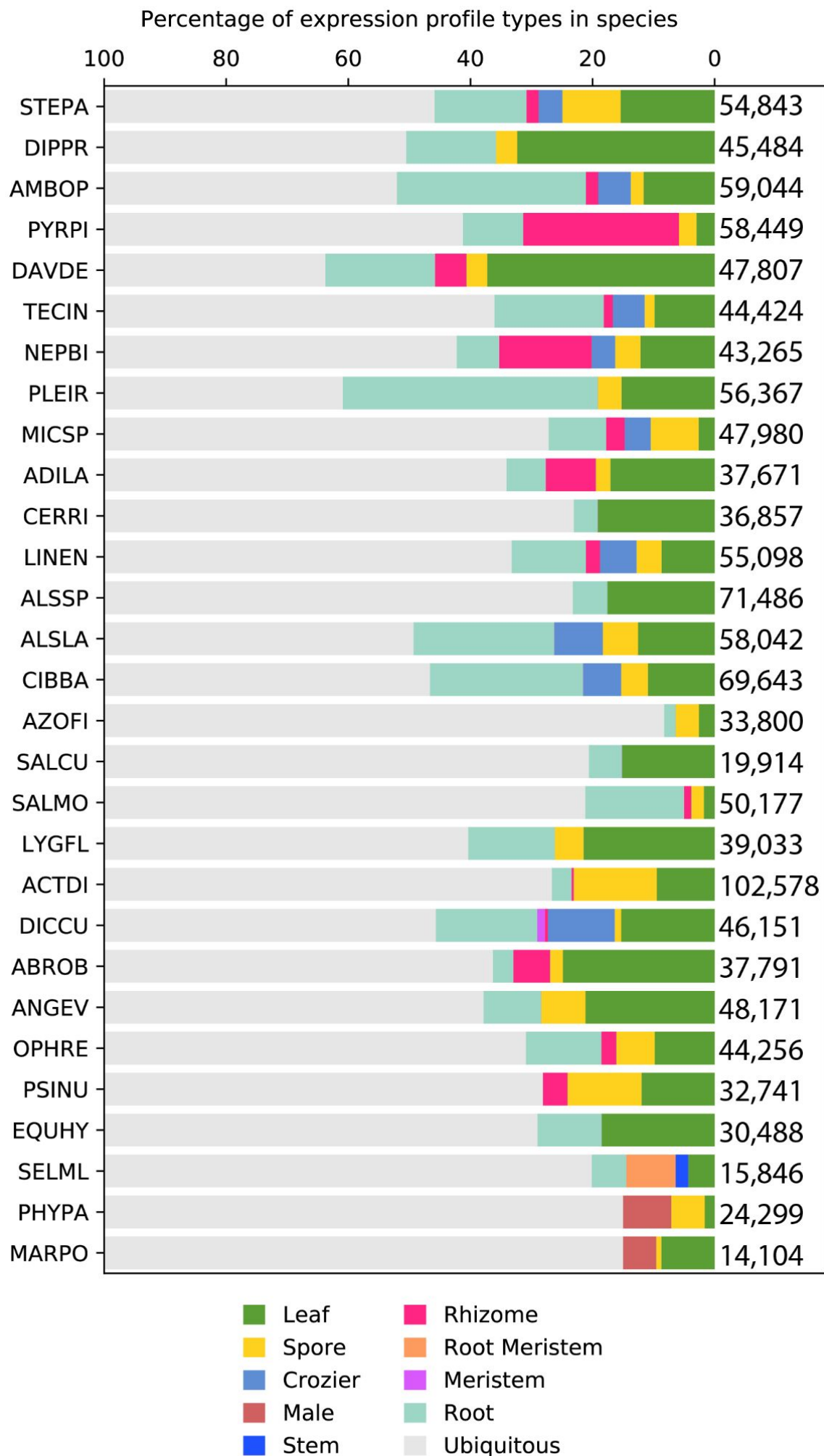
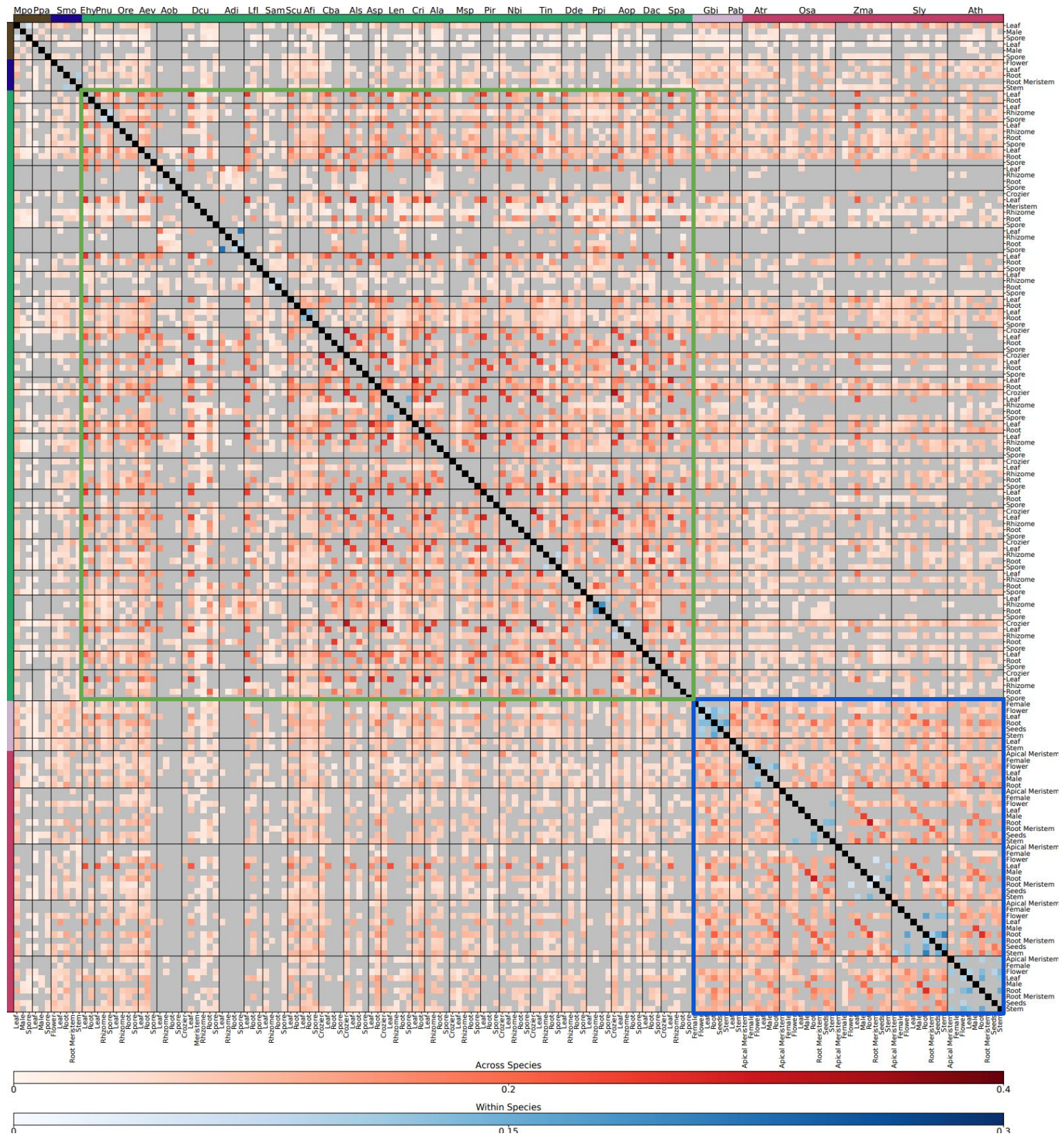
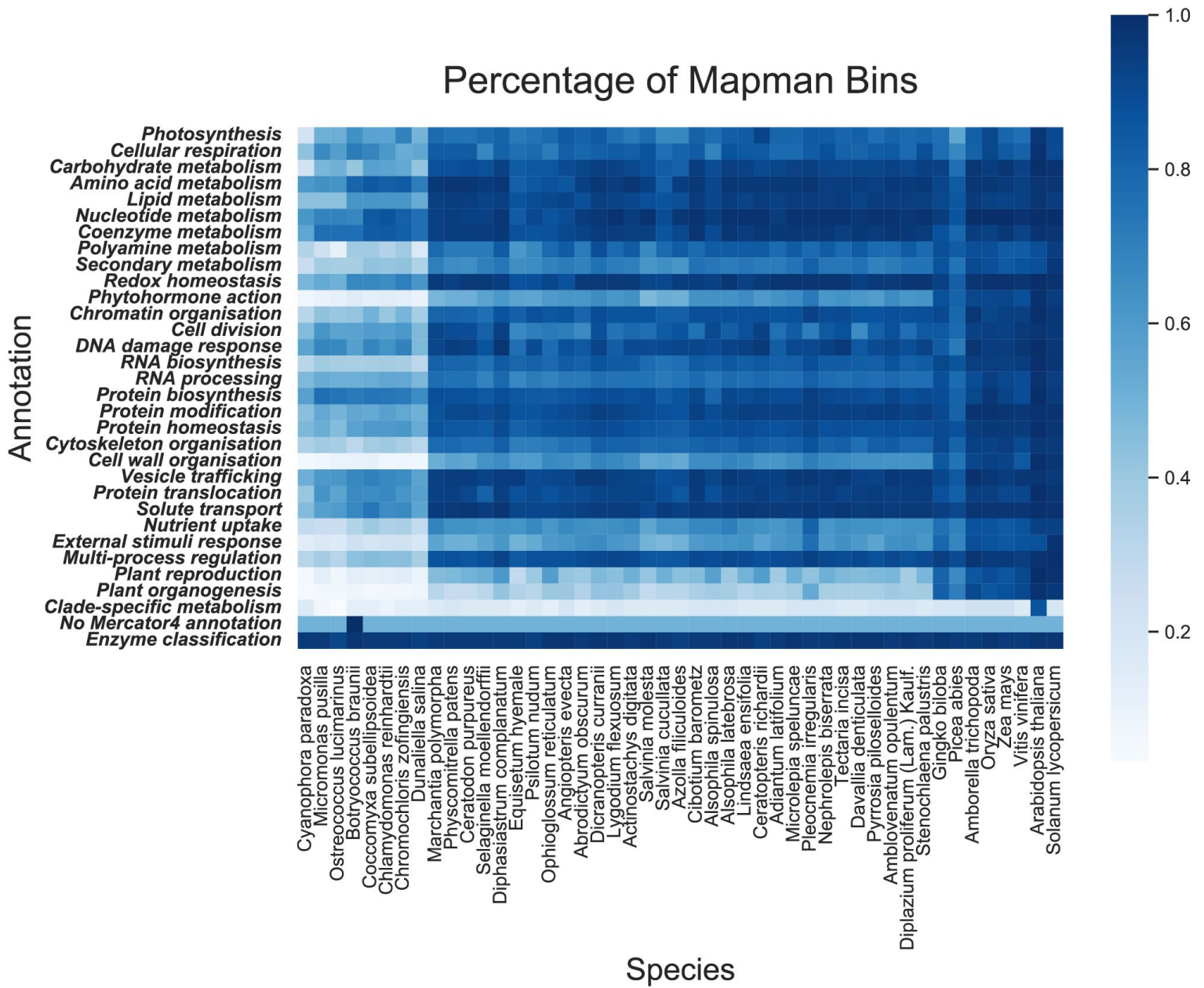


Figure S7. Expression profiles for species-specific genes.

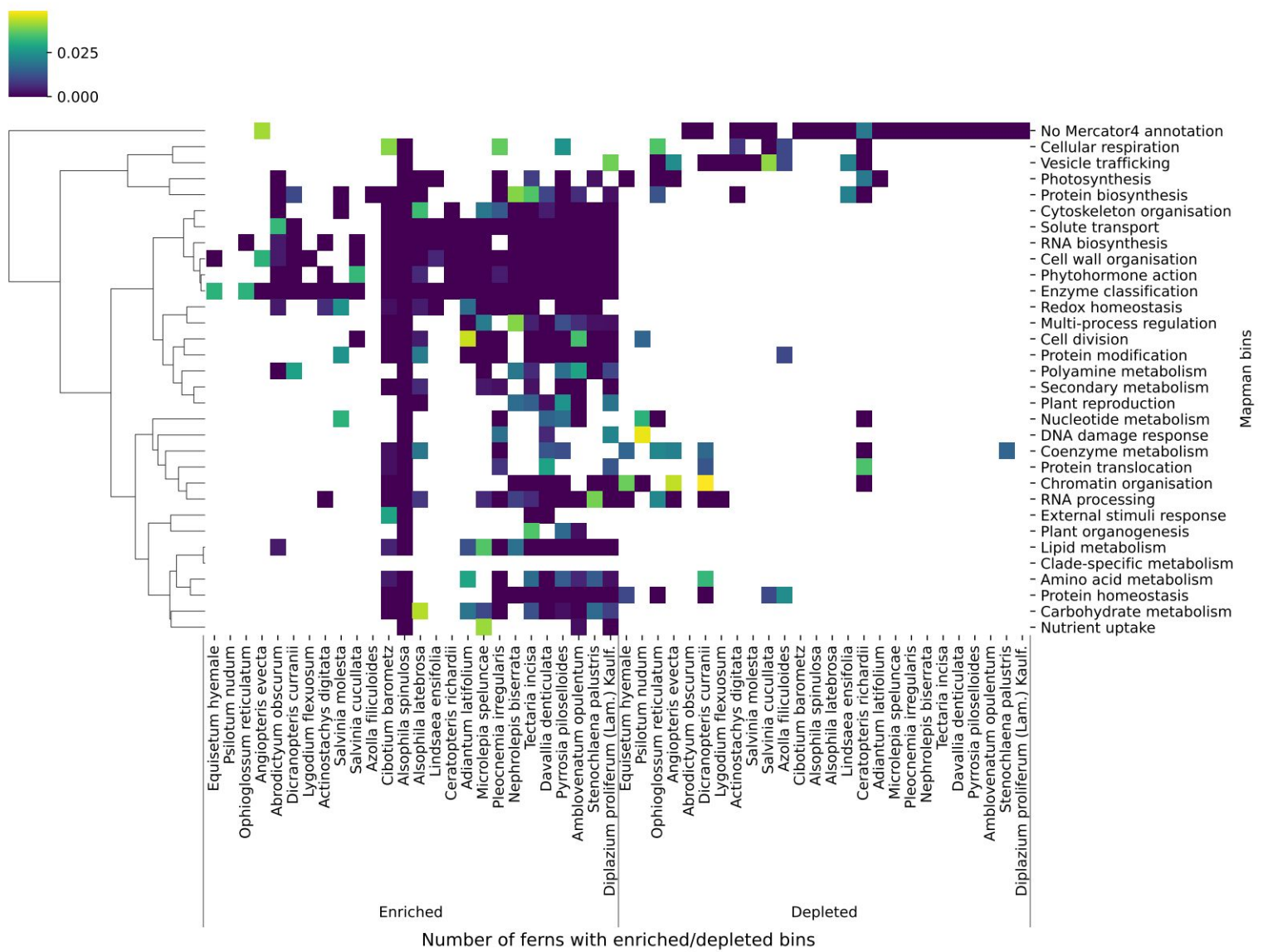


**Figure S8. Transcriptome similarity comparison of Archaeplastida.** The heatmap shows the conservation of organ-specific orthogroups across the species. The jaccard index of across species similarities are indicated by red shades, and for within species similarly with blue shades.

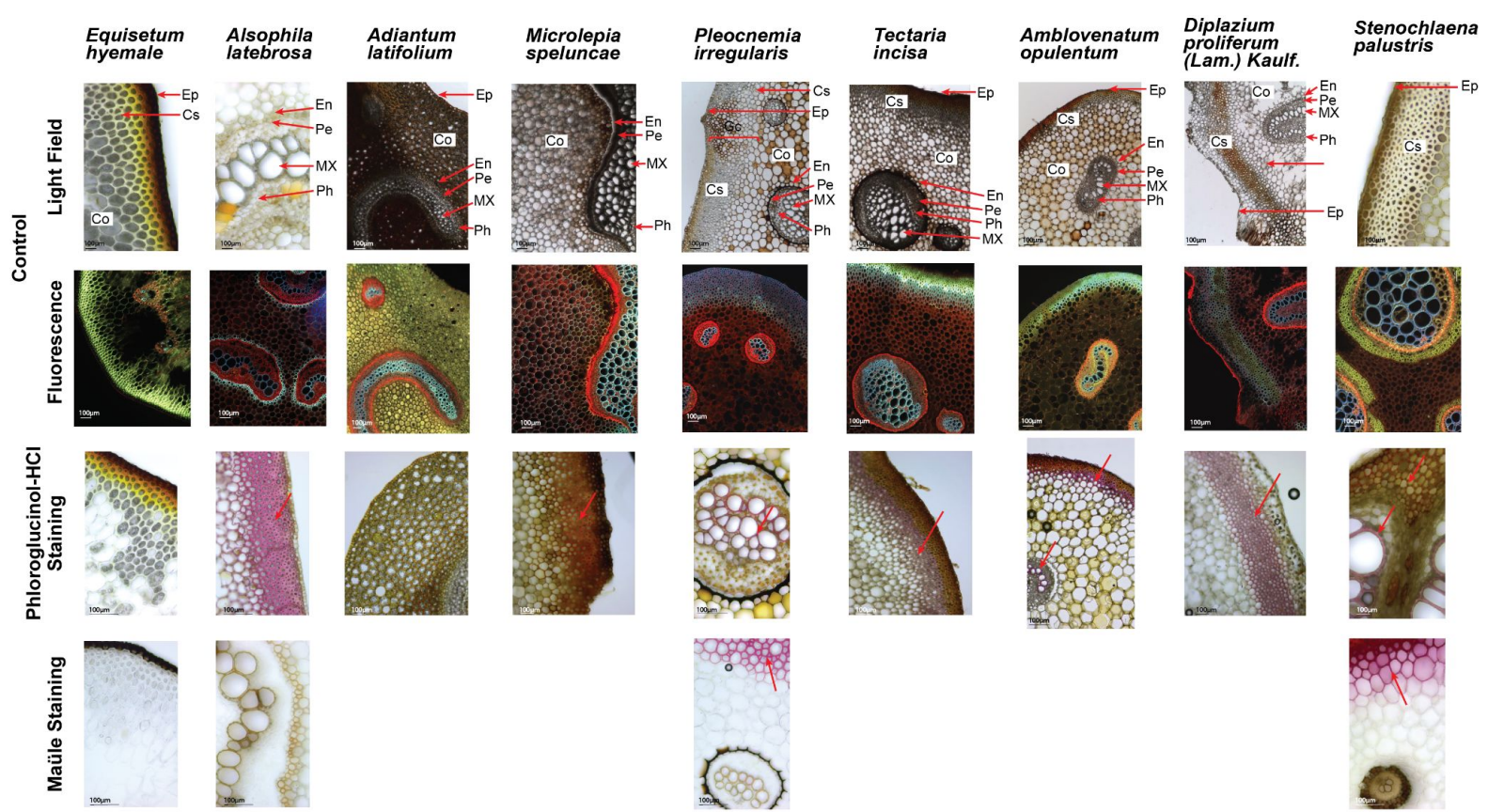
## Percentage of Mapman Bins



**Figure S9. Gene functions found in Archaeplastida.** Mapman bins (rows) found in the different species (columns). The colours indicate the fraction of found bins in a given species, where 1 indicates that all genes in a given bin are present, while 0 indicates complete absence.



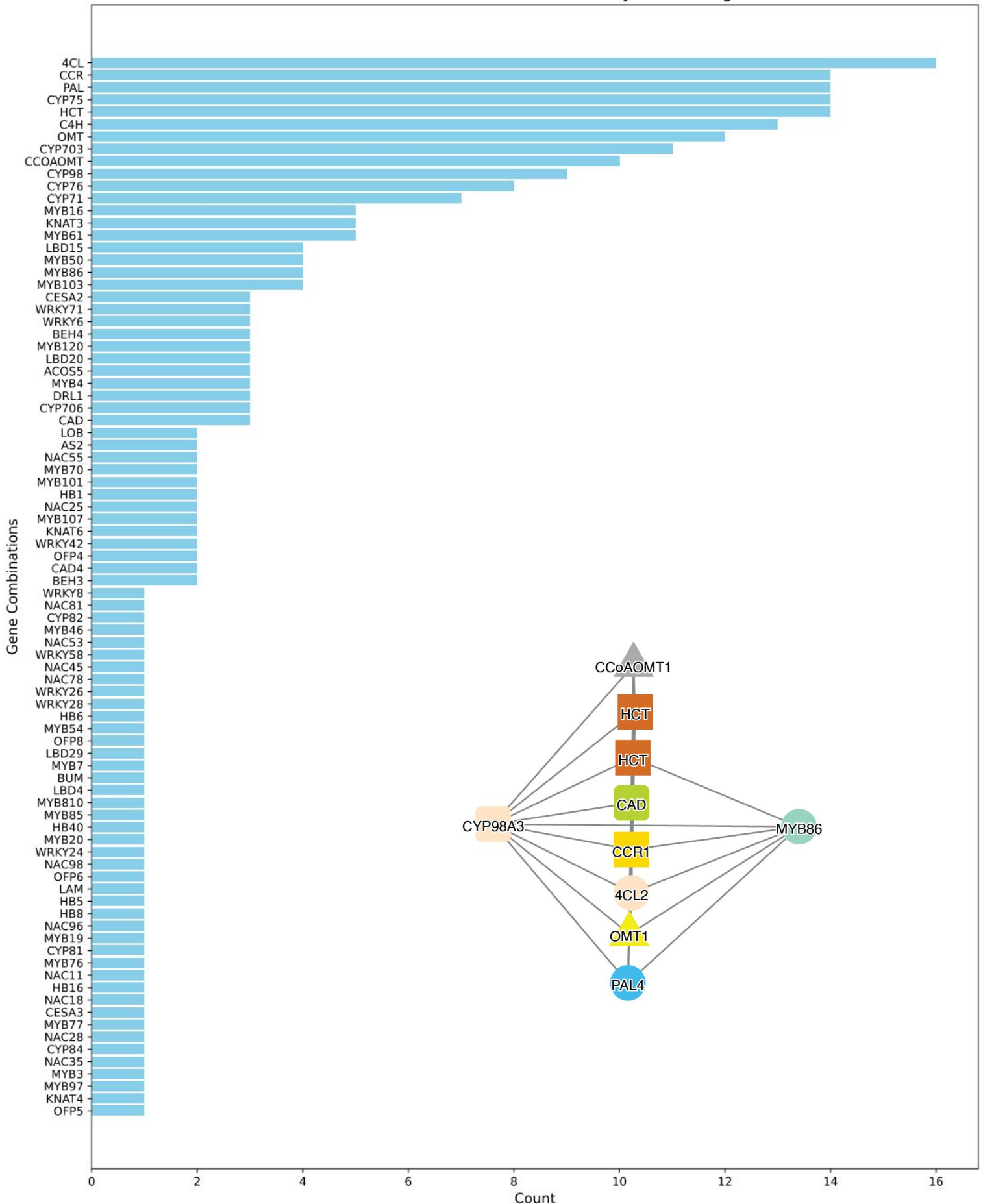
**Figure S10. Enrichment and depletion of biological processes in neighbours of fern-specific genes.** Clustermap showing significantly enriched and depleted primary Mapman bins (y-axis) in neighbours of fern-specific genes across ferns. The colour map indicates the significance of the biological processes, with yellow representing a p-value of 0.05 and blue representing a p-value of 0.00. P-values above 0.05 are masked.



Ep: Epidermis, Cs: Cortical Sclerenchyma, Co: Cortex, En: Endodermis, Pe: Pericycle, Ph: Phloem, MX: Metaxylem

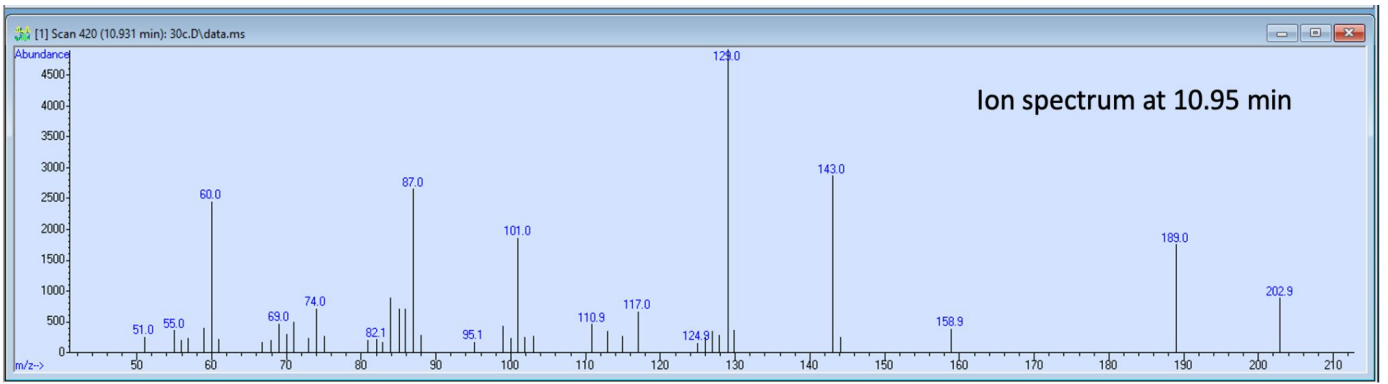
**Figure S11. Light field, fluorescence, phloroglucinol and Maüle staining of the selected ferns.**

Bar Chart of Gene Combinations Sorted by Descending Order

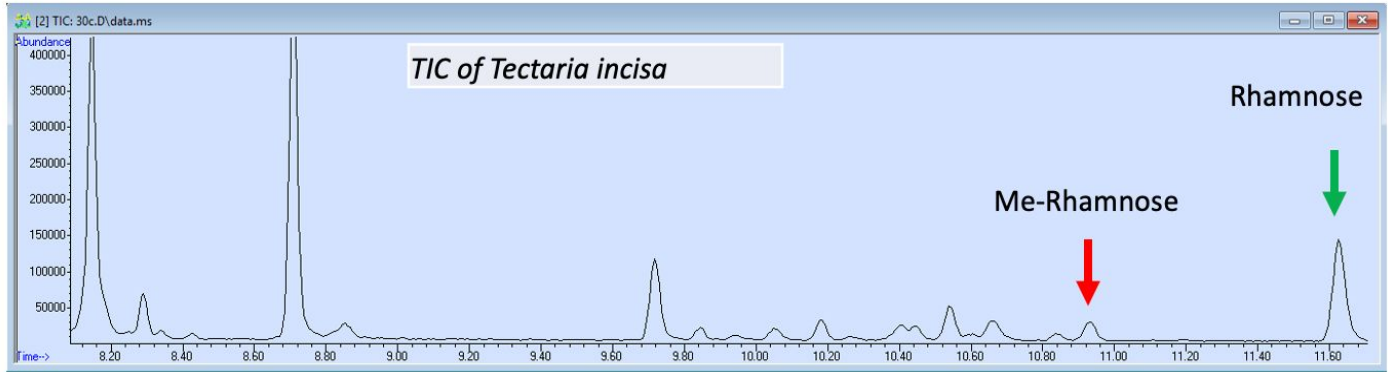


**Figure S12.** The number (x-axis) of the transcription factors, CYP450s enzyme families and lignin-related enzymes co-expressed with at least two lignin biosynthetic enzymes in the analyzed ferns. Co-expression network of PAL4 (blue circle) from *Dicranopteris curranii*. CYP96A3 and MYB86 are connected to seven and five enzymes involved in lignin biosynthesis.

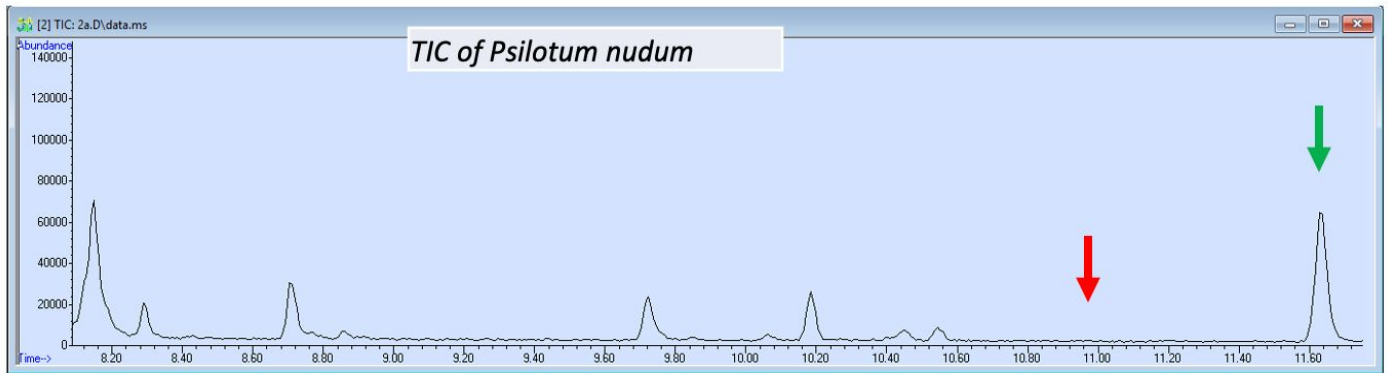
a



b

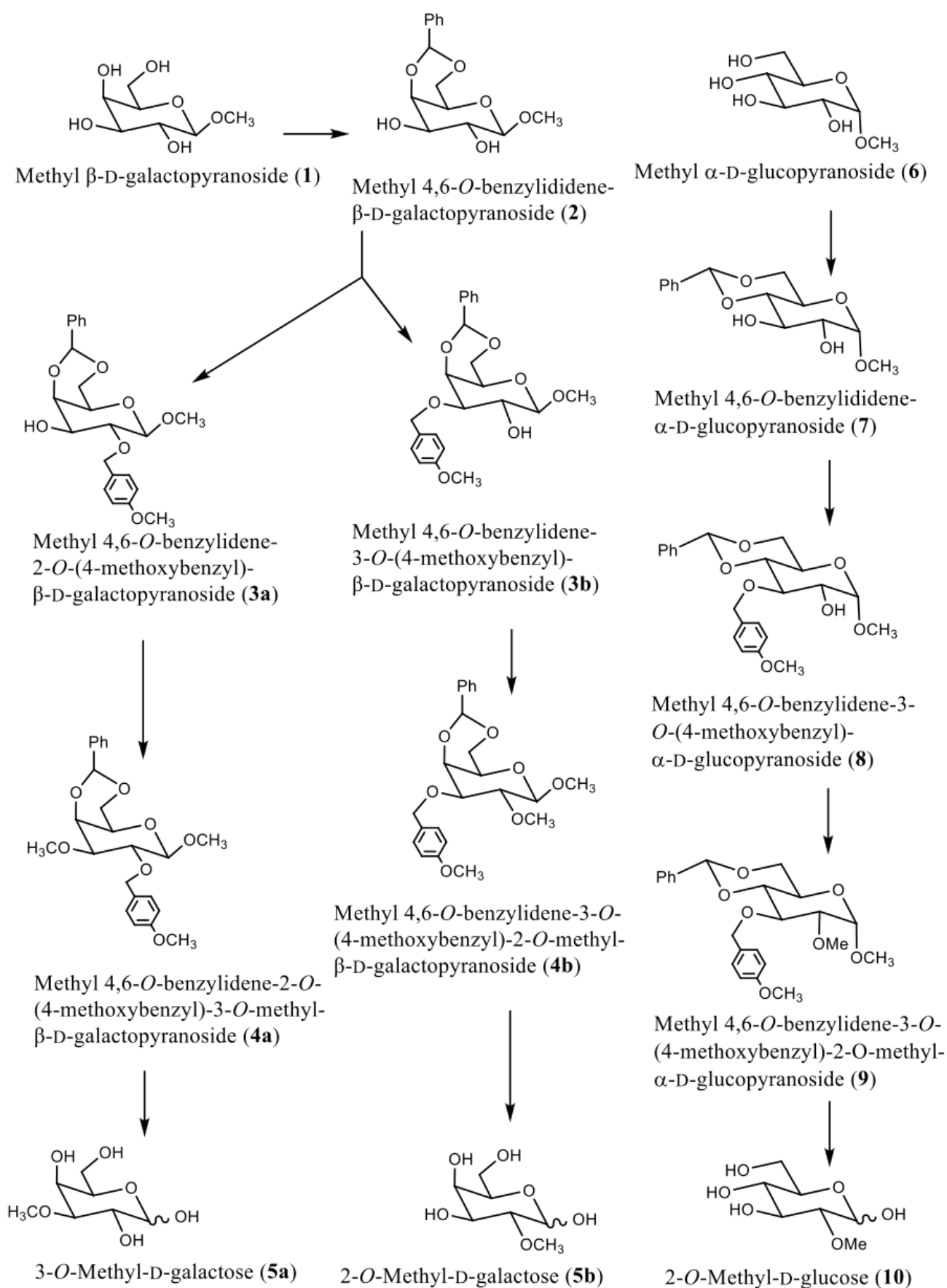


c

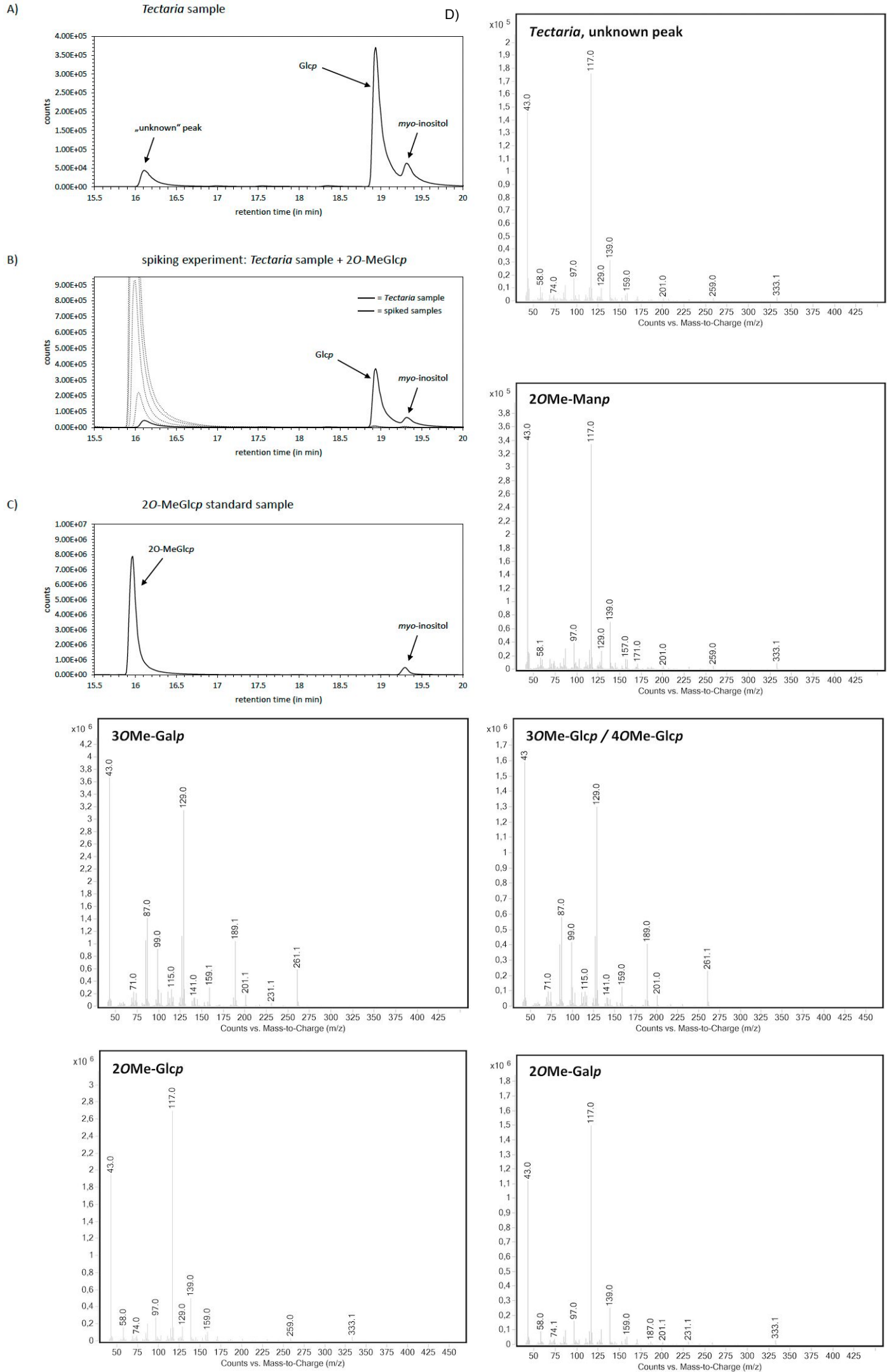


**Figure S13. GC-MS analysis of 3O-MeRhap.** a) MS profile of 3O-MeRhap. b) *Tectaria incisa* contains 3O-MeRhap (red arrow) and rhamnose. c) *Psilotum nudum* contains rhamnose but no 3O-MeRhap.

**Chart 1:** The protocol of chemical synthesis of 2-*O*-methyl- and 3-*O*-methyl- $\alpha,\beta$ -D-galactopyranose **5a** and **5b** and 2-*O*-methyl-D-glucopyranose **10**



**Figure S14.** The protocol of chemical synthesis of 2-*O*-methyl- and 3-*O*-methyl- $\alpha,\beta$ -D-galactopyranose **5a** and **5b** and 2-*O*-methyl-D-glucopyranose **10**.

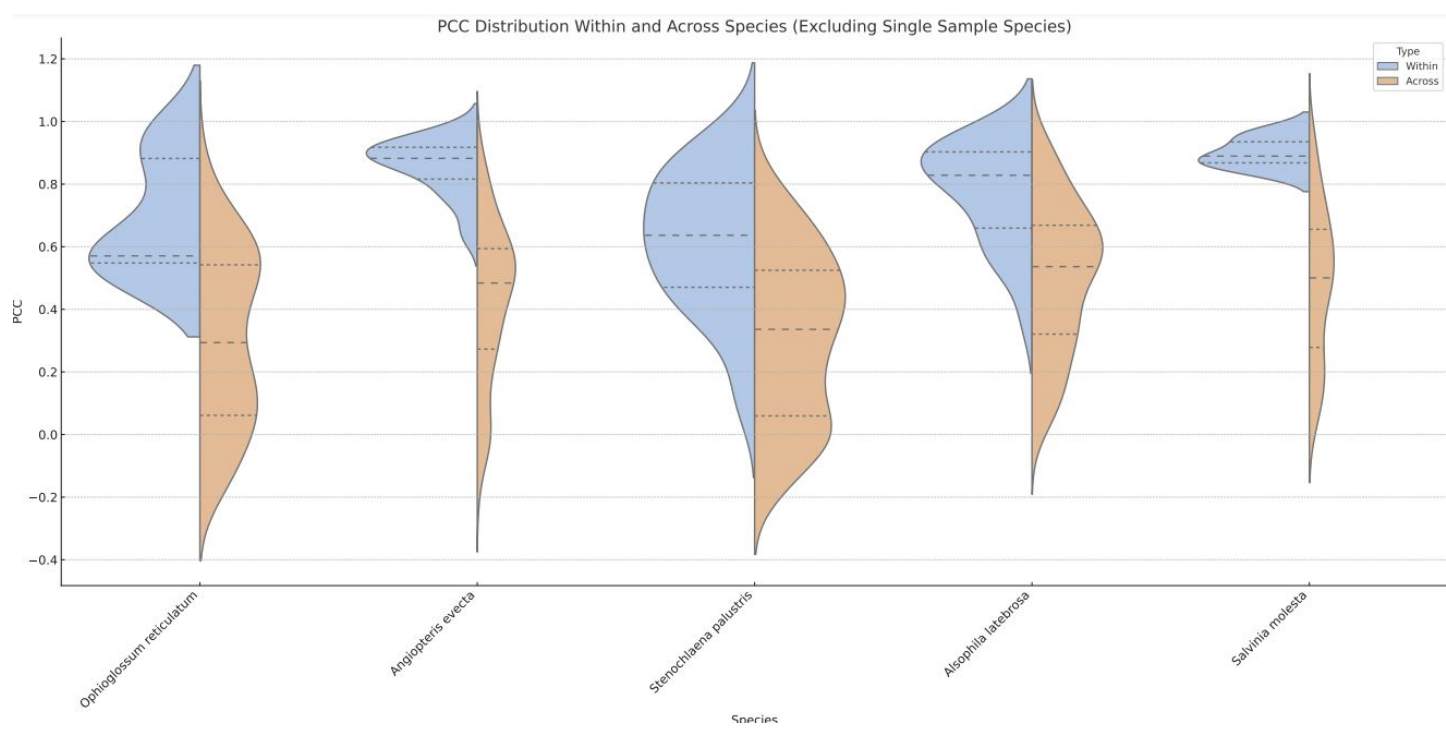
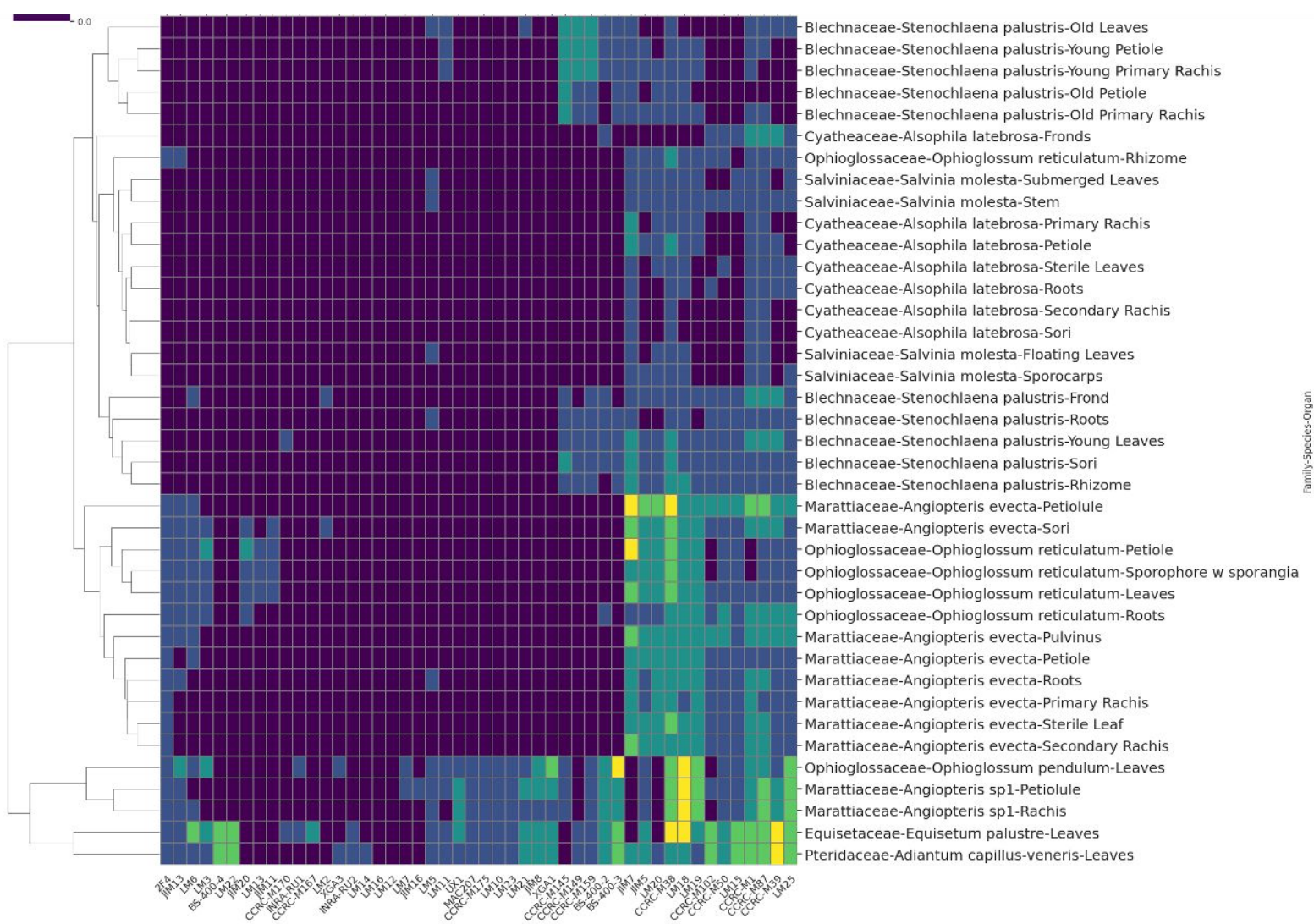


**Figure S15. GC-MS analysis of cell wall sugars.** a) Profile of *Tectaria incisa*, b) *Tectaria incisa* and 2O-MeGlcp standard and c) 2O-MeGlcp standard. d) GC-MS spectra of the unknown peak and the methylated sugar standards.

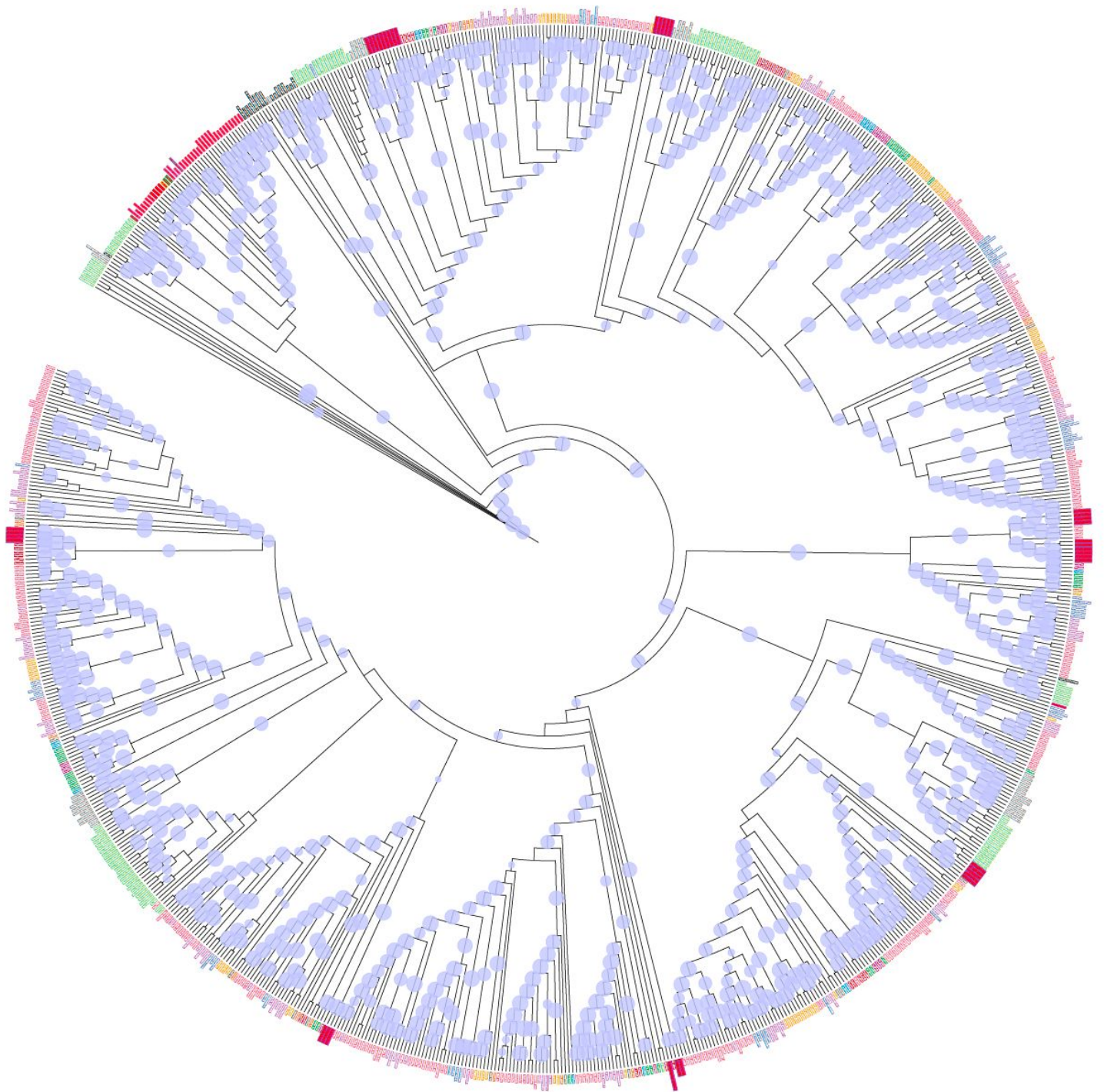
# HRGP Counts by Species

1KP Groups	M. polymorpha	P. patens	C. purpureus	S. moellendorffii	D. complanatum	E. hyemale	P. nudum	O. reticulatum	A. eVecta	A. obscurum	D. curranii	L. flexuosum	A. digitata	S. molesta	S. cucullata	A. filiculoides	C. barometz	A. spinulosa	A. latebrosa	L. ensifolia	C. richardii	A. latifolium	M. speluncae	P. irregularis	N. biserrata	T. incisa	D. denticulata	P. piloselloides	A. opulentum	S. palustris	D. proliferum	G. biloba	P. abies	A. trichopoda	O. sativa	Z. mays	S. lycopersicum	V. vinifera	A. thaliana			
GPI-AGPs	20	16	6	3	2	1	2	7	2	11	6	1	14	4	3	0	22	3	3	9	9	3	10	13	10	0	1	1	8	6	4	5	9	3	12	20	11	0	17	class 1		
CL-extensins	1	0	1	6	5	2	0	2	1	4	2	2	2	0	1	0	2	1	3	4	5	0	3	4	1	2	3	1	6	7	3	8	3	4	0	0	5	0	18	class 2		
PRPs	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	class 3	
non-GPI-AGPs	9	18	28	8	13	9	7	11	5	11	10	9	38	13	4	6	18	15	11	13	29	4	9	8	8	23	15	28	26	16	9	8	3	5	21	25	14	1	7	class 4		
Hybrid AGPs	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	0	1	2	3	3	0	1	class 5		
Hybrid AGPs	1	1	0	0	2	0	1	1	1	0	1	0	0	0	0	1	1	0	0	1	3	0	0	2	1	1	2	1	0	0	0	0	0	0	0	0	0	0	0	1	class 6	
Hybrid AGPs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	class 7	
Hybrid AGPs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	class 8	
GPI-extensins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	class 9	
Hybrid extensins	0	0	4	1	0	1	1	0	1	2	0	0	0	0	1	0	2	4	6	5	0	1	5	3	2	1	0	10	3	8	3	0	4	0	0	0	10	0	1	class 10		
Hybrid extensins	2	0	1	0	0	1	0	0	0	0	0	2	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	3	2	0	2	0	0	0	0	0	0	0	0	0	class 11	
Hybrid extensins	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	2	0	2	0	0	0	0	1	0	1	0	10	0	0	class 12		
Hybrid extensins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	class 13	
Hybrid PRPs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	class 14	
Hybrid PRPs	0	0	0	0	0	0	0	0	1	0	1	2	5	2	0	0	1	0	1	1	1	0	0	4	0	0	2	0	0	4	0	0	0	0	0	0	0	2	1	0	0	class 15
Hybrid PRPs	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	class 16	
Hybrid PRPs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	class 17	
Hybrid PRPs	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	4	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	class 18	
Shared	1	2	2	1	0	0	1	0	1	2	2	0	6	0	0	0	1	1	0	3	0	0	2	0	0	2	2	3	3	1	0	1	3	1	0	1	2	0	1	class 19		
Shared	0	0	0	1	1	1	0	1	2	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	2	7	2	0	0	1	0	6	class 20		
Shared	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	class 21		
Shared	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	2	0	0	0	1	0	1	2	1	1	0	0	0	0	0	0	1	0	0	class 22		
Shared	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	class 23	
< 15% motif	13	1	6	2	3	10	2	7	9	15	9	3	19	10	2	1	13	0	6	9	12	3	8	13	4	1	5	18	18	11	7	8	3	3	25	16	6	0	7	class 24		

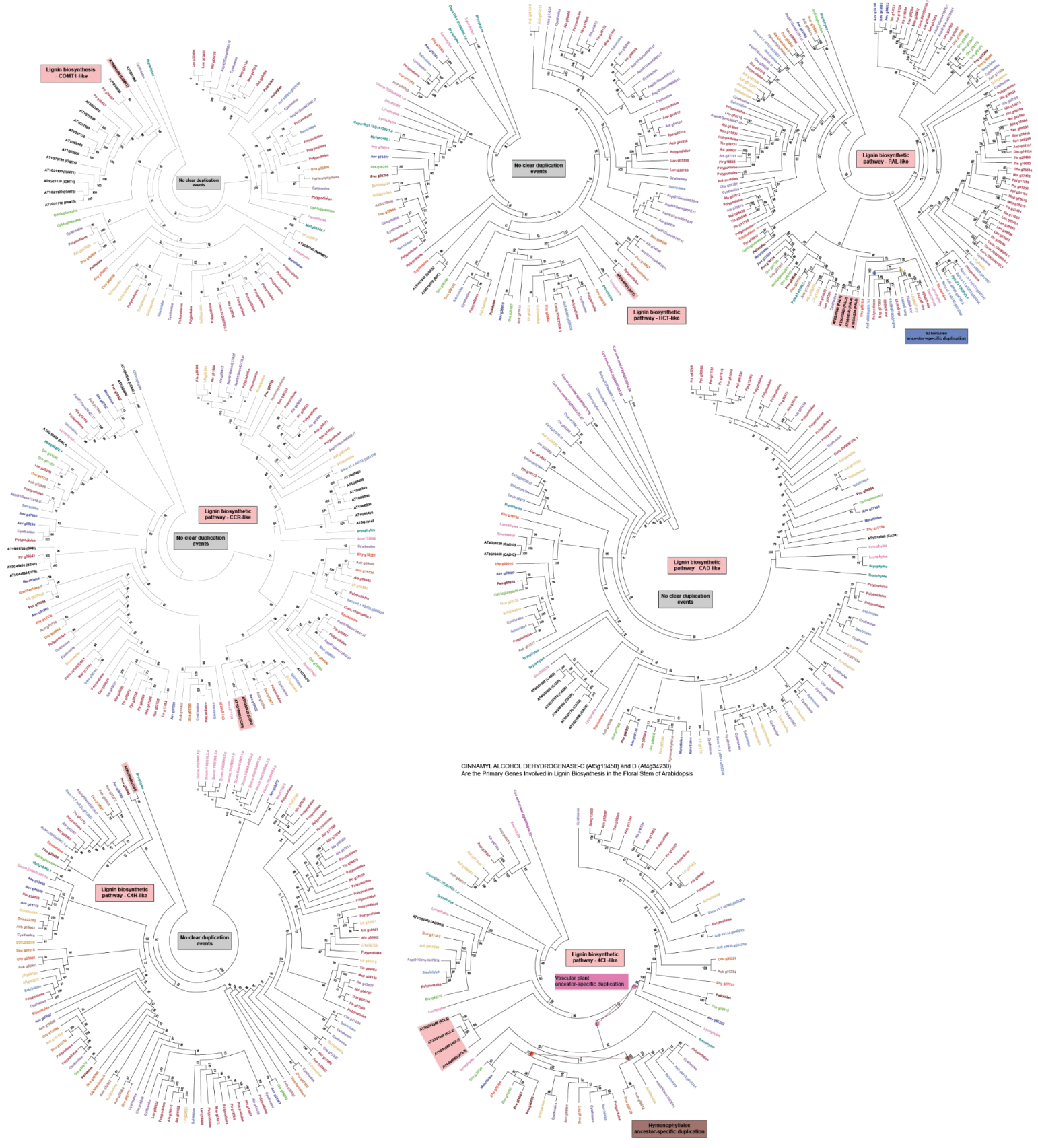
**Figure S16. Gene copy number analysis of hydroxyproline-rich glycoproteins (HRGPs).** Columns represent species, while rows correspond to a given class of HRGP. Red and blue numbers indicate that a given species contains significantly more/less genes than others.



**Figure S17. CoMPP analysis of ferns.** a) The clustermap shows fern samples (rows) and antibodies (columns). The cells indicate the signal scaled from 0 (dark blue) to 1 (bright yellow). b) Pearson Correlation Coefficient (PCC) distribution of CoMPP profiles within (blue) and across (brown) species.



**Figure S18. Phylogenetic analysis of CESA genes.** The blue circles represent bootstrap values (value <50 are not indicated by a circle). The leaf colors represent the different species and orders.



**Figure S19. Phylogenetic analysis of lignin biosynthetic genes.** Any inferred duplication events are indicated.