
Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment

In the format provided by the authors and unedited

1 **Supplementary Notes, Tables, and Figures**

2 **TABLE OF CONTENTS**

3 **1. SAMPLING SPECIES, METADATA, AND DNA AND RNA PREPARATION**5

4 **Supplementary Figure 1.1** Phylogenetic tree of Alismatales, including seagrasses.5

5 **Supplementary Table 1.1** Plant material metadata.6

6 **2. GENOME SEQUENCING AND ASSEMBLIES**7

7 **2.1 Nuclear genomes**..... 7

8 **Supplementary Table 2.1.1** Genomic libraries included in each seagrass genome assembly and their

9 respective assembled sequence coverage levels in the final release.7

10 **Supplementary Figure 2.1.1** Genome assembly pipeline used for *T. testudinum*, *P. oceanica* and *C. nodosa* .7

11 **Supplementary Table 2.1.2** Summary statistics of the initial output of the primary RACON polished HiFiAsm

12 assembly.8

13 **Supplementary Note 2.1** Final Primary assemblies for main and alternate haplotypes.9

14 **Supplementary Table 2.1.3** Final summary primary and alternate assembly statistics for each chromosome-

15 scale assembly (see for further details on *Zostera marina* v3.1).10

16 **Supplementary Figure 2.1.2** Distribution of the genomic features for the 26 largest scaffolds of *P.*

17 *acutifolius*.11

18 **Supplementary Table 2.1.4.** Primary genome assembly, annotation statistics and BUSCO completeness

19 assessment of protein coding sequences. See Supplementary Table 2.1.3 for additional details for the

20 alternate haplotypes12

21 **2.2 Chloroplast genomes** 13

22 **Supplementary Note 2.2** Chloroplast genome assemblies and annotations13

23 **Supplementary Figure 2.2.1** The complete *C. nodosa* chloroplast genome.....13

24 **Supplementary Figure 2.2.2** The complete *Z. marina* chloroplast genome.14

25 **Supplementary Figure 2.2.3** The complete *T. testudinum* chloroplast genome.15

26 **Supplementary Figure 2.2.4** The complete *P. oceanica* chloroplast genome.16

27 **2.3 Mitochondrial genomes** 16

28 **Supplementary Note 2.3** Mitochondrial genome assemblies and annotations16

29 **Supplementary Figure 2.3** *C. nodosa* mitochondrial chromosomes according to genome recombination

30 activity.....17

31 **2.4 Nuclear-mitochondria and nuclear-chloroplast transfer** 18

32 **Supplementary Note 2.4** Nuclear-mitochondria (NUMTs) and nuclear-chloroplast (NUPTs) integrants.18

33 **Supplementary Table 2.4** NUMTs and NUPTs.....18

34 **3. GENOME ANNOTATION** 19

35 **3.1 Non-protein coding RNA annotations** 19

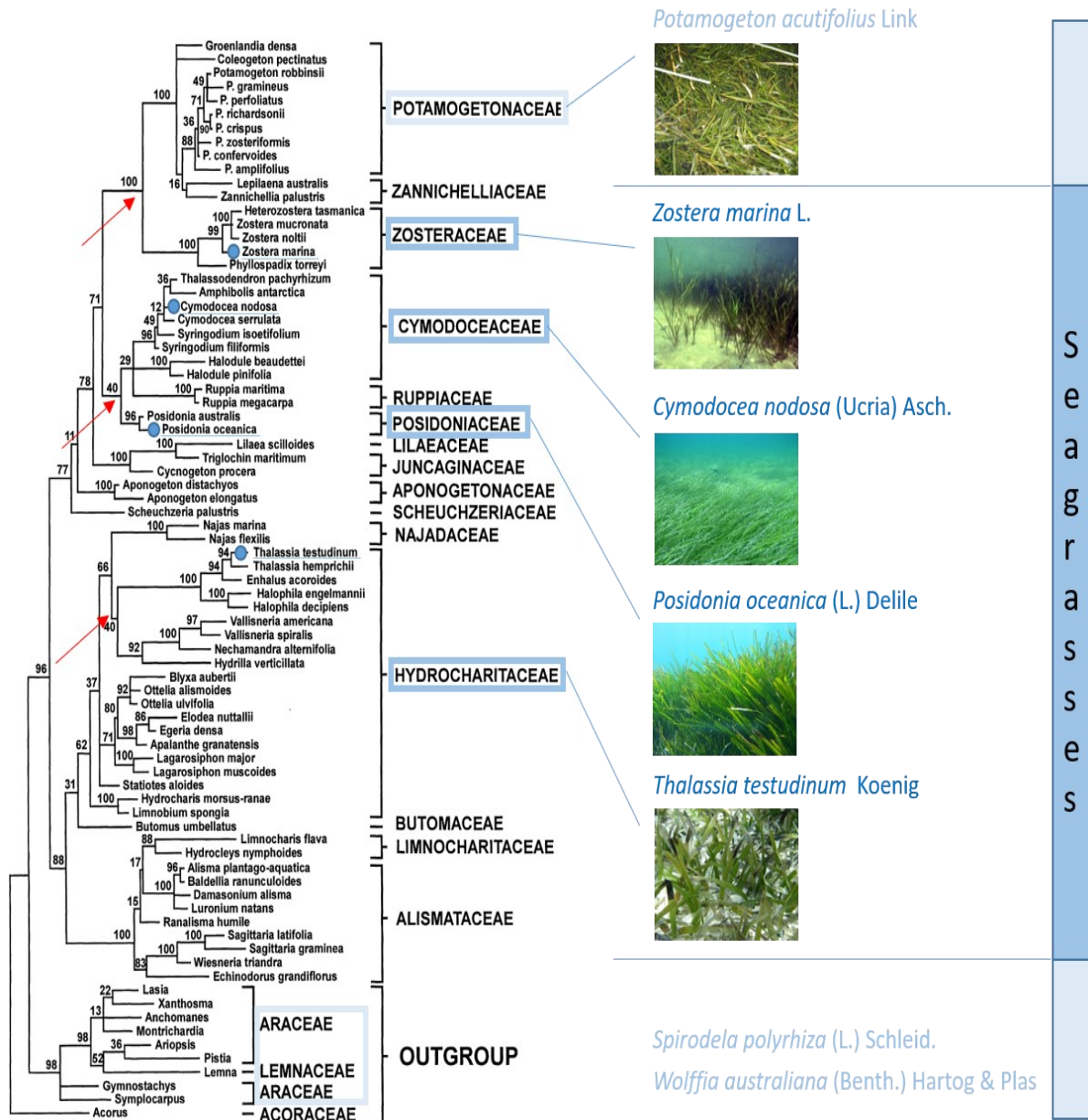
36 **Supplementary Note 3.1** rRNA, tRNA and snRNAs19

37	Supplementary Table 3.1 Number of loci of major families of non-protein coding RNAs detected in	
38	seagrasses.	20
39	Supplementary Figure 3.1 Organization of 18S, 5.8S and 28S rRNA repeat units and in the clusters on	
40	seagrasses chromosomes.	21
41	3.2 Transcriptome libraries, sequencing, and assembly	22
42	Supplementary Note 3.2 Transcriptome libraries	22
43	Supplementary Table 3.2.1 Transcriptome sequencing data for <i>Thalassia testudinum</i>	22
44	Supplementary Table 3.2.2 Transcriptome sequencing data for <i>Posidonia oceanica</i>	23
45	Supplementary Table 3.2.3 Transcriptome sequencing data for <i>Cymodocea nodosa</i>	23
46	Supplementary Table 3.2.4 Transcriptome sequencing data for <i>Potamogeton acutifolius</i>	24
47	4. GENOME EVOLUTION	25
48	4.1 Transposable elements	25
49	Supplementary Table 4.1 Statistics on transposable elements (TE).....	25
50	Supplementary Figure 4.1 Insertion time distributions of LTR/Gypsy and LTR/Copia in <i>T. testudinum</i> , <i>P.</i>	
51	<i>oceanica</i> , <i>C. nodosa</i> and <i>Z. marina</i> . Note differences in the scale of the y-axis. See main text for details.....	25
52	4.2 Identifying Whole Genome Duplications (WGD)	26
53	Supplementary Note 4.2.1 <i>K_s</i> age distributions.....	26
54	Supplementary Figure 4.2.1 <i>K_s</i> distributions for anchor pair duplicates (duplicates laying in duplicated,	
55	colinear blocks) and the whole paranome of four seagrasses, as well as for <i>P. acutifolius</i> and <i>S. polyrhiza</i> ,	
56	generated by the wgd software (see Methods).	27
57	Supplementary Figure 4.2.2 Comparison of <i>Z. marina</i> , <i>P. oceanica</i> , and <i>T. testudinum</i> with a reconstructed	
58	ancestral monocot karyotype (AMK) (Murat et al. 2017).	28
59	Supplementary Figure 4.2.3 Comparison of the ancestral monocot karyotype (AMK) (Murat et al. 2017) with	
60	<i>C. nodosa</i> and comparison of <i>P. oceanica</i> and <i>C. nodosa</i>	29
61	Supplementary Figure 4.2.4 Comparison of <i>P. acutifolius</i> with the ancestral monocot karyotype (AMK)	
62	(Murat et al. 2017), <i>Z. marina</i> , <i>P. oceanica</i> , and <i>C. nodosa</i> , respectively.	30
63	Supplementary Figure 4.2.5 <i>K_s</i> Distributions for paralogs and the whole paranome of four seagrasses and <i>P.</i>	
64	<i>acutifolius</i> generated by KSRATES software.	31
65	Supplementary Note 4.2.2 Gene tree-species tree reconciliation.....	32
66	Supplementary Figure 4.2.6 Bayesian inference of retention rates (q) of 11 hypothetical WGD models in	
67	WHALE (Zwaenepoel and Van de Peer 2019a).	33
68	Supplementary Note 4.2.3 Absolute dating of WGDs.....	34
69	Supplementary Figure 4.2.7 Estimation of the 'absolute age' of the WGT/WGD events in seagrasses and <i>P.</i>	
70	<i>acutifolius</i> by phylogenomic dating.	35
71	Supplementary Table 4.2 The absolute of WGD events taken from literature	36
72	Supplementary Figure 4.2.8 Estimation of the 'absolute age' of seven independent WGD events	
73	experienced by <i>E. guineensis</i> , <i>A. officinalis</i> , <i>R. apiculata</i> , <i>A. marina</i> and <i>U. gibba</i> respectively by	
74	phylogenomic dating of corresponding paralogues.	37
75	4.3 Phylogenetic tree construction and estimation of divergence time	38
76	Supplementary Note 4.3 Species selection and construction of time-calibrated phylogeny.	38

77	5. ADAPTATION TO THE MARINE ENVIRONMENT	39
78	5.1. Use it or lose it	39
79	Supplementary Figure 5.1 Normalized gene counts for each species	39
80	5.2 Pathogen resistance (R-) genes	40
81	Supplementary Note 5.2 Pathogen resistance gene	40
82	Supplementary Table 5.2 <i>NLR</i> gene counts by domain architecture and completeness in seagrasses and <i>P.</i>	
83	<i>acutifolius</i>	41
84	Supplementary Figure 5.2.1 Phylogenetic tree of seagrass <i>NLR</i> genes based on NBS domain.	42
85	Supplementary Figure 5.2.2 Distribution of seagrass <i>NLR</i> genes across chromosomes.	43
86	5.3 Heat Shock factor (HSF) gene family evolution	43
87	Supplementary Note 5.3 <i>HSF</i> gene family	43
88	Supplementary Table 5.3 Average (\pm SD) number of total <i>HSFs</i> and number of <i>HSFs</i> from the three main	
89	classes (<i>HSFA</i> , <i>HSFB</i> and <i>HSFC</i>) in the analyzed plant genomes.	45
90	5.4 Cellular salt tolerance	46
91	Supplementary Figure 5.4.1 Sequence alignment showing amino acid substitutions in regulatory domains of	
92	<i>SOS1</i> orthologs of seagrasses, indicating a diverged but convergent regulation of <i>SOS1/NHX7</i> in these	
93	species.	46
94	Supplementary Figure 5.4.2 Sequence alignment of <i>AKT5/6/1</i> showing the loss of Shaker-type K^+ channels	
95	with a TTGYGD-selectivity filter in all seagrasses.	47
96	5.5 Hypoxia	48
97	Supplementary Figure 5.5.1 Differential expression of <i>ERF-VII</i> s in the four seagrass species.	48
98	Supplementary Figure 5.5.2 Syntenic relationship of genes mentioned in the main text for <i>P. oceanica</i> , <i>T.</i>	
99	<i>testudinum</i> , <i>Z. marina</i> , <i>C. nodosa</i> , and <i>P. acutifolius</i>	49
100	5.6 Light perception, circadian clock, and photosynthetic carbon acquisition	50
101	Supplementary Note 5.6.1 CO_2 -concentrating mechanisms (CCMs) and photosynthetic carbon acquisition 50	
102	Supplementary Table 5.6 Prediction of sub-cellular localizations of α -Carbonic Anhydrases (α -CA) in the	
103	studied seagrass species and <i>P. acutifolius</i>	51
104	52
105	Supplementary Figure 5.6.1 Differential expression of α -CA, β -CA and λ -CA in root, rhizome, flower, and leaf	
106	tissues of the studied seagrass species.	52
107	Supplementary Note 5.6.2 Photosynthesis	53
108	Supplementary Figure 5.6.2 Gene families containing photosystems I and II components and Light-	
109	harvesting chlorophyll protein complex.	54
110	Supplementary Figure 5.6.3 Differential expression of Cyt b6/f complex, <i>LHCB</i> and electron transport genes	
111	in <i>Z. marina</i> and <i>C. nodosa</i>	55
112	Supplementary Note 5.6.3 Light Signaling & Circadian Clock	55
113	Supplementary Figure 5.6.4 Gene families containing photoreceptors and the main integration of light	
114	signalling toolkit genes.	57
115	Supplementary Figure 5.6.5 The N terminus alignment of UVB-Resistance 8.	58
116	Supplementary Figure 5.6.6 Phylogenetic tree of phytochromes obtained from the 84 proteins sequences	
117	included in the orthogroups OG0007303, OG0003273, OG0011336 and OG0026441.	59

118	5.7 NAC transcriptional factors	60
119	Supplementary Note 5.7 NAC transcriptional factors	60
120	Supplementary Figure 5.7 Evolutionary analysis by Maximum Likelihood method of <i>JUB1</i> in seagrasses.....	60
121	5.8 Nitrogen metabolism	61
122	Supplementary Note 5.8 Nitrogen metabolism.....	61
123	5.9 Flower and pollen development	61
124	Supplementary Note 5.9 Flower and pollen development.....	61
125	Supplementary Table 5.9. The MADS-box genes in seagrasses and <i>P. acutifolius</i>	62
126	Supplementary Figure 5.9 Normalized gene copy numbers for flower and pollen development genes and	
127	gene families for 96 species, including 6 genomic data and 90 transcriptomic data	63
128		
129		

1. Sampling species, metadata, and DNA and RNA preparation



131
132

Supplementary Figure 1.1 Phylogenetic tree of Alismatales, including seagrasses.

133
134
135
136
137

The tree is based on a Maximum parsimony analysis of *rbcL* genes, adapted from (Les et al. 1997). The four seagrass species discussed in the present study are marked by blue dots. The one freshwater species sequenced in the present study, *Potamogeton acutifolius*, is missing from the tree, but close relatives are included (sister group of Zosteraceae). Solid red arrows indicate nodes denoting the aquatic-marine splits.

138
139

140 **Supplementary Table 1.1 Plant material metadata.**

141 See Materials and Methods for details on DNA and RNA preparation. Additional information for *Zostera marina*
 142 can be found in (Olsen et al. 2016) and (Ma et al. 2021a).

Species	Code	Location	Date	Collectors	Depth	Lat	Long	Tissue	HMW DNA extraction method /RNA extraction method
<i>Thalassia testudinum</i> Diploid 2N=18	Tt-101	Coconut Grove Dog Park, Crocodile Point, Miami, FL USA	12 June 2019	JL Olsen & JE Campbell	-1 m	40.78735	14.08974	Leaf Leaf Rhizome	CTAB via Arizona Genomics Institute. /NucleoSpin RNA Plant and Fungi Kit (Machery Nagel, USA)
<i>Posidonia oceanica</i> Diploid 2N=20	Po-2	Gulf of Pozzuoli, Napoli, Italy	15 May 2019	G Procaccini	-7 m	40.78735	14.08974	Leaf Leaf Rhizome Root	As above As above
<i>Cymodocea nodosa</i> Diploid 2N=18 Known polyploidy in some populations	Cn-1	Miseno Cape, Napoli, Italy	15 May 2019	G Procaccini & L Marin Guirao	-10 m	40.78396	14.07458	Leaf Leaf Rhizome Root Flowers	As above As above
<i>Potamogeton acutifolius</i> Diploid 2N=26 Known polyploidy in other species Freshwater	Pa	Baláta-lake, Hungary, Somogy county, near Kaszó	18 Sept 2020	A. Mesterházy	-1.2 m	14.31388	17.205	Leaf Leaf Turion Root	Max Planck-Genome with a NucleoBond HMW DNA kit (Macherey Nagel). RNAeasy Plant Kit (Qiagen)

143

144

145 **2. Genome sequencing and assemblies**

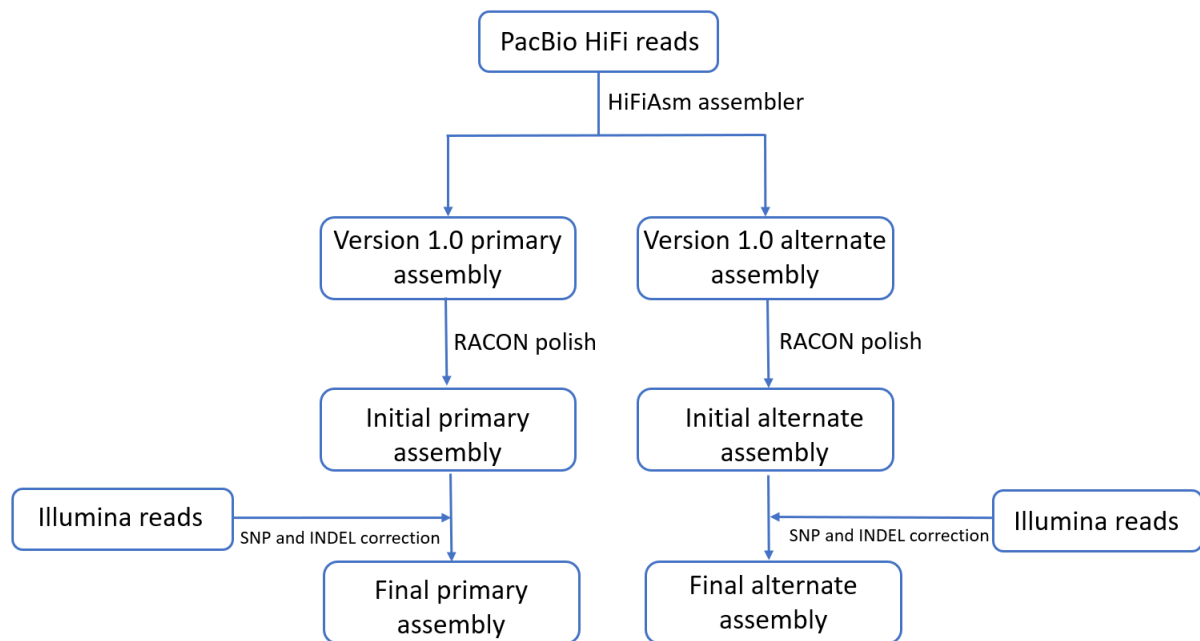
146 **2.1 Nuclear genomes**

147 **Supplementary Table 2.1.1** Genomic libraries included in each seagrass genome assembly and their
 148 respective assembled sequence coverage levels in the final release.

	Library	Sequencing Platform	Average Read/Insert Size	Read Number	Assembled Sequence Coverage (x)
<i>Thalassia testudinum</i>	JDRM	Illumina	400	2,454,318,746	92.04
	IYTB	Illumina/HiC	N/A	1,829,959,374	68.62
	Total Illumina			4,284,278,120	161
		PacBio HiFi	18,712*	11,884,855	51.53
<i>Posidonia oceanica</i>	JDRD	Illumina	400	2,700,014,314	135
	JDMC	Illumina/HiC	N/A	3,843,643,266	192.1
	Total Illumina			6,543,657,580	327
		PacBio HiFi	19,006*	12,018,008	79.44
<i>Cymodocea nodosa</i>	JLSQ	Illumina	400	397,732,894	119.3
	JHNL	Illumina/HiC	N/A	296,170,716	88.9
	Total Illumina			693,903,610	208
		PacBio HiFi	23,472*	1,698,568	79.24
<i>Potamogeton acutifolius</i>	N/A	Illumina/Tell-Seq	N/A	54,401,190	13.4
	N/A	PacBio HiFi	14,000*	1,900,000	43.5

149 *Average read length of PacBio reads

150



151

152

153 **Supplementary Figure 2.1.1** Genome assembly pipeline used for *T. testudinum*, *P. oceanica* and *C.*
 154 *nodosa*

155

156
157
158
159

Supplementary Table 2.1.2 Summary statistics of the initial output of the primary RACON polished HiFiAsm assembly.

The table shows number of contigs and total numbers of assembled base pairs for each set of scaffolds greater than the size listed in the left-hand column.

	Minimum Scaffold Length	Number of Scaffolds	Number of Contigs	Scaffold Size	Basepairs	% Non-gap Basepairs
<i>Thalassia testudinum</i>	5 Mb	14	14	4,757,383,018	4,757,383,018	100.00%
	2.5 Mb	14	14	4,757,383,018	4,757,383,018	100.00%
	1 Mb	17	17	4,761,201,586	4,761,201,586	100.00%
	500 Kb	21	21	4,764,120,924	4,764,120,924	100.00%
	250 Kb	29	29	4,766,780,278	4,766,780,278	100.00%
	100 Kb	152	152	4,784,190,954	4,784,190,954	100.00%
	50 Kb	609	609	4,813,980,144	4,813,980,144	100.00%
	25 Kb	1,965	1,965	4,865,616,055	4,865,616,055	100.00%
	10 Kb	1,987	1,987	4,866,121,113	4,866,121,113	100.00%
	5 Kb	1,987	1,987	4,866,121,113	4,866,121,113	100.00%
	2.5 Kb	1,987	1,987	4,866,121,113	4,866,121,113	100.00%
	1 Kb	1,987	1,987	4,866,121,113	4,866,121,113	100.00%
0 bp	1,987	1,987	4,866,121,113	4,866,121,113	100.00%	
<i>Posidonia oceanica</i>	5 Mb	15	15	3,003,306,251	3,003,306,251	100.00%
	2.5 Mb	15	15	3,003,306,251	3,003,306,251	100.00%
	1 Mb	21	21	3,012,259,886	3,012,259,886	100.00%
	500 Kb	29	29	3,017,253,515	3,017,253,515	100.00%
	250 Kb	52	52	3,024,690,979	3,024,690,979	100.00%
	100 Kb	201	201	3,045,962,177	3,045,962,177	100.00%
	50 Kb	1,061	1,061	3,100,982,328	3,100,982,328	100.00%
	25 Kb	3,383	3,383	3,190,011,804	3,190,011,804	100.00%
	10 Kb	3,468	3,468	3,191,950,648	3,191,950,648	100.00%
	5 Kb	3,468	3,468	3,191,950,648	3,191,950,648	100.00%
	2.5 Kb	3,468	3,468	3,191,950,648	3,191,950,648	100.00%
	1 Kb	3,470	3,470	3,191,952,950	3,191,952,950	100.00%
0 bp	3,470	3,470	3,191,952,950	3,191,952,950	100.00%	
<i>Cymodocea nodosa</i>	5 Mb	21	21	370,838,187	370,838,187	100.00%
	2.5 Mb	22	22	375,803,086	375,803,086	100.00%
	1 Mb	26	26	381,781,051	381,781,051	100.00%
	500 Kb	27	27	382,524,700	382,524,700	100.00%
	250 Kb	46	46	388,996,552	388,996,552	100.00%
	100 Kb	148	148	403,137,124	403,137,124	100.00%
	50 Kb	678	678	437,596,903	437,596,903	100.00%
	25 Kb	1,362	1,362	465,959,990	465,959,990	100.00%
	10 Kb	1,362	1,362	465,959,990	465,959,990	100.00%
	5 Kb	1,362	1,362	465,959,990	465,959,990	100.00%
	2.5 Kb	1,362	1,362	465,959,990	465,959,990	100.00%
	1 Kb	1,362	1,362	465,959,990	465,959,990	100.00%
0 bp	1,362	1,362	465,959,990	465,959,990	100.00%	

160

161 **Supplementary Note 2.1 Final Primary assemblies for main and alternate haplotypes.**

162 The final primary assembly (see Methods) of *T. testudinum* contains 4,261.9 Mb of sequence, consisting of 184
163 contigs with a contig N50 of 371.6 Mb and a total of 98.92% of assembled bases in 9 chromosomes. The final
164 primary assembly of *P. oceanica* contains 2,963.0 Mb of sequence, consisting of 19 contigs with a contig N50 of
165 355.8 Mb and a total of 99.98% of assembled bases in 10 chromosomes. The final primary assembly of *C. nodosa*
166 contains 379.5 Mb of sequence, consisting of 22 contigs with a contig N50 of 21.9 Mb and a total of 100% of
167 assembled bases in 18 chromosomes (Supplementary Table 2.1.3).

168 Correspondingly, the final alternative release of *T. testudinum* contains 4,177.2 Mb of sequence, consisting of
169 3,364 contigs with a contig N50 of 6.1 Mb and a total of 85.18% of assembled bases in 9 chromosomes. The final
170 alternative release of *P. oceanica* contains 2,496.3 Mb of sequence, consisting of 826 contigs with a contig N50
171 of 8.7 Mb and a total of 99.67% of assembled bases in 10 chromosomes. The final alternative assembly for *C.*
172 *nodosa* contains 374.7 Mb of sequence, consisting of 28 contigs with a contig N50 of 20.8 Mb and a total of
173 99.91% of assembled bases in 18 chromosomes (Supplementary Table 2.1.3). The final assembly of *P. acutifolius*
174 is in the Table 1.

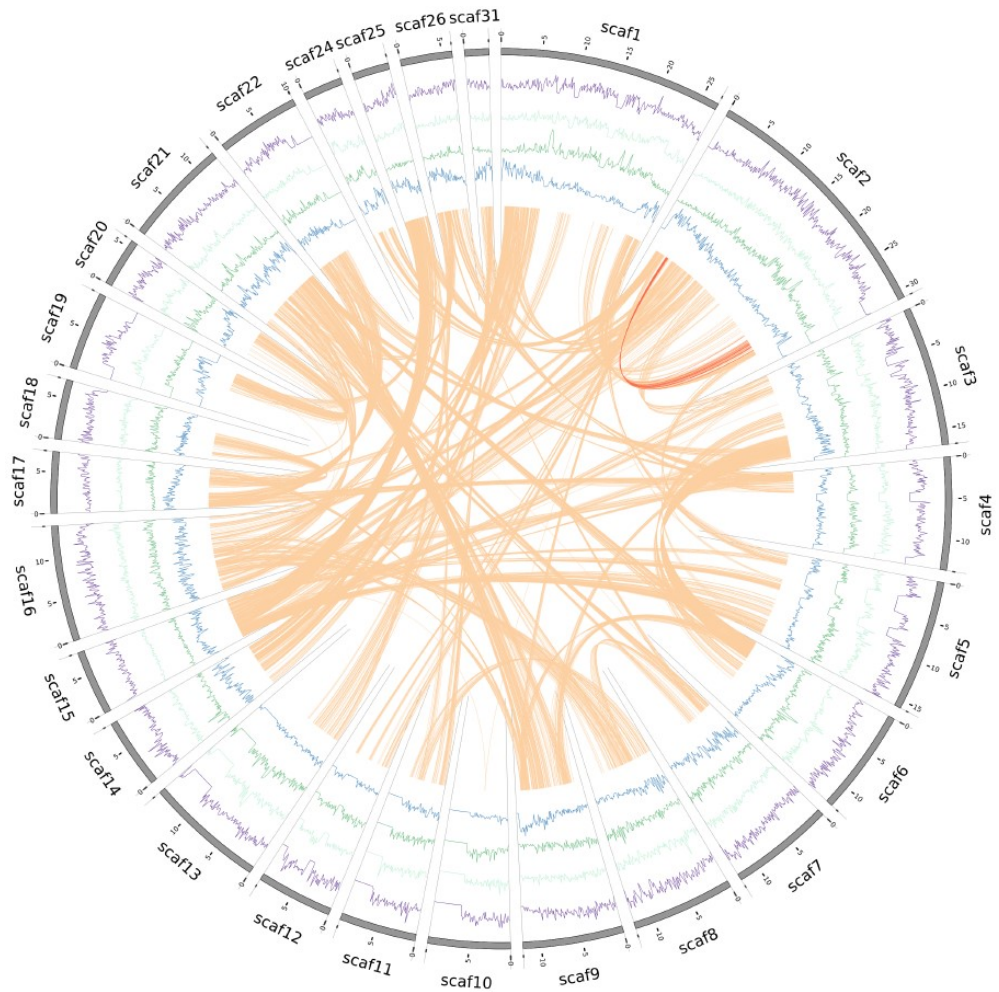
175

176
177

Supplementary Table 2.1.3 Final summary primary and alternate assembly statistics for each chromosome-scale assembly (see for further details on *Zostera marina* v3.1).

Final Primary Assembly Release Haplotype 1	genome release stats:	<i>Thalassia testudinum</i>	<i>Posidonia oceanica</i>	<i>Cymodocea nodosa</i>	<i>Zostera marina</i> v3.1
	Number of chromosomes (1N)	9	10	18	6
	Number of main genome scaffold total:	169	13	18	310
	Number of main genome contig total:	184	19	22	432
	Length of main genome scaffold sequence total: (MB)	4262.1	2963.1	379.6	260.5
	Length of main genome contig sequence total:	4261.9	2963	379.5	259.3
	Main genome scaffold L/N50: (MB)	4/523.9	4/355.8	8/22.6	4/34.6
	Main genome contig L/N50: (MB)	5/371.6	4/355.8	8/21.9	12/7.0
	Number of scaffolds > 50 KB:	49	13	18	217
	% main genome in scaffolds > 50 KB:	99.9	100	100	98.9
Final Alternative Assembly Release Haplotype 2	Number of chromosomes (1N)	9	10	18	N/A
	Number of alternate genome scaffold total:	2264	29	22	N/A
	Number of alternate genome contig total:	3364	826	28	N/A
	Length of alternate genome scaffold sequence total: (MB)	4188.2	2504.3	374.8	N/A
	Length of alternate genome contig sequence total: (MB)	4177.2	2496.3	374.7	N/A
	Genome scaffold L/N50: (MB)	5 / 458.8 Mb	3/352.0	8/22.1	N/A
	Genome contig L/N50: (MB)	162 / 6.1 Mb	85/8.7	8/20.8	N/A
	Number of scaffolds > 50 KB	805	29	22	N/A
	% main genome in scaffolds > 50 KB	98.7	100	100	N/A

178
179



180

181 **Supplementary Figure 2.1.2** Distribution of the genomic features for the 26 largest scaffolds of *P.*
 182 *acutifolius*.

183 Tracks from the inner to outer side correspond to gene density (blue); LTR/Gypsy density (green); LTR/Copia
 184 (orange); DNA transposable elements (pink) and chromosomes (with length in Mb). Curved lines through the
 185 center denote synteny between different scaffolds.

186

187 **Supplementary Table 2.1.4. Primary genome assembly, annotation statistics and BUSCO**
 188 **completeness assessment of protein coding sequences.** See Supplementary Table 2.1.3 for additional
 189 details for the alternate haplotypes.

Statistics	<i>T. testudinum</i>	<i>P. oceanica</i>	<i>C. nodosa</i>	<i>Z. marina v 3.1</i> (<i>Ma et al. 2021b</i>)	<i>P. acutifolius</i>
Assembly					
Haploid-chromosome number	9	10	18	6	13
Genome size, Mb	4261.9	2963.0	379.5	260.5	612
Contig N50, (Mb)	371.6	355.8	21.9	7.0	3.09
Scaffold N50 (Mb)	523.9	355.8	22.6	34.6	4.45
Genome assembly BUSCO					
Complete %	73.3	93.6	92.2	93.4	96.9
single-copy %	72.1	92.2	90.8	90.9	93.7
duplicated %	1.2	1.4	1.4	2.5	3.2
Fragmented %	6.9	2.7	2.7	1.1	1.1
Missing %	19.8	3.7	5.1	5.5	2.0
Annotation					
Protein coding genes	25,665	23,306	20,563	22,256	21,277
Mean gene length, bp	19,151	7,017	5,866	3,237	3,448
Mean CDS length, bp	1,077	1,210	1,207	1,241	1,290
Mean exon length, bp	218	222	214	248	243
Mean exon per gene	4.95	5.46	5.65	5.01	5.3
Mean intron length, bp	4,576	1,303	1,003	499	503
Number of introns > 1 kb %	36.2	24.2	22.5	9.5	9.7
Number of introns > 10 kb %	13.5	2.6	1.4	0.6	0.3
Number of introns > 20 kb %	7.1	0.7	0.2	0.06	0.01
Longest intron, bp	283,604	224,280	89,280	46,497	34,817
Transcriptome support					
TPM > 0 %	87.4	84.7	97.1	91.5	82.7
TPM > 1 %	72.9	70.1	86.6	80.2	75.6
BUSCO					
Complete %	94.2	97.4	95.8	95.7	97.5
single-copy %	92.1	95.2	94.4	93.2	94.6
duplicated %	2.1	2.2	1.4	2.5	2.9
Fragmented %	1.1	0.2	0.3	80.5	0.2
Missing %	4.7	2.4	3.9	3.8	2.3
Functional annotation					
	92.3%	96.8%	91.8%	96.2%	98.9%

190

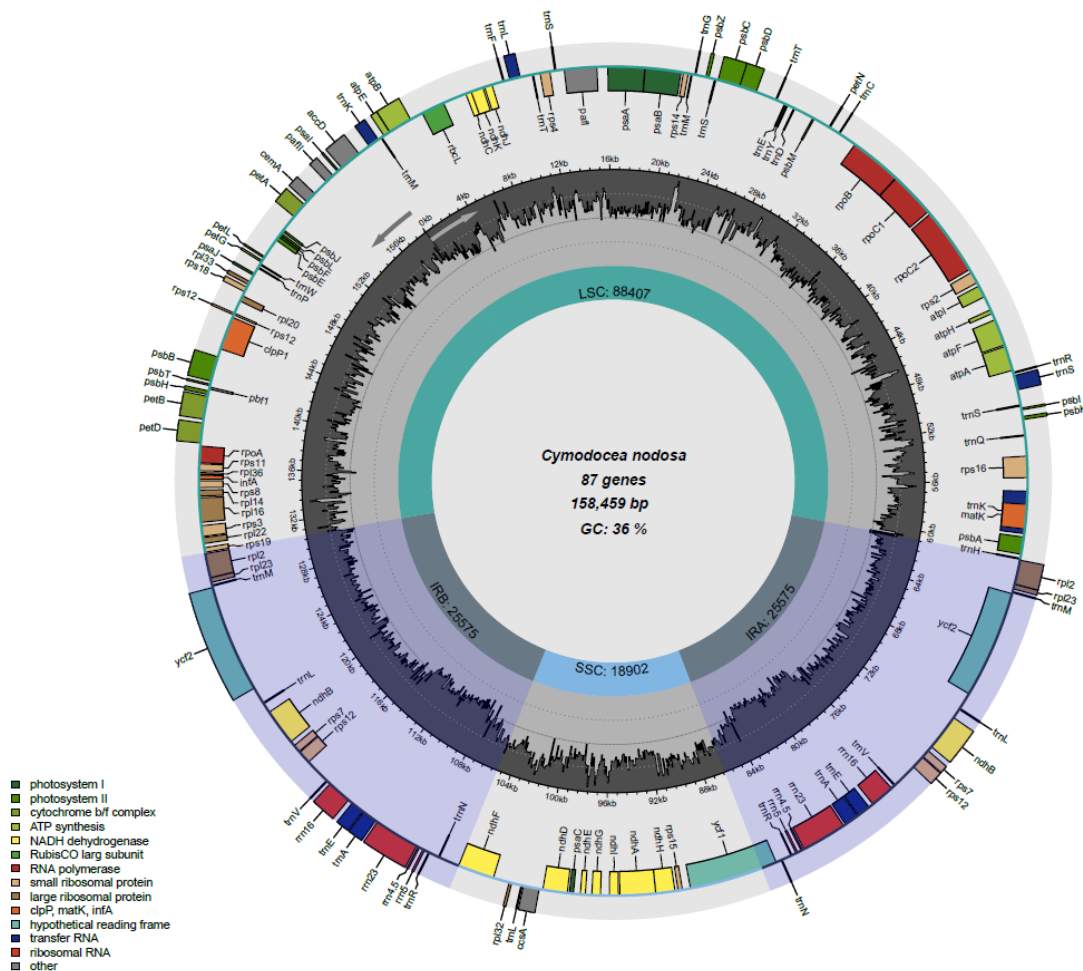
191

192 2.2 Chloroplast genomes

193 **Supplementary Note 2.2 Chloroplast genome assemblies and annotations**

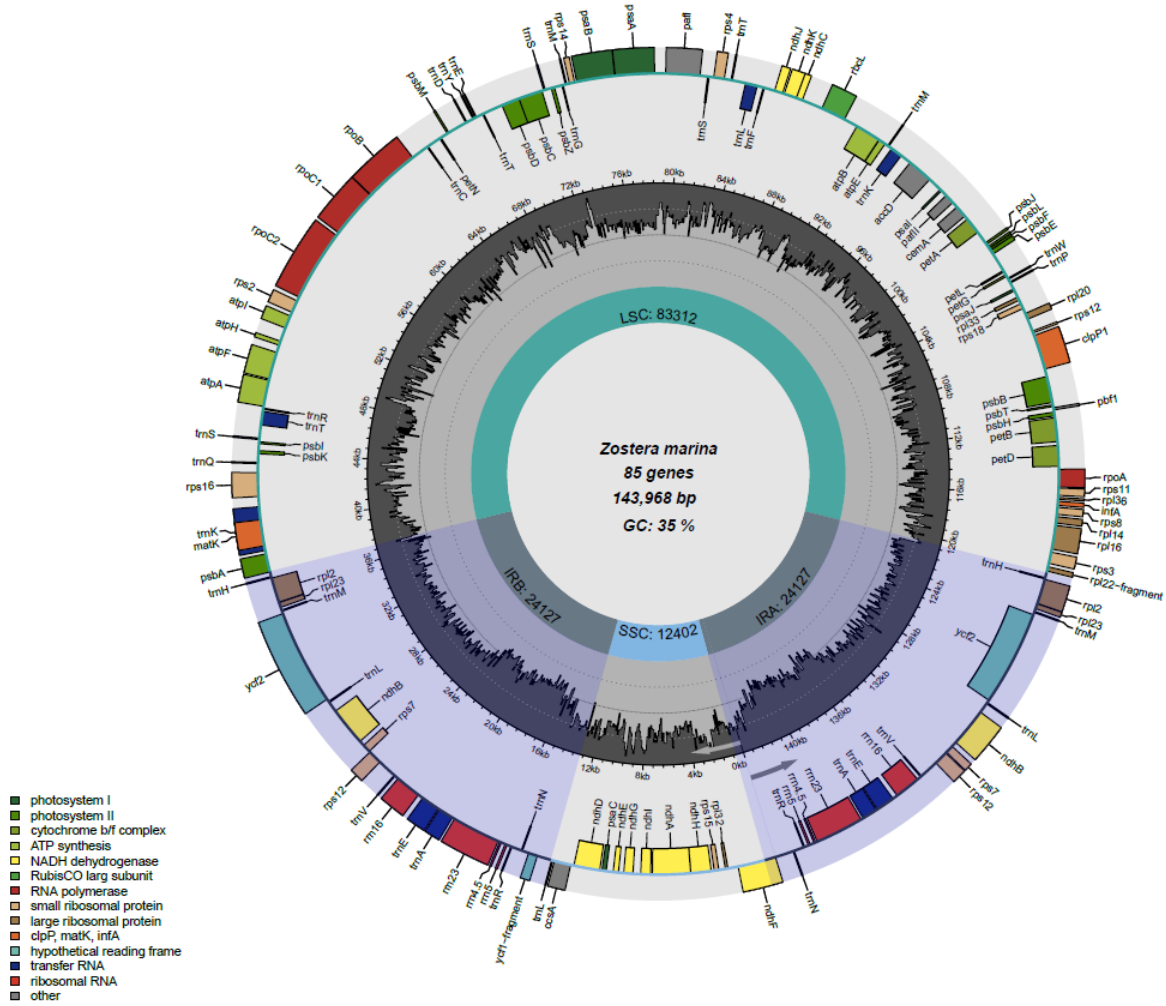
194 Complete chloroplast genomes were assembled *de novo* from Illumina short read data with NOVOPlasty using
195 the *rbcL* gene as a seed (Dierckxsens et al. 2017). *Thalassia testudinum* cpDNA was additionally, manually
196 curated according to the chloroplast derived PacBio contigs from the main genome assembly. The *Z. marina*
197 genome was obtained from (Ma et al. 2021a). All chloroplast genomes were polished with pilon (Walker et al.
198 2014) in addition to manual curation. The gene content of the chloroplast genomes was annotated using the
199 software GeSeq (Tillich et al. 2017) and with Chloë (Zhong 2020). As the preferred annotator for CDS and rRNA,
200 ARAGORN (Laslett and Canback 2004) was used for tRNA annotation. Genes that were much shorter than
201 expected are marked as fragmented. Figures were drawn with Chloroplot (Zheng et al. 2020).

202 We identified 79, 78, 75, and 74 protein-coding genes (only counting genes in the inverted repeats once) in the
203 *C. nodosa*, *Z. marina*, *T. testudinum*, and *P. oceanica* chloroplast genomes, respectively, of which *Z. marina* lost
204 one gene (*rps19*), *T. testudinum* lost four genes (*accD*, *infA*, *ndhB*, and *ndhF*) and *P. oceanica* lost five genes
205 (*ndhG*, *ndhH*, *ndhI*, *ndhJ*, and *ndhK*) (Supplementary Figure 2.2.1 – Supplementary Figure 2.2.4). We note that
206 the chloroplast NADH dehydrogenase complex, encoded by *ndh* genes, has been lost in *P. oceanica* and *T.*
207 *testudinum*. The loss has been proposed to have happened independently in various Alismatales lineages,
208 correlating with the submerged fresh or marine habitat (Ross et al. 2016). Notably, the complex is lost in the
209 two species that demonstrate the highest levels of genome perturbation.

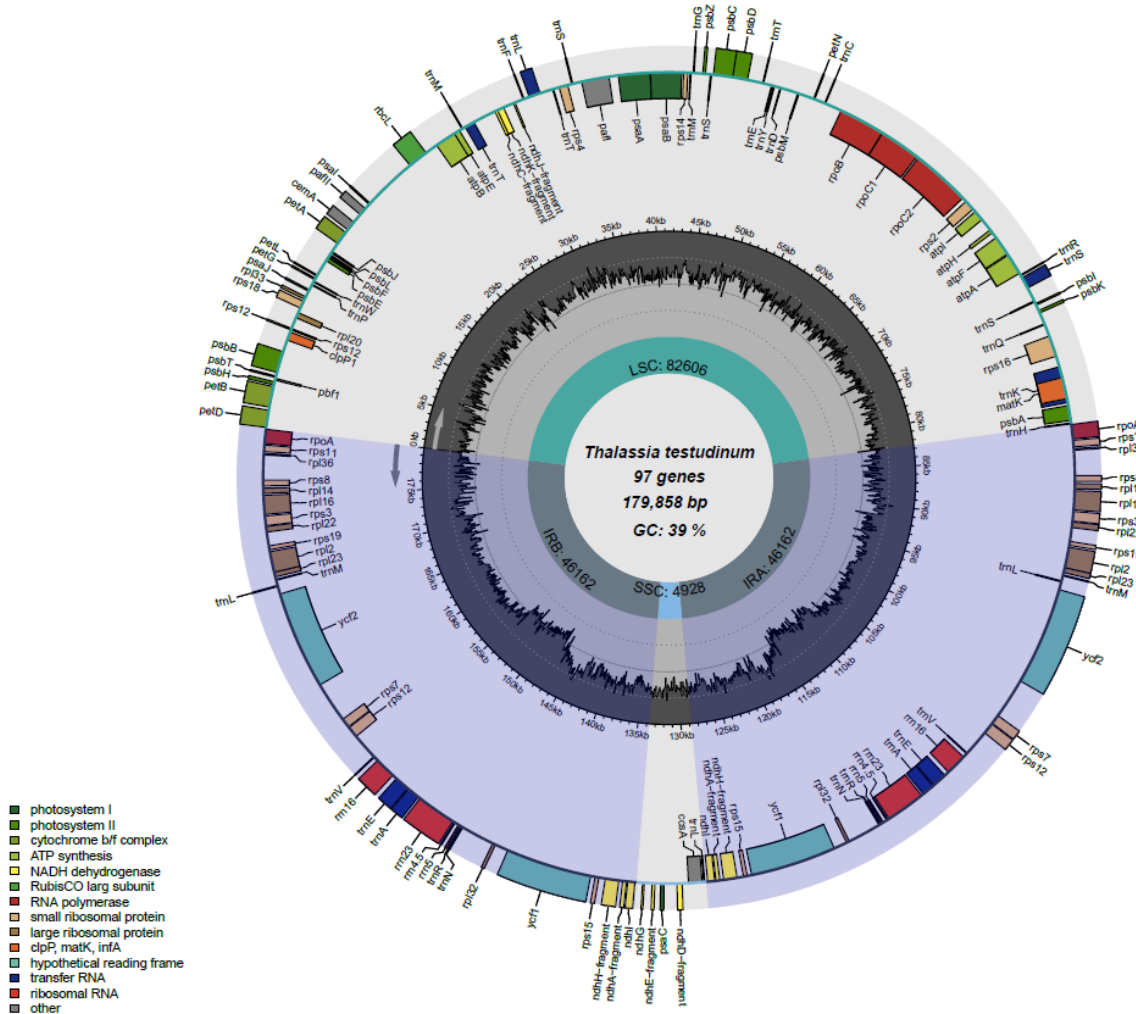


210

211 **Supplementary Figure 2.2.1 The complete *C. nodosa* chloroplast genome.**



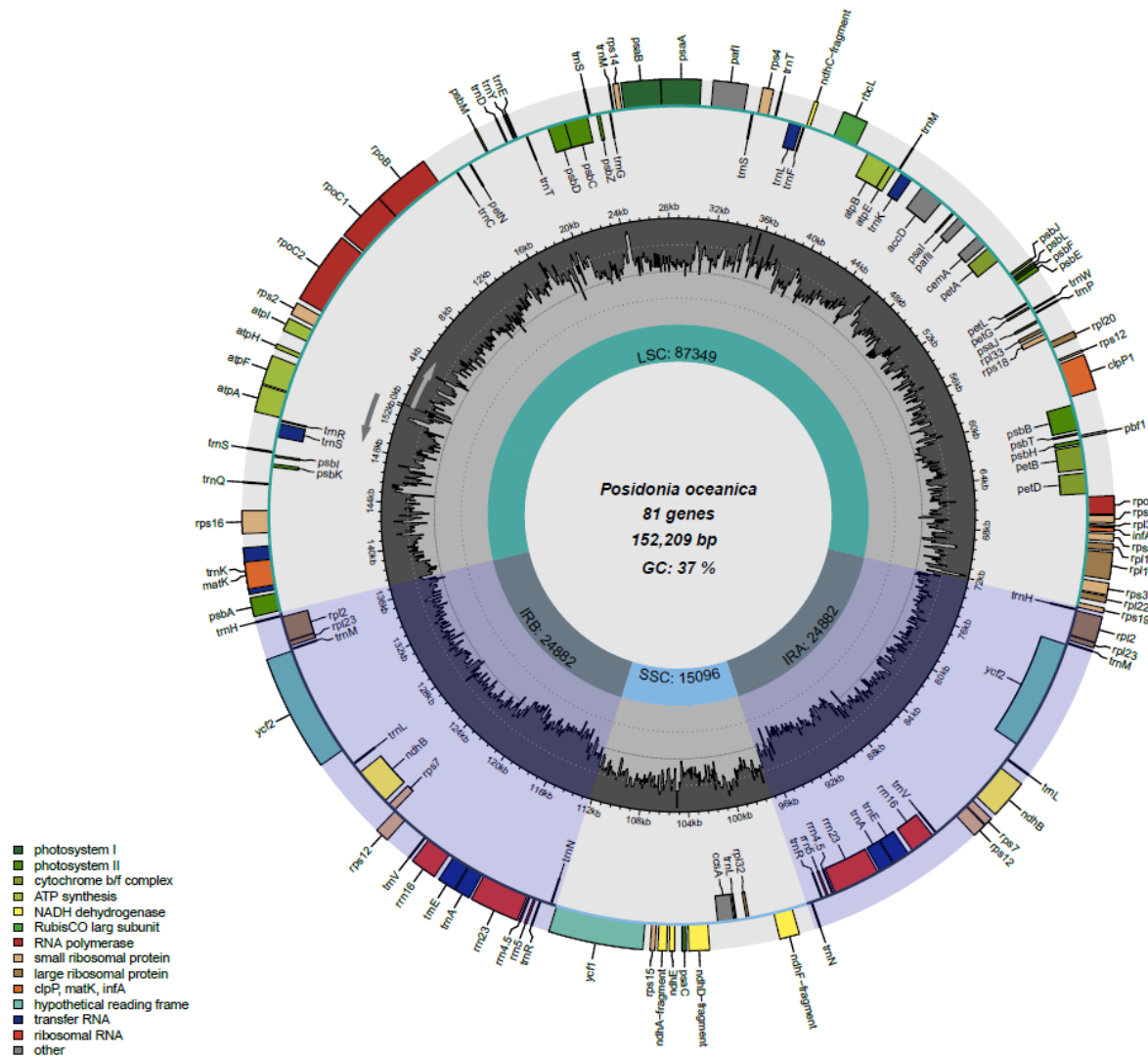
214 **Supplementary Figure 2.2.2** The complete *Z. marina* chloroplast genome.



217

218 **Supplementary Figure 2.2.3** The complete *T. testudinum* chloroplast genome.

219



220

221 **Supplementary Figure 2.2.4** The complete *P. oceanica* chloroplast genome.

222

223 2.3 Mitochondrial genomes

224 **Supplementary Note 2.3** Mitochondrial genome assemblies and annotations

225 Mitochondrial genomes were manually assembled de novo from PacBio contigs containing at least five
 226 mitochondrial genes. All mitochondrial genomes were polished with pilon. Genes were identified by BLAST and
 227 TBLASTN (Altschul et al. 1990) using the mitochondrial gene collection in (Petersen et al. 2017).

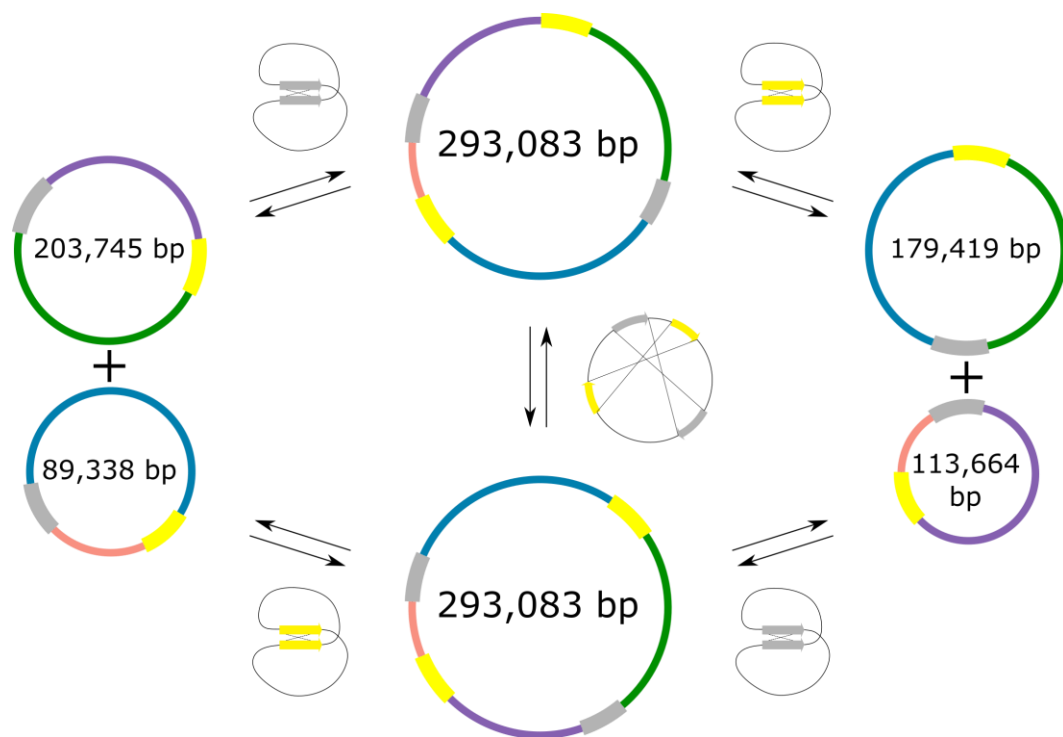
228 Unlike animal mtDNA, plant mtDNA genomes vary enormously in size and do not exist as a single stable circle
 229 but rather as a dynamic combination of linear, branched and small circular loops called isoforms (Morley and
 230 Nielsen 2017). The additional DNA is dominated by repeats and noncoding regions. The various isoforms are due
 231 to recombination (Kozik et al. 2019). Therefore, the classic circular representation is inappropriate. Accordingly,
 232 we found varying degrees of genome size, completeness, and fragmentation in the seagrass mitochondrial
 233 genomes.

234 The *Z. marina* mitochondrial genome is complete (Marina et al. 2023) as is *C. nodosa*. *T. testudinum* is incomplete
 235 and only a single mt-contig was recovered for *P. oceanica*.

236 The complete assembly of the *C. nodosa* mitochondrial genome resulted in one circular chromosome of 293,083
 237 bp with two recombinationally active direct repeats of 2,683 bp and 2,476 bp. Given the possible recombination
 238 events, the genome has six chromosomes (Supplementary Figure 2.3) All PacBio assembly contigs containing
 239 fragments of five or more mitochondrial genes agree with at least one of the six chromosomes. In total, 36 of
 240 44 mitochondrial protein and rRNA coding genes (or their fragments), were detected by BLAST in the assembled
 241 genome, thus providing additional evidence of completeness.

242 The incomplete assembly of the *T. testudinum* mitochondrial genome (not shown) comprises two chromosomes
 243 with no shared DNA between them: one circular of 192,371 bp and the other one of 136,585 bp has an open
 244 loop shape. The open loop shape of the second chromosome is formed by an inverted repeat of 10,932 bp. At
 245 least one additional scaffold of 90,412 bp is likely to have originated from one of the chromosomes although the
 246 exact configuration is unclear; the scaffold has an open loop shape and overlaps with the 192,371 bp circle. The
 247 two chromosomes were further used in the analysis. In total 30 of 44 mitochondrial protein and rRNA coding
 248 genes (or their fragments) are encoded in the two chromosomes. However, a few additional genes detected in
 249 other small scaffolds and in nuclear chromosomes strongly suggests that the assembly status of *T. testudinum*
 250 mitochondrial genome remains incomplete.

251 Only a single mitochondrial contig (132,748 bp) was found for *P. oceanica* and was used for further analysis (not
 252 shown). This contig contains 20 of 44 mitochondrial protein and rRNA coding genes (or their fragments)
 253 confirming the fragmented status of the assembly. Additional mitochondrial genes were detected on nuclear
 254 chromosomes.



255

256 **Supplementary Figure 2.3** *C. nodosa* mitochondrial chromosomes according to genome
 257 recombination activity.

258 See text for explanation and discussion. The two yellow and two grey segments represent two pairs of 2,683 bp
 259 and 2,476 bp respectively. Other colored segments represent parts of the mitogenome that remain stable during
 260 recombination. The two smaller circles (a.k.a subgenomes) are formed when one repeat pair recombines, while
 261 the mirrored conformation is the result of recombination via both repeat pairs simultaneously.

262 **2.4 Nuclear-mitochondria and nuclear-chloroplast transfer**

263 **Supplementary Note 2.4 Nuclear-mitochondria (NUMTs) and nuclear-chloroplast (NUPTs) integrants.**

264 Nuclear-mitochondria (NUMTs) and nuclear-chloroplast (NUPTs) integrants were identified by BLAST using the
 265 parameters suggested in (Smith et al. 2011). Shared-overlapped regions in nuclear chromosomes were joined
 266 into “joined NUMTs” and “joined NUPTs”. Because of the repeats in organellar genomes and non-linear
 267 organellar genome assemblies, the same chromosome region might appear multiple time in the BLAST output.
 268 Therefore, we joined overlapping blast hits into one and called them “joined NUMTs” and “joined NUPTs”. This
 269 way we ensure that we do not overestimate the number (and total length) of the NUMTs and NUPTs in the
 270 nuclear chromosomes.

271 The integrated organellar DNAs within the nuclear genome have been shown to play important roles. Several
 272 processes can enhance DNA transfer including biotic and abiotic stress, increased organelle copy number or
 273 large gene-free regions composed of multiple repeats (Zhao et al. 2019; Ma et al. 2020; Zhang et al. 2020).

274 We assessed the intensity of intracellular DNA transfer of shared DNA segments between the nucleus and
 275 mitochondria (NUMTs), and the nucleus and chloroplasts (NUPTs). *Thalassia testudinum* revealed large
 276 uninterrupted NUMTs that point towards relatively recent mitochondrial-DNA transfer, whereas *Z. marina* has
 277 very few (Supplementary Table 2.4). The recent LTR/Gypsy burst in *T. testudinum* (Supplementary Figure 4.1) is
 278 probably the main cause for *T. testudinum*'s extreme genome size and intron expansion. It is further correlated
 279 with increased intracellular DNA transfer from organellar to nuclear genome. Transposable elements have been
 280 proposed to contribute to the post-insertion dynamics of the transferred DNA rather than the insertion rate
 281 (Michalovova et al. 2013). Thus, multiple insertions of organellar DNA may be another consequence of a general
 282 genome instability caused by TE expansion.

283 **Supplementary Table 2.4 NUMTs and NUPTs.**

284 Intracellular DNA transfer and the relative age of shared DNA segments between the nucleus and mitochondria
 285 (NUMTs), and the nucleus and chloroplasts (NUPTs). Recent insertions are longer (e.g., *T. testudinum*). Over time,
 286 segments become shorter (e.g., *Z. marina*).

Species	Mitogenome status	Total length	Max joined NUMT length	Mean joined NUMT length	Joined NUMT >10000bp	Percent shared positions with nucleus
<i>Cymodocea nodosa</i>	complete	292,353	71,177	796	3	59
<i>Posidonia oceanica</i>	fragment	132,748	49,857	880	12	99
<i>Thalassia testudinum</i>	incomplete	328,953	95,806	1,275	108	100
<i>Zostera marina v3.1</i>	complete	187,048	9,623	391	0	26
Species	Chloroplast genome status	Total length	Max joined NUPT length	Mean joined NUPT length	Joined NUPT >3000bp	Percent shared positions with nucleus
<i>Cymodocea nodosa</i>	complete	158,459	11,621	387	18	94
<i>Posidonia oceanica</i>	complete	152,209	30,959	289	29	99
<i>Thalassia testudinum</i>	complete	180,136	129,692	527	181	100
<i>Zostera marina v3.1</i>	complete	143,968	6,039	396	1	44

287

288 3. Genome annotation

289 3.1 Non-protein coding RNA annotations

290 **Supplementary Note 3.1** rRNA, tRNA and snoRNAs

291 The prediction of non-protein coding RNA families (i.e., rRNAs, tRNA and snoRNAs) in *Z. marina*, *C. nodosa*, *P.*
292 *oceanica*, *T. testudinum* and *P. acutifolius* highlighted an overall expansion in the number of loci in *T. testudinum*
293 when compared to other species (Supplementary Table 3.1). The 5S rRNA (119 nucleotides in length), the minor
294 component of the large subunit of the ribosome, was detected in 10,923 distinct loci in the *T. testudinum*
295 genome, and in 3,566 loci in *P. acutifolius*, 2,605 in *C. nodosa*, 730 in *P. oceanica*, and 364 in *Z. marina*. Clusters
296 of the repetitive units of the large and small ribosomal subunits, organized by the 18S followed by the 5.8S and
297 the 25S rRNAs, of 154, 3,401 and 1,851 nucleotides in length, respectively, were found on a single chromosome
298 in *C. nodosa* (chr 18) and *P. oceanica* (chr 6), on three different chromosomes in *T. testudinum* (chr4, chr5 and
299 chr8) and in different scaffolds in *Z. marina* and *P. acutifolius* (Supplementary Table 3.1 and Supplementary
300 Figure 3.1). All the repetitive units show an almost conserved organization and the length of the two internal
301 transcribed sequences (ITSs) within a species is conserved too, even in the repeated loci along the three different
302 chromosomes in *T. testudinum*.

303 Of note, *C. nodosa* does not show duplications of the represented RNA families, while an evident expansion, also
304 accompanied by the snoRNAs class, is revealed for *T. testudinum* (Supplementary Table 3.1). This expansion
305 appears also evident in *P. acutifolius* scaffolds, although the scaffold nature of this genome did not allow to
306 confirm the real trend. Also transfer RNA (tRNA) sequences, with a length of 71 nt, were more abundant in *T.*
307 *testudinum* (1,200 hits) and *P. acutifolius* (4,954), compared to *C. nodosa* (228), *P. oceanica* (367), and *P.*
308 *oceanica* (478) species. snoRNAs, the RNA class that guides chemical modifications of other rRNAs, show a strong
309 overrepresentation in *T. testudinum* (12,846 hits), compared to the other species (224 in *P. acutifolius*, 99 in *C.*
310 *nodosa*, 238 in *P. oceanica*, 155 in *Z. marina*), although the number of loci detected in the *P. acutifolius* are lower
311 than *P. oceanica* (Supplementary Table 3.1). This will require deeper investigations when the *P. acutifolius*
312 chromosomes will be defined.

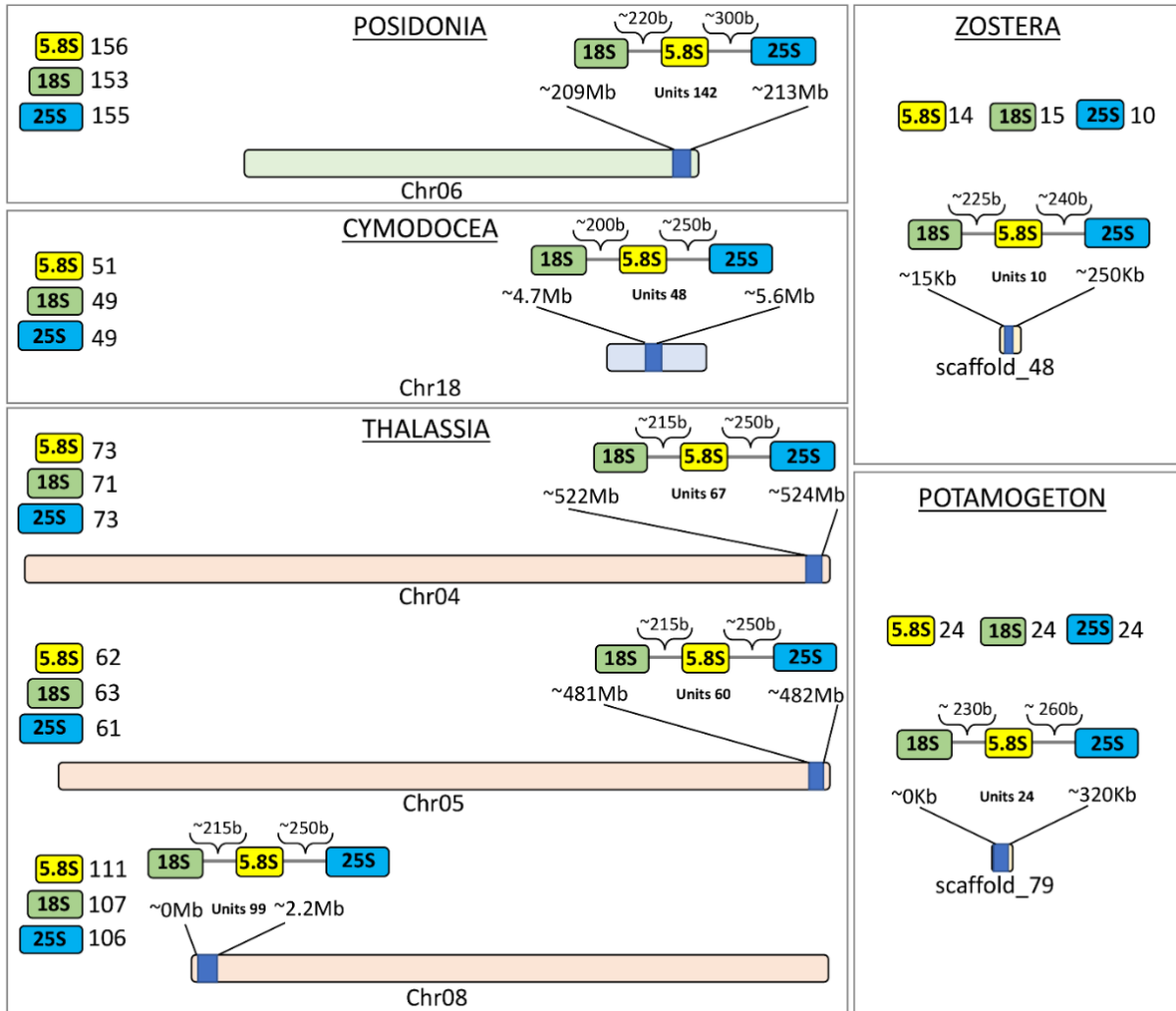
313

314
315

Supplementary Table 3.1 Number of loci of major families of non-protein coding RNAs detected in seagrasses.

Species	Genomic element	5S	5.8S	LSU_EUK(28S)	SSU_EUK(18S)	tRNA	snoRNA
<i>Posidonia oceanica</i>	Chr1	3	-	-	-	68	57
	Chr2	3	-	-	-	51	33
	Chr3	2	-	-	-	27	24
	Chr4	-	-	-	-	32	16
	Chr5	18	-	-	1	53	21
	Chr6	-	160	155	153	43	19
	Chr7	2	1	1	-	35	7
	Chr8	1	-	-	-	12	15
	Chr9	2	-	-	-	19	19
	Chr10	699	1	1	-	20	25
	Scaffolds	-	-	-	-	7	2
	TOTAL		730	162	157	154	367
<i>Cymodocea nodosa</i>	Chr1	2,195	-	-	1	8	10
	Chr2	17	-	-	-	20	6
	Chr3	21	-	-	-	19	10
	Chr4	20	-	-	-	23	4
	Chr5	15	-	-	-	22	11
	Chr6	18	-	-	-	8	9
	Chr7	17	-	-	-	14	6
	Chr8	19	1	-	-	21	3
	Chr9	17	-	-	-	6	3
	Chr10	13	-	-	-	13	1
	Chr11	16	-	-	-	16	4
	Chr12	17	-	-	-	1	4
	Chr13	146	-	-	-	3	6
	Chr14	23	-	-	-	13	3
	Chr15	16	-	-	-	7	3
	Chr16	15	-	-	-	6	7
	Chr17	13	-	-	-	15	6
	Chr18	7	51	49	49	13	3
	TOTAL		2,605	52	49	50	228
<i>Zostera marina</i>	Chr1	6	-	-	1	55	14
	Chr2	5	-	-	1	79	7
	Chr3	2	-	-	-	59	33
	Chr4	5	1	1	-	48	39
	Chr5	4	1	-	-	47	30
	Chr6	1	-	-	-	55	20
	Scaffolds	341	120	90	124	135	12
	TOTAL		364	122	91	126	478
<i>Thalassia testudinum</i>	Chr1	27	-	-	1	120	1,619
	Chr2	41	-	-	-	137	1,712
	Chr3	28	-	-	-	94	1,729
	Chr4	33	74	73	73	126	1,553
	Chr5	16	67	61	64	125	1,387
	Chr6	3,435	-	-	3	113	1,459
	Chr7	41	8	5	6	84	1,306
	Chr8	7,294	111	106	107	75	1,189
	Chr9	8	-	-	-	40	885
	Scaffolds	-	-	-	-	286	7
TOTAL		10,923	260	245	254	1,200	12,846
<i>Potamogeton acutifolius</i>	Scaffolds	3,566	1,548	1,665	1,693	4,954	224

316



317

318 **Supplementary Figure 3.1** Organization of 18S, 5.8S and 28S rRNA repeat units and in the clusters on
 319 seagrasses chromosomes.

320 Considering *Zostera marina* and *Potamogeton acutifolius*, the organization of the longest clusters and the
 321 scaffold in which they have been found. The length of the ITSs and the number of repeats in a cluster (units
 322 number) are reported.

323

324 **3.2 Transcriptome libraries, sequencing, and assembly**

325 See Methods and Supplementary Table 1.1 for RNA preparation

326 **Supplementary Note 3.2 Transcriptome libraries**

327 Strand-specific RNASeq library(s) were created and quantified by qPCR. RNA sequencing was performed using
 328 an Illumina instrument (Supplementary Table 3.2.1 – Supplementary Table 3.2.4). Raw fastq file reads were
 329 filtered and trimmed using the JGI QC pipeline resulting in the filtered fastq file. Using BBduk
 330 (<https://sourceforge.net/projects/bbmap/>), raw reads were evaluated for artifact sequence by kmer matching
 331 (kmer=25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads. RNA spike-in
 332 reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred
 333 trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed
 334 (minimum length 25 bases or 1/3 of the original read length - whichever is longer).

335 **Supplementary Table 3.2.1 Transcriptome sequencing data for *Thalassia testudinum***

	Sequencing	Tissue	library ID	Replicate	Number of raw Reads	Number of clean Reads	Number of mapped reads	% of aligned reads
Summary of transcriptome sequencing in <i>Thalassia</i>	RNA-seq (NovaSeq S4)	Rhizome	GZGXW	1	121,598,766	118,237,066	114,398,701	96.75
			GZGXX	2	73,636,900	70,598,156	67,892,538	96.17
			GZGXY	3	132,873,546	129,273,100	124,854,697	96.58
		Leaf	GZGXT	1	93,743,754	90,902,870	87,180,663	95.91
			GZGXU	2	62,515,064	61,189,816	58,999,410	96.42
		Sequencing	Tissue	library ID	Replicate	Number of full-length ccs reads	Number of high-quality isoforms	Number of mapping reads
	Iso-seq (SEQUELII)	Leaf	GYNYG	1	3,645,896	217,994	347,411	98.4
			GYNYH	2	3,267,856	135,187		
		Rhizome	GNYNO	1	3,002,957	177,124	351,956	99.1
			GNYNY	2	2,801,968	178,044		

336

Supplementary Table 3.2.2 Transcriptome sequencing data for *Posidonia oceanica*

	Sequencing	Tissue	library ID	Replicate	Number of raw Reads	Number of clean Reads	Number of mapping reads	% of aligned reads
Summary of transcriptome sequencing in <i>Posidonia</i>	RNA-seq (NovaSeq S4)	Leaf	GZGXH	1	118,219,220	110,207,116	106,925,423	97.02
			GYOZB	2	103,334,962	89,188,206	85,643,795	96.03
			GZGXN	3	112,093,828	106,876,636	103,508,597	96.85
		Rhizome	GZGXO	1	90,749,690	87,505,084	84,244,810	96.27
			GZGXP	2	97,902,404	95,308,522	92,137,451	96.67
	Sequencing	Tissue	library ID	Replicate	Number of full-length ccs reads	Number of high-quality isoforms	Number of mapping reads	% of aligned reads
	Iso-seq (SEQUELII)	Leaf	GZHTC	1	1,852,186	93,731	265,793	99.7
			GZHTB	2	2,542,984	172,871		
		Rhizome	GZHTH	1	2,525,643	183,914	335,745	99.6
			GZHTG	2	1,941,033	152,859		

Supplementary Table 3.2.3 Transcriptome sequencing data for *Cymodocea nodosa*

	Sequencing	Tissue	library ID	Replicate	Number of raw Reads	Number of clean Reads	Number of mapping reads	% of aligned reads
Summary of transcriptome sequencing in <i>Cymodocea</i>	RNA-seq (NovaSeq S4)	Leaf	GZGWX	1	229,014,654	221,616,172	215,735,219	97.35
			GZGWY	2	85,687,350	81,851,552	78,999,864	96.52
		Rhizome	GZGWZ	1	152,179,412	142,836,610	138,185,544	96.74
		Root	GZGXA	1	77,449,636	75,252,242	72,717,889	96.63
			GZGXB	2	103,516,046	102,007,494	99,073,508	97.12
		Flower	GZGXC	1	167,193,082	157,226,850	153,290,632	97.50
			GZGXC	2	170,015,134	160,442,432	156,397,982	97.48
	Sequencing	Tissue	library ID	Replicate	Number of full-length ccs reads	Number of high-quality isoforms	Number of mapping reads	% of aligned reads
	Iso-seq (SEQUELII)	Mixed pool of Rhizome, root and flower	GWWWG	1	1,103,846	130,315	568,012	98.5
			GWWWG	2	2,318,964	171,811		
		Leaf	GZHTN	1	2,102,451	157,082	117,547	
			GZHTO	2	1,982,756	117,547		

340

Supplementary Table 3.2.4 Transcriptome sequencing data for *Potamogeton acutifolius*.

	Sequencing	Tissue	library ID	Replicate	Number of raw reads	Number of clean reads	Number of mapping reads	% of aligned reads
Summary of transcriptome sequencing in <i>Potamogeton</i>	RNA-seq (NovaSeq 6000 S4)	Leaf	PALF	1	85,063,302	85,063,302	35,512,097	41.7
		Root	PARO	1	68,366,930	68,366,930	35,807,278	52.4
		Turion	PATU	1	64,107,268	64,107,268	56,978,015	88.9

341

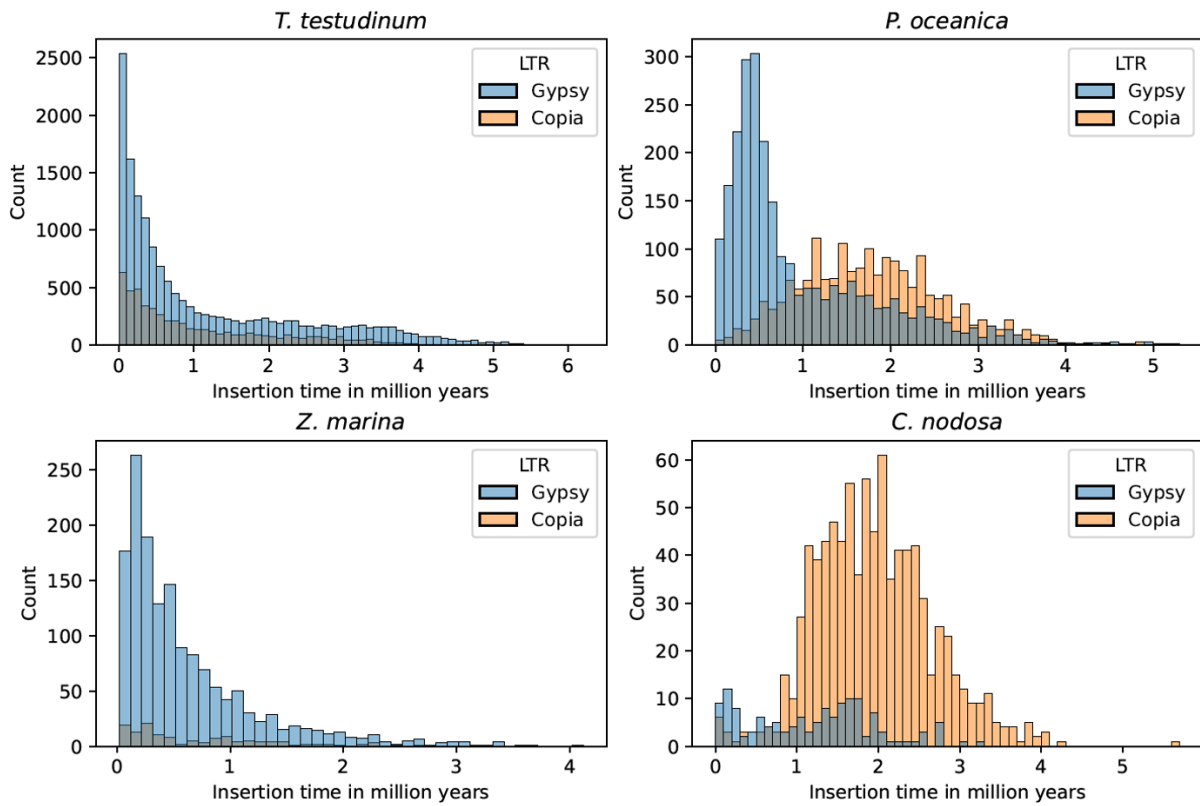
342 4. Genome Evolution

343 4.1 Transposable elements

344 **Supplementary Table 4.1** Statistics on transposable elements (TE).

TE Statistics	<i>T. testudinum</i>	<i>P. oceanica</i>	<i>C. nodosa</i>	<i>Z. marina</i>	<i>P. acutifolius</i>
Overall TE content (%)	87.36	85.18	64.65	65.57	40.76
Overall LTR (%)	72.27	65.89	45.72	41.72	12.41
LTR/Gypsy (%)	63.18	57.8	17.43	32.11	3.39
LTR/Copia (%)	8.28	4.09	28.29	9.49	6.11
LINE (%)	5.06	2.63	2.38	4.05	0.59
SINE (%)	0.03	0.23	0.00	0.09	0.23
DNA transposon (%)	7.78	8.39	2.93	5.71	7.14
Unclassified (%)	2.21	2.76	13.61	14.00	18.95

345



Supplementary Figure 4.1 Insertion time distributions of LTR/Gypsy and LTR/Copia in *T. testudinum*, *P. oceanica*, *C. nodosa* and *Z. marina*. Note differences in the scale of the y-axis. See main text for details.

346

347 4.2 Identifying Whole Genome Duplications (WGD)

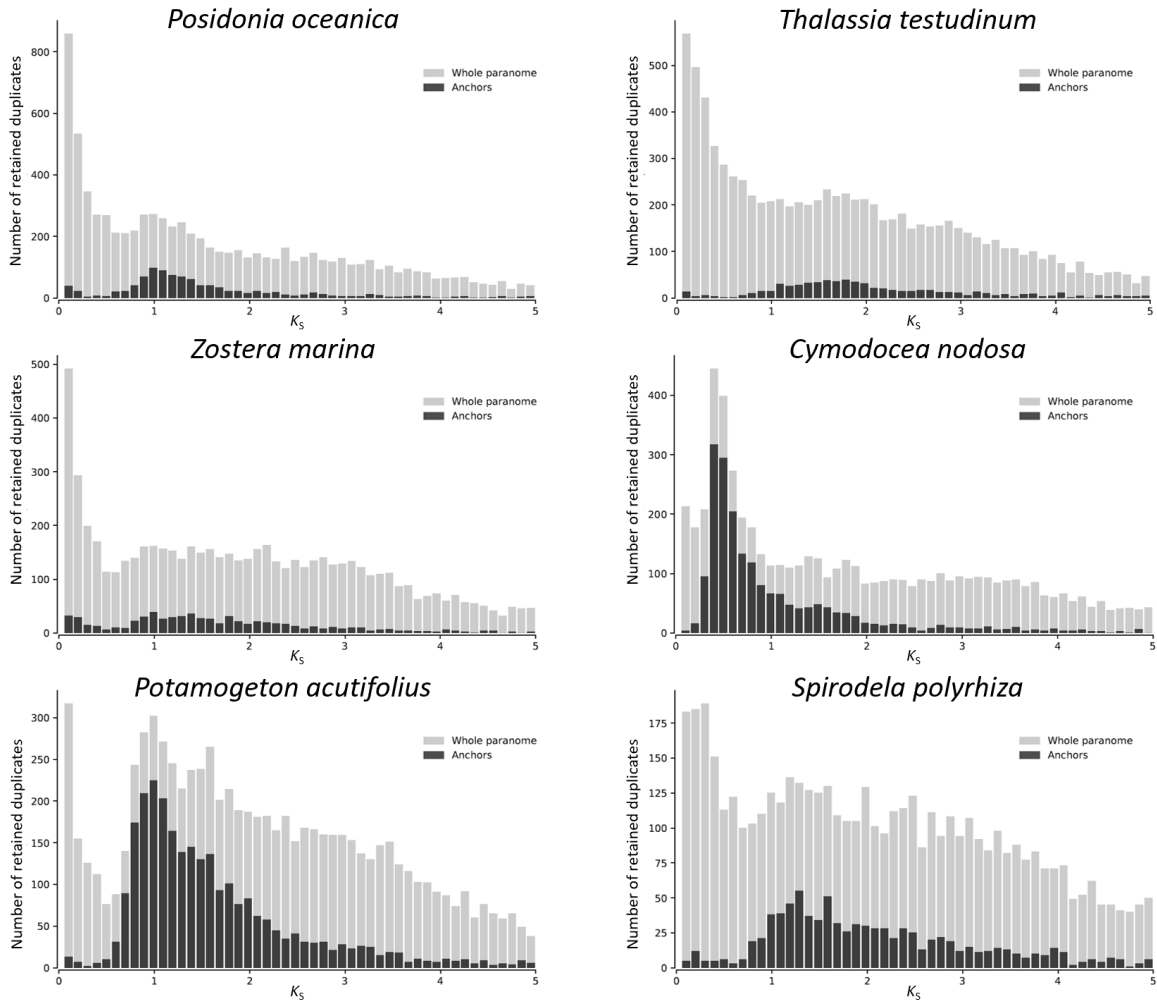
348 **Supplementary Note 4.2.1** K_s age distributions

349 K_s age distribution analysis was performed using the wgd package (Zwaenepoel and Van de Peer 2019b). The
350 paranome (the entire collection of duplicated genes) was obtained with 'wgd mcl' using all-against-all BlastP and
351 MCL clustering. Anchor pairs (i.e., paralogous genes lying in collinear or syntenic regions of the genome) were
352 obtained using i-ADHoRe (Simillion et al. 2008), employing the default settings in 'wgd syn'. K_s distribution
353 analysis was also performed using the KSRATES software (Sensalari et al. 2021), which locates ancient
354 polyploidization events with respect to speciation events within a phylogeny. It compares paralog and ortholog
355 K_s distributions, while correcting for substitution rate differences across the involved lineages. First, an all-
356 versus-all amino acid level similarity search for the set of protein-coding sequences was conducted using BLAST
357 (v2.6.0+) with an E-value cut-off of $1e-10$. The resulting sequence similarity graph was clustered by the Markov
358 clustering algorithm mcl (v10-201) (Dongen 2008) with default inflation factor to identify the paralogous gene
359 families. Second, a codon-level, multiple sequence alignment (MSA) was obtained by inferring an amino acid
360 MSA using the MUSCLE (v3.8.31) (Edgar 2004) under default parameters, which was then back-translated to a
361 codon-level nucleotide MSA. The gene families which had >200 members or an MSA length less <100 amino
362 acids were filtered out. A maximum likelihood estimate of each gene pair in the family (for the pairwise
363 synonymous distance (K_s)) was obtained using the CODEML program or the PAML (Yang 2007b) package (v4.9j).
364 Third, an estimate of the phylogenetic tree topology of each paralogous gene family was obtained using Fasttree
365 (v2.1.7) (Price et al. 2010) and rooted using midpoint rooting.

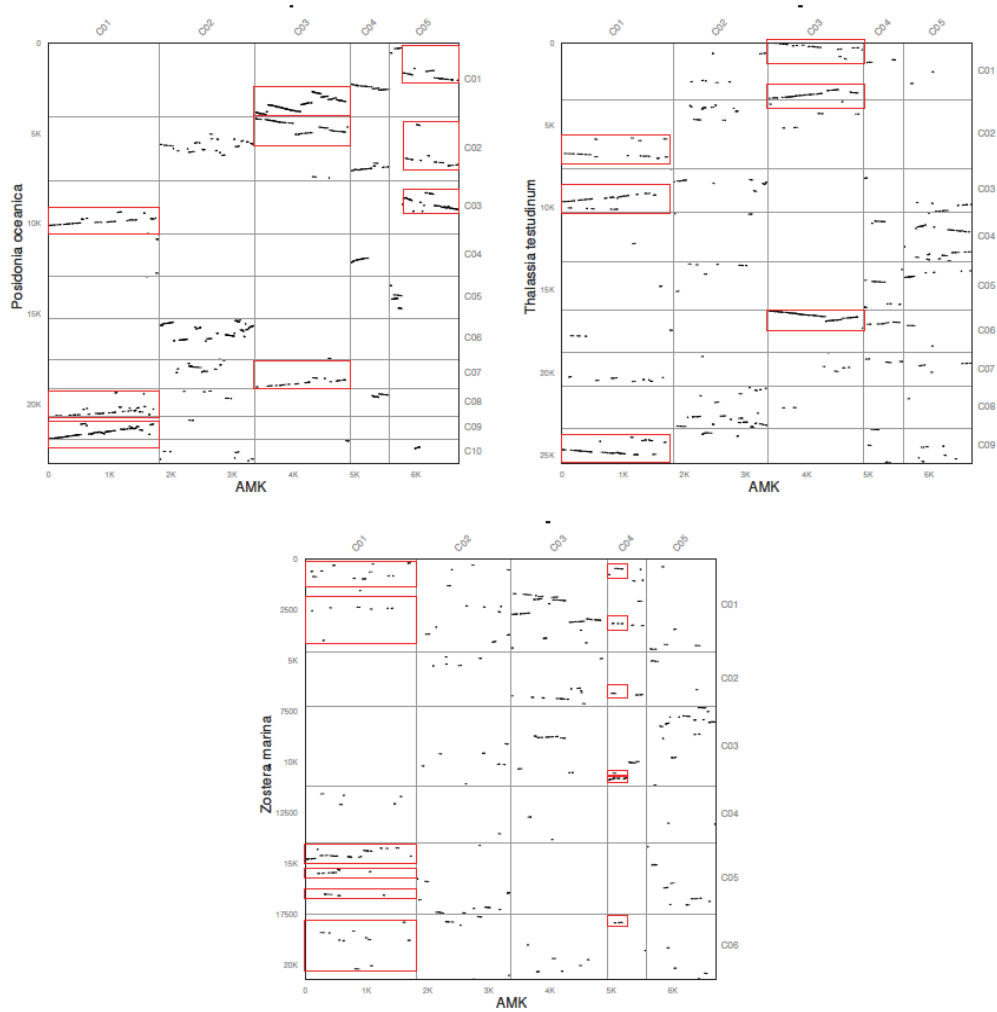
366 The most recent common ancestor (MRCA) node depth, of each gene pair, was associated with the pairwise K_s
367 estimate as a weight, in which for each duplication node in the phylogenetic tree, all n pairwise K_s estimates of
368 the descendant clades were added to the K_s distribution with a weight of $1/n$ to reduce redundancy. The K_s
369 ortholog age distributions were based on the one-to-one orthologs (reciprocal best hits) between two species
370 using BLAST (v2.6.0+) with an E-value cut-off of $1e-10$ following the same K_s estimation process as for the
371 paranome age distributions. The substitution rate correction across different species was achieved as follows:
372 1) The original divergence times between the focal species and the other species (represented as K_s distance)
373 were obtained by the mode of the ortholog K_s distributions using bootstrapped kernel density estimation; 2) A
374 substitution-rate-adjusted K_s estimate was calculated by transforming the original K_s distance into branch-
375 specific K_s distance using the reference of outgroup species and then rescaling the K_s distance with the diverged
376 species into the K_s timescale of the focal species for each divergence event. The corrected divergence times in
377 K_s timescale were then compared with the paranome K_s distribution of focal species after above processes. The
378 maximum number of outgroup species/trios selected to correct each divergent species pair was set as 6 and the
379 consensus peak for multiple outgroups was set as best outgroup in KSRATES. Other parameters were set as
380 default for rate correction using KSRATES. The species tree adopted consists of *Thalassia testudinum*,
381 *Cymodocea nodosa*, *Posidonia oceanica*, *Potamogeton acutifolius*, *Zostera marina*, *Spirodela polyrhiza*, *Wolffia*
382 *Australiana*, with *Brachypodium distachyon* and *Oryza sativa* as outgroup species, covering all the four families
383 of seagrasses (Posidoniaceae, Zosteraceae, Hydrocharitaceae and Cymodoceaceae).

384 The syntenic analysis was performed by MCscan (Python version) with default parameters (Tang et al. 2008).
385 Collinear blocks containing fewer than five orthologous gene pairs were filtered out. The collinearity information
386 of the gene set and other genomic features in the seagrass's genomes were visualized by Tbtools and
387 Circos (Krzywinski et al. 2009; Chen et al. 2020).

388

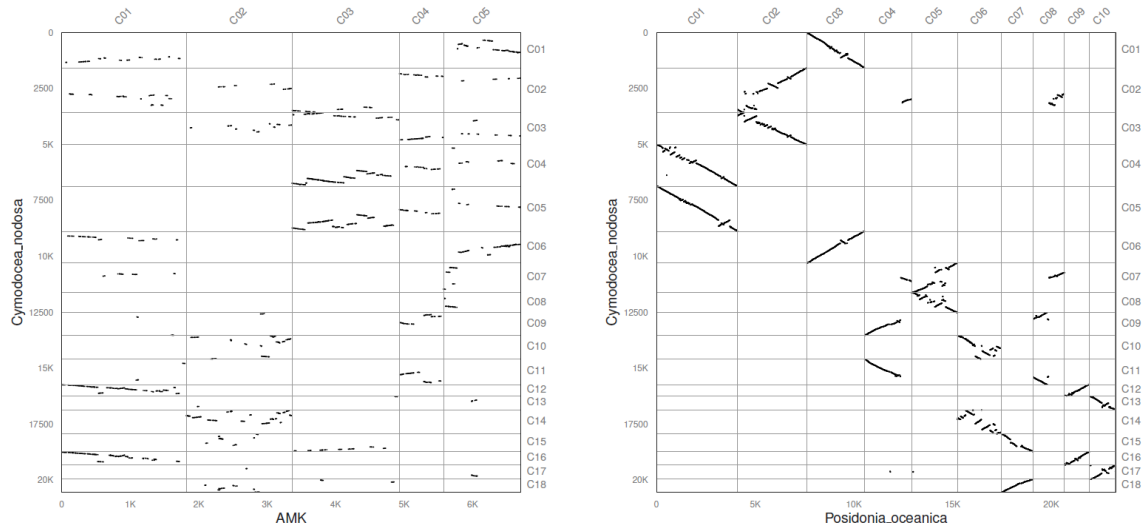


Supplementary Figure 4.2.1 K_s distributions for anchor pair duplicates (duplicates laying in duplicated, colinear blocks) and the whole paranome of four seagrasses, as well as for *P. acutifolius* and *S. polyrhiza*, generated by the wgd software (see Methods).



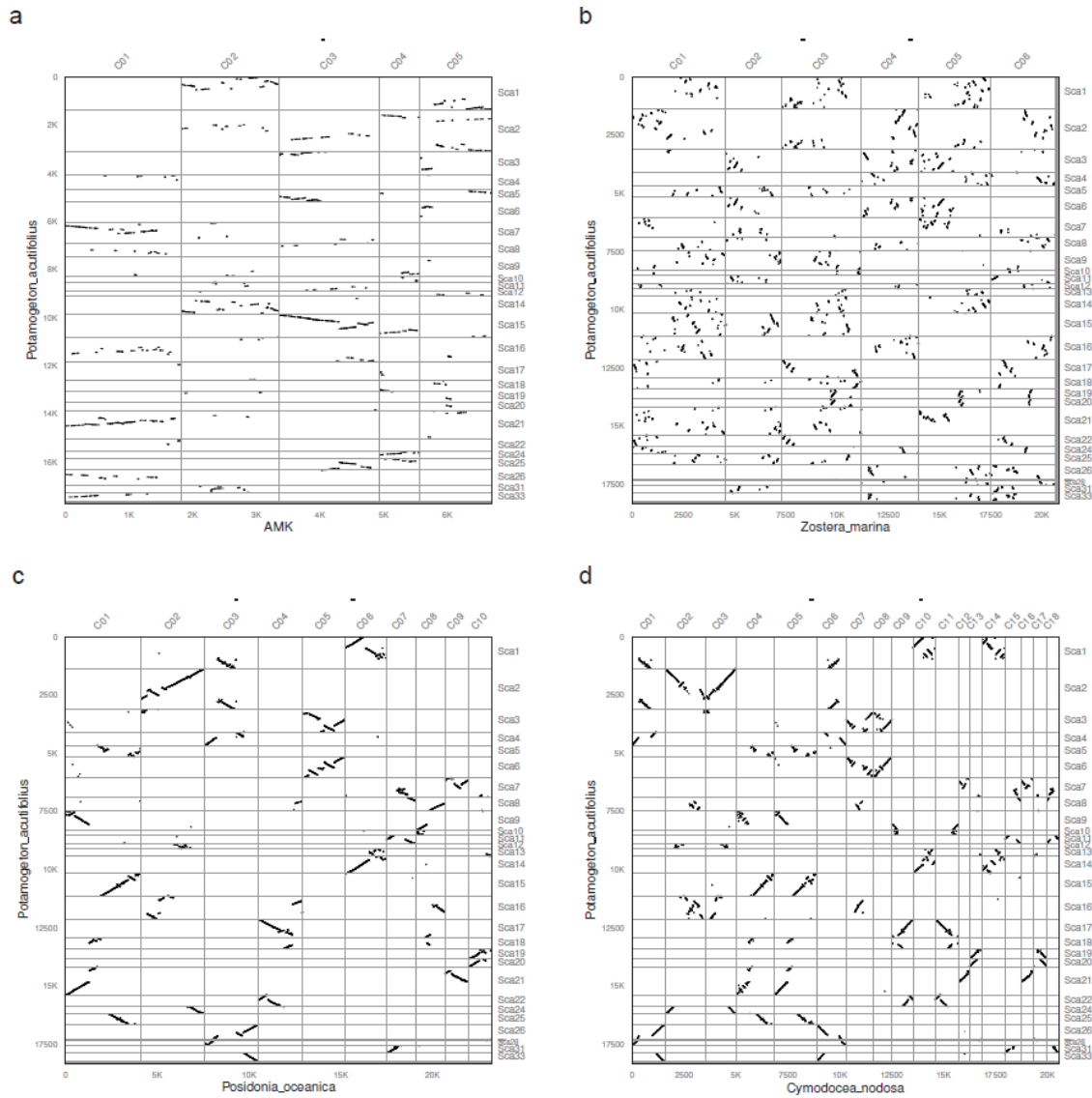
Supplementary Figure 4.2.2 Comparison of *Z. marina*, *P. oceanica*, and *T. testudinum* with a reconstructed ancestral monocot karyotype (AMK) (Murat et al. 2017).

Comparison between the AMK and *P. oceanica*, *T. testudinum* shows a clear 1:3 relationship, while the dot plot between the AMK and *Z. marina* shows a probable 1:6 relationship.



Supplementary Figure 4.2.3 Comparison of the ancestral monocot karyotype (AMK) (Murat et al. 2017) with *C. nodosa* and comparison of *P. oceanica* and *C. nodosa*.

Comparison between the AMK and *C. nodosa* shows a 1:6 synteny relationship. Comparison of *P. oceanica* and *C. nodosa* shows a 1:2 synteny relationship. This supports an extra WGD in *C. nodosa*, after its divergence with *P. oceanica*.

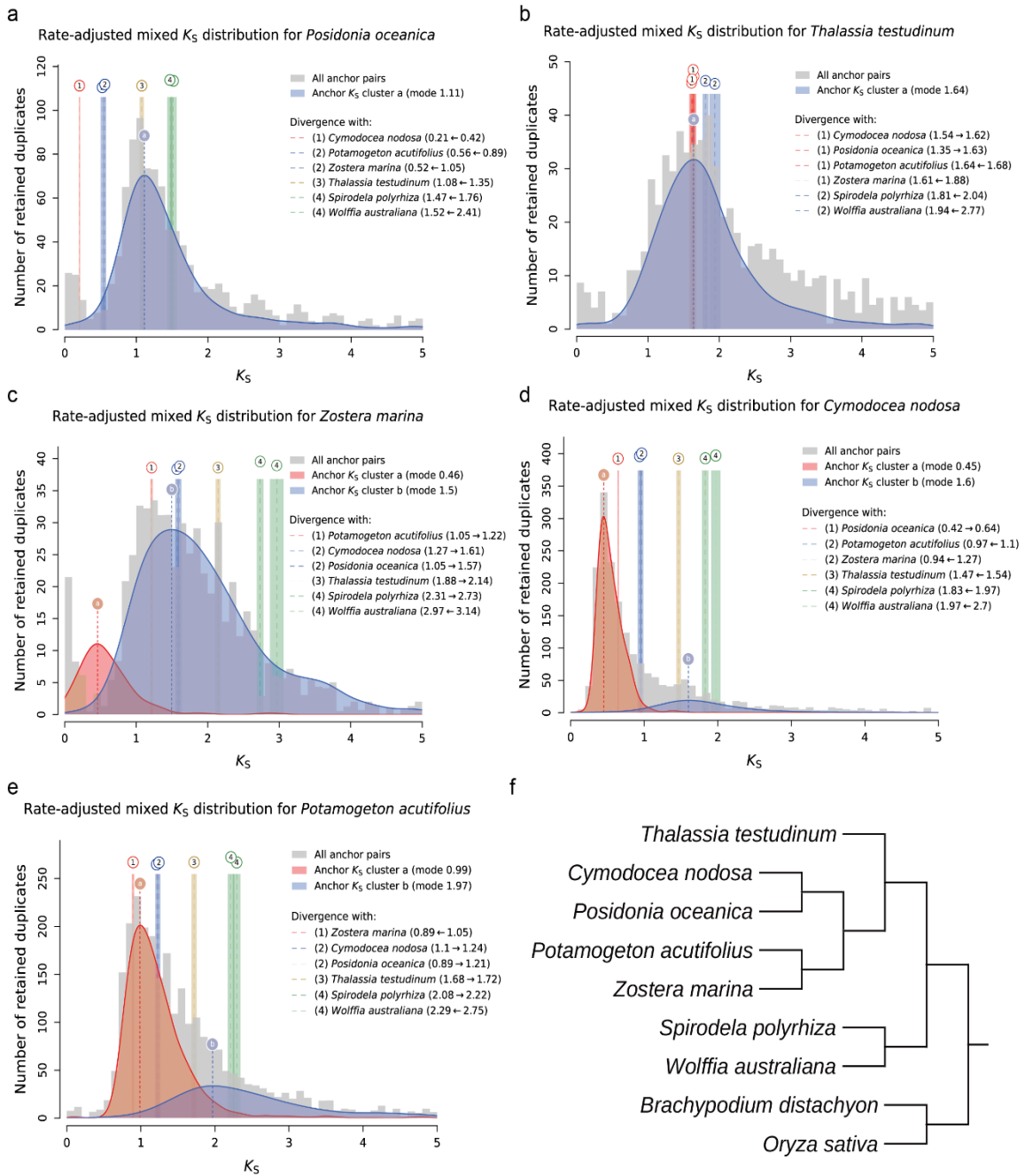


Supplementary Figure 4.2.4 Comparison of *P. acutifolius* with the ancestral monocot karyotype (AMK) (Murat et al. 2017), *Z. marina*, *P. oceanica*, and *C. nodosa*, respectively.

a) Comparison between *P. acutifolius* and the AMK shows a 1:6 syntenic relationship. B) Comparison of *P. acutifolius* and *Z. marina* and shows a non-obvious syntenic relationship c) *P. acutifolius* shows a 2:1 relationship with *P. oceanica*. D) *P. acutifolius* shows a 2:2 relationship with *C. nodosa*. See text for details.

397

398



Supplementary Figure 4.2.5 K_S Distributions for paralogs and the whole panome of four seagrasses and *P. acutifolius* generated by KSRATES software.

a – e, K_S distributions in *P. oceanica*, *T. testudinum*, *Z. marina*, *C. nodosa* and *P. acutifolius*. F, topology used in KSRATES analysis.

399

400

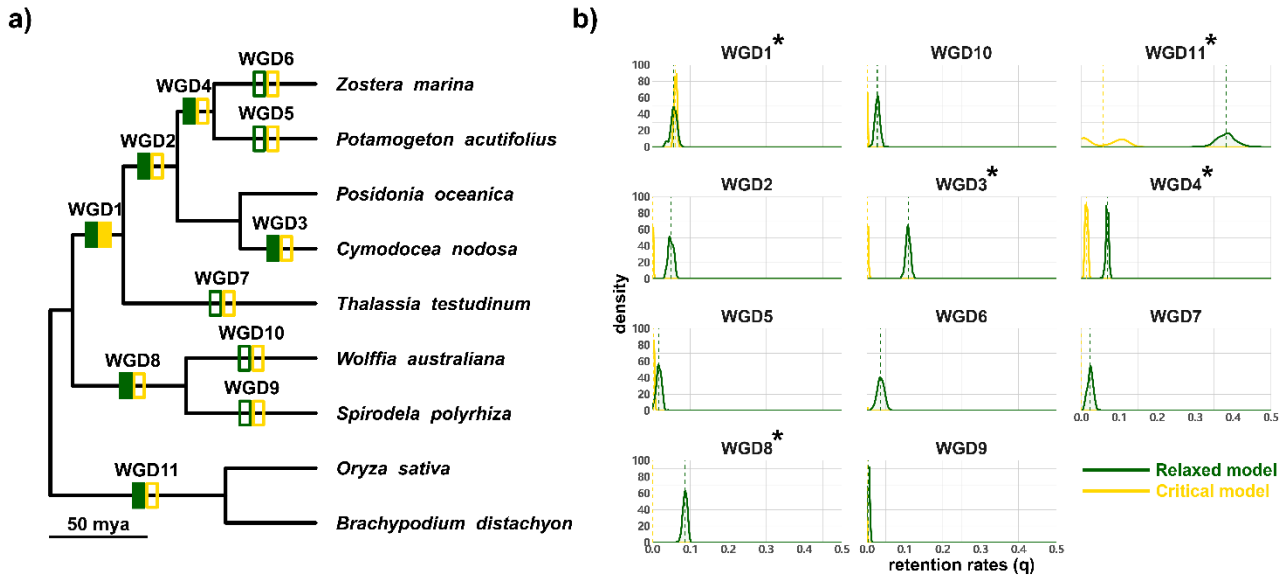
401 **Supplementary Note 4.2.2 Gene tree-species tree reconciliation**

402 OrthoFinder (Emms and Kelly 2019) (v2.3.3) was used to build orthologous gene families with the inflation factor
403 set to 3.0 and the remaining parameters set as default. Gene families that did not have at least one gene from
404 both clades at the root or have a family size exceeding 2 times the median of the square root of the family size,
405 based on a Poisson outlier criterion, were filtered out (Zwaenepoel and Van de Peer 2019a). An amino acid,
406 multiply sequence alignment (MSA) was obtained using PRANK (Loytynoja and Goldman 2005) for each gene
407 family and the resulting MSA was then used as input for the Markov Chain Monte Carlo (MCMC) analysis in
408 mrbayes (Huelsenbeck and Ronquist 2001) (v3.2.6) to sample from the posterior probability distribution. The
409 rate matrix for amino acid data (Aamodelpr) was set to a fixed (LG) and the rate was set as gamma-distributed
410 approximating four rate categories. The sampling frequency was set to 10 and the number of generations was
411 set to 110,000 to reach a total of 11,000 posterior samples. The ALExamine (Szöllösi et al. 2013) was then used
412 to construct the conditional clade distribution (CCD) containing marginal clade frequencies with a 'burn-in' of
413 1000, based on the 11,000 posterior samples for each gene family. The topology of the species tree was set as
414 shown in Figure 3). The time-calibrated species tree was inferred by MCMCtree from the PAML package (Yang
415 2007a), using reference divergence times of 42-52 million years ago (MYA) for the most common ancestor of
416 *Oryzae sativa* and *Brachypodium distachyon*, 118-129 MYA for that between *Spirodela polyrhiza* and *Zostera*
417 *marina* and 130-140 for that between *Spirodela* and other terrestrial monocots (An et al. 2019).

418 The duplication-loss (DL)+WGD model, under critical and relaxed branch-specific rates, was implemented for the
419 inference of the significance and corresponding retention rates of the assumed WGD events under Bayesian
420 inference (BI) (Zwaenepoel and Van de Peer 2019a). In the critical-branch-specific DL+WGD model, the prior η ,
421 which denotes the parameter of the geometric prior distribution on the number of genes at the root, was set to
422 follow a truncated-univariate-Beta distribution with shape parameters as (3,1) in the interval [0.01, 0.99]; the
423 prior r , which denotes the mean of the branch rates distribution, was set to follow a flat distribution; the prior
424 σ , which denotes the deviation of the branch rates distribution, was set to follow an exponential distribution
425 with scale 0.1, λ (which denotes the duplication rate of each branch) was set to follow a multivariate normal
426 distribution. For each branch, the loss rate μ was set to be equal to λ , while in the relaxed branch specific model,
427 λ and μ were independent with the rates variation parameter τ set to follow an exponential distribution with
428 scale 1 (Zwaenepoel and Van de Peer 2019a). To estimate duplication and loss rate λ and μ per branch
429 incorporating both small-scale gene duplications and WGDs, another model without WGD nodes where all
430 branch lengths were set as 1, the prior η was set to follow a Beta distribution with shape as (3,1) and the
431 duplication and loss rate were set respectively to follow a normal distribution with mean as 0 and standard
432 deviation as 5 were implemented. The Bayes Factor was calculated using the "bfact.jl" script within the public
433 github repository of WHALE to measure the strength of evidence in favor of the assumed WGD models using the
434 Savage-Dickey density ratio.

435 In total, 9 WGD(T) models were set on the branches leading to the MRCA of Potamogetonaceae, Zosteraceae,
436 Posidoniaceae, Cymodoceaceae and Hydrocharitaceae (labelled as WGD1 or WGT1), the MRCA of
437 Potamogetonaceae, Zosteraceae, Posidoniaceae and Cymodoceaceae (WGD2 or WGT2), the *C. nodosa* lineage
438 (WGD3), the *P. acutifolius* lineage (WGD4), the *T. testudinum* lineage (WGD5 or WGT5), the MRCA of *S. polyrhiza*
439 and *W. australiana* (WGD6), the *S. polyrhiza* lineage (WGD7), the *W. australiana* lineage (WGD8), the MRCA of
440 *B. distachyon* and *O. sativa* (WGD9), respectively (Supplementary Figure 4.2.6). Posterior mean of duplication
441 (left) and loss (right) rates estimated under DL+WGD modelling colored on the time-calibrated species tree. In
442 the left panel, green squares indicate the significantly supported WGDs under the relaxed branch-specific model
443 while empty squares indicate the WGDs that are not significantly supported under the relaxed branch-specific
444 model. In the right panel, light green squares indicate the significantly supported WGDs under the critical
445 branch-specific model while empty squares indicate the WGDs that are not significantly supported under the
446 critical branch-specific model.

447



Supplementary Figure 4.2.6 Bayesian inference of retention rates (q) of 11 hypothetical WGD models in WHALE (Zwaenepoel and Van de Peer 2019a).

a) Summary representation of the support from ‘relaxed’ and ‘critical’ models on a ‘divergence time tree’, in which filled rectangles denote support in terms of Bayes Factor, while outlined rectangles denote lack of support. The green rectangles show the results from the relaxed model, while the yellow rectangles show the results using the critical model. B) Posterior distribution of retention rates of the 11 hypothetical WGD models. Model categories follow the same color code as in a). Hypothetical WGD models which gained significant support and ‘accepted’ in this study are marked with asterisks.

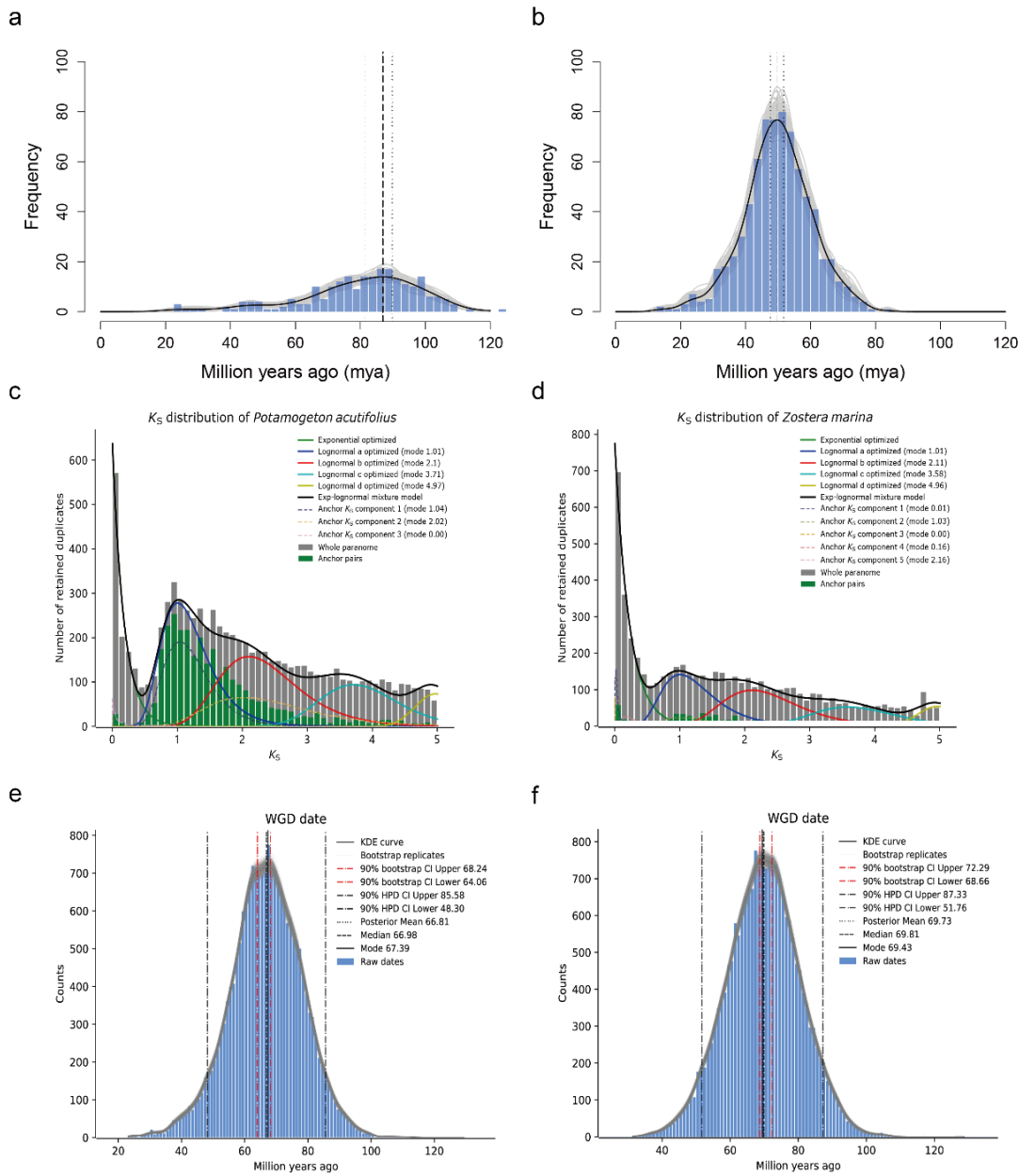
448

449

450 **Supplementary Note 4.2.3 Absolute dating of WGDs**

451 Absolute dating of WGD events was done as described previously described for *Zostera marina* (Olsen et al.
452 2016). Paralogous gene pairs located in duplicated segments (so called anchors) and duplicated pairs lying under
453 the WGD peak (so-called peak-based duplicates) were collected for phylogenetic dating. Anchors, which are
454 assumed to correspond to the most recent WGD, were detected using i-ADHoRe 3.0 (Simillion et al. 2008). For
455 each WGD paralogous pair, an orthogroup was created that included the two paralogues plus several
456 orthologues from other plant species, as identified by InParanoid (v4.1), using a broad taxonomic sampling, i.e.,
457 one representative from the order Cucurbitales, two from the Rosales, two from the Fabales, two from the
458 Malpighiales, two from the Brassicales, one from the Malvales, one from the Solanales, two from the Poales,
459 one orthologue from *Musa acuminata* (Zingiberales), and one orthologue from *Spirodela polyrhiza* (Alismatales).
460 WGT/WGD paralogues were then dated using the BEAST v1.7 package under an uncorrelated relaxed clock
461 model with the LG+G (four rate categories) evolutionary model. A starting tree with branch lengths satisfying all
462 fossil-prior-constraints was created according to the consensus APGIII phylogeny. Fossil calibrations were
463 implemented using log-normal calibration priors on the following nodes: the node uniting the Malvaceae based
464 on the fossil *Dressiantha bicarpellate* (Gandolfo et al. 1998) with prior offset = 82.8, mean = 3.8528, and s.d. =
465 0.5), the node uniting the Fabaceae based on the fossil *Paleoclusia chevalieri* (Crepet and Nixon 1998) with prior
466 offset = 82.8, mean = 3.9314, and s.d. = 0.5, the node uniting the Alismatales (including *Zostera marina* and
467 *Spirodela polyrhiza*) with the other monocots based on the oldest fossil monocot pollen, *Liliacidites* (Doyle et al.
468 2008; Iles et al. 2015) from the Trent's Reach locality, with prior offset = 125, mean = 2.0418, and s.d. = 0.5
469 (Janssen and Bremer 2004; Nauheimer et al. 2012) and the root with prior offset = 124, mean = 4.0786, and s.d.
470 = 0.5 (Smith et al. 2010). A run without data was performed to ensure proper placement of the marginal
471 calibration prior distributions. The Markov chain Monte Carlo (MCMC) for each orthogroup was run for 10^7
472 generations, sampling every 1,000 generations, resulting in a sample size of 10^4 . The resulting trace files of all
473 orthogroups were evaluated manually using Tracer v1.570 with a burn-in of 1,000 samples to ensure proper
474 convergence (minimum ESS for all statistics at least 200). To resolve the absolute dates of other involved WGDs
475 in our analysis (Figure 2), we also redated the WGDs of *Elaeis guineensis*, *Asparagus officinalis*, *Rhizophora*
476 *apiculata*, *Avicennia marina* and *Utricularia gibba* using the same pipeline as above. Moreover, the fossil of
477 *Sabalites carolinensis* (Berry 1914) were also chosen as calibrations as previous study (Vanneste et al. 2014). The
478 final WGD dates were shown in Supplementary Table 4.2 and Supplementary Figure 4.2.8.

479

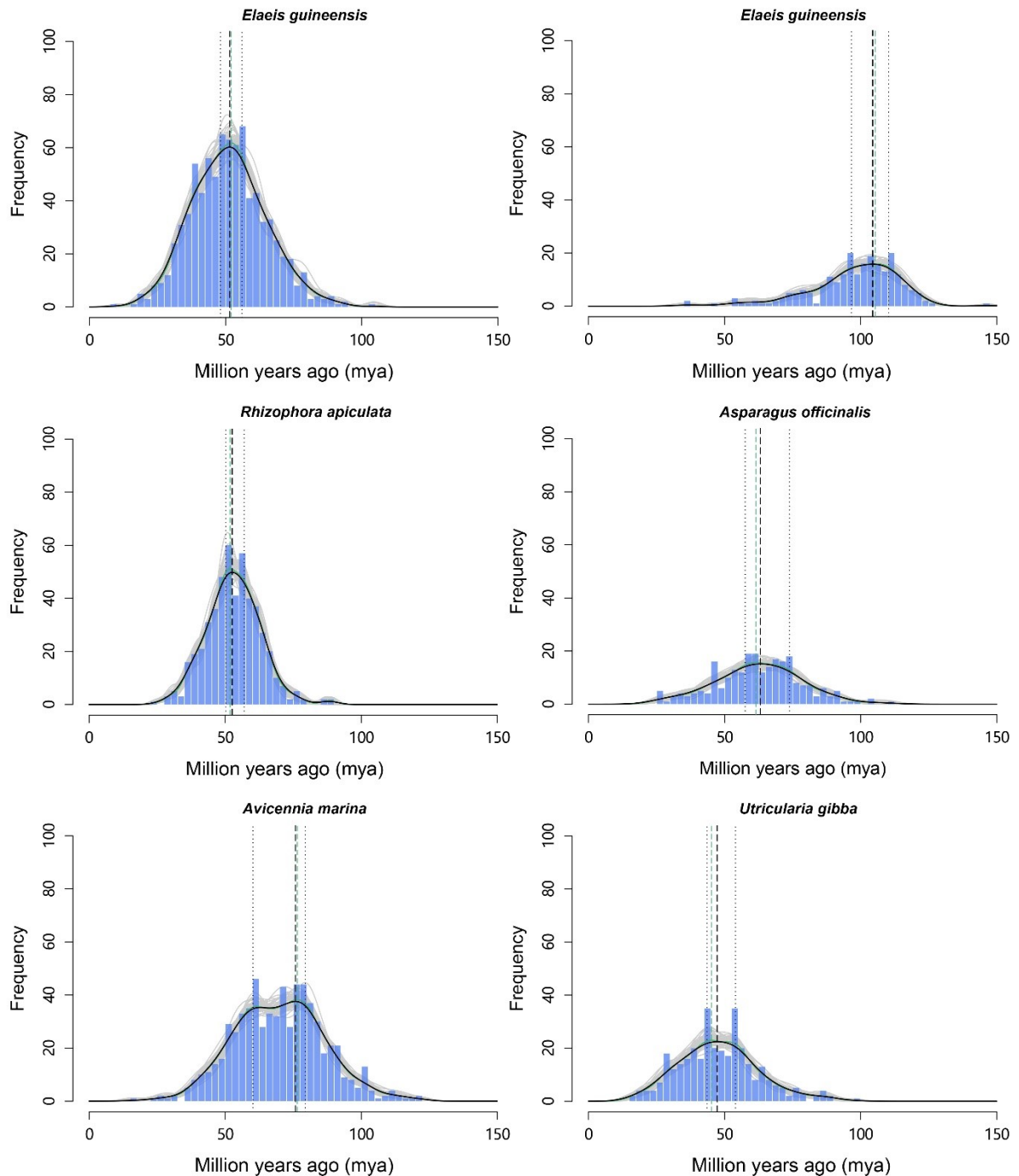


Supplementary Figure 4.2.7 Estimation of the 'absolute age' of the WGT/WGD events in seagrasses and *P. acutifolius* by phylogenomic dating.

a) Estimation of the 'absolute age' of the *P. oceanica* WGT event by phylogenomic dating. The solid black line represents the KDE (Kernel Density Estimation) of the dated paralogues, while the vertical dashed black line represents its peak at 86.96 Mya, which was used as the consensus WGT age estimate. The grey lines represent density estimates from 2,500 bootstrap replicates (after we obtained the age estimates for each accepted orthogroup that satisfied the condition that the minimum ESS for every parameter is larger than 200, we calculated the bootstrap 90% confidence interval for the mode of fitted kernel density estimation (KDE) using the "boot" library in R upon these age estimates), while the vertical black dotted lines represent the corresponding 90% confidence interval for the WGD age estimate, 89.89 - 79.81 Mya. The histogram shows the raw distribution of dated paralogues. b) Estimation of the 'absolute age' of the recent *C. nodosa* WGD event. Interpretation is as in a). c) The mixture modeling results of whole paraneome and anchor K_s distribution of *P. acutifolius*. d) The mixture modeling results of whole paraneome and anchor K_s distribution of *Z. marina*. e) Estimation of the 'absolute age' of the recent *P. acutifolius* WGD event. Interpretation is as in a). f) Estimation of the 'absolute age' of the recent *Z. marina* WGD event. Interpretation is as in a).

Supplementary Table 4.2 The absolute of WGD events taken from literature

Species name	Date	Family	Order	Clade	Phylogenetic position of WGD	Data Source
<i>Spirodela polyrhiza</i>	83-94, 95-100	Araceae	Alismatales	Monocots	Araceae	(Wang et al. 2014; An et al. 2019)
<i>Oryza sativa</i>	63.08-69.89, 100-120, 110-135	Poaceae	Poales	Commelinids	Poaceae, Poales, non-Alismatales monocots	(Tang et al. 2010; Vanneste et al. 2014; Ming et al. 2015)
<i>Elaeis guineensis</i>	48.25-55.97, 96.71-110.48	Arecaceae	Arecales	Commelinids	Partial Arecaceae, Arecales	Our dating, see Supplementary Figure 4.2.8
<i>Asparagus officinalis</i>	59.24-72.43	Asparagaceae	Asparagales	Monocots	Asparagus	Our dating, see Supplementary Figure 4.2.8
<i>Arabidopsis thaliana</i>	49.27-50.99	Brassicaceae	Brassicales	Rosids	Partial Brassicaceae	(Guo et al. 2018)
<i>Populus trichocarpa</i>	60-65, 125-140	Salicaceae	Malpighiales	Rosids	Specific to Populus and Salix, Core eudicots	(Tuskan et al. 2006)
<i>Rhizophora apiculata</i>	50.36-56.95	Rhizophoraceae	Malpighiales	Rosids	Rhizophoraceae	Our dating, see Supplementary Figure 4.2.8
<i>Vitis vinifera</i>	125-140	Vitaceae	Vitales	Rosids	Core eudicots	(International Peach Genome et al. 2013)
<i>Avicennia marina</i>	60.46-78.78	Acanthaceae	Lamiales	Asterids	<i>Avicennia marina</i>	Our dating, see Supplementary Figure 4.2.8
<i>Utricularia gibba</i>	43.37-53.99	Lentibulariaceae	Lamiales	Asterids	Utricularia	Our dating, see Supplementary Figure 4.2.8
<i>Solanum lycopersicum</i>	62.64-64.84	Solanaceae	Solanales	Asterids	Solanaceae	(Vanneste et al. 2014)



483

484 **Supplementary Figure 4.2.8** Estimation of the 'absolute age' of seven independent WGD events
 485 experienced by *E. guineensis*, *A. officinalis*, *R. apiculata*, *A. marina* and *U. gibba* respectively by
 486 phylogenomic dating of corresponding paralogues.

487 The solid black line represents the KDE (Kernel Density Estimation) of the dated paralogues, and the vertical
 488 dashed black line represents the peak which was used as the consensus WGD age estimate. The grey lines
 489 represent density estimates from 2,500 bootstrap replicates and the vertical black dotted lines represent the
 490 corresponding 90% confidence interval for the WGD age estimate. The histogram shows the raw distribution of
 491 dated paralogues.

492

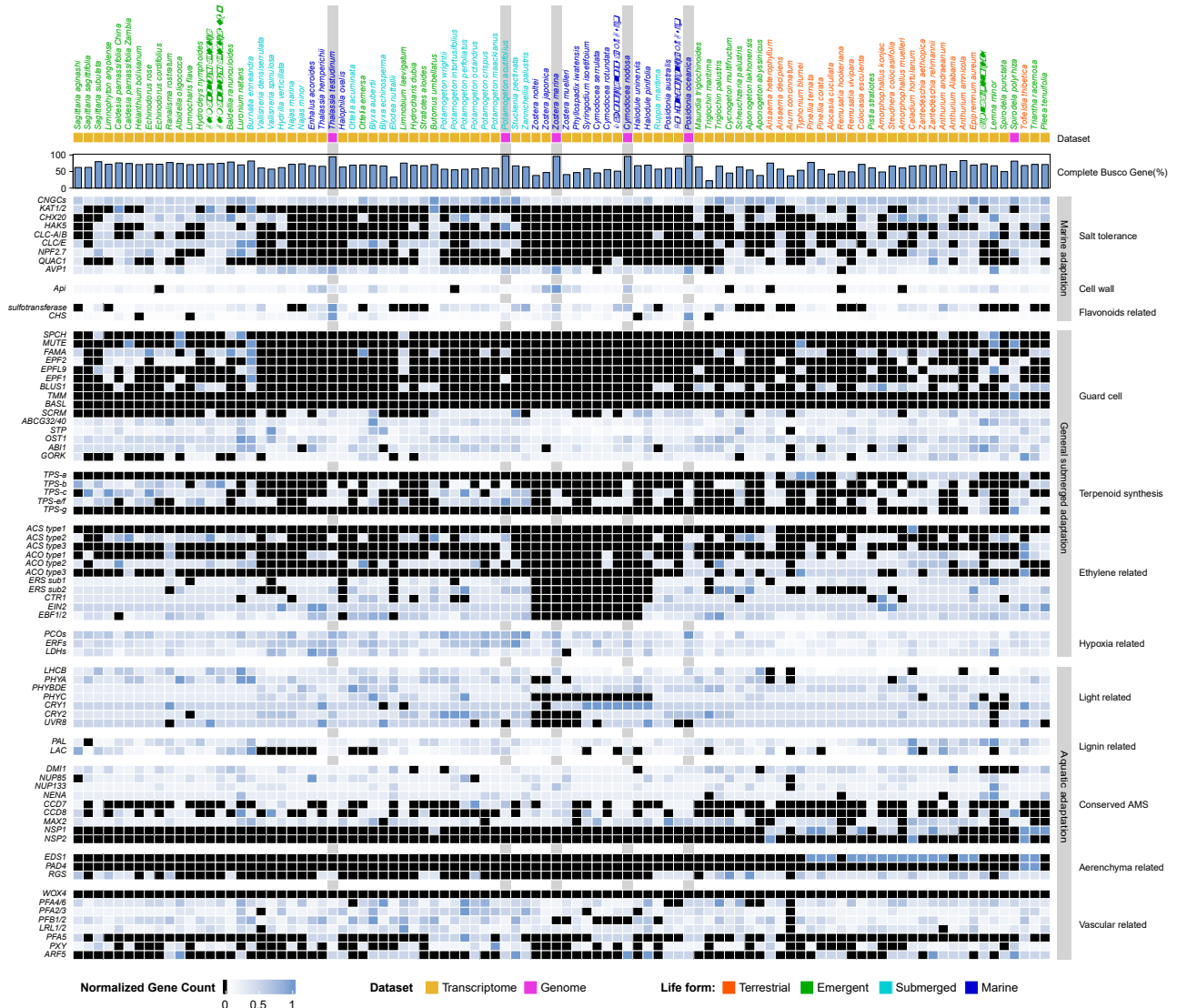
493 4.3 Phylogenetic tree construction and estimation of divergence time

494 **Supplementary Note 4.3** Species selection and construction of time-calibrated phylogeny.

495 Protein sets were collected for 23 species, including *Oryza sativa* (PLAZA 5.0), *Brachypodium distachyon* (PLAZA
496 5.0), *Ananas comosus* (PLAZA 5.0), *Elaeis guineensis* (PLAZA 5.0), *Asparagus officinalis* (PLAZA 5.0), *Beta vulgaris*
497 (PLAZA 5.0), *Utricularia gibba* (PLAZA 5.0), *Solanum lycopersicum* (PLAZA 5.0), *Coffea canephora* (PLAZA 5.0),
498 *Vitis vinifera* (PLAZA 5.0), *Populus trichocarpa* (PLAZA 5.0), *Arabidopsis thaliana* (PLAZA 5.0), *Theobroma cacao*
499 (PLAZA 5.0), *Avicennia marina* (PLAZA 5.0), *Spirodela polyrhiza* (PLAZA 5.0), *Amborella trichopoda* (PLAZA 5.0),
500 *Wolffia australiana* (<https://duckweeds.plantprofile.net/>), *Rhizophora apiculata* (from the author), our four
501 seagrasses and *Potamogeton acutifolius* (from ORCAE, <https://bioinformatics.psb.ugent.be/orcae/>). These
502 species were selected as representatives for monocots and eudicots, and representing different habitats from
503 terrestrial, freshwater-floating, freshwater-submerged, to marine-submerged. Orthofinder v2.3 (Emms and
504 Kelly 2015) was used to delineate gene families with mcl inflation factor 3.0. All-versus-all Diamond blast with
505 an E-value cutoff of 1e-05 was performed and orthologous genes were clustered using OrthoFinder. Single-copy
506 orthologous genes were extracted from the clustering results. MAFFT (Rozewicki et al. 2019) with default
507 parameters was used to perform multiple sequence alignment of protein sequences for each set of single-copy
508 orthologous genes, and to transform the protein sequence alignments into codon alignments after removing
509 the poorly aligned or divergent regions using trimAl (Capella-Gutiérrez et al. 2009). The resulting codon
510 alignments from all single-copy orthologs were then concatenated into one supergene for species phylogenetic
511 analysis. A maximum-likelihood phylogenetic tree of single-copy protein alignments and codon alignments was
512 constructed using IQ-TREE (Minh et al. 2020) with the GTR+G model and 1,000 bootstrap replicates. Divergence
513 times between the 23 plant species were estimated using MCMCtree from the PAML package under the GTR+G
514 with reference divergence times of 124-170 MYA for the common ancestor of monocots and eudicots, 118-129
515 MYA for the divergence between *Spirodela* and *Zostera* and 130-140 MYA between *Spirodela* and other
516 terrestrial monocots (An et al. 2019). We used MCMCTree to obtain 10,000 trees from the posterior sampling
517 every 150 iterations after a burn-in of 500,000 iterations. We compared two independent runs with each other
518 to verify convergence and with a run of the MCMC algorithm under the prior alone to compare the posterior
519 distribution for the node ages to the effective prior implied by the fossil calibrations.

520

521 5. Adaptation to the marine environment
 522 5.1. Use it or lose it



523 **Supplementary Figure 5.1** Normalized gene counts for each species. Species with light grey
 524 backgrounds denote seagrass species and one freshwater relative, *Potamogeton acutifolius*, discussed in the
 525 present study. Other species are discussed in (Chen et al. 2022). Taxon order is phylogenetic. Normalization
 526 for each gene family was obtained by dividing the number of genes in that gene family for a particular species
 527 by the largest gene copy number within that family (considering all species). Genes in black are absent.

528

529 5.2 Pathogen resistance (R-) genes

530 **Supplementary Note 5.2 Pathogen resistance gene**

531 Effector-triggered immunity is one of the two main arms of the plant immune system and allows angiosperms
532 to specifically detect pathogen effectors or their impact on host proteins. The detection is guided by nucleotide-
533 binding leucine rich repeat receptors (*NLRs*), which is one of the largest gene families in plants, and under
534 diversifying selection (Jacob et al. 2013). Of the two main domains of *NLR* resistance genes, the nucleotide
535 binding site (NBS) domain is responsible for downstream signaling, and the leucine rich repeat (LRR) domain
536 binds the target. *NLR* genes are often difficult to identify in genomes. Therefore, we used two software packages,
537 *NLR-Annotator* (Steuernagel et al. 2020) and *NLGenomeSweeper* (Toda et al. 2020), followed by manual curation
538 in a genome browser.

539 The number of *NLR* genes is strongly reduced (N=44) in *Z. marina* (Olsen et al. 2016) and similar reductions have
540 since been found in many freshwater species (N= 100 range) (Liu et al. 2021), which is far less than in terrestrial
541 species (N=100-300-500 [2300 in wheat]). Thus, we expected to see a similar extreme reduction in our new
542 seagrass species, but this was not the case. While reduced in comparison to terrestrial species the number of
543 *NLR* genes was markedly higher than in *Z. marina*; *C. nodosa* (N=87), *P. oceanica* (N=95) and *T. testudinum* (N=54).

544 Confirming our general hypothesis of convergent evolution at the genomic level, these seagrass species have
545 low counts of *NLR* gene copies (Supplementary Table 5.2). Further, *NLRs* with a TIR domain are completely
546 absent, which is typical of many monocots, and a few genes are missing the LRR domain. We also found that 30-
547 40% of gene copies are non-functional in *C. nodosa*, *P. oceanica*, *T. testudinum* and *P. acutifolius*, either because
548 of stop mutations or partial copies. By contrast, only 8% are non-functional in *Z. marina* (Supplementary Table
549 5.2).

550 The *NLR* gene copies occur in clusters towards the terminal ends of the chromosomes consistent with findings
551 in other plants (Jacob et al. 2013). These clusters are made up of tandem copies as evidenced by their
552 relationship shown in the NBS-domain-based phylogenetic tree and chromosomal location (Supplementary
553 Figure 5.2.1 and Supplementary Figure 5.2.2). Here, they cluster into several clades, each including all of the
554 species, thus indicating that the ancestor also contained these gene lineages. Similarly, when incorporating other
555 more distantly related species in the NBS tree (not shown), the seagrass-genes-branches are distributed
556 throughout, so old *NLR* gene lineages are still maintained, despite the reduction in total number compared to
557 other plants. Single lineages are expanded into clusters of dozens of copies at the species level, especially in *C.*
558 *nodosa* and *P. acutifolius* (Supplementary Figure 5.2.1). From an evolutionary perspective, clustering is
559 considered as a reservoir of genetic variation (Jacob et al. 2013).

560

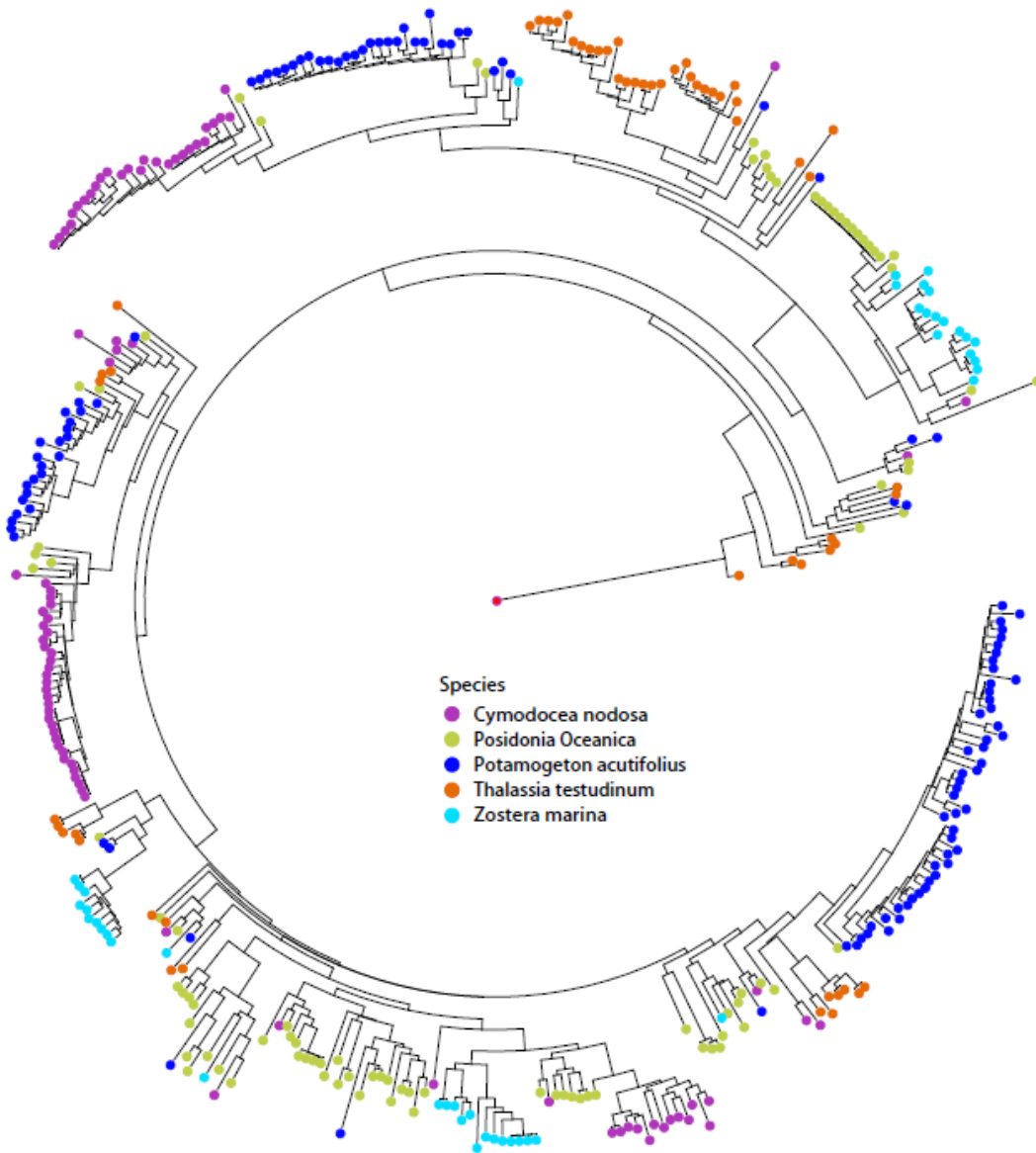
561 **Supplementary Table 5.2** *NLR* gene counts by domain architecture and completeness in seagrasses
 562 and *P. acutifolius*.

563 Completeness was determined by the NLR-Annotator software. Complete/partial refers to the set of motifs
 564 needed for a functional gene; pseudogene refers to loci that do not have complete open reading frames. Gene
 565 class abbreviations in parentheses are a more general designation used in some papers.

566 Note: Due to their partly fragmented nature not all *NLR* genes are present in the genome annotation with gene
 567 IDs. The genomic coordinates of all identified *NLR* genes can however be found in Extended Data Table 5.

<i>NLR</i> gene class	<i>Cymodocea nodosa</i>	<i>Posidonia oceanica</i>	<i>Thalassia testudinum</i>	<i>Zostera marina v3.1</i>	<i>Potamogeton acutifolius</i>
CC-NBS (CN)	2	0	0	0	2
CC-NBS-LRR (CNL)	59	58	27	21	71
NBS-LRR (NL)	26	37	27	23	40
NBS (N)	0	0	0	0	2
TIR-NBS (TN)	0	0	0	0	0
TIR-NBS-LRR (TNL)	0	0	0	0	0
Total	87	95	54	44	115
NLR gene completeness					
Complete	69	62	42	26	87
Complete pseudogene	20	28	15	18	27
Partial	13	14	12	0	25
Partial pseudogene	44	41	13	4	17
% Complete	61	62	70	92	73

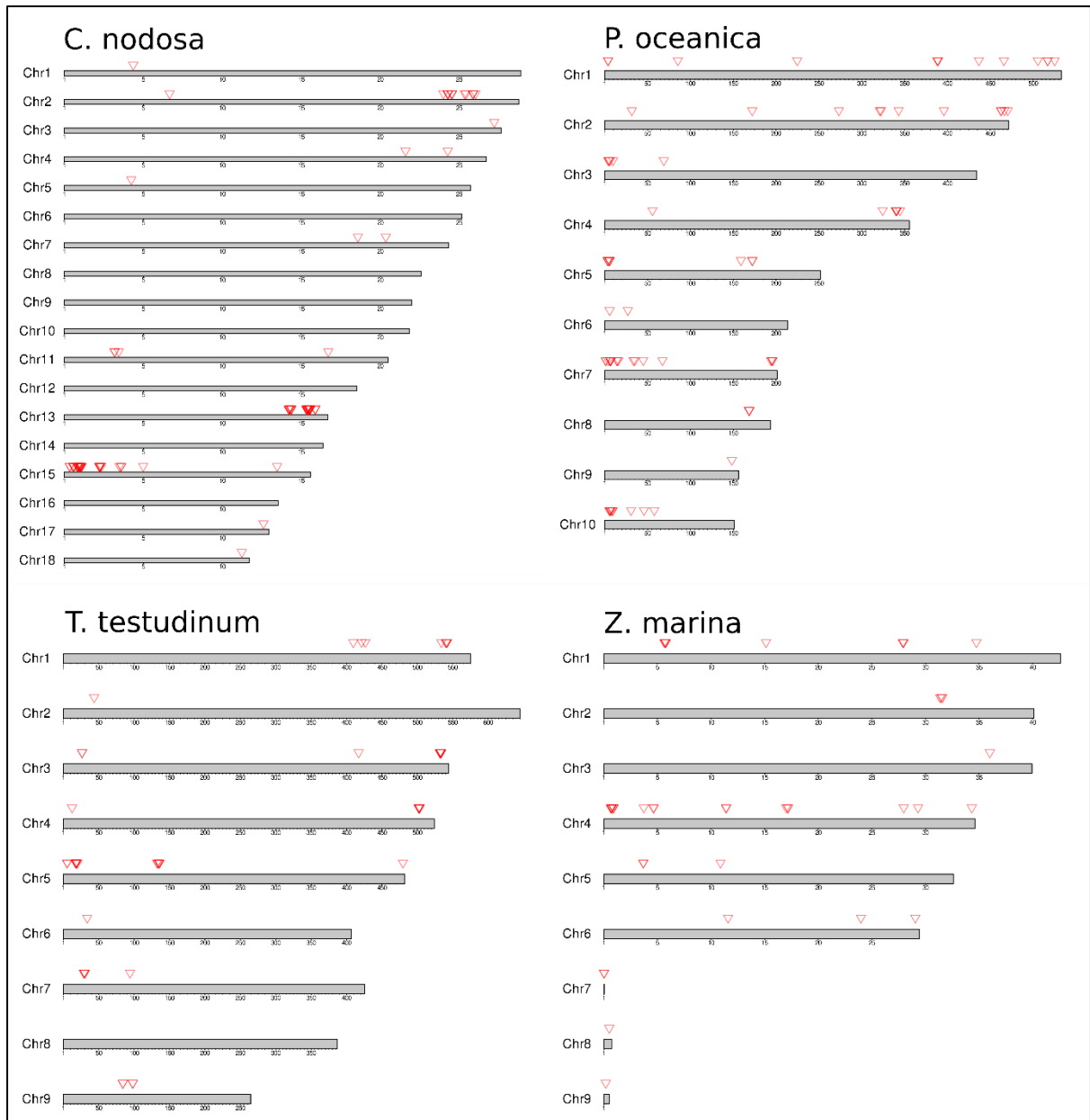
568



570

571 **Supplementary Figure 5.2.1** Phylogenetic tree of seagrass *NLR* genes based on NBS domain.

572



573

574 **Supplementary Figure 5.2.2** Distribution of seagrass *NLR* genes across chromosomes.

575 *NLR* genes are indicated by red arrows and predominantly occur in the distal regions.

576

577 5.3 Heat Shock factor (*HSF*) gene family evolution

578 **Supplementary Note 5.3** *HSF* gene family

579 Heat shock transcription factors (*HSFs*) are a family of DNA-binding proteins that activate a cascading network
 580 of genes that act together to enhance plant tolerance to abiotic stress conditions, including heat, cold, drought,
 581 hypoxia, salinity, toxicity and excessive irradiance (Scharf et al. 2012). Based on the topology of their domains,
 582 *HSFs* are classified into three major classes (*HSFA*, *HSFB* and *HSFC*), which are further subdivided into 16
 583 subfamilies: *HSFA1-HSFA9*, *HSFB1-HSFB5* and *HSFC1-HSFC2*. Individual *HSFs* have unique functions as part of
 584 different signal transduction pathways operating in response to environmental stress and during plant
 585 development (von Koskull-Döring et al. 2007). To examine the different composition of *HSF* gene families in
 586 seagrasses, *HSF* families of *Cymodocea nodosa*, *Posidonia oceanica*, *Thalassia testudinum* and *Zostera marina*,

587 were compared with those of the freshwater plant *Potamogeton acutifolius*, ten terrestrial eudicots (*Amborella*
588 *trichopoda*, *Arabidopsis thaliana*, *Beta vulgaris*, *Solanum lycopersicum*, *Populus trichocarpa*, *Vitis vinifera*, *Coffea*
589 *canephora*, *Theobroma cacao* and the mangroves *Avicennia marina* and *Rhizophora apiculata*), five terrestrial
590 monocots (*Oryza sativa*, *Brachypodium distachyon*, *Elaeis guineensis*, *Ananas comosus* and *Asparagus officinalis*)
591 and three other freshwater plant species (one eudicot: *Utricularia gibba*; two monocots: *Spirodela polyrhiza* and
592 *Wolffia australiana*). *HSF* sequences were searched in the Plant Transcriptional Factor Data Base (PlantTFDB,
593 <http://plantfdb.gao-lab.org/>) and in Phytozome 13 (<https://phytozome-next.jgi.doe.gov/>). Protein sequences
594 were subsequently downloaded from PLAZA (<https://bioinformatics.psb.ugent.be/plaza/>), the PlantTFDB or,
595 when needed, from the species genome web page (*Wolffia australiana*: <https://duckweeds.plantprofile.net/>;
596 *Utricularia gibba*: <http://genomevolution.org/CoGe/>). Sequences were then uploaded to the HEATSTER
597 platform to check their identity as *HSFs* and to use a single criterion for their classification within the 16 *HSF* sub-
598 classes. Classification and annotation were performed in HEATSTER via two successive steps of repeated
599 searches in a motif database (motifs: DBD and OD with HR-A and HR-B region). In those cases where a specified
600 sequence did not contain all *HSF*-associated motifs, classification was based on the recognition of the most
601 conserved domain (DBD) with an E-value < 1e-20.

602 In our analysis, the average number of *HSFs* in land plants (Supplementary Table 5.3) was similar to the values
603 recently reported in a comprehensive study involving 29 eudicots and 10 monocots (Wang et al. 2018). We found
604 that aquatic plants have a lower number of *HSFs* than terrestrial plants (54% on average), with no clear
605 differences between marine and freshwater species. The four studied seagrass species have on average 11.8 (\pm
606 1.0) sequences recognized as *HSFs*, with all three types of *HSFs* (A, B and C) showing a strong contraction. The
607 average number of class A and B members is reduced by 67.4% and 43% respectively in seagrasses compared to
608 terrestrial monocots (Supplementary Table 5.3). In addition, some subclasses are completely absent in
609 seagrasses (Extended Data Table 5). In the *HSFA* class, subclass A3, A7, A8 and A9 are lacking in seagrasses but
610 are common in terrestrial monocots (except A9 found only in eudicots). Along with their role in the response to
611 heat stress, these subfamilies have specialized functions in the response of plants and seeds to different abiotic
612 stresses, mainly dehydration, drought stress and oxidative stress (Sakuma et al. 2006; von Koskull-Döring et al.
613 2007; Scharf et al. 2012; Personat et al. 2014). The emergence of new functionalities has been associated with
614 the weak purifying selection of these subfamilies in terrestrial plants (Wang et al. 2018). Contrarily, subfamilies
615 A1, A2, A5 and A6, which are subjected to more severe selection pressure and directly involved in the heat stress
616 response (Heerklotz et al. 2001; Mishra et al. 2002; Scharf et al. 2012; Xue et al. 2014), are represented in
617 seagrasses. These results reveal that marine plants have lost several subclasses of *HSFA* previously acquired to
618 cope with the stress conditions associated with a terrestrial lifestyle.

619 Regarding the *HSFB* class, subclasses B3 and B5 are not present in seagrasses, but neither are they present in
620 terrestrial monocots as they presumably arose after the split of monocots and eudicots (Scharf et al. 2012; Guo
621 et al. 2016). Seagrasses retained B1 and B2 subclasses (Extended Data Table 5), both of which are involved in
622 promoting the activity of *HSFA1*, which is the master regulator of the heat stress response in *Arabidopsis* (Ikeda
623 et al. 2011; Scharf et al. 2012). Moreover, *HSFB1* form a triad with *HSFA1* and *HSFA2* in tomato, acting as
624 synergistic coactivator of heat stress responsive genes during the exposure and recovery to high temperatures
625 (Mishra et al. 2002; Scharf et al. 2012). It is therefore evident that marine plants have conserved all major *HSF*
626 subclasses that functionally cooperate in the heat stress response and thermotolerance of plants (A1, A2, A5,
627 A6, B1 and B2).

628 *HSFC* family members (i.e., C1 and C2) are completely absent in seagrasses. Although the function of class C *HSFs*
629 is the least known, they appear to be integrated into signaling pathways not directly related to the heat stress
630 response (Scharf et al. 2012). *HSFC2* act as transcriptional activator of heat shock protein (*HSP*) genes in wheat
631 during heat, drought and salt stress (Xue et al. 2014), and is up-regulated in rice by oxidative and heat stress
632 (Mittal et al. 2012). Similarly, *HSFC1* showed altered expression levels in *Arabidopsis* under several stress
633 conditions, including cold stress, freeze stress and dehydration stress (Lee et al. 2005; Xin et al. 2007; Ding et al.

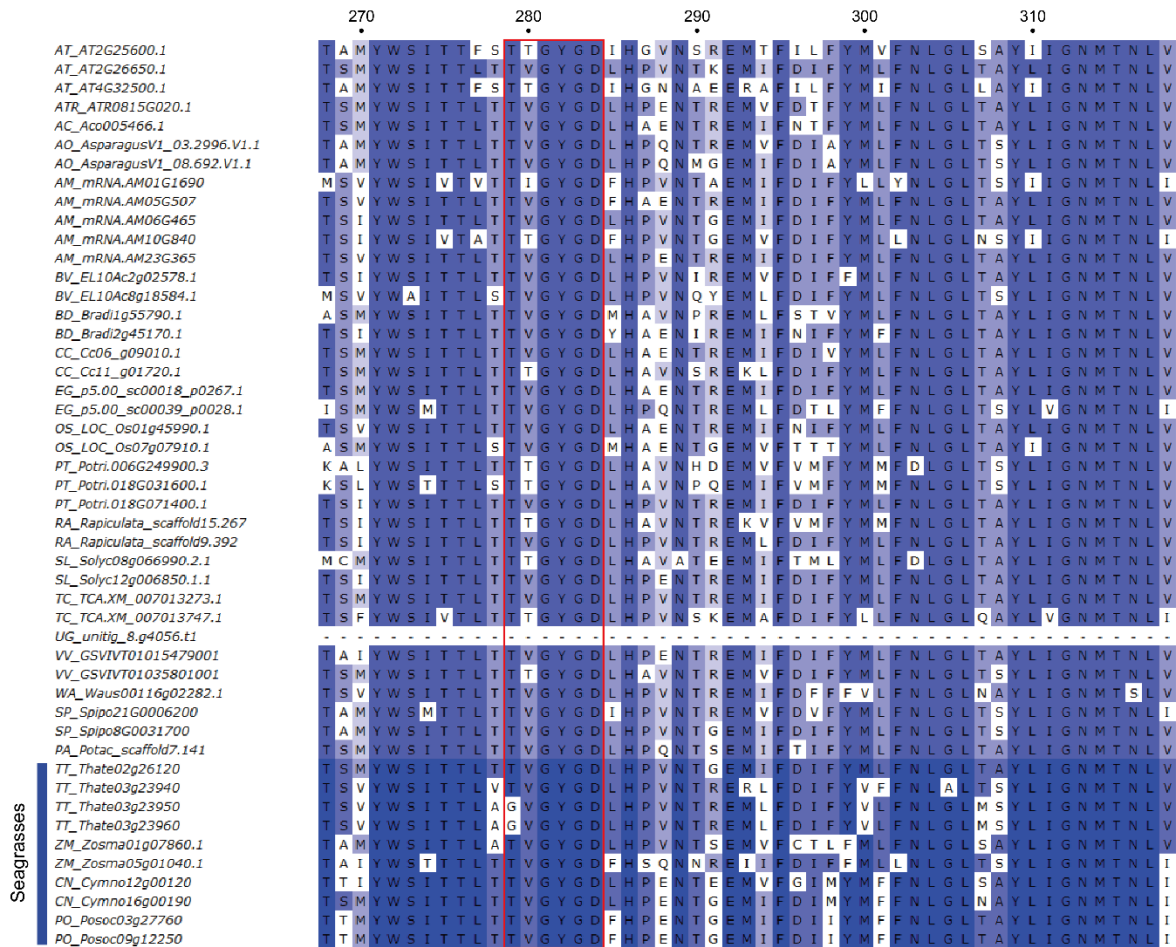
634 2013; Zhuang et al. 2018). These results indicate that the functionally specialized *HSFC* family, which emerged
 635 during the evolution of plants towards a terrestrial lifestyle, has been lost when plants returned to the sea.

636 In summary, plants with an aquatic lifestyle, including those adapted to the marine environment (seagrasses),
 637 have a reduced number of *HSFs* compared to terrestrial plants. Despite having a small number of members,
 638 seagrasses have retained those *HSF* subfamilies under strong purifying selection in land plants, probably to
 639 maintain important biological functions. Among them is the main group of *HSFs* directly related to heat stress
 640 response and thermotolerance in plants. In contrast, subfamilies of *HSFs* with specialized functionalities for plant
 641 adaptation to terrestrial habitats and not directly related to heat stress response but to other types of abiotic
 642 stresses (e.g., drought) have been lost in seagrasses. The greater homogeneity and stability of environmental
 643 conditions at sea relative to those on land is the most likely cause of these changes. Finally, only tropical
 644 seagrasses retained some of the key heat stress-related *HSFs* from WGD and WGT events (*C. nodosa*: *HSFA1* and
 645 *HSFB4*; *T. testudinum*: *HSFB2*), which could be related to their warmer native environment and higher heat stress
 646 tolerance compared to temperate seagrasses (*P. oceanica* and *Z. marina*).

647 **Supplementary Table 5.3 Average (\pm SD) number of total *HSFs* and number of *HSFs* from the three**
 648 **main classes (*HSFA*, *HSFB* and *HSFC*) in the analyzed plant genomes.**

649 The basal angiosperm *Amborella trichopoda* (11 sequences) and the freshwater eudicot plant *Utricularia gibba*
 650 (21 sequences) were not included in the table.

Plant group	No. species	Total <i>HSFs</i>	<i>HSFA</i>	<i>HSFB</i>	<i>HSFC</i>
TERRESTRIAL	14	28.0 \pm 10.0	16.5 \pm 6.6	8.3 \pm 3.7	2.1 \pm 1.8
Eudicots	9	26.1 \pm 10.3	15.1 \pm 6.5	8.3 \pm 4.3	1.2 \pm 0.9
Monocots	5	31.4 \pm 9.7	19.2 \pm 6.5	8.4 \pm 2.5	3.8 \pm 1.8
AQUATIC (monocots)	7	12.9 \pm 3.3	7.4 \pm 2.6	5.4 \pm 1.3	0.0 \pm 0.0
Freshwater	3	14.3 \pm 5.1	9.0 \pm 3.6	5.3 \pm 1.5	0.0 \pm 0.0
Seagrasses	4	11.8 \pm 1.0	6.3 \pm 0.5	5.5 \pm 1.3	0.0 \pm 0.0



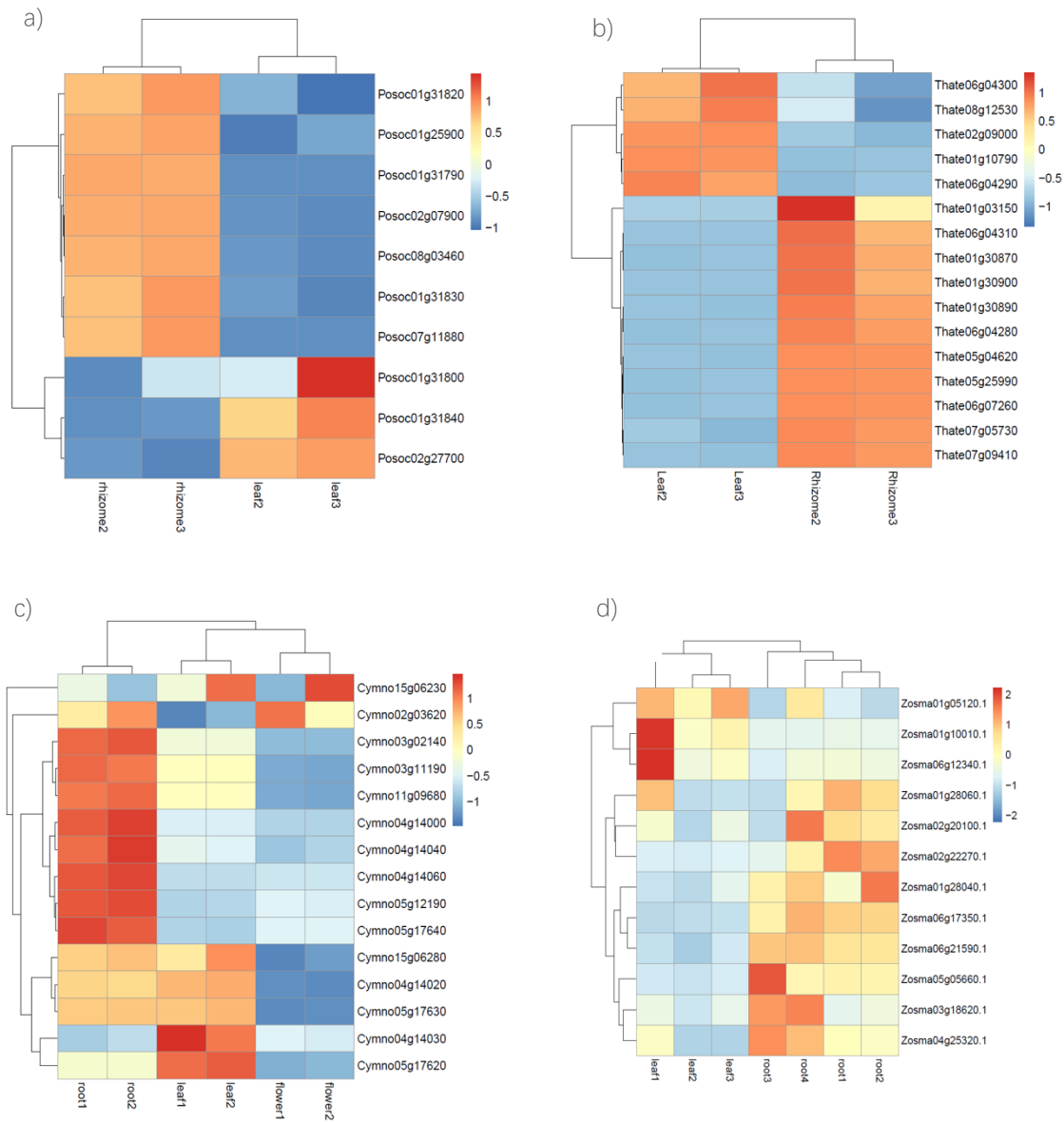
656

657 **Supplementary Figure 5.4.2** Sequence alignment of *AKT5/6/1* showing the loss of Shaker-type K^+
 658 channels with a TTYGVD-selectivity filter in all seagrasses.

659 A TTYGVD-selectivity filter is indicated by the red box; Full list of abbreviation of the species names used in the
 660 figure: AT *Arabidopsis thaliana*; SL *Solanum lycopersicum*; ATR *Amborella trichopoda*; AC *Ananas comosus*; AO
 661 *Asparagus officinalis*; AM *Avicennia marina*; BV *Beta vulgaris*; BD *Brachypodium distachyon*; CC *Coffea*
 662 *canephora*; CN *Cymodocea nodosa*; EG *Elaeis guineensis*; OS *Oryza sativa*; PO *Posidonia oceanica*; PT *Populus*
 663 *trichocarpa*; PA *Potamogeton acutifolius*; RA *Rhizophora apiculata*; SP *Spirodela polyrhiza*; TT *Thalassia*
 664 *testudinum*; TC *Theobroma cacao*; VV *Vitis vinifera*; UG *Utricularia gibba*; WA *Wolffia australiana*; ZM *Zostera*
 665 *marina*.

666

667

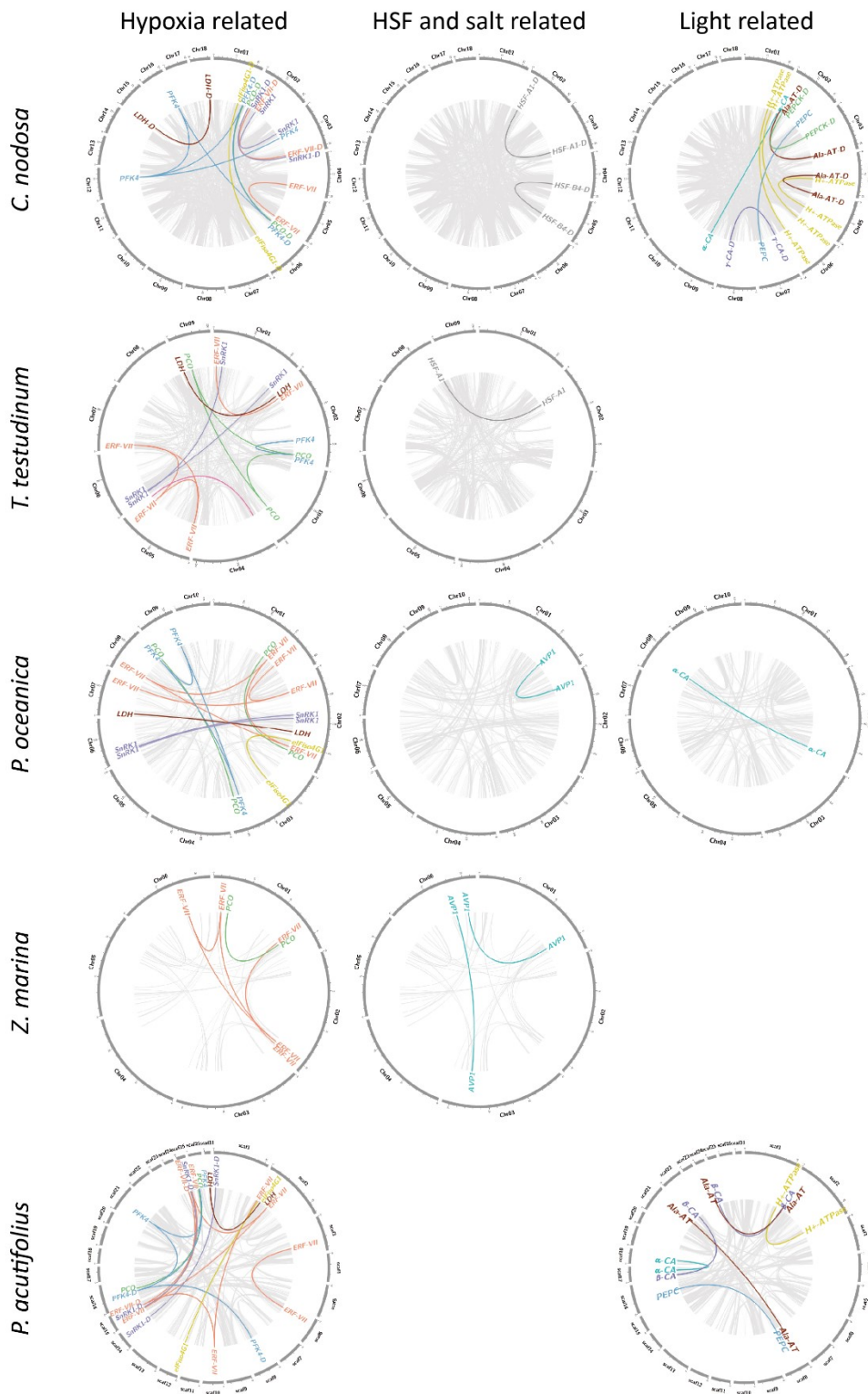


671 **Supplementary Figure 5.5.1** Differential expression of *ERF-VII*s in the four seagrass species.

672 a. Leaf vs. rhizome in *P. oceanica*, b. Leaf vs. rhizome in *T. testudinum*, c. leaf vs. rhizome vs. flower in *C. nodosa*;

673 d. leaf vs. root in *Z. marina*. Most *ERF-VII*s had higher expression in rhizomes and roots as compared to leaves in

674 four seagrasses.



Supplementary Figure 5.2 Syntenic relationship of genes mentioned in the main text for *P. oceanica*, *T. testudinum*, *Z. marina*, *C. nodosa*, and *P. acutifolius*.

Different colors represent different families derived from the whole-genome triplication event, as well as from the whole-genome duplication events for *P. oceanica*, *T. testudinum*, *Z. marina*, *C. nodosa*, and *P. acutifolius*. Gene names without the -D suffix means they are derived from the WGT event; Gene names with -D suffix in *C. nodosa*, and *P. acutifolius* means they are derived from the WGD event.

676 5.6 Light perception, circadian clock, and photosynthetic carbon acquisition

677 **Supplementary Note 5.6.1** CO₂-concentrating mechanisms (CCMs) and photosynthetic carbon 678 acquisition

679 One of the major challenges that face seagrasses is the acquisition of inorganic carbon (Ci) for photosynthesis.
680 Photosynthetic carbon limitation in the marine environment results from several physicochemical factors that
681 restrict the supply of Ci to the leaf surface of seagrasses. Of the DIC (dissolved inorganic carbon) pool in seawater,
682 the bicarbonate ion (HCO₃⁻) accounts for nearly 90%, while the primary Ci source for RuBisCO (CO_{2(aq)}) is in
683 limited supply (roughly 1% of the DIC pool) (Campbell and Fourqurean 2013). Many algae and terrestrial plants
684 have thus evolved various carbon-concentrating mechanisms (CCMs) to enhance their photosynthetic
685 capabilities under Ci limitation. The occurrence of true CCMs in seagrasses is currently still a matter of debate
686 (Larkum et al. 2017; Larkum et al. 2018), although recent findings point to the existence of biophysical CCMs
687 and demonstrate an evolutionary adaptation of RuBisCO kinetics across submerged angiosperms (Capó-Bauçà
688 et al. 2022). This closely resembles what seen in eukaryotic algae rather than that of terrestrial C4 plants
689 (biochemical CCMs). The acquisition of HCO₃⁻ for photosynthesis can occur via two (non-exclusive) basic models:
690 (a) apoplastic conversion of HCO₃⁻ to CO₂ and OH⁻, catalyzed by external carbonic anhydrases (CA); (b) direct
691 uptake of HCO₃⁻ by anion transporters or an H⁺-HCO₃⁻ symport based on H⁺-ATPase pumps (Larkum et al. 2018).
692 In *P. oceanica*, a direct HCO₃⁻ uptake via a fusicoccin-sensitive H⁺-ATPase pump has recently been demonstrated
693 (Rubio et al. 2017).

694 The four seagrass species studied, as well as *Potamogeton*, possess genes encoding for all three carbonic
695 anhydrase gene families present in higher plants (i.e., α , β and λ) (DiMario et al. 2017). Six orthogroups (OGs)
696 encode for α -CA, the most abundant family, two are associated to β -CA and two to λ -CA, respectively (Extended
697 Data Table 9). Overall, 87% of seagrass α -CA are identified as extracellular/soluble proteins (Supplementary
698 Table 5.6), which could be excreted from epidermal cells for catalyzing the conversion of HCO₃⁻ to CO₂ and OH⁻,
699 likely contributing to CCMs (Larkum et al. 2017). 64% of seagrass β -CA are targeted to the chloroplast, generally
700 to the chloroplast stroma or thylakoid membranes, while 100% of λ -CA are targeted to the mitochondria (data
701 not shown). This seems to confirm that seagrasses do have a requirement for an extracellular CA activity for
702 adequate photosynthesis, as previously demonstrated by using chemical inhibitors (Larkum et al. 2017; Larkum
703 et al. 2018; Capó-Bauçà et al. 2022). α -CA OG0013954 is exclusive of seagrasses (except for *T. testudinum*) and
704 *P. acutifolius* (Extended Data Table 9). All sequences within this OG possess the typical α -CA domain, but most
705 of them have variable mutations of the canonical His at the active site (data not shown), with unknown effects.
706 Almost all OG0013954 members, as well as other α -CA transcripts, are highly expressed in leaf tissue
707 (Supplementary Figure 5.6.1), which could support the hypothesis that they constantly increase the CO₂
708 concentration in the periplasmic space, thus enhancing its diffusive transfer to RuBisCO and the photosynthetic
709 rate in seagrass leaves. α -CA OG0028785 appears unique to *Z. marina*. A slight expansion of α -CA genes is
710 evident in seagrasses with respect to terrestrial and freshwater/brackish-water species, when considering the
711 average number of genes per species (terrestrial: 7; freshwater/brackish: 6; marine: 8). This increase in copy
712 number of α -CA in *P. oceanica* and *P. acutifolius* results from the WGT event as well as specific tandem
713 duplications.

714 A C4 or C3-C4 intermediate photosynthetic metabolism could also contribute to CCMs in seagrasses. We
715 screened the components of the C4 photosynthesis pathway as outlined in Rao et al. (2016). The analysis of C4-
716 pathway related genes revealed that all genes typical of terrestrial plants are present in seagrasses (Extended
717 Data Table 9). However, this is not diagnostic of a functional C4 biochemical pathway, as they have functions
718 other than C4 photosynthesis. In addition, none of the studied species possesses the Ser residue characteristic
719 of C4 Phosphoenolpyruvate carboxylase (PEPC) (data not shown). In *C. nodosa*, which has been previously
720 hypothesized to be a C4 species (Koch et al. 2013), there were 15 C4-related genes retained specifically after
721 WGT or WGD events, including two encoding for PEPC (retained after WGD). Similarly, in *P. acutifolius*, 17 C4-
722 related genes were retained following WGT or WGD (Extended Data Table 9). We cannot exclude the presence

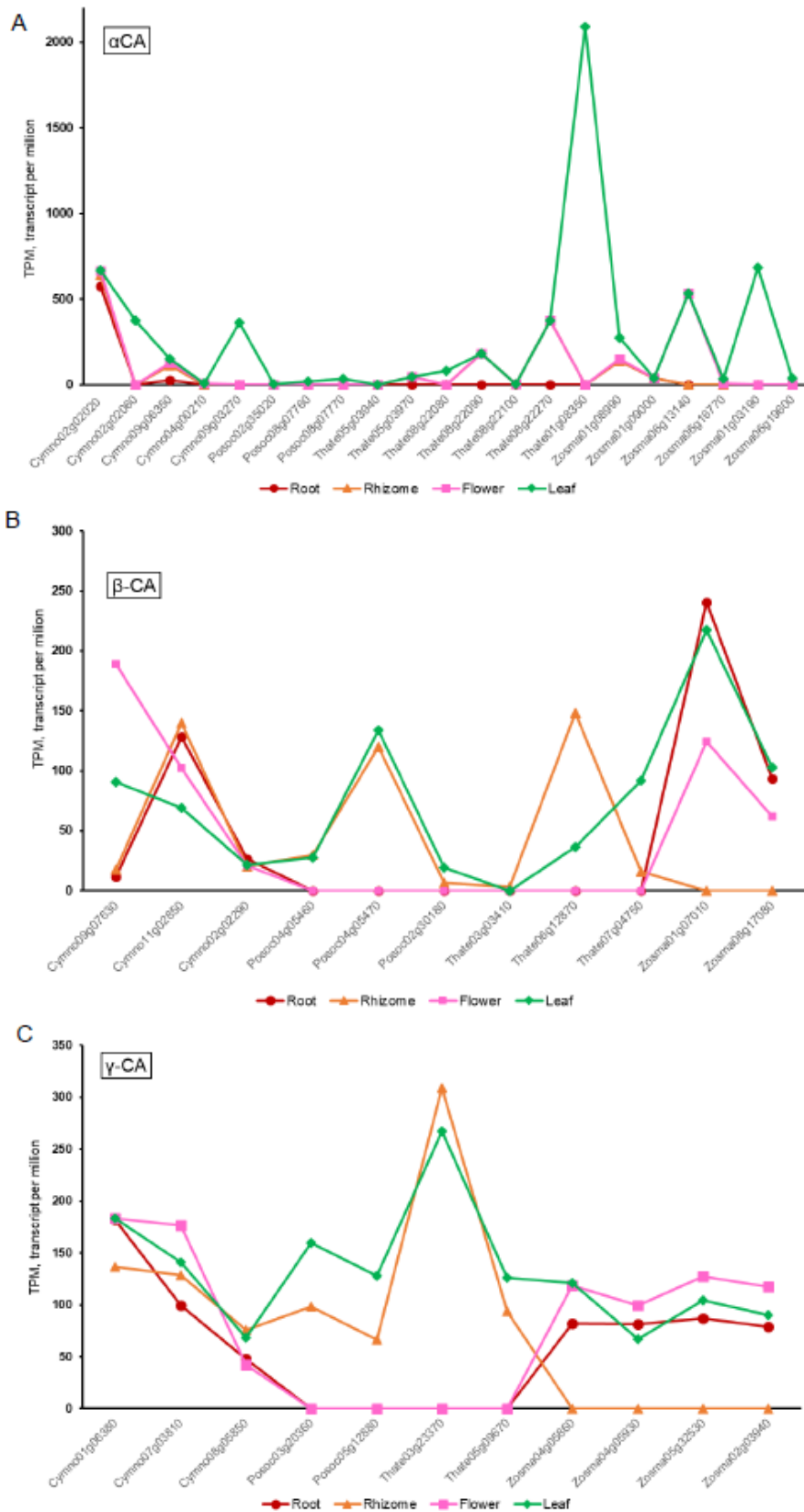
723 of some kind of C3-C4 intermediate metabolism at least in these species, similar to what observed in the
 724 freshwater hydrophyte *Hydrilla verticillata* (Hydrochariraceae), where facultative single cell C4 type
 725 photosynthesis is known to occur (Rao et al. 2002).

726 Four orthogroups are annotated as boron transporters (HCO₃⁻ transporter family) in the studied species
 727 (Extended Data Table 9). One orthogroup is completely absent in aquatic species (OG0012730). Regarding
 728 proton pumps (H⁺-ATPases), only three OGs encoded for plasma membrane H⁺-ATPases (Extended Data Table
 729 9), whose role could be associated to bicarbonate symport for photosynthesis. Overall, seagrasses have less
 730 genes, on average per species, than terrestrial ones, but several of them were retained following WGT or WGD
 731 events (7 in *C. nodosa*, 6 in *P. oceanica* and 3 in *P. acutifolius*) (Extended Data Table 9). This confirms that ATPase
 732 pumps could still play a role in seagrasses for driving symport of different substances across the plasma
 733 membrane, as demonstrated by Rubio et al. (2017) in *P. oceanica* for bicarbonate uptake.

734 **Supplementary Table 5.6 Prediction of sub-cellular localizations of α -Carbonic Anhydrases (α -CA) in**
 735 **the studied seagrass species and *P. acutifolius*.**

736 The presence of transit and signal peptides for specific cellular compartments and the determination of the
 737 protein type (soluble vs. membrane) was inferred by using multiple bioinformatics approaches. Genes in red are
 738 retained after WGT or WGD (in red) events.

Orthogroup	<i>C. nodosa</i>		<i>P. oceanica</i>		<i>T. testudinum</i>		<i>Z. marina</i>		<i>P. acutifolius</i>	
OG0000299	Cymno02g02020	Endoplasmic reticulum, Membrane	Posoc02g30740	Extracellular, Soluble	Thate05g03940	Sec/SPI; Extracellular, Soluble	Zosma01g08990	Sec/SPI; Extracellular, Soluble	Potac_scaff old2.170	Sec/SPI; Extracellular, Soluble
	Cymno02g02030	Sec/SPI; Extracellular, Soluble	Posoc02g30790	Sec/SPI; Extracellular, Soluble	Thate05g03970	Sec/SPI; Extracellular, Soluble	Zosma01g09000	Sec/SPI; Extracellular, Soluble	Potac_scaff old2.171	Sec/SPI; Extracellular, Soluble
	Cymno02g02040	Sec/SPI; Extracellular, Soluble	Posoc02g35020	Sec/SPI; Extracellular, Soluble	Thate08g22080	Sec/SPI; Extracellular, Soluble	Zosma06g13140	Sec/SPI; Extracellular, Soluble	Potac_scaff old2.173	Sec/SPI; Extracellular, Soluble
	Cymno02g02050	Extracellular, Soluble	Posoc06g06700	Endoplasmic reticulum, Membrane	Thate08g22090	Sec/SPI; Extracellular, Soluble	Zosma06g16770	Sec/SPI; Extracellular, Soluble		
	Cymno02g02060	Lysosome/ Vacuole, Soluble			Thate08g22100	Sec/SPI; Extracellular, Soluble				
	Cymno09g06350	Sec/SPI; Extracellular, Soluble			Thate08g22270	Sec/SPI; Extracellular, Soluble				
OG0002316	Cymno04g00210	Extracellular, Soluble	Posoc01g00410	Sec/SPI; Extracellular, Soluble	Thate01g08320	Cytoplasm, Soluble			Potac_scaff old2.352	Cytoplasm, Soluble
					Thate01g08350	Sec/SPI; Extracellular, Soluble			Potac_scaff old9.158	Sec/SPI; Extracellular, Soluble
					Thate06g11790	Sec/SPI; Extracellular, Soluble				
OG0013954	Cymno09g03270	Sec/SPI; Extracellular, Soluble	Posoc08g07760	Sec/SPI; Extracellular, Soluble			Zosma06g19600	Sec/SPI; Extracellular, Soluble	Potac_scaff old17.661	Sec/SPI; Extracellular, Soluble
			Posoc08g07770	Sec/SPI; Extracellular, Soluble			Zosma06g20930	Sec/SPI; Extracellular, Soluble	Potac_scaff old17.663	Sec/SPI; Extracellular, Soluble
									Potac_scaff old18.302	Sec/SPI; Extracellular, Soluble
OG0028785							Zosma01g03190	Sec/SPI; Extracellular, Soluble		



739
740
741
742

Supplementary Figure 5.6.1 Differential expression of α -CA, β -CA and λ -CA in root, rhizome, flower, and leaf tissues of the studied seagrass species.

743 **Supplementary Note 5.6.2 Photosynthesis**

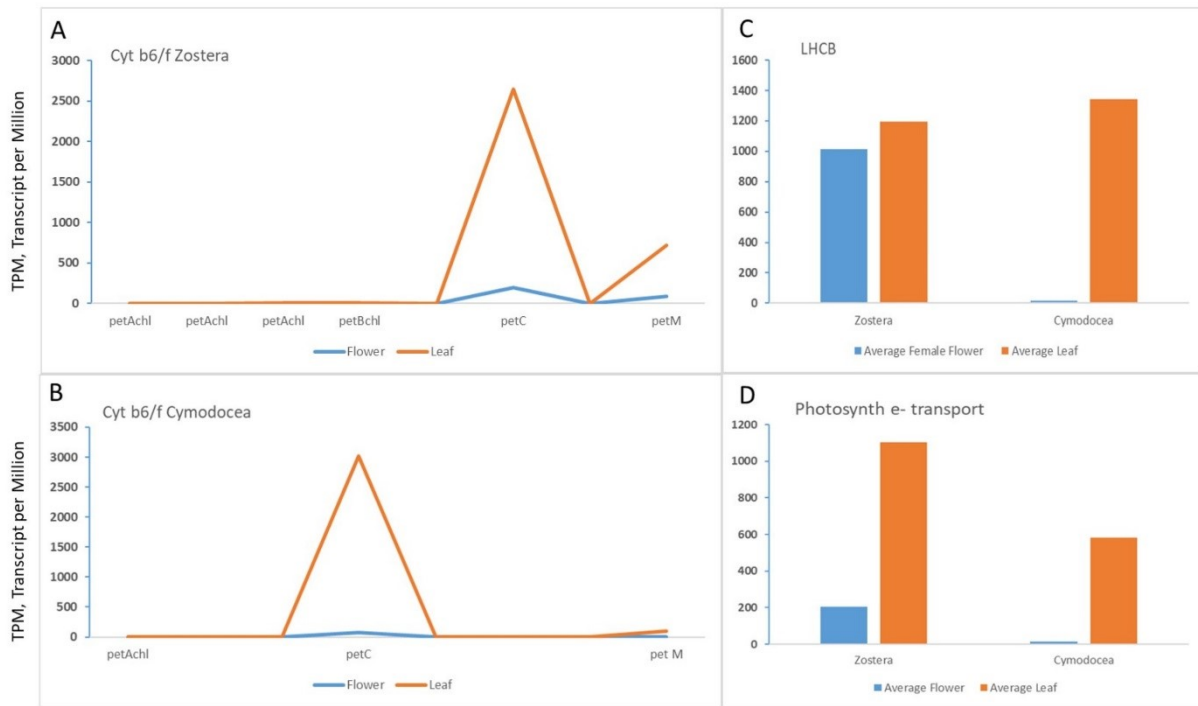
744 Seagrasses face different light environments according to the depth and latitude where they live. Irradiance
745 decreases with depth, and light quality is also altered along the water column. In order to investigate if the
746 adaptation to the light environment experienced by seagrasses in their submerged marine life imposed
747 evolutionary adaptations resulting in expansion/reduction of gene families, the following categories of genes
748 have been analyzed (from KEGG Photosynthesis Proteins - *Arabidopsis thaliana*): Photosystem and electron
749 transport system: Photosystem II (P680 chlorophyll a); Photosystem I (P700 chlorophyll a); Cytochrome b6/f
750 complex; Photosynthetic electron transport; F-type ATPase; Antenna proteins: Light-harvesting chlorophyll-
751 protein complex A and B (*LHCA*, *LHCB*). The total number of genes present in the orthology groups for the
752 investigated gene families are presented in the Supplementary Figure 5.6.2. In comparison with the other
753 Alismatales, seagrasses have a noteworthy expansion in genes of the *psA/psB* complexes and of the *LHCB*.
754 Looking at the last group (*LHCB*) in particular, *Posidonia*, *Thalassia* and *Zostera* have a clear expansion in
755 comparison to *Cymodocea*, which could be related to the larger depth gradient experienced by the formers.
756 Genes for the *psA/psB* complexes are coded by the plastid genome, as well as for the Cytochrome b6/f complex
757 and the F-type ATPase. Nevertheless, in most of case for seagrass species and for some of the other species
758 included in the analysis there are also extra copies in the nuclear genome (numbers in red in the Supplementary
759 Figure 5.6.2). *T. testudinum*, for example, has 6 nuclear copies of *psbA*. Nevertheless, nuclear copies of
760 chloroplast genes are not expressed. See for example genes for the Cyt b6/f complex in *C. nodosa* and *Z. marina*
761 leaf and flower tissues (Supplementary Figure 5.6.3 A and B).

762 Seagrass species analyzed feature different floral morphology. For two species (*Z. marina* and *C. nodosa*) gene
763 expression data have been obtained also from floral tissue. Results indicate that *Zostera* flowers are
764 photosynthetically active, while *Cymodocea* flowers are not. *LHCs* seem to be active at the same levels of leaf
765 tissue while electron transport genes express at a lower level (Supplementary Figure 5.6.3 C and D).

766 In *Z. marina*, photosynthesis occurs mainly in pistillate (“female”) flowers. This could be due to the maturation
767 timing of staminate (“male”) and pistillate flowers, as suggested for *P. oceanica*, where it was also found that
768 only pistillate flowers expressed photosynthetic genes (Entrambasaguas et al. 2017).

769

	Orthogroup	Gene name	Seagrasses				Freshwater species			Mangrove		Moconots					Eudicots							ATR	
			CN	PO	TT	ZM	PA	SP	WA	AM	RA	OS	BD	AC	EG	AO	BV	UG	SL	CC	VV	PT	AT		TC
Psa	OG0001413	<i>PsaA</i>	2	4	2	2	0	0	0	0	1	5	3	1	0	0	5	4	16	1	2	6	1	0	0
	OG0004325	<i>PsaB</i>	1	1	3	1	0	1	0	0	3	3	2	2	0	0	7	2	4	0	2	1	1	0	0
	OG0012100	<i>PsaC</i>	1	1	1	1	0	0	0	0	0	3	2	1	0	0	1	0	1	0	1	1	1	0	0
	OG0004391	<i>PsaD2</i>	1	1	1	1	4	1	1	2	2	1	1	1	0	1	1	2	1	0	1	2	2	2	1
	OG0004272	<i>PsaE</i>	2	1	2	1	4	2	0	1	2	1	1	1	1	0	1	3	2	1	1	1	2	2	1
	OG0006616	<i>PsaF</i>	1	1	1	2	2	2	2	2	2	1	1	1	0	0	1	2	1	0	1	1	1	1	1
	OG0012149	<i>PsaI</i>	1	1	1	1	0	2	0	0	0	0	4	0	0	0	0	0	2	0	1	1	1	0	0
	OG0012852	<i>PsaJ</i>	1	1	1	1	0	1	0	0	0	0	3	0	0	0	0	0	2	0	0	1	1	0	0
	OG0006947	<i>PsaO</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	2
	OG0008605	<i>PsaK</i>	1	1	1	1	1	1	1	1	1	1	1	2	1	0	1	1	1	1	1	2	1	1	1
	OG0006761	<i>PsaL</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	2	1	2	1
	OG0007957	<i>PsaG</i>	1	1	1	1	2	1	1	1	1	1	1	0	2	0	3	2	1	0	1	2	1	0	1
	OG0005369	<i>PsaH</i>	1	1	1	1	1	1	1	1	2	2	1	1	2	1	1	1	3	1	1	1	2	1	1
	OG0010572	<i>PsaN*</i>	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	2	1	1	1
	OG0010507	<i>PsaN*</i>	0	1	1	1	1	1	1	1	1	0	1	2	1	1	1	0	0	1	1	2	1	1	1
	OG0003908	<i>PsaP</i>	2	2	1	1	2	2	1	1	2	2	2	2	2	2	1	1	1	1	1	2	1	1	1
	Psb	OG0001201	<i>psbA</i>	2	1	7	2	0	2	3	0	3	5	3	4	1	0	6	3	11	1	1	2	1	0
OG0001495		<i>psbB</i>	1	3	1	2	0	1	1	0	4	5	7	6	1	1	9	1	3	0	2	3	1	0	1
OG0003337		<i>psbC</i>	1	1	2	2	0	1	2	0	0	3	2	1	0	0	6	2	7	0	2	2	1	0	1
OG0008341		<i>psbD*</i>	1	1	2	2	0	1	0	0	0	3	2	0	0	0	3	2	2	0	2	2	1	0	1
OG0022354		<i>psbD*</i>	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
OG0008317		<i>psbE</i>	1	1	1	2	0	1	0	0	0	3	5	1	0	0	5	0	1	0	1	1	1	0	1
OG0013128		<i>psbF</i>	1	1	1	1	0	1	0	0	0	0	3	0	0	0	0	0	1	0	0	1	1	0	0
OG0013809		<i>psbL</i>	1	1	1	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	1	0	0
OG0013127		<i>psbJ</i>	1	1	1	1	0	0	0	0	0	0	4	0	0	1	0	0	1	0	0	0	1	0	0
OG0010941		<i>psbH</i>	1	1	1	3	0	1	0	0	0	4	1	0	0	0	0	0	1	0	0	3	1	0	0
OG0013442		<i>psbI</i>	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	2	1	0	0
OG0011920		<i>psbK</i>	1	1	1	2	0	0	0	0	0	4	0	4	0	0	0	0	0	0	1	1	1	0	0
OG0013443		<i>psbM</i>	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	2	1	0	0
OG0005164		<i>psbO</i>	2	1	1	1	1	1	1	2	1	1	1	1	2	1	1	2	2	1	1	2	2	1	1
OG0009830		<i>psbP*</i>	1	1	1	0	1	1	1	1	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1
OG0007493		<i>psbP*</i>	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1
OG0003480		<i>psbP*</i>	2	1	1	2	3	1	1	3	0	1	1	3	3	2	1	2	1	1	1	2	1	1	1
OG0011018		<i>psbQ*</i>	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
OG0010574		<i>psbQ*</i>	1	0	0	1	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
OG0005668		<i>psbQ*</i>	1	1	1	1	2	3	1	1	1	1	1	1	2	0	1	2	1	1	1	2	2	1	1
OG0003547		<i>psbR</i>	1	2	2	1	2	1	1	2	1	3	3	1	2	1	1	2	1	1	4	1	1	1	1
OG0005051		<i>psbS</i>	1	2	1	1	2	2	1	1	2	3	1	1	1	1	2	1	1	1	1	1	1	1	1
OG0012853		<i>psbT</i>	1	1	1	1	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	3	1	0	0
OG0003259		<i>psbW</i>	1	1	1	1	2	2	2	2	2	2	2	1	2	2	1	2	1	2	2	2	1	3	1
OG0007350		<i>psbY</i>	1	1	1	1	1	2	1	0	2	1	1	0	2	1	1	3	1	1	0	2	1	1	1
OG0011749		<i>psbZ</i>	1	1	1	1	0	1	0	0	0	4	1	2	0	0	0	0	3	0	0	1	1	0	0
OG0007852		<i>psbZ7</i>	1	1	1	1	1	1	0	2	1	1	1	2	1	1	0	2	1	1	1	2	1	1	1
OG0008772		<i>psbZ8</i>	1	1	1	1	3	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cytochrome b6/f complex		OG0002741	<i>petA</i>	2	2	2	4	0	1	0	0	0	5	6	1	0	0	6	1	3	1	2	2	1	0
	OG0006791	<i>petB</i>	1	1	1	2	0	0	0	0	0	3	5	4	0	0	2	0	3	0	0	4	1	0	0
	OG0004917	<i>petC</i>	1	1	2	1	4	1	2	1	1	1	1	1	1	1	1	1	1	1	3	2	1	1	1
	OG0008192	<i>petD</i>	1	1	1	1	0	2	0	0	0	4	3	2	0	0	4	0	4	0	0	2	1	0	0
	OG0012851	<i>petG</i>	1	1	1	1	0	1	0	0	0	0	4	0	0	0	0	0	1	0	0	1	1	0	0
	OG0013129	<i>petL</i>	1	1	1	1	0	1	0	0	0	0	2	0	0	0	0	0	2	0	1	0	1	0	0
	OG0005169	<i>petM</i>	1	1	1	1	0	2	1	1	2	1	1	1	2	1	1	3	1	1	1	4	1	1	1
OG0013068	<i>petN</i>	1	1	1	1	0	1	0	0	0	0	3	1	0	0	1	0	0	0	0	0	1	0	0	
Photosynthetic e-transport	OG0003691	<i>petE</i>	0	3	1	2	2	2	1	2	1	1	1	1	2	3	1	1	1	2	1	2	2	1	1
	OG0000289	<i>petF*</i>	2	3	2	1	12	2	12	5	2	5	6	5	5	3	3	3	7	4	1	7	4	2	3
	OG0008036	<i>petF*</i>	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
	OG0009442	<i>petF*</i>	1	1	1	1	1	0	2	1	1	1	1	1	1	0	1	1	1	1	1	2	1	1	1
	OG0004729	<i>petH</i>	1	1	1	1	2	1	1	1	1	2	2	1	1	2	1	1	2	1	2	2	2	1	1
OG0006612	<i>petJ</i>	1	1	1	1	1	1	2	2	1	1	1	4	1	1	0	0	2	1	1	1	1	1	1	
F-type ATPase	OG0000400	<i>atpA</i>	1	6	9	1	1	1	3	0	7	7	2	9	1	1	13	3	11	2	1	4	3	1	1
	OG0001351	<i>atpB</i>	2	2	1	7	0	0	1	0	2	6	5	3	0	0	7	2	3	1	0	2	1	0	9
	OG0005914	<i>atpD</i>	2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	1	1	1	2	1	1	1
	OG0006895	<i>atpE</i>	1	2	1	3	0	0	0	0	0	5	5	1	0	1									



782

783 **Supplementary Figure 5.6.3** Differential expression of Cyt b6/f complex, LHCb and electron transport 784 genes in *Z. marina* and *C. nodosa*.

785 A, B: expression of nuclear copies of chloroplast genes (*chl*) and of nuclear genes in leaf and flower tissue. C, D:
786 comparison between gene expression in leaf and flower tissue.

787

788 **Supplementary Note 5.6.3** Light Signaling & Circadian Clock

789 Seagrasses mostly conserved the full repertoire of orthologous genes for photosensory proteins and signalling
790 systems, evolved in the green lineage during the different stages of plant terrestrialization (Supplementary
791 Figure 5.6.4) (Han et al. 2019; Jing and Lin 2020). Seagrasses conserved genes for all the three classes of
792 photoreceptors UV-A/Blue, UV-B and RED/FAR-RED typical of higher plants, with few exceptions (Supplementary
793 Figure 5.6.4) as well as at least one ortholog for each of the three main classes of UVA/Blue photoreceptors, that
794 are Phototropins (*PHOT*), Cryptochromes (*CRY*) and the LOV/F-box protein (*FKF/LKP/ZTL*) (Supplementary Figure
795 5.6.4).

796 The UV-B receptor (UVB-Resistance 8) that is already known to be absent in *Z. marina* is still present in the other
797 three species although the predicted protein sequence of *UVR8* in *P. oceanica* has a shorter N terminus
798 compared to other species (data not shown) and lacks the C27 domain (Supplementary Figure 5.6.5). This is a
799 region of 27 amino acids from the C terminus that mediates the interaction with proteins repressor of
800 photomorphogenesis 1, 2 (*RUP1* and *RUP2*) (Yin et al. 2015), two proteins belonging to UV-B signalling including
801 UV-B acclimation and tolerance. *RUP1* and *RUP2* are missing in the *Z. marina* genome (Supplementary Figure
802 5.6.4). These observations indicate that UV-B tolerance and the downstream regulation signalling pathways vary
803 among species and are related to the relative light habitat features.

804 Red/far-red photoreceptors (Phytochromes) are present in a variety of organisms (Rockwell and Lagarias 2020).
805 The phytochrome structure in plants is highly conserved, showing the same domain architecture in all members
806 of the streptophyte (charophyte algae and land plants) PHY1/2 lineage, having originated in a common ancestor

807 (Li et al. 2015; Rockwell and Lagarias 2020). In some algal lineages, such as Zygnematales and Coleochaetales,
808 phytochromes also show non-canonical forms (Li et al.2015). In seed plants, phytochromes underwent lineage-
809 specific gene duplications, leading to three main forms (phyA, phyB, phyC), plus two additional forms (phyD and
810 phyE), which are restricted to some taxa (Mathews 2010). The number of phytochromes varies among species:
811 eudicots tend to have two or three phytochromes genes (up to five *PHYA-PHYE* in *A.thaliana*), while monocots
812 tend to have a lower number of genes, generally only one gene for *PHYA* and one for *PHYB*. The transition to a
813 submerged marine environment did not lead to a general reduction of phytochrome genes (Supplementary
814 Figure 5.6.4); indeed, all the four species investigated possess at least one gene for phytochrome A (*PHYA*) and
815 phytochrome B (*PHYB*) while *P. oceanica* and *T. testudinum* also possess an orthologous gene for *PHYC*.
816 Furthermore, *P. oceanica* and *C. nodosa* have a unique orthologous cluster (OG0026441) for a Phytochrome E
817 (*PHYE*) (Supplementary Figure 5.6.6) often absent in monocotyledonous plants (Smith 2000; Mathews 2006).
818 After WDG and WGT, *P. oceanica* and *C. nodosa* as well as *P. acutifolius* retained duplicated genes for *CRY* or
819 *PHOTs* (Supplementary Figure 5.6.4) while *P. oceanica* and *P. acutifolius* also for *ZTL/FKF1*.

820 Components of the downstream light signaling pathways such phytochrome interacting factors (*PIFs*),
821 constitutive photomorphogenic protein 1 (*COP1*) and elongated hypocotyl 5 (*HY5*) are still present
822 (Supplementary Figure 5.6.4) with several orthologous genes comparable with other aquatic and land species.
823 Also, the repertoire of transcription factors essential for photomorphogenesis and seed emergence, like the far-
824 red elongated hypocotyl 1,3 (*FHY1/3*), far-red-impaired response1 (*FAR1*) and long after far-red light 1 (*LAF1*) is
825 the one typical of land angiosperms. However, further functional studies must investigate if those genes have
826 also conserved the same pattern of expression of land plants, especially during critical stages of seed setting and
827 plant development.

828 Perception of surrounding light cues is critical also for the entrainment of the circadian clock system. The
829 circadian clock regulates a plethora of processes that affect physiology and life cycle in plants, such as daily
830 water and carbon availability and hormone signalling pathways (McClung 2019). All seagrass species, apart from
831 *T. testudinum*, lost ortholog genes for *Timing of Cab* (*TOC1*) (Supplementary Figure 5.6.4). *TOC1* is one of the
832 key clockwork components of the evening transcriptional-translational loop (Harmer 2009) belonging to the
833 PSEUDO RESPONSE REGULATOR (*PRR*) family with a crucial function in the integration of light signals to the
834 circadian control (Pokhilko et al. 2013). *TOC1* has also a central role in adapting plant physiology to drought
835 (Legnaioli et al. 2009; Wang et al. 2020) and in regulating the day-night energy metabolism (Cervela-Cardona et
836 al. 2021). Remarkably, *TOC1* is also lost in the freshwater *P. acutifolius* and *W. australiana*, the latter showing a
837 reduced circadian time control of gene expression in comparison with *Arabidopsis* (Michael et al. 2021). The loss
838 of some genes related to the circadian system in a large part of marine and freshwater species can suggest that,
839 in the aquatic environment, the absence of some environmental stressors typical of land habitats, such as water
840 deficit, has led to a reduction of the regulative constraints for daily management of some metabolic and
841 developmental plant processes. Further functional studies could highlight changes in regulative networks
842 mediated by circadian clock genes and their implication for seagrass adaptation to marine environments.

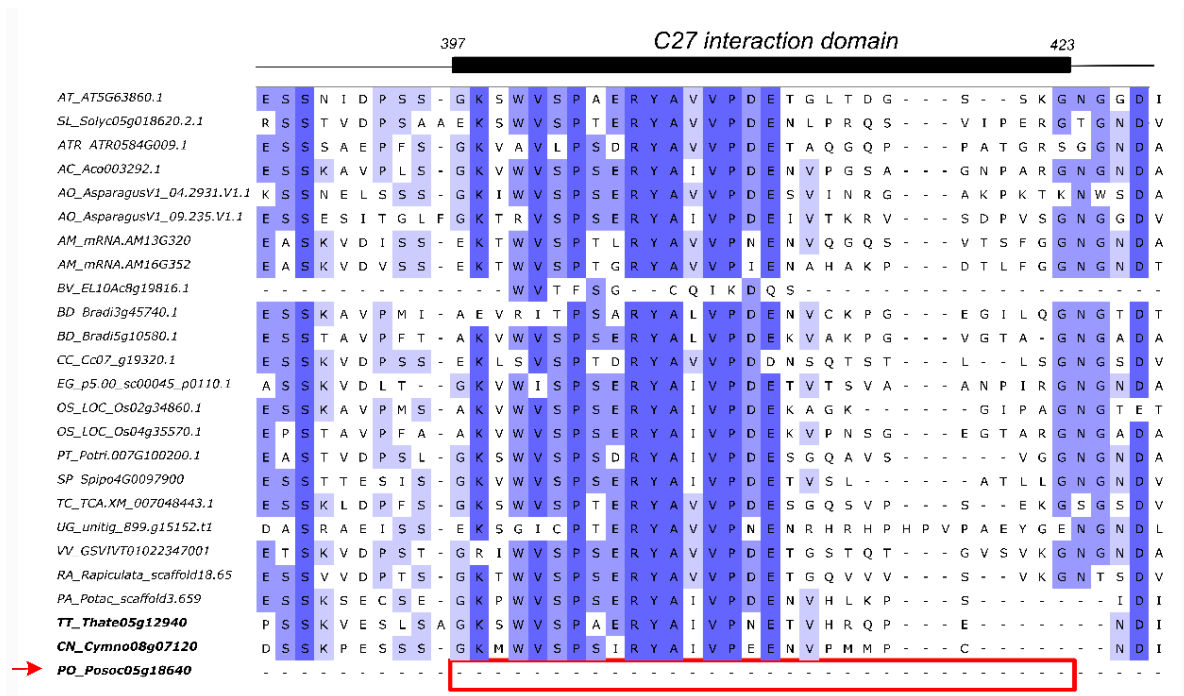
843 *C. nodosa*, *P. oceanica* and *T. testudinum* retained, after WGT and WGD events, one gene each related to the
844 circadian clock and photoperiodism, respectively *LNK1*, *ZTL* and *GI* (Supplementary Figure 5.6.4).

845

	Orthogroup	Gene name	Seagrasses				Freshwater species			Mangrove		Moconots					Eudicots								ATR		
			CN	PO	TT	ZM	PA	SP	WA	AM	RA	OS	BD	AC	EG	AO	BV	UG	SL	CC	VV	PT	AT	TC			
Gene families containing photoreceptors	OG0007303	<i>PHYA</i>	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	
	OG0003273	<i>PHYB/E</i>	1	1	1	1	1	1	1	2	3	1	1	1	2	1	1	4	3	1	2	2	3	2	1	1	
	OG0026441	<i>PHYE</i>	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	OG0011336	<i>PHYC</i>	0	1	1	0	1	1	1	2	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	
	OG0008432	<i>UVR8</i>	1	1	1	0	0	1	1	2	1	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	
	OG0004552	<i>CRY1</i>	2	1	1	1	1	1	1	3	2	2	2	1	2	1	1	1	2	1	1	2	1	1	1	1	
	OG0006828	<i>CRY2</i>	1	2	2	0	1	1	2	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	
	OG0010438	<i>CRY3</i>	1	1	1	0	1	1	1	0	1	1	1	1	1	2	1	0	1	2	1	1	1	1	1	1	
	OG0001443	<i>PHOT1, PHOT2</i>	3	3	2	2	2	2	4	2	2	3	2	2	2	2	2	2	2	2	2	3	2	2	2	2	
	OG0006742	<i>PAS/LOV</i>	1	1	1	1	1	1	1	2	1	1	1	2	2	1	1	0	2	1	1	2	1	1	1	1	
	OG0003641	<i>ZTL, ADO1</i>	1	2	2	1	2	2	1	4	1	2	2	1	2	2	1	1	1	1	1	2	1	1	1	1	
	OG0017621	<i>ADO2</i>	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	OG0010351	<i>FKF1, ADO3</i>	1	1	1	0	2	1	2	1	1	1	1	0	1	0	1	0	1	1	1	2	1	1	1	1	
	OG0006492	<i>RUP1, RUP2</i>	1	3	1	0	1	1	1	4	1	1	1	1	1	1	0	1	1	1	1	2	2	1	1	1	
	OG0002079	<i>HYS</i>	3	2	1	1	2	3	2	3	2	2	3	3	3	2	1	1	1	1	2	3	2	1	1	2	
	OG0003554	<i>COP1</i>	1	1	1	2	1	1	1	3	2	1	1	1	1	1	2	2	2	2	2	3	1	2	1	1	
	OG0004587	<i>SPA1</i>	1	1	1	1	1	1	1	2	1	1	1	1	2	1	2	0	2	2	2	3	2	2	1	1	
	OG0002845	<i>COR28</i>	2	2	2	1	0	0	2	2	2	3	2	2	3	1	1	2	1	1	3	2	2	1	1	1	
	OG0008799	<i>CSU2</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	
	OG0010907	<i>COP10</i>	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	
	OG0009317	<i>DET1</i>	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	
	OG0004432	<i>DDB1B</i>	1	1	1	1	6	1	1	2	1	2	2	1	2	1	2	0	1	0	1	1	2	1	1	1	
	OG0003768	<i>PIF3/15</i>	2	1	1	1	0	1	2	3	3	0	1	2	3	3	1	1	1	1	1	3	1	2	1	1	
	OG0016987	<i>PIF-like 15</i>	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	OG0009778	<i>PIF-like 15</i>	2	2	2	3	4	1	0	0	0	2	2	3	0	0	0	0	0	0	1	0	1	0	1	0	
	OG0009607	<i>PIF-like 13</i>	2	2	0	0	1	2	1	0	0	4	2	1	0	2	1	1	0	0	1	0	3	0	1	1	
	OG0004930	<i>PIF-like 13</i>	2	2	2	0	0	1	2	1	1	1	2	3	2	4	0	1	1	1	1	2	2	1	0	1	
	OG0007588	<i>UNE10</i>	1	1	1	0	0	0	1	3	2	1	1	1	2	1	1	0	2	1	2	2	1	1	1	1	
	OG0003522	<i>SRR1</i>	1	1	5	3	1	1	1	1	1	1	1	1	3	1	0	2	1	2	1	3	1	1	1	1	
	OG0006894	<i>FHY1</i>	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	0	2	1	1	1	1	1	1	2	
	OG0000045	<i>FAR1, FHY3</i>	8	8	7	6	6	11	6	10	19	22	17	1	0	1	10	0	16	0	9	26	12	17	0	1	
	OG0003983	<i>FAR1-related</i>	1	2	1	1	1	2	1	1	2	1	1	0	0	3	2	0	2	0	3	4	1	2	3	2	
	OG0000000	<i>LAF1 (MYB-related)</i>	31	31	25	20	15	20	27	25	10	40	40	38	3	29	26	19	41	28	37	70	44	43	22	3	
	OG0000541	<i>FT</i>	1	2	3	4	5	3	4	2	2	8	9	5	6	5	2	2	3	3	0	2	2	1	1	1	
	OG0006964	<i>FRI</i>	0	1	0	0	0	2	1	1	1	1	1	1	1	1	0	2	6	1	2	2	1	1	1	1	
	OG0002154	<i>FRI-like</i>	1	1	1	1	1	2	2	3	2	3	2	2	3	2	2	2	3	1	2	3	1	3	2	2	
	OG0004911	<i>VRNS, VILL1</i>	1	1	1	1	1	1	1	3	2	2	2	1	1	1	1	2	1	1	2	2	1	1	1	1	
	Gene families involved in circadian clock toolkit genes	OG0003641	<i>ZTL, ADO1</i>	1	2	2	1	2	2	1	4	1	2	2	1	2	2	1	1	1	1	1	2	1	1	1	1
		OG0017621	<i>ADO2</i>	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
		OG0010351	<i>FKF1, ADO3</i>	1	1	1	0	2	1	2	1	1	1	1	0	1	0	1	0	1	1	1	2	1	1	1	1
		OG0003077	<i>LHY/CCA1</i>	2	1	2	1	2	2	2	1	2	1	1	2	3	1	1	5	1	1	1	2	2	1	1	1
		OG0010029	<i>TOC1</i>	0	0	1	0	0	1	0	2	2	1	1	2	1	1	1	2	2	1	1	1	1	1	1	1
		OG0010705	<i>PRR2</i>	1	1	1	0	1	1	1	0	2	0	0	1	1	1	1	0	2	1	1	2	1	1	1	1
		OG0001611	<i>PRR9/5</i>	1	3	2	1	2	3	1	3	3	1	1	3	1	5	2	5	2	1	2	4	2	2	1	1
		OG0003230	<i>PRR7/3</i>	1	1	1	1	1	1	1	2	2	2	2	2	2	1	2	0	2	2	3	2	2	3	1	1
OG0004910		<i>GI</i>	1	1	2	1	1	1	1	2	2	1	1	1	2	2	1	2	2	1	1	2	1	1	1	1	
OG0004704		<i>LUX</i>	2	2	3	1	0	1	2	1	1	1	1	1	1	0	2	2	2	1	1	2	2	1	1	1	
OG0004484		<i>CO</i>	1	1	2	0	0	1	1	2	2	2	2	2	2	1	1	0	3	1	1	2	3	2	1	1	
OG0000552		<i>ELF4</i>	1	2	1	2	1	1	2	5	4	3	3	6	5	0	3	6	6	3	3	7	5	4	2	2	
OG0003752		<i>LNK1</i>	2	2	1	0	1	0	2	1	5	1	1	2	4	1	1	3	2	1	2	1	1	1	1	1	
OG0022062		<i>LNK1</i>	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
OG0003157		<i>LNK2</i>	1	1	2	0	2	1	1	2	2	1	1	1	3	5	1	1	0	4	1	7	1	1	0	1	
OG0011416		<i>LNK3/4</i>	0	1	0	0	1	1	1	3	0	0	0	1	0	0	1	0	2	0	1	2	2	1	1	1	
OG0002371		<i>TIC</i>	1	1	1	0	1	3	1	4	7	1	1	1	2	1	0	4	1	1	2	5	2	2	2	2	

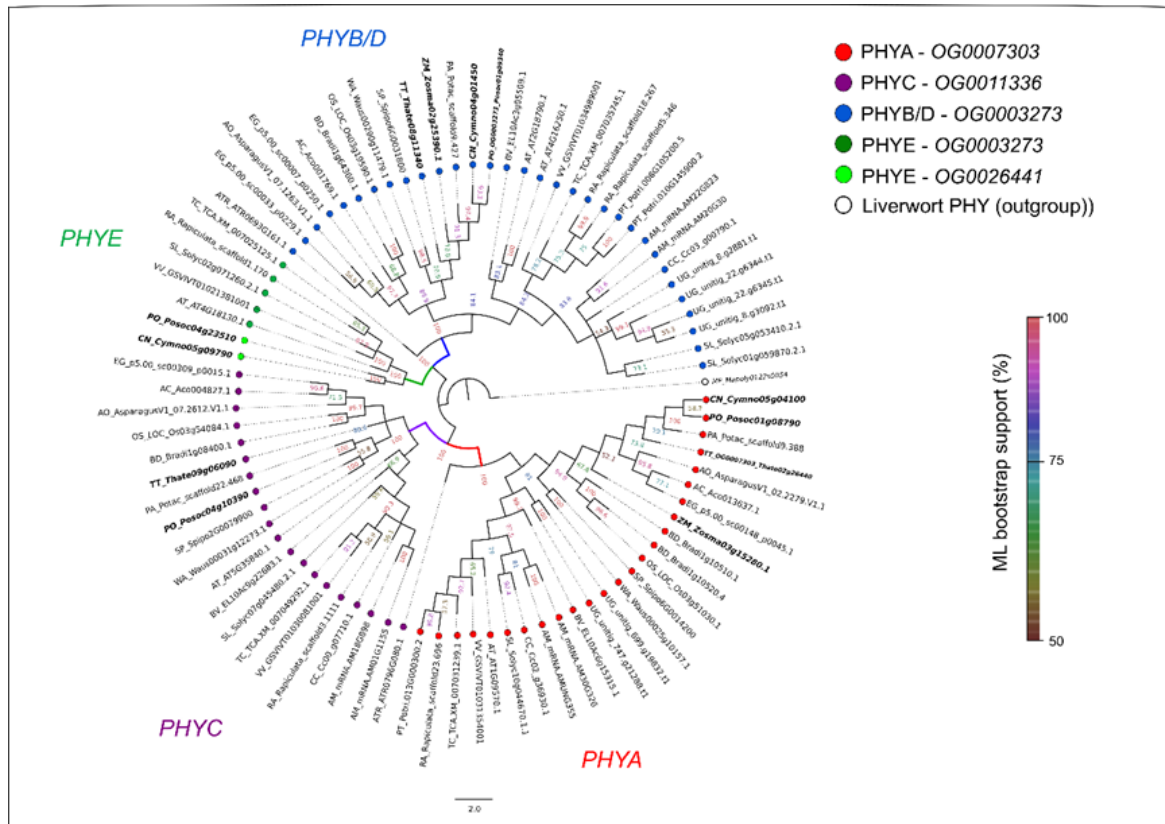
846
847
848
849
850
851
852
853
854
855

Supplementary Figure 5.6.4 Gene families containing photoreceptors and the main integration of light signalling toolkit genes. Species sequenced in this work are in bold. Full list of abbreviation of the species names used in the figure: CN *Cymodocea nodosa*; PO *Posidonia oceanica*; TT *Thalassia testudinum*; ZM *Zostera marina*; PA *Potamogeton acutifolius*; SP *Spirodela polyrhiza*; WA *Wolffia australiana*; AM *Avicennia marina*; RA *Rhizophora apiculata*; OS *Oryza sativa*; BD *Brachypodium distachyon*; AC *Ananas comosus*; EG *Elaeis guineensis*; AO *Asparagus officinalis*; BV *Beta vulgaris*; UG *Utricularia gibba*; SL *Solanum lycopersicum*; CC *Coffea canephora*; VV *Vitis vinifera*; PT *Populus trichocarpa*; AT *Arabidopsis thaliana*; TC *Theobroma cacao*; ATR *Amborella trichopoda*.



858 **Supplementary Figure 5.6.5 The N terminus alignment of UVB-Resistance 8.**

859 For the alignment, proteins sequences of the orthogroup OG0008432 were used. Sequences of seagrasses are
 860 in bold. Full list of abbreviation of the species names used in the figure: AT *Arabidopsis thaliana*; SL *Solanum*
 861 *lycopersicum*; ATR *Amborella trichopoda*; AC *Ananas comosus*; AO *Asparagus officinalis*; AM *Avicennia marina*;
 862 BV *Beta vulgaris*; BD *Brachypodium distachyon*; CC *Coffea canephora*; CN *Cymodocea nodosa*; EG *Elaeis*
 863 *guineensis*; OS *Oryza sativa*; PO *Posidonia oceanica*; PT *Populus trichocarpa*; PA *Potamogeton acutifolius*; RA
 864 *Rhizophora apiculata*; SP *Spirodela polyrhiza*; TT *Thalassia testudinum*; TC *Theobroma cacao*; VV *Vitis vinifera*;
 865 UG *Utricularia gibba*.



868

869 **Supplementary Figure 5.6.6** Phylogenetic tree of phytochromes obtained from the 84 proteins
 870 sequences included in the orthogroups OG0007303, OG0003273, OG0011336 and OG0026441.

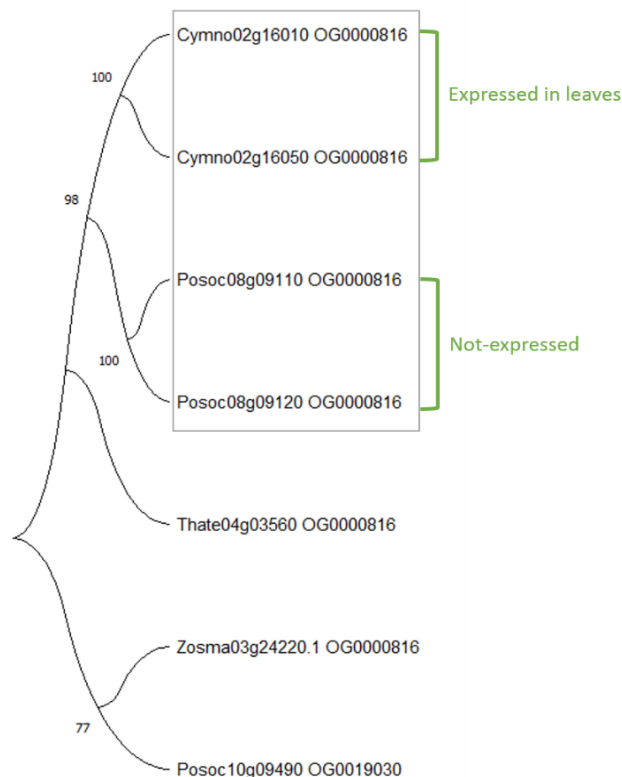
871 The 84 protein sequences of the four OGs (OG0007303, OG0003273, OG0011336 and OG0026441), which were
 872 functionally annotated as putative phytochromes, were aligned using Muscle (MEGA5), with the phytochrome
 873 sequence of the liverwort *Marchantia polymorpha* (MP_Mapoly0122s0054 white dot; selected from PLAZA 5.0
 874 (ORTHO05M001074 https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_monocots/)
 875 as an outgroup. The resulting multiple sequence alignment was trimmed with TRIMAL using the Automated 1 method. The ML
 876 phylogenetic tree was generated using the JTT + I +G model with MEGA5. Bootstrap values (expressed as
 877 percentages) were calculated over 1,000 replications. Only branches with a bootstrap support over 52% are
 878 shown. Protein sequences of seagrasses are indicated in bold, while *Arabidopsis thaliana* orthologs of PHYA-E
 879 were in italic (i.e., AT_AT1G09570.1 phyA, AT_AT5G35840.1 phyC, AT_AT2G18790.1 phyB, AT_AT4G16250.1
 880 phyD, AT_AT4G18130.1 phyE; according to TAIR classification). PHYA (red branch) and PHYC (violet branch) were
 881 encoded by proteins sequences which were included respectively in the OG0007303 (red dots) and OG0011336
 882 (violet dots). PHYB/D (blue branch) groups proteins sequences of OG0003273 (blue dots) while the PHYE form
 883 (green branch) is represented by five proteins included in the OG0003273 (dark-green dots) plus two proteins
 884 of Orthogroup OG0026441 (light-green dots) which is an OG exclusive of seagrasses, suggesting that the two
 885 seagrasses *P. oceanica* and *C. nodosa* have retained a gene copy of PHYE even if this phytochrome form is often
 886 missing in other monocotyledonous plants (Mathews 2006). Complete list of species names' abbreviations used
 887 in the figure: AM *Avicennia marina*; AC *Ananas comosus*; AT *Arabidopsis thaliana*; AO *Asparagus officinalis*; ATR
 888 *Amborella trichopoda*; BV *Beta vulgaris*; BD *Brachypodium distachyon*; CC *Coffea canephora*; CN *Cymodocea*
 889 *nodosa*; EG *Elaeis guineensis*; OS *Oryza sativa*; PO *Posidonia oceanica*; PT *Populus trichocarpa*; PA *Potamogeton*
 890 *acutifolius*; RA *Rhizophora apiculata*; SL *Solanum lycopersicum*; SP *Spirodela polyrhiza*; TT *Thalassia testudinum*;
 891 TC *Theobroma cacao*; VV *Vitis vinifera*; UG *Utricularia gibba*.

892 5.7 NAC transcriptional factors

893 **Supplementary Note 5.7** NAC transcriptional factors

894 NAC proteins (NAM, ATAF1–2 and CUC2 transcription factors) are one of the largest family of transcriptional
895 factors that are involved in different developmental processes as well as in the regulation of signaling pathways
896 in response to abiotic stressors, especially salt stress (Puranik et al. 2012). A comparable number of sequences
897 were found in seagrasses with respect to land plants, freshwater and mangroves species. However, specific
898 orthogroups were found for seagrasses. One of them is annotated as JUNGBRUNNEN 1 (JUB1), and a specific
899 functional analysis revealed that while *P. oceanica* retained JUNGBRUNNEN 1 (*JUB1*) genes with low expression
900 values, *C. nodosa* expressed *JUB1* genes in leaves. *JUB1* is a central longevity regulator as well as a regulator of
901 responses to abiotic stressors enhancing salt stress tolerance (Wu et al., 2012) regulating plant responses to
902 environmental factors.

903 In addition, other sequences annotated as *JUB1* were found across all species belonging to different orthogroups.
904 The OG0000816 was the most representative orthogroup counting a total of 67 sequences across all species
905 (land plants, freshwater and mangroves species). Here too, in *P. oceanica* only Posoc08g09120 and
906 Posoc08g09110 were weakly expressed in leaves. Contrarily, sequences found for the other seagrasses
907 (*Cymno02g16010*, *Cymno02g16050*, *Zosma03g24220.1*) showed higher expression values especially in leaves.
908 In *T. testudinum*, one single gene copy was found (*Thate04g03560*) specifically expressed in root and leaf. Thus,
909 a phylogenetic tree was built including these sequences to visualize relationships of the *JUB1* sequences
910 between seagrasses (Supplementary Figure 5.7). Sequences of *P. oceanica* (*Posoc08g09120* and *Posoc08g09110*)
911 and *C. nodosa* (*Cymno02g16010*, *Cymno02g16050*) belonging to the same orthogroup (OG0000816) formed a
912 single clade, apart from *T. testudinum* (*Thate04g03560*), *Z. marina* (*Zosma03g24220.1*) and *Posoc10g09490*.
913 Considering that *P. oceanica* and *C. nodosa* are phylogenetically closely related, the different expression levels



Supplementary Figure 5.7 Evolutionary analysis by Maximum Likelihood method of *JUB1* in seagrasses.

914 observed for JUB1 sequences could suggests a functional re-organization that could be related to the different
915 ecological requirements of these species, modulating stress tolerance in seagrasses, including response to
916 salinity. *P. oceanica*, in fact, colonizes open coastal habitats with a very narrow range of salinity, contrary to *C.*
917 *nodosa* which can also be present in estuarine dynamic environmental conditions and highly variable salinities.

918 5.8 Nitrogen metabolism

919 **Supplementary Note 5.8 Nitrogen metabolism**

920 Seagrass meadows act as an important nitrogen (N) filter in the coastal environments. In this context, seagrasses
921 assimilate large amounts of N, and exude oxygen and labile carbon into sediments, which stimulate other
922 processes in the nitrogen metabolism pathway, including nitrification-denitrification process that
923 counterbalances the net N loads through microbial transformation (Zarnoch et al. 2017; Aoki et al. 2020). The
924 key genes linked to nitrogen uptake/transport and assimilation were retained in all of the plants examined
925 (Extended table 10). This corresponds to at least 66 and 25 orthogroups which function in uptake/transport and
926 assimilation, respectively. Their existence is essential because an efficient N metabolism process is required to
927 ensure normal growth and development in plants, regardless of their diversity, habitat or nature. Moreover,
928 seagrasses may have acquired a more effective nitrogen metabolism in N-deficient marine environments
929 through symbiotic N₂-fixing bacteria, that could have facilitated the migration of flowering plants back to the
930 sea some 100 million years ago (Mohr et al. 2021).

931 As compared to non-seagrass genomes, the nitrate transporter (*NRT*) gene families of seagrasses were
932 contracted (40.71%), indicating that seagrasses may have evolved alternative mechanisms to utilize nitrogen
933 sources more effectively (Extended table 10). Other than nitrate, seagrasses rely on ammonium as a primary
934 source of nitrogen (Touchette and Burkholder 2001; Xu et al. 2020), particularly when exposed to anoxic
935 conditions in marine sediment where nitrate is scarce due to disrupted ammonium-to-nitrate oxidation. In this
936 case, ammonium is metabolized directly via GOGAT pathway, which catalyzes the formation of glutamine from
937 glutamate and ammonium, instead of converting nitrate to ammonium prior to glutamine formation (Wang et
938 al. 2021).

939 Nitrate reductase (*NR*) was expressed in all parts of seagrasses (flower, root, vegetative, rhizome), with *NR*
940 robustly expressed in the root of *Zostera marina* and the leaves of *Cymodocea nodosa* and *Thalassia testudinum*.
941 *NR* activity is widely influenced by light (Touchette and Burkholder 2001) and given that *NR* activity is highest
942 during photosynthetic periods and lowest in the dark, nitrite reduction occurs more frequently in leaf tissues
943 than in root tissues in most seagrasses, suggesting the importance of light on *NR* response (Manassa et al. 2017;
944 Wang et al. 2021; Jiménez-Ramos et al. 2022). *Zostera marina* is unique from the other seagrasses as its *NR*
945 activity can be maintained in the dark, provided that the environment is nitrate-enriched and tissue
946 carbohydrate levels are high (Touchette and Burkholder 2001). The intensity and duration of *NR* activity is
947 directly parallel with the soluble carbohydrate supplies (Touchette and Burkholder 2007). On the other hand,
948 seagrass leaves such as those in *T. testudinum* tend to have higher efficiency of nitrogen assimilation compared
949 to root, given that the NH₄⁺ in the water column is relatively lower than the sediments where seagrass inhabit
950 (Lee and Dunton 1999; Cornelisen and Thomas 2004).

951 5.9 Flower and pollen development

952 **Supplementary Note 5.9 Flower and pollen development**

953 MADS-box genes encode transcription factors that play a crucial role in controlling various developmental
954 programs including the development of floral organs. We predicted considerable re-arrangements in seagrasses
955 as petals and sepals are completely reduced. Type II MADS-box genes have been extensively studied for their
956 role in specifying floral organ development.

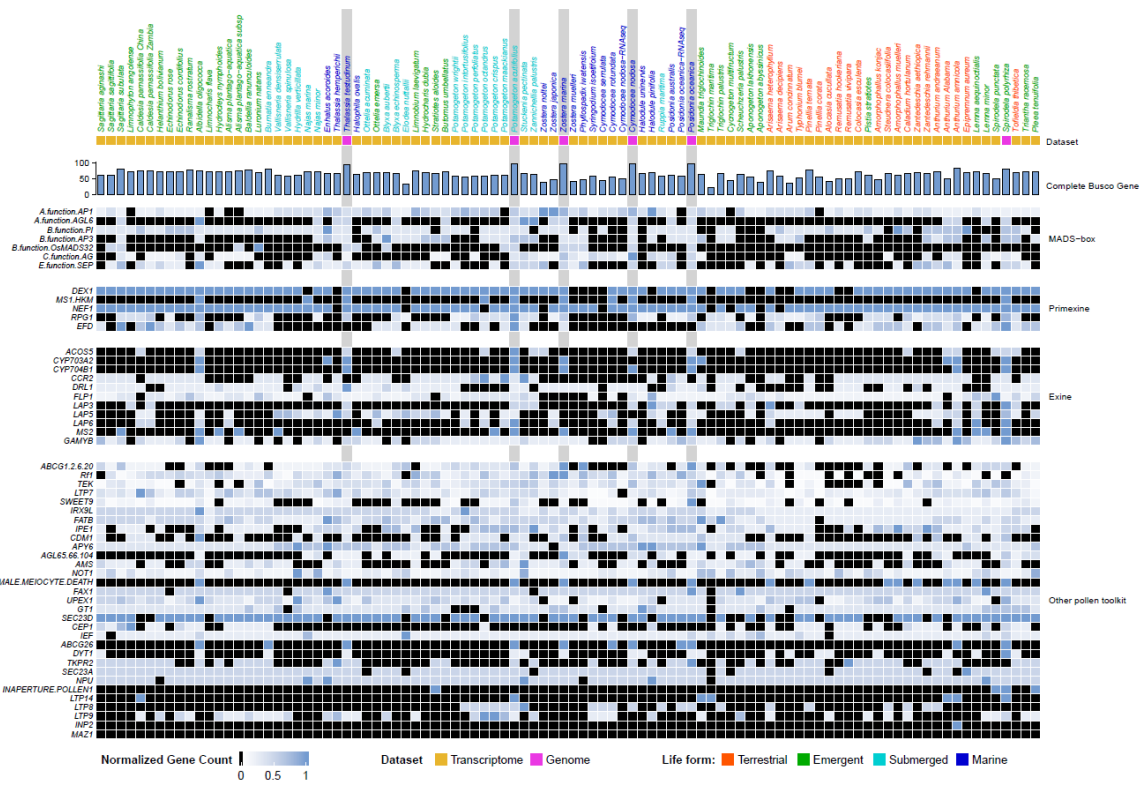
957 To identify MADS-box genes in the genomes of *Z. marina*, *C. nodosa*, *P. oceanica*, *T. testudinum*, and *P. acutifolius*,
 958 we employed a hidden Markov model (HMM). The HMM profile for the SRF-TF domain (PF00319) was obtained
 959 from the Pfam database (Mistry et al. 2021). This profile was used to search against the local protein database
 960 using the HMMER software, with an E-value threshold of $< 1e-5$. Using the same method described above,
 961 MADS-box genes of *A. thaliana* and *O. sativa* were obtained. Subsequently, all candidate proteins from the seven
 962 species mentioned above were aligned using MAFFT, resulting in a concatenated dataset, which was then used
 963 for phylogenetic analysis to further identify the type II MADS-box genes. We identified 22, 26, 29, 24, and 29
 964 Type II MADS-box genes in *Z. marina*, *P. oceanica*, *C. nodosa*, *T. testudinum*, and *Potamogeton*, respectively
 965 (Supplementary Table 5.9 and Figure 4a). Among these, several are homologues of genes defining the well-
 966 known ABCE model (Lohmann and Weigel 2002; Krizek and Fletcher 2005): AP1 and AGL6 (A function for sepals
 967 and petals), PI and AP3 (B function for petals and stamen), as well as OsMADS32 (B function in rice), AG (C
 968 function for stamen and carpel), and SEP (E function for interacting with ABC function proteins).

969 We also analyzed the expression profile of these genes in various tissues. For the flower tissue of *Posidonia*
 970 *oceanica*, RNA-seq data was obtained from the NCBI Short Read Archive under BioProject ID PRJNA375717
 971 (Entrambasaguas et al. 2017). Subsequently, the data was aligned to the *P. oceanica* assembly.

972 **Supplementary Table 5.9. The MADS-box genes in seagrasses and *P. acutifolius***

		<i>Z. marina</i>	<i>P. oceanica</i>	<i>C. nodosa</i>	<i>T. testudinum</i>	<i>P. acutifolius</i>
MADS-box genes in total		48	46	44	34	38
Type II		22	26	29	24	29
A function	AP1	4	2	4	2	3
	AGL6	1	1	1	1	1
B function	PI	1	1	1	6	1
	AP3	1	1	1	1	1
	OsMADS32	1	1	0	1	1
C function	AG	4	3	6	1	3
E function	SEP	2	2	2	2	2

973



974

975 **Supplementary Figure 5.9** Normalized gene copy numbers for flower and pollen development genes
 976 and gene families for 96 species, including 6 genomic data and 90 transcriptomic data. The light grey
 977 background denotes our genomic data of four seagrasses and one freshwater relative, *Potamogeton acutifolius*.
 978 Others are the transcriptomic data from Chen et al. (2022) and *Spirodela polyrhiza* genomic data. Normalization
 979 for each gene family was obtained by dividing the number of genes in that gene family for a particular species
 980 by the largest gene copy number within that family (considering all species). Genes in black are absent.

981

References

- 982
983
984 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-
985 410.
- 986 An D, Zhou Y, Li C, Xiao Q, Wang T, Zhang Y, Wu Y, Li Y, Chao DY, Messing J et al. 2019. Plant evolution and
987 environmental adaptation unveiled by long-read whole-genome sequencing of *Spirodela*. *Proc Natl*
988 *Acad Sci U S A* **116**: 18893-18899.
- 989 Aoki LR, McGlathery KJ, Oreska MPJ. 2020. Seagrass restoration reestablishes the coastal nitrogen filter through
990 enhanced burial. *Limnology and Oceanography* **65**: 1-12.
- 991 Berry EW. 1914. The Upper Cretaceous and Eocene floras of South Carolina and Georgia. doi:10.3133/pp84.
- 992 Campbell JE, Fourqurean JW. 2013. Mechanisms of bicarbonate use influence the photosynthetic carbon dioxide
993 sensitivity of tropical seagrasses. *Limnology and Oceanography* **58**: 839-848.
- 994 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in
995 large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.
- 996 Capó-Bauçà S, Iñiguez C, Aguiló-Nicolau P, Galmés J. 2022. Correlative adaptation between Rubisco and CO₂-
997 concentrating mechanisms in seagrasses. *Nature Plants* **8**: 706-716.
- 998 Cervela-Cardona L, Yoshida T, Zhang Y, Okada M, Fernie A, Mas P. 2021. Circadian Control of Metabolism by the
999 Clock Component TOC1. *Frontiers in Plant Science* **12**.
- 1000 Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: An Integrative Toolkit Developed for
1001 Interactive Analyses of Big Biological Data. *Mol Plant* **13**: 1194-1202.
- 1002 Chen L-Y, Lu B, Morales-Briones DF, Moody ML, Liu F, Hu G-W, Huang C-H, Chen J-M, Wang Q-F. 2022.
1003 Phylogenomic Analyses of Alismatales Shed Light into Adaptations to Aquatic Environments. *Molecular*
1004 *Biology and Evolution* **39**.
- 1005 Cornelisen CD, Thomas FIM. 2004. Ammonium and nitrate uptake by leaves of the seagrass *Thalassia testudinum*:
1006 impact of hydrodynamic regime and epiphyte cover on uptake rates. *Journal of Marine Systems* **49**:
1007 177-194.
- 1008 Crepet W, Nixon K. 1998. Fossil Clusiaceae from the late Cretaceous (Turonian) of New Jersey and implications
1009 regarding the history of bee pollination. *Am J Bot* **85**: 1122.
- 1010 Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole
1011 genome data. *Nucleic Acids Res* **45**: e18.
- 1012 DiMario RJ, Clayton H, Mukherjee A, Ludwig M, Moroney JV. 2017. Plant Carbonic Anhydrases: Structures,
1013 Locations, Evolution, and Physiological Roles. *Molecular Plant* **10**: 30-46.
- 1014 Ding Y, Liu N, Virilouvet L, Riethoven J-J, Fromm M, Avramova Z. 2013. Four distinct types of dehydration stress
1015 memory genes in *Arabidopsis thaliana*. *BMC Plant Biology* **13**: 229.
- 1016 Dongen SV. 2008. Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and*
1017 *Applications* **30**: 121-141.
- 1018 Doyle JA, Endress PK, Upchurch GR. 2008. Early Cretaceous monocots: a phylogenetic evaluation. *Sborník*
1019 *Národního muzea v Praze Acta Musei nationalis Pragae* **64**: 61-87.
- 1020 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids*
1021 *Res* **32**: 1792-1797.
- 1022 Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically
1023 improves orthogroup inference accuracy. *Genome Biology* **16**: 157.
- 1024 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*
1025 **20**: 238.
- 1026 Entrambasaguas L, Jahnke M, Biffali E, Borra M, Sanges R, Marin-Guirao L, Procaccini G. 2017. Tissue-specific
1027 transcriptomic profiling provides new insights into the reproductive ecology and biology of the iconic
1028 seagrass species *Posidonia oceanica*. *Mar Genomics* **35**: 51-61.
- 1029 Gandolfo M, Nixon K, Crepet W. 1998. A new fossil flower from the Turonian of New Jersey: *Dressiantha*
1030 *bicarpellata* gen. et sp. nov. (Capparales). *Am J Bot* **85**: 964.
- 1031 Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, Teodor R, Lu Y, Bowser TA, Graham IA et al. 2018. The opium poppy
1032 genome and morphinan production. *Science* **362**: 343-347.
- 1033 Guo M, Liu JH, Ma X, Luo DX, Gong ZH, Lu MH. 2016. The Plant Heat Stress Transcription Factors (HSFs): Structure,
1034 Regulation, and Function in Response to Abiotic Stresses. *Front Plant Sci* **7**: 114.
- 1035 Han X, Chang X, Zhang Z, Chen H, He H, Zhong B, Deng XW. 2019. Origin and Evolution of Core Components
1036 Responsible for Monitoring Light Environment Changes during Plant Terrestrialization. *Molecular Plant*
1037 **12**: 847-862.

1038 Harmer SL. 2009. The circadian system in higher plants. *Annual review of plant biology* **60**: 357-377.

1039 Heerklotz D, Döring P, Bonzelius F, Winkelhaus S, Nover L. 2001. The balance of nuclear import and export
1040 determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2.
1041 *Mol Cell Biol* **21**: 1759-1768.

1042 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-
1043 755.

1044 Ikeda M, Mitsuda N, Ohme-Takagi M. 2011. Arabidopsis HsfB1 and HsfB2b act as repressors of the expression
1045 of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiol* **157**: 1243-
1046 1254.

1047 Iles WJ, Smith SY, Gandolfo MA, Graham SW. 2015. Monocot fossils suitable for molecular dating analyses.
1048 *Botanical Journal of the Linnean Society* **178**: 346-374.

1049 International Peach Genome I, Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori
1050 MT, Grimwood J et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique
1051 patterns of genetic diversity, domestication and genome evolution. *Nat Genet* **45**: 487-494.

1052 Jacob F, Vernaldi S, Maekawa T. 2013. Evolution and Conservation of Plant NLR Functions. *Front Immunol* **4**: 297.

1053 Janssen T, Bremer K. 2004. The age of major monocot groups inferred from 800+ rbcL sequences. *Botanical
1054 Journal of the Linnean Society* **146**: 385-398.

1055 Jiménez-Ramos R, Villazán B, Egea LG, Cantero R, Pérez-Lloréns JL, Vergara JJ, Brun FG. 2022. Differential
1056 ecophysiological responses to inorganic nitrogen sources (ammonium versus nitrate) and light levels in
1057 the seagrass *Zostera noltei*. *Marine Ecology Progress Series* **702**: 57-70.

1058 Jing Y, Lin R. 2020. Transcriptional Regulatory Network of the Light Signaling Pathways. *The New phytologist*.

1059 Koch M, Bowes G, Ross C, Zhang X-H. 2013. Climate change and ocean acidification effects on seagrasses and
1060 marine macroalgae. *Global Change Biology* **19**: 103-132.

1061 Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. 2019. The alternative reality
1062 of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet* **15**: e1008373.

1063 Krizek BA, Fletcher JC. 2005. Molecular mechanisms of flower development: an armchair guide. *Nature Reviews
1064 Genetics* **6**: 688-698.

1065 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an
1066 information aesthetic for comparative genomics. *Genome Res* **19**: 1639-1645.

1067 Larkum AWD, Davey PA, Kuo J, Ralph PJ, Raven JA. 2017. Carbon-concentrating mechanisms in seagrasses.
1068 *Journal of Experimental Botany* **68**: 3773-3784.

1069 Larkum AWD, Pernice M, Schliep M, Davey P, Szabo M, Raven JA, Lichtenberg M, Brodersen KE, Ralph PJ. 2018.
1070 Photosynthesis and Metabolism of Seagrasses. In *Seagrasses of Australia: Structure, Ecology and
1071 Conservation*, doi:10.1007/978-3-319-71354-0_11 (ed. AWD Larkum, et al.), pp. 315-342. Springer
1072 International Publishing, Cham.

1073 Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide
1074 sequences. *Nucleic Acids Res* **32**: 11-16.

1075 Lee BH, Henderson DA, Zhu JK. 2005. The Arabidopsis cold-responsive transcriptome and its regulation by ICE1.
1076 *Plant Cell* **17**: 3155-3175.

1077 Lee K-S, Dunton KH. 1999. Inorganic nitrogen acquisition in the seagrass *Thalassia testudinum*: Development of
1078 a whole-plant nitrogen budget. *Limnology and Oceanography* **44**: 1204-1215.

1079 Legnaioli T, Cuevas J, Mas P. 2009. TOC1 functions as a molecular switch connecting the circadian clock with
1080 plant responses to drought. *The EMBO Journal* **28**: 3745-3757.

1081 Les DH, Cleland MA, Waycott M. 1997. Phylogenetic Studies in Alismatidae, II: Evolution of Marine Angiosperms
1082 (Seagrasses) and Hydrophily. *Systematic Botany* **22**: 443.

1083 Li F-W, Melkonian M, Rothfels CJ, Villarreal JC, Stevenson DW, Graham SW, Wong GK-S, Pryer KM, Mathews S.
1084 2015. Phytochrome diversity in green plants and the origin of canonical plant phytochromes. *Nature
1085 Communications* **6**: 7852.

1086 Liu Y, Zeng Z, Zhang YM, Li Q, Jiang XM, Jiang Z, Tang JH, Chen D, Wang Q, Chen JQ et al. 2021. An angiosperm
1087 NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal
1088 transduction component deletion. *Mol Plant* **14**: 2015-2031.

1089 Lohmann JU, Weigel D. 2002. Building beauty: the genetic control of floral patterning. *Dev Cell* **2**: 135-142.

1090 Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions.
1091 *Proc Natl Acad Sci U S A* **102**: 10557-10562.

1092 Ma X, Fan J, Wu Y, Zhao S, Zheng X, Sun C, Tan L. 2020. Whole-genome de novo assemblies reveal extensive
1093 structural variations and dynamic organelle-to-nucleus DNA transfers in African and Asian rice. *Plant J
1094* **104**: 596-612.

1095 Ma X, Olsen JL, Reusch TBH, Procaccini G, Kudrna D, Williams M, Grimwood J, Rajasekar S, Jenkins J, Schmutz J
1096 et al. 2021a. Improved chromosome-level genome assembly and annotation of the seagrass, *Zostera*
1097 *marina* (eelgrass). *F1000Research* **10**.

1098 Ma X, Olsen JL, Reusch TBH, Procaccini G, Kudrna D, Williams M, Grimwood J, Rajasekar S, Jenkins J, Schmutz J
1099 et al. 2021b. Improved chromosome-level genome assembly and annotation of the seagrass, *Zostera*
1100 *marina* (eelgrass). *F1000Res* **10**: 289.

1101 Manassa RP, Smith TM, Beardall J, Keough MJ, Cook PLM. 2017. Capacity of a temperate intertidal seagrass
1102 species to tolerate changing environmental conditions: Significance of light and tidal exposure.
1103 *Ecological Indicators* **81**: 578-586.

1104 Marina K, Thorsten BHR, Tal D. 2023. Worldwide population genomics reveal long-term stability of the
1105 mitochondrial chromosome composition in a keystone marine plant. *bioRxiv*
1106 doi:10.1101/2023.04.21.537793: 2023.2004.2021.537793.

1107 Mathews S. 2006. Phytochrome-mediated development in land plants: red light sensing evolves to meet the
1108 challenges of changing light environments. *Molecular ecology* **15**: 3483-3503.

1109 Mathews S. 2010. Evolutionary Studies Illuminate the Structural-Functional Model of Plant Phytochromes. *The*
1110 *Plant Cell* **22**: 4-16.

1111 McClung CR. 2019. The plant circadian oscillator. Vol 8.

1112 Michael TP, Ernst E, Hartwick N, Chu P, Bryant D, Gilbert S, Ortleb S, Baggs EL, Sree KS, Appenroth KJ et al. 2021.
1113 Genome and time-of-day transcriptome of *Wolffia australiana* link morphological minimization with
1114 gene loss and less growth control. *Genome Research* **31**: 225-238.

1115 Michalovova M, Vyskot B, Kejnovsky E. 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus
1116 (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. *Heredity* **111**:
1117 314-320.

1118 Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E et al. 2015.
1119 The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* **47**: 1435-1442.

1120 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2:
1121 New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology*
1122 *and Evolution* **37**: 1530-1534.

1123 Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf KD. 2002. In the complex family of heat
1124 stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato.
1125 *Genes Dev* **16**: 1555-1567.

1126 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S,
1127 Richardson LJ et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**: D412-d419.

1128 Mittal D, Madhyastha DA, Grover A. 2012. Gene expression analysis in response to low and high temperature
1129 and oxidative stresses in rice: combination of stresses evokes different transcriptional changes as
1130 against stresses applied individually. *Plant Sci* **197**: 102-113.

1131 Mohr W, Lehnen N, Ahmerkamp S, Marchant HK, Graf JS, Tschitschko B, Yilmaz P, Littmann S, Gruber-Vodicka H,
1132 Leisch N et al. 2021. Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine
1133 bacterium. *Nature* **600**: 105-109.

1134 Morley SA, Nielsen BL. 2017. Plant mitochondrial DNA. *Front Biosci (Landmark Ed)* **22**: 1023-1032.

1135 Murat F, Armero A, Pont C, Klopp C, Salse J. 2017. Reconstructing the genome of the most recent common
1136 ancestor of flowering plants. *Nature Genetics* **49**: 490-496.

1137 Nauheimer L, Metzler D, Renner SS. 2012. Global history of the ancient monocot family Araceae inferred with
1138 models accounting for past continental positions and previous ranges based on fossils. *New Phytologist*
1139 **195**: 938-950.

1140 Olsen JL, Rouze P, Verhelst B, Lin YC, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F et al. 2016.
1141 The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**: 331-
1142 335.

1143 Personat JM, Tejedor-Cano J, Prieto-Dapena P, Almoguera C, Jordano J. 2014. Co-overexpression of two Heat
1144 Shock Factors results in enhanced seed longevity and in synergistic effects on seedling tolerance to
1145 severe dehydration and oxidative stress. *BMC Plant Biol* **14**: 56.

1146 Petersen G, Cuenca A, Zervas A, Ross GT, Graham SW, Barrett CF, Davis JI, Seberg O. 2017. Mitochondrial genome
1147 evolution in Alismatales: Size reduction and extensive loss of ribosomal protein genes. *PLoS One* **12**:
1148 e0177606.

1149 Pokhilko A, Mas P, Millar AJ. 2013. Modelling the widespread effects of TOC1 signalling on the plant circadian
1150 clock and its outputs. *BMC Systems Biology* **7**: 23-23.

1151 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments.
1152 *PLoS One* **5**: e9490.

1153 Puranik S, Sahu PP, Srivastava PS, Prasad M. 2012. NAC proteins: regulation and role in stress tolerance. *Trends*
1154 *in Plant Science* **17**: 369-381.

1155 Rao SK, Magnin NIC, Reiskind JB, Bowes G. 2002. Photosynthetic and Other Phosphoenolpyruvate Carboxylase
1156 Isoforms in the Single-Cell, Facultative C4System of Hydrilla verticillata *Plant Physiology* **130**: 876-886.

1157 Rockwell NC, Lagarias JC. 2020. Phytochrome evolution in 3D: deletion, duplication, and diversification. *New*
1158 *Phytologist* **225**: 2283-2300.

1159 Ross TG, Barrett CF, Soto Gomez M, Lam VKY, Henriquez CL, Les DH, Davis JI, Cuenca A, Petersen G, Seberg O et
1160 al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* **32**: 160-178.

1161 Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and
1162 structural alignment. *Nucleic Acids Research* **47**: W5-W10.

1163 Rubio L, García D, García-Sánchez MJ, Niell FX, Felle HH, Fernández JA. 2017. Direct uptake of HCO₃⁻ in the
1164 marine angiosperm *Posidonia oceanica* (L.) Delile driven by a plasma membrane H⁺ economy. *Plant,*
1165 *Cell & Environment* **40**: 2820-2830.

1166 Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi-Shinozaki K. 2006. Dual function of an
1167 Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene
1168 expression. *Proc Natl Acad Sci U S A* **103**: 18822-18827.

1169 Scharf KD, Berberich T, Ebersberger I, Nover L. 2012. The plant heat stress transcription factor (Hsf) family:
1170 structure, function and evolution. *Biochim Biophys Acta* **1819**: 104-119.

1171 Sensalari C, Maere S, Lohaus R. 2021. ksrates: positioning whole-genome duplications relative to speciation
1172 events in KS distributions. *Bioinformatics* doi:10.1093/bioinformatics/btab602.

1173 Simillion C, Janssens K, Sterck L, Van de Peer Y. 2008. i-ADHoRe 2.0: an improved tool to detect degenerated
1174 genomic homology using genomic profiles. *Bioinformatics* **24**: 127-128.

1175 Smith DR, Crosby K, Lee RW. 2011. Correlation between nuclear plastid DNA abundance and plastid number
1176 supports the limited transfer window hypothesis. *Genome Biol Evol* **3**: 365-371.

1177 Smith H. 2000. Phytochromes and light signal perception by plants-an emerging synthesis. *Nature* **407**: 585-591.

1178 Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for
1179 flowering plants. *Proceedings of the National Academy of Sciences* **107**: 5897-5902.

1180 Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, Baggs E, Witek AI, Yadav I,
1181 Krasileva KV et al. 2020. The NLR-Annotator Tool Enables Annotation of the Intracellular Immune
1182 Receptor Repertoire. *Plant Physiol* **183**: 468-482.

1183 Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013. Efficient exploration of the space of reconciled
1184 gene trees. *Syst Biol* **62**: 901-912.

1185 Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes.
1186 *Science* **320**: 486-488.

1187 Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the
1188 monocot lineage. *Proc Natl Acad Sci U S A* **107**: 472-477.

1189 Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and
1190 accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6-W11.

1191 Toda N, Rustenholz C, Baud A, Le Paslier MC, Amselem J, Merdinoglu D, Faivre-Rampant P. 2020.
1192 NLGenomeSweeper: A Tool for Genome-Wide NBS-LRR Resistance Gene Identification. *Genes (Basel)*
1193 **11**.

1194 Touchette BW, Burkholder J. 2001. Nitrate reductase activity in a submersed marine angiosperm: Controlling
1195 influences of environmental and physiological factors. *Plant Physiology and Biochemistry* **39**: 583-593.

1196 Touchette BW, Burkholder JM. 2007. Carbon and nitrogen metabolism in the seagrass, *Zostera marina* L.:
1197 Environmental control of enzymes involved in carbon allocation and nitrogen assimilation. *Journal of*
1198 *Experimental Marine Biology and Ecology* **350**: 216-233.

1199 Tuskan GA Difazio S Jansson S Bohlmann J Grigoriev I Hellsten U Putnam N Ralph S Rombauts S Salamov A et al.
1200 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604.

1201 Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful
1202 genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res* **24**: 1334-
1203 1347.

1204 von Koskull-Döring P, Scharf KD, Nover L. 2007. The diversity of plant heat stress transcription factors. *Trends*
1205 *Plant Sci* **12**: 452-457.

1206 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK et
1207 al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly
1208 improvement. *PLoS One* **9**: e112963.

1209 Wang H, Tang X, Chen J, Shang S, Zhu M, Liang S, Zang Y. 2021. Comparative studies on the response of *Zostera*
1210 *marina* leaves and roots to ammonium stress and effects on nitrogen metabolism. *Aquatic Toxicology*
1211 **240**.

1212 Wang J, Du Z, Huo X, Zhou J, Chen Y, Zhang J, Pan A, Wang X, Wang F, Zhang J. 2020. Genome-wide analysis of
1213 PRR gene family uncovers their roles in circadian rhythmic changes and response to drought stress in
1214 *Gossypium hirsutum* L. *PeerJ* **8**: e9936-e9936.

1215 Wang W, Haberer G, Gundlach H, Glasser C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA,
1216 Shanklin J et al. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenenous reduction fast
1217 growth and aquatic lifestyle. *Nat Commun* **5**: 3311.

1218 Wang X, Shi X, Chen S, Ma C, Xu S. 2018. Evolutionary Origin, Gradual Accumulation and Functional Divergence
1219 of Heat Shock Factor Gene Family with Plant Evolution. *Front Plant Sci* **9**: 71.

1220 Xin Z, Mandaokar A, Chen J, Last RL, Browse J. 2007. Arabidopsis ESK1 encodes a novel regulator of freezing
1221 tolerance. *Plant J* **49**: 786-799.

1222 Xu S, Xu S, Zhou Y, Yue S, Qiao Y, Liu M, Gu R, Song X, Zhang Y, Zhang X. 2020. Sonar and in situ surveys of eelgrass
1223 distribution, reproductive effort, and sexual recruitment contribution in a eutrophic bay with intensive
1224 human activities: Implication for seagrass conservation. *Mar Pollut Bull* **161**: 111706.

1225 Xue GP, Sadat S, Drenth J, McIntyre CL. 2014. The heat shock factor family from *Triticum aestivum* in response
1226 to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J Exp*
1227 *Bot* **65**: 539-557.

1228 Yang Z. 2007a. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

1229 Yang ZH. 2007b. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:
1230 1586-1591.

1231 Yin R, Arongaus AB, Binkert M, Ulm R. 2015. Two distinct domains of the UVR8 photoreceptor interact with COP1
1232 to initiate UV-B signaling in arabidopsis. *Plant Cell* **27**: 202-213.

1233 Zarnoch CB, Hoellein TJ, Furman BT, Peterson BJ. 2017. Eelgrass meadows, *Zostera marina* (L.), facilitate the
1234 ecosystem service of nitrogen removal during simulated nutrient pulses in Shinnecock Bay, New York,
1235 USA. *Mar Pollut Bull* **124**: 376-387.

1236 Zhang GJ, Dong R, Lan LN, Li SF, Gao WJ, Niu HX. 2020. Nuclear Integrants of Organellar DNA Contribute to
1237 Genome Structure and Evolution in Plants. *Int J Mol Sci* **21**.

1238 Zhao N, Grover CE, Chen Z, Wendel JF, Hua J. 2019. Intergenomic gene transfer in diploid and allopolyploid
1239 *Gossypium*. *BMC Plant Biol* **19**: 492.

1240 Zheng S, Poczai P, Hyvonen J, Tang J, Amiryousefi A. 2020. Chloroplast: An Online Program for the Versatile
1241 Plotting of Organelle Genomes. *Front Genet* **11**: 576124.

1242 Zhong X. 2020. Assembly, annotation and analysis of chloroplast genomes. doi:10.26182/5f333d9ac2bee.

1243 Zhuang L, Cao W, Wang J, Yu J, Yang Z, Huang B. 2018. Characterization and Functional Analysis of FaHsfC1b
1244 from *Festuca arundinacea* Conferring Heat Tolerance in Arabidopsis. *Int J Mol Sci* **19**.

1245 Zwaenepoel A, Van de Peer Y. 2019a. Inference of Ancient Whole-Genome Duplications and the Evolution of
1246 Gene Duplication and Loss Rates. *Mol Biol Evol* **36**: 1384-1404.

1247 Zwaenepoel A, Van de Peer Y. 2019b. wgd-simple command line tools for the analysis of ancient whole-genome
1248 duplications. *Bioinformatics* **35**: 2153-2155.

1249

1250