Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment

Editor summary:

Newly sequenced seagrass genomes unveil a hexaploid ancestry for seagrasses. The transition to marine environments involved fine-tuning of many processes that all had to happen in parallel, likely explaining why adaptation to a marine lifestyle has been rare.

Reviewer Recognition:

Nature Plants thanks Sean Graham and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

10

Figure or Table # Please group Extended Data items by type, in sequential order. Total number of items (Figs. + Tables) must not exceed 10.	Figure/Table title One sentence only	Filename Whole original file name including extension. i.e.: Smith_ED_Fig1.jpg	Figure/Table Legend If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Table. 1	The copy number of guard cell toolkit genes.	Guard_cell_toolkit_ED_Fig1.jpg	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 apiculate*; 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*.
Extended Data Table. 2	The copy number of genes involved in triterpenes, MeSA and ethylene biosynthesis and signaling pathways.	Volatile_metabolites_ED_Fig2.jpg	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 apiculate*; 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*.

Extended Data Table. 3	The copy number of genes involved in vascular development.	Vascular_development_ED_Fig3.jpg	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 apiculate*; 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*.
Extended Data Table. 4	The copy number of genes involved in lignin biosynthesis.	Lignin_biosynthesis_ED_Fig4.jpg	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 apiculate*; 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*.
Extended Data Table. 5	The copy number of resistance (NRL) genes and heat shock factots (HSF)	NRL_ED_Fig5.jpg	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 apiculate*; 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*.
Extended Data Table. 6	The copy number of genes involved in flavonoids biosynthesis.	Flavonoids_biosynthesis_ED_Fig6.jpg	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 apiculate*; 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*.
Extended Data Table. 7	The copy number of genes involved in salt stress	Salt_tolerance_ED_Fig7.jpg	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 apiculate*; 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*.
Extended Data Fig. 8	The copy number of genes involved in hypoxia tolerance	Hypoxia_tolerance_ED_Fig8.jpg	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 apiculate*; 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*.
Extended Data Table. 9	The copy number of CO2-concentrating mechanisms (CCM)-related genes (Carbonic anhydrases, Boron transporters and proton pumps, C4-metabolism, Rubisco activase)	CCM_ED_Fig9.jpg	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 apiculate*; 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*.
Extended Data Table. 10	The copy number of gene families involved in nitrogen metabolism	Nitrogen_ED_Fig10.jpg	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 apiculate*; 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*.

11

Item	Present?	Filename Whole original file name including extension. i.e.:	A brief, numerical description of file contents.

		Smith_SI.pdf. The extension must be .pdf	i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	Yes	Xiao_et_al_Supplementary_Information.pdf	5 sections and Supplementary Notes, Supplementary Tables 1.1–5.9, and Figures 1.1–5.9
Reporting Summary	Yes	**[F]NPLANTS-230314716B.pdf**	

12

13

Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment

Xiao Ma[1,2,25], Steffen Vanneste[3,25], Jiyang Chang[1,2,25], Luca Ambrosino[4], Kerrie Barry[5], Till Bayer[6], Alexander A. Bobrov[7], LoriBeth Boston[8], Justin E Campbell[9], Hengchi Chen[1,2], Maria Luisa Chiusano[4,10], Emanuela Dattolo[11], Jane Grimwood[8], Guifen He[5], Jerry Jenkins[8], Marina Khachaturyan[6,12], Lázaro Marín-Guirao[11,13], Attila Mesterházy[14], Danish-Daniel Muhd[15], Jessica Pazzaglia[11,16], Chris Plott[8], Shanmugam Rajasekar[17], Stephane Rombauts[1,2], Miriam Ruocco[11,18,19], Alison Scott[20], Min Pau Tan[15], Jozefien Van de Velde[20], Bartel Vanholme[1,2], Jenell Webber[8], Li Lian Wong[15], Mi Yan[5], Yeong Yik Sung[15], Polina Novikova[20], Jeremy Schmutz[5,8], Thorsten. B. H. Reusch[21,*], Gabriele Procaccini[11,16*], Jeanine L. Olsen[22,*], Yves Van de Peer[1,2,23,24*]

[25] These authors contributed equally

*e-mail: treusch@geomar.de, gpro@szn.it, j.l.olsen@rug.nl, yves.vandepeer@psb.ugent.be

Affiliations

1. Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium.
2. VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium.
3. Department Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
4. Department of Research Infrastructure for Marine Biological Resources, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy
5. DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Mail Stop: 91R183, 1 Cyclotron Road, Berkeley, CA 94720, USA.
6. GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Marine Evolutionary Ecology, Wischhofstr. 1-3, 24148 Kiel, Germany
7. Papanin Institute for Biology of Inland Waters RAS, Borok, Nekouz Distr., Yaroslavl Reg., 152742, Russia
8. Genome Sequencing Center, Hudson Alpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA.
9. Coastlines and Oceans Division, Institute of Environment, Florida International University-Biscayne Bay Campus, 3000 NE 151st St., Miami, Florida 33181, USA.
10. Department of Agricultural Sciences, University Federico II of Naples, Naples, Italy
11. Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy.
12. Institute of General Microbiology, University of Kiel, Kiel, Germany
13. Seagrass Ecology Group, Oceanographic Center of Murcia, Spanish Institute of Oceanography (IEO-CSIC), Murcia, Spain
14. Centre for Ecological Research, Wetland Ecology Research Group, Bem tér 18/C, Debrecen, H-4026, Hungary
15. Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Terengganu, Malaysia
16. National Biodiversity Future Centre (NBFC), Palermo, Italy
17. Arizona Genomics Institute, University of Arizona, 1657 E. Helen St, Tucson, AZ 85721, USA.
18. University of Bologna, Department of Biological, Geological and Environmental Sciences, Bologna, Italy.
19. Fano Marine Center, Fano, Italy.
20. Max Planck Institute for Plant Breeding Research (MPIPZ), Department of Chromosome Biology, Carl-von-Linne-Weg 10, 50829 Köln, Germany
21. Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany.
22. Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Nijenborgh 7, 9747AG, Groningen, Netherlands.
23. Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa.
24. College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, China.

ABSTRACT

We present chromosome-level genome assemblies from representative species of each of three independently evolved seagrass lineages, namely *Posidonia oceanica*, *Cymodocea nodosa*, *Thalassia testudinum*, and *Zostera marina.* We also include a draft genome of *Potamogeton acutifolius*, belonging to a freshwater sister lineage to Zosteraceae. All seagrass species share an ancient whole genome triplication, while additional whole genome duplications were uncovered for *C. nodosa*, *Z. marina* and *P. acutifolius*. Comparative analysis of selected gene families suggests that the transition from submerged-freshwater to submerged-marine environments mainly involved fine-tuning of multiple processes, e.g., osmoregulation, salinity, light capture, carbon acquisition and temperature, that all had to happen in parallel, likely explaining why adaptation to a marine lifestyle has been exceedingly rare. Major gene losses related to stomata, volatiles, defense, and lignification, are likely a consequence of the return to the sea rather than the cause of it. These new genomes will accelerate functional studies and solutions — as continuing losses of the 'savannas of the sea' are of major concern in times of climate change and loss of biodiversity.

KEYWORDS: Alismatales, convergent evolution, *Cymodocea nodosa,* hexaploidy, *Posidonia oceanica*, *Potamogeton acutifolium*, seagrasses, *Thalassia testudinum,* whole genome duplication (WGD), whole genome triplication (WGT), *Zostera marina*

INTRODUCTION

Seagrasses are unique flowering plants, adapted to a fully submerged existence in the highly saline environment of the ocean, where they must root in reducing sediments, endure chronic light limitation, and withstand considerable hydrodynamic forces. In spite of these obstacles, the 80 or so species are among the most widely distributed flowering plants [1-3] with recently measured estimates of coverage ranging from 600,000 km^2 [4] to a modeled value of 1,6 million km^2 [5,6]. Seagrasses fulfill many critical ecosystem functions and services including carbon sequestration, nutrient cycling, bacterial suppression, and coastal erosion protection [7-11]. Along with mangroves, saltmarshes, and coral reefs, seagrass meadows are among the most biologically productive ecosystems on Earth. They act as breeding and nursery grounds for a huge variety of organisms including juvenile and adult fish, epiphytic and free-living algae, mollusks, bristle worms, nematodes, and other invertebrates such as scallops, crabs, and shrimp. Their importance for marine megafauna such as sea turtles, dugongs and manatees is unrivalled and their disappearance an important driver of the decline of these marine animals [12]. Seagrasses also rank amongst the most efficient natural carbon sinks on Earth, sequestering CO_2 through photosynthesis and storing organic carbon in sediments for millennia [13]. While occupying only 0.1% of the ocean surface, seagrasses have been estimated to bury 27–44 Tg C_{org} per year globally, accounting for 10-18% of the total C burial in the oceans and being up to 40 times more efficient at capturing organic carbon than land-forests soils [14].

Previous work on *Zostera marina* [15,16] uncovered several unique gene family losses, as well as metabolic pathway losses and gains, that collectively underly novel structural and physiological traits, along with evidence for ancient polyploidy. Here, we expand on this work and present new chromosome-scale, high-quality reference genomes to understand the specific morphological and physiological adaptations that have enabled their worldwide distribution , except for Antarctica [1]. These included *Posidonia oceanica* (L.) Delile (Posidoniaceae), *Cymodocea nodosa* (Ucria) Ascherson (Cymodoceaceae), and *Thalassia testudinum* K. D. Koenig (Hydrocharitaceae) to chromosome level assemblies, and a closely related freshwater-submerged alismatid, *Potamogeton acutifolius* Link (Potamogetonaceae), to draft level. Representative seagrass species within each family (Supplementary Figure 1.1) were chosen based on ecological importance, susceptibility to anthropogenic pressure, and availability of an extensive ecological literature. Briefly, *Posidonia oceanica* is the iconic Mediterranean seagrass and the largest in terms of plant size and physical biomass. It is a climax species characterized by extreme longevity and carbon storage capacity. *Thalassia testudium* (turtle grass) is a climax tropical species unique to the greater Caribbean region, with a single sister species endemic to the Indo-Pacific. *Cymodocea nodosa* is restricted mainly to the Mediterranean, Black and Caspian Seas, with an Atlantic extension along the Canary Island archipelago and along the subtropical Atlantic coast of Africa. It is the only temperate species of an otherwise disjunct tropical genus from the Indo-Pacific. The curly pondweed *Potamogeton acutifolius* belongs to the sister family of Zosteraceae and was chosen as its closest submerged freshwater sister taxon. We also included the recently upgraded genome of *Zostera marina* L. [17], which is found throughout the northern hemisphere and is arguably the most widespread species on the planet [18]. To distinguish between adaptations to an aquatic lifestyle, and those unique to the ocean environment, our comparative analysis also included genomes of two recently sequenced emergent freshwater alismatids (which are rooted in underwater substrate, but have leaves and stems extending out of the water), along with the genomes of two distantly related salt-water tolerant mangrove species. In addition, representative transcriptomic data [16] of 89 Alismatales species was utilized to gain a more comprehensive view of shared and unique seagrass and freshwater adaptations within the order Alismatales (Supplementary Figure 1.1).

To better understand the extremely rare transition from a freshwater environment to a submerged saline environment, we compared gene family and pathway evolution across species, considering gene loss, as well as gene birth through small and large-scale gene duplication events, and investigated their effect on plant body structure (cell walls, stomata, hypolignification) and also investigated their relationship to physiological adaptations (hypoxia, plant defense, secondary metabolites, light perception, carbon acquisition, heat shock factors and especially salt tolerance mechanisms).

RESULTS AND DISCUSSION

Genome assemblies and gene annotations

We assembled the genomes of *T. testudinum*, *P. oceanica*, and *C. nodosa* to chromosomal level using a combination of short sequence reads, PacBio HiFi, PacBio long reads, and Hi-C chromosome mapping. The novel seagrass genomes varied in haploid chromosome number from 6 to 18 and were very different in size, while containing approximately the same number of gene models (Supplementary Table 2.1.4). Further details of genome assembly and annotation, based on a combination of *ab initio* prediction, homology searches, RNA-aided evidence, and manual curation can be found in Methods, Supplementary Table 2.1.4., Supplementary Note 2.1, and Supplementary Table 2.1.3. BUSCO scores of >95% demonstrate the high level of completeness in the genomes. The prediction of non-protein coding RNA families (i.e., rRNAs, tRNAs, snoRNAs) for *Z. marina, C. nodosa, P. oceanica, T. testudinum*, and *P. acutifolius* can be found in Supplementary Note 3.1 and Supplementary Table 3.1.). Figure 1 shows the distribution of different genomic features along the reconstructed pseudochromosomes for the different seagrass species. Information on plastid and mitochondrial genomes can be found respectively in Supplementary Note 2.2 and Supplementary Note 2.3.

Information on Nuclear-mitochondria (NUMTs) and nuclear-chloroplast (NUPTs) integrants can be found in Supplementary Note 2.4 and Supplementary Table 2.4.

Genome Evolution

Transposable elements

Transposable elements (TE) comprise more than 85% of the genomes of *T. testudinum* and *P. oceanica*, as compared to only 65% for *C. nodosa* and *Z. marina*, and 40% for *P. acutifolius* (Supplementary Table 4.1). Long terminal-repeat retrotransposons (LTR-REs) are the major class of TEs and account for 72%, 66%, 46% and 42% in *T. testudinum, P. oceanica, C. nodosa* and *Z. marina*, respectively. LTR/Gypsy elements account for 63.18% in *T. testudinum*, 57.8% in *P. oceanica* and 32.11% in *Z. marina*, whereas the proportion of LTR/Copia elements was higher than that of LTR/Gypsy in *C. nodosa* and *P. acutifolius*. Bursts of TEs (especially LTRs) create new genetic variation that may be adaptive under conditions of stress. Over evolutionary time, different TE loads and distributions among species provide clues related to habitat differences and stress resistance [19,20]. The insertion times of LTRs in the seagrass genomes (Methods) indicates a massive LTR/Gypsy burst around 200 thousand years ago (Kya) in *T. testudinum* (see y-axis), a moderate burst around 400 Kya in *P. oceanica* and *Z. marina*, but not in *C. nodosa*. By contrast, an expansion in Copia-elements happened around 2 Mya in *C. nodosa* but was weaker in *P. oceanica,* and nearly absent in *T. testudinum* and *Z. marina*. The recent TE gypsy burst (200 Kya) and older Copia burst (2 Mya median) coincide with drastic environmental fluctuations during Pleistocene ice ages (Supplementary Figure 4.1) and the timing of the trans-Arctic dispersal of *Z. marina* to the Atlantic from the Pacific [18]. The Gypsy bursts at 400 and 200 Kya correspond to Marine Isotope Stage MIS12 and MIS6, two heavy glaciations that were followed by rapid warming [21].

Whole genome duplication, ancient (hexa)polyploidy and dating

Next, we revisited the established whole genome duplication (WGD) in *Z. marina* [15] and investigated whether evidence for ancient polyploidy could be found in the other seagrasses, which are all behaving as functional diploids [22]. To this end, we used inferred age distributions of synonymous substitution rate (K_S) for paralogs retained in collinear regions (anchor pairs), along with gene-tree/species-tree reconciliation methods (see Methods, Supplementary Note 4.2.1 and Supplementary Note 4.2.2). First, K_S distributions of all seagrass species showed peaks indicative of ancient WGDs (Supplementary Figure 4.2.1)[16]. This was supported by intra- and inter-genomic collinearity analysis (see Supplementary Note 4.2.1.). Comparison of *P. oceanica* and *T. testudinum* with a reconstructed ancestral monocot karyotype (AMK [23]) shows a clear 3:1 synteny relationship, while a comparison of *Z. marina* with the AMK exhibits a 1:6 synteny relationship (Supplementary Figure 4.2.2). *Cymodocea nodosa* was also found to show a 6:1 relationship compared to the AMK, while showing a 2:1 relationship with its sister

species *P. oceanica* (Supplementary Figure 4.2.3), providing strong support for an additional WGD in *C. nodosa* after diverging from the *P. oceanica* lineage. Likewise, the freshwater species *P. acutifolius* was found to show a collinear relationship of 6:1 with the AMK and a 2:1 relationship with *P. oceanica*, and a 2:2 relationship with *C. nodosa,* while the colinearity relationship with its sister species *Z. marina* was more obscure (Supplementary Figure 4.2.4). However, these findings provide evidence that also *P. acutifolius* experienced an additional WGD event after its divergence with *P. oceanica* and *C. nodosa*. Of note, the overall 1:3 or 1:6 synteny relationships with the AMK suggested a hexaploid rather than a tetraploid ancestry for seagrasses and relatives.

Second, based on a K_S analysis using ksrates [24], we were able to confirm that this paleohexaploidy is shared by *P. oceanica*, *C. nodosa*, *Z. marina,* and *P. acutifolius,* while the analysis was inconclusive for *T. testudinum* (Supplementary Figure 4.2.5). To resolve this issue, we applied a gene-tree/species-tree reconciliation approach using WHALE [25], which confirmed that the ancient whole genome triplication (WGT) event is shared by all seagrasses, and *P. acutifolius.* WHALE also supported the younger WGD in *Z. marina* is shared with *P. acutifolius* (Supplementary Note 4.2.2 and Supplementary Figure 4.2.6). Phylogenomic dating of the WGT (see Methods and Supplementary Note 4.2.3) further shows that most gene duplicates are reconciled on the branch leading to the most recent common ancestor (MRCA) of Potamogetonaceae, Zosteraceae, Posidoniaceae, Cymodoceaceae and Hydrocharitaceae, at approximately 86.96 (89.89 - 79.81) Mya (Figure 2 and Supplementary Figure 4.2.7a). Recently, Chen et al. [16] also reported a WGD shared by all core Alismatales (Supplementary Figure 1.1). However, these authors suggested a WGD rather than a WGT, which can be attributed to the lack of structural data, since their study was based solely on transcriptome data. Independent absolute dating of the shared WGD for *P. acutifolius* and *Z. marina* confirmed an earlier obtained date for the *Zostera* WGD of approximately 65 Mya (Supplementary Figure 4.2.7c-f), coinciding with the K/Pg boundary [15], which was also used to date a recent within-species phylogeographic study for *Z. marina* [18].

Adaptation to the Marine Environment

All three seagrass lineages characterized in this study share many specific morphological and physiological adaptations to their specific environment. Historically, a number of features were proposed as prerequisites for marine angiosperm life, such as tolerance to submergence, tolerance to salinity, hydrophilous pollination, and a capacity for vegetative anchorage [26,27]. Previous studies have already reported genes potentially linked to the adaptation to the marine environment [15], while a recent study that conducted a broad transcriptome-based sampling of Alismatales uncovered some patterns of gene loss and gain also likely associated with aquatic and/or marine adaptation [16]. Discrimination between aquatic (i.e., freshwater) and marine adaptations is not necessarily easy. To achieve greater insights into both adaptations, we used a common set of species for which full genome information is available (four seagrasses, three freshwater alismatids, and 16 other angiosperms, Figure 2 and Supplementary Note 4.3). We also utilized the extensive transcriptome dataset of Chen et al. ([16]) and broadly assessed commonalities and differences in gains and losses across gene families (further referred to as orthogroups, see Methods and Extended Data Table 1-10). The most important findings on adaptation to both aquatic-submerged, and marine conditions are summarized in Figure 3 (and Supplementary Figure 5.1).

Use it or lose it - convergence and specificity of gene losses

Under water, stomata are not required and may even be harmful for a submerged lifestyle because of the intrusion of water. Hence, seagrasses, and to a limited extent also freshwater alismatids, e.g., *P. acutifolius,* have reduced the number of genes involved in their development. Specifically, out of 30 orthogroups containing guard cell toolkit genes [28], eleven have been convergently and completely lost in seagrasses, while six others were significantly contracted compared to non-seagrass genomes (Figure 3a and Extended Data Table 1). Lost gene families include positive (SMF transcription factors), negative (EPIDERMAL PATTERNING FACTOR1 AND 2 (encoded by *EPF1*, *EPF2*), and TOO MANY MOUTHS (encoded by *TMM*)) regulators of stomatal development, as well as stomatal function (encoded by *BLUS1*, *KAT1/2* and *CHX20*) (Figures 3a and 3c). Gene losses and

contractions in the guard cell toolkit are also seen in the submerged freshwater alismatid *P. acutifolius* studied here, and to a less extreme degree in the floating alismatid *S. polyrhiza* (Figure 3a and Extended Data Table 1).

The aqueous habitat of seagrasses is also not conducive to emitting volatile substances as signals. Accordingly, we observed a convergent loss of orthogroups associated with volatile metabolites and signals. This includes the biosynthesis of triterpenes, and the volatile systemic acquired resistance signal, methyl salicylate [29] (Extended Data Table 2). Probably a more dramatic gene loss relates to ethylene biosynthesis and signaling (Extended Data Table 2). Two species, *C. nodosa* and *Z. marina*, do not contain *ACS* or *ACO* genes and hence, are not expected to produce ethylene or its precursor 1-aminocyclopropane 1-carboxylic acid (ACC). Moreover, they seem to have lost the ability to respond to ethylene, as indicated by a severe contraction of the early ethylene signal transduction components (Figure 3a and 3d) [15,16,30]. In contrast, the downstream ethylene transcription factors (encoded by *EIN3*/*EIL1/2*) have been retained in all seagrasses, suggesting they can still exert ethylene-independent functions. Remarkably, and unlike *C. nodosa* and *Z. marina*, *T. testudinum* and *P. oceanica,* as well as freshwater submerged species, retained some components for functional ethylene biosynthesis and signaling, as was also reported by Chen et al. [16]. As diffusion of ethylene into water is extremely slow compared to diffusion in the air, ethylene rapidly accumulates in submerged organs. Such accumulation typically serves as a signal for submergence, and activating adaptive responses, such as formation of aerenchyma, adventitious rooting, shoot elongation, quiescence and priming the metabolism for efficient low-oxygen responses [31,32]. However, while the accumulation of ethylene can be considered beneficial for the flooding tolerance of land plants, high levels and prolonged exposure to ethylene can have detrimental effects, such as stunted growth, senescence and abscission of leaves and flowers, root growth inhibition, and increased stress sensitivity [33]. One possible mechanism that may prevent the accumulation of deleterious levels of ethylene, and thus explain its retention in *T. testudinum* and *P. oceanica,* is via epiphytic and endophytic bacteria that express ACC deaminases. This hypothesis is supported by the presence of multiple ACC deaminases in the metagenome of *P. oceanica* sediments [34], but needs further study.

Seagrasses increase their morphological flexibility to withstand hydrodynamic wave and current forces by a reduction in vascular tissues, the main site of lignification [35], consistent with the absence of vascular proliferation factor encoded by *WOX4*, and a contraction of the number of pericycle cell identity transcription factors (Figure 3a and Extended Data Table 3). This finding seems a more general adaption to aquatic lifestyles, as also suggested by analysis of the transcriptomes of different Alismatales (Supplementary Figure 5.1, this study, and ref ([16])). The most severe reduction of the vascular bundle is seen in *Z. marina* which even lacks a pericycle[36], a finding that correlates with the loss and divergence of the vascular proliferation regulators encoded by *PXY* and *MONOPTEROS/ARF5* (Figure 3a and Extended Data Table 3). Notably, the lack of *MONOPTEROS/ARF5* in *Z. marina* is further reflected in its inability to form an embryonic primary root [37]. The general cellular hypolignification in seagrasses is reflected in the reduction in the number of *LACCASE*s encoding the final enzymes in the lignin pathway, which oxidize monolignols to facilitate their polymerization into lignin [38,39] (Figures 3a, 3e and Extended Data Table 4). The reduced need for the monolignol production is matched by a reduction of respectively PHENYLALANINE AMMONIA LYASE (encoded by *PAL*), and HYDROXYCINNAMOYL-COA SHIKIMATE/QUINATE HYDROXYCINNAMOYL TRANSFERASE (encoded by *HCT*) genes, which constitute entrance points into phenylpropanoid biosynthesis [40] (Figures 3a, 3f and Extended Data Table 4). Gene family contractions in lignin biosynthesis are also observed for the submerged freshwater species *P. acutifolius* and the freshwater floating species *S. polyrhiza* (Figure 3a).

Arbuscular mycorrhizal symbiosis (AMS) were and are critical for plant terretrialization [41,42] and are found in salt marsh plants, mangrove forests, and freshwater ecosystems [43-45]. There is currently no evidence for any seagrass species to form mycorrhizal associations [46], which is reflected in the absence (secondary loss) of AMS-specific genes, with the sole exception of DMI3 in *P. oceanica* (Figure 3a). Gene loss of AMS-specific genes is also seen in freshwater submerged and floating species (Figure 3a). We also investigated so-called AMS-conserved genes, which have non-symbiotic roles [47] and discovered that seagrasses and *P. acutifolius* consistently retained a specific

set of these conserved genes (DMI1, NUP85, NUP133, NENA, CCD7, CCD8 and MAX2) (Figure 3a). The absence of NSP1 and NSP2 is not unique to seagrasses but seems to be rather a common adaptation observed in aquatic environments (Supplementary Figure 5.1) and Proteales species [48].

The pathogen landscape of the marine environment is associated with a different composition of plant resistance (R-genes) genes. In the seagrasses, there are fewer genes containing nucleotide-binding leucine-rich repeat receptors (NLRs) as compared to most other plants (Extended Data Table 5, Supplementary Note 5.2, Supplementary Table 5.2 and Supplementary Figure 5.2.1). As in many monocots, NLRs with a Toll/interleukin-1 receptor/resistance protein (TIR) domain are also completely absent in all seagrass lineages, as well as a few other NLR genes from the leucine rich repeat (LRR) domain. It is currently unclear what selective pressure was responsible for the unique R-gene composition of the seagrasses. Lower counts of disease resistance genes have also been observed for other aquatic plants [49].

Temperature fluctuations are much slower and show a lower amplitude in the marine compared to terrestrial environment [50]. Accordingly, we observed a reduction in the number of plant heat shock transcription factors (*HSF*s) that are involved in the rapid activation of stress-responsive genes upon temperature changes, and which have been linked to the evolutionary adaptation of plants to the terrestrial environment [50]. Seagrasses contain only about half the number of *HSF*s as compared with terrestrial plants (Extended Data Table 5, Supplementary Note 5.3 and Supplementary Table 5.3). Notably, only seagrasses belonging to the tropical genera retained some of the key heat stress-related *HSF*s from WGD and WGT events (Extended Data Table 5), which is consistent with their warmer native environment and higher heat stress tolerance compared to temperate seagrasses (*P. oceanica* and *Z. marina*).

Multi-level "tweaking" to adapt to the marine environment

Protective flavonoids and phenolics

Most seagrasses, except *C. nodosa,* seem to have greatly expanded the number of CHALCONE SYNTHASEs, which channel *p*-coumaroyl-CoA into flavonoid biosynthesis at the expense of monolignol biosynthesis (Figure 3a, 3f, and Extended Data Table 6). Flavonoids provide protection against UV and fungi, while enhancing recruitment of N-fixing bacteria [34,51,52]. Flavonoids and other phenolics in seagrasses can be sulphated by the activity of cytosolic sulphotransferases to increase their water solubility and bioactivity in the marine environment [53,54]. For example, the sulphated monolignol, zosteric acid (*O*-sulfonated *p*-coumaric acid) is an antifouling agent that prevents biofilm formation at the leaf surface [55]. Cytosolic sulphotransferases are expanded in seagrasses, but significantly contracted in *Potamogeton*. However, flavonoid glycosyltransferases and flavonoid beta-glucosidases are contracted in both (Figure 3a, 3f, and Extended Data Table 6). Jointly, these data illustrate how rerouting precursors of the lignin biosynthesis pathway likely facilitated two traits, i.e., reduced rigidity, which appears to be a general aquatic adaptation, and sulphated protection, which contributes to the evolution of the marine lifestyle of seagrasses [34,54]. In the case of *P. oceanica*, secreted phenolic compounds, together with anoxia, both inhibit microbial consumption of sucrose from root exudates [34].

Diverse mechanisms of cellular salt tolerance

Salt tolerance in flowering plants is a complex trait that involves multiple cellular processes [56]. In the extreme case of invasion of highly-saline, marine environments, one might assume wholesale changes in salt tolerance mechanisms and/or the evolution of specialized features, such as salt glands in mangrove species. To date, no obvious specialized structures involved in salt tolerance have been identified in seagrasses. Instead, it seems that canonical salt tolerance mechanisms have been fine-tuned or "tweaked" towards higher efficiency on multiple levels. A major challenge associated with the marine environment is to prevent the accumulation of noxious levels of Na^+ and Cl^-, while allowing the efficient uptake of the essential ion K^+. Angiosperms employ secondary Na^+ transport mechanisms based on Na^+/H^+ antiporters fueled by a strong electrochemical H^+ gradient. Surprisingly, no notable gene gains or losses were observed among the putative sodium transporting NHXs (*NHX1* and *SOS1/NHX7*), except for *C. nodosa,* which contains a few extra copies of *NHX1* and *SOS1* orthologs (Extended Data

Table 7). Instead of an increased number of genes, we observed similar amino-acid substitutions in regulatory domains of *SOS1* orthologs in all four species (Supplementary Figure 5.4.1), indicating the possibility of altered regulation of *SOS1/NHX7* in these species, a notion that is also supported by the loss of *SOS3*, a key regulator of *SOS1* activity in *C. nodosa* (Extended Data Table 7). The electrochemical H^+ gradients that fuel Na^+ transport is established via H^+ ATPases (encoded by *AHA*), V-ATPases and vacuolar H^+-PPases (encoded by *AVP1*). Of these genes, only the *AVP1* genes were obviously expanded in all the seagrasses, containing almost twice the number of *AVP1* genes found on average in other angiosperms (Figure 3a and 3b). Interestingly, the expansion of *AVP1*-like genes can, at least partly, be linked to the ancient WGT followed by their specific retention, suggesting that these additional AVP copies were co-opted for adaptation to a marine lifestyle (Supplementary Figure 5.5.2). Indeed, overexpression of such PPases has been shown to improve salt tolerance in several angiosperms (e.g., *Arabidopsis*, poplar, sugar cane) [57-59], by enhancing Na^+ sequestration in the vacuole [60]. Analysis of the K^+-channel repertoire in seagrasses reveals the loss of Shaker-type K^+ channels (Supplementary Figure 5.4.2) [61], and a greatly reduced number of CYCLIC NUCLEOTIDE GATE CATION CHANNELs (Figure 3a, 3b and Extended Data Table 7). Moreover, the constant high K+ concentrations in seawater (9.7mM) renders high-affinity K+ transport systems superfluous, explaining the absence of *AtHAK5* in all seagrass genomes (Figure 3a and 3b). Also, the Cl^- transporter repertoire is reduced in seagrasses (Figure 3a and 3b), and seagrasses lack orthologs for NPF2.4 and ALMT12/QUAC1, CLC-A, B and CLC-E, likely reflecting their adaptation to a marine lifestyle (Figure 3a and 3b).

Maintaining the elasticity of the cell wall is another critical component of salt tolerance. The elasticity and structural strength of the cell wall are mainly dictated by components such as cellulose and pectins that cross-link the cellulose microfibrils. The bivalent cation Ca^{2+} stiffens the cell wall by establishing electrostatic bond between pectin strands. The excess of monovalent Na+ in seawater may displace the divalent calcium and hinder dimerization of homogalacturonan chains that are present in canonical pectin [62]. In addition to the canonical pectin polysaccharides, seagrasses deposit apiogalacturonan in their cell walls [63]. The borate-bridges that cross-link apiogalacturonan chains are less sensitive to sodium displacement, providing an advantage to plants grown under high salt condition [64]. One of the few known key enzymes in the synthesis of apiogalaturonan is UDP-D-apiose/UDP-D-xylose synthase (encoded by *Api*), which converts UDP-D-glucuronate into UDP-D-apiose [65]. Its expansion in seagrasses (in particular in *Zostera* and *Cymodocea*) is reflected in the cell-wall composition of seagrasses and therefore likely contributes to salt tolerance (Figure 3a). In addition, the apiogalacturonan could provide a way to incorporate boron into the cell wall, and protect seagrasses against its toxic effects.

Compared to terrestrial lineages, no major changes were observed for cellulose and hemi-cellulose biosynthesis (Extended Data Table 7). Notably, most of the salt related evolutionary changes in seagrasses are not reflected in the genomes of mangrove species (*Avicinea marina* and *Rhizophora apiculate*), which is consistent with the independent evolution of salt tolerance in mangrove species [66,67].

Coping with hypoxic sediments

The solubility of oxygen in seawater is limited (typically around 10 mL O_2 L^{-1}), while the sediments in which seagrasses grow are oxygen-free and reducing below a sediment depth of a few mm. This increases the O_2 demand/draw-down by extensive belowground root-rhizome tissues that often comprise >50% of total plant biomass. Consistent with the increased risk of hypoxia, all seagrasses have expanded their repertoire of Plant Cysteine Oxidases (encoded by PCOs) and group VII Ethylene Responsive (*ERF-VII*s) genes, for direct sensing and transcriptional adjustment to hypoxia (Figure 3a, 3e and Extended Data Table 8). As expected, most *ERF-VII*s had higher expression in rhizomes and roots as compared to leaves (Supplementary Figure 5.5.1). Also, *P. acutifolius* contains an expanded hypoxia response machinery, reflecting its adaptation to submergence (Figure 3a). This is also supported by the transcriptome data of other Alismatales (Supplementary Figure 5.1) [16]. Again, many, if not most, *ERF-VII* members reside within syntenic blocks retained from the WGT event in seagrasses, especially for *P. oceanica* and *T. testudinum* (Supplementary Figure 5.5.2). Such increases in the number of genes through whole genome duplication is also true for multiple hypoxia-related genes. Some examples are: (1) the *PFK4* gene family, which encodes the rate-limiting enzyme in the glycolysis pathway (including enolases), expanded in both

seagrasses and *P. acutifolius*, and derived from the WGT event (Supplementary Figure 5.5.2); (2) Lactate dehydrogenase, a rate-limiting enzyme in lactate fermentation, that is also expanded in seagrasses (Figure 3a and Extended Data Table 8) and has been shown to provide higher waterlogging tolerance in *Arabidopsis* upon overexpression [68]; and (3) genes encoding the energy-sensing sucrose nonfermenting kinase *SnRK1* [69] and *eIFiso4G1* (the dominant regulator in translational regulation by *SnRK1* under hypoxia [70]) (Extended Data Table 8) are increased as a result of the WGT (Supplementary Figure 5.5.2). In conclusion, we speculate that the increase and specific retention of many hypoxia responsive genes, subsequent to the WGT (dated at ~86 Mya), might have coincided with the Cenomanian-Turonian anoxic event (~91± 8.6 Mya, [71,72]); if true, this low oxygen period may have helped to select for hypoxia tolerance in submerged species. In *C. nodosa* and *P. acutifolius*, additional recent lineage specific WGDs and tandem duplications may have also contributed to further expansion of the hypoxia responsive genes as a possible adaptation to submergence.

Light perception and photosynthetic carbon acquisition

Seagrass growth and zonation are constrained by light availability, as ocean waters rapidly attenuate photosynthetic active radiation with depth and modify its spectral quality, enriching blue while reducing red wavelengths [73]. Most seagrass species grow in shallow water and even in the clearest waters, only a few species reach depths of 40 m or more. Dissolved inorganic carbon (DIC) is mainly available as bicarbonate (HCO_3^-) in seawater (nearly 90% DIC at normal pH) that needs to be exploited via special acquisition systems, as it cannot diffuse passively across the cell plasma membrane [74]. The availability of dissolved CO_2 for photosynthesis is instead limited to ~1% of the DIC pool, hence submerged plants and algae evolved CO_2-concentration and convergent evolution of HCO_3- to CO_2 mechanisms (CCMs) to overcome this low availability. A recent report identified an evolutionary adaptation of RuBisCO kinetics across submerged angiosperms from marine, brackish-water and freshwater environments that correlates with the development and effectiveness of CCMs [75].

The analysis of genes related to inorganic carbon (Ci) acquisition revealed a slight increase in extracellular *α-CA* (encoding Carbonic Anhydrase α-type) copy number across the studied species (Supplementary Note 5.6.1). In *P. oceanica* and *P. acutifolius*, extra genes again have been specifically retained following the WGT event, although some copies also evolved local tandem duplications. *α-CA* OG0013954 was found to be specific to seagrasses (except for *T. testudinum*) and *P. acutifolius* (Extended Data Table 9 and Supplementary Table 5.6), and most of the corresponding genes are highly expressed in leaves (Supplementary Figure 5.6.1). This supports their involvement in Ci acquisition and possibly CCMs, as the presence of external CAs catalyzing the apoplastic dehydration of HCO_3^- to the RuBisCO substrate CO_2, together with a higher activity of the extrusion proton pumps [76], likely evolved to alleviate dissolved inorganic carbon limitation in most seagrass species [77].

Our findings of a retention of 15 C4-related genes after WGT or WGD events (of which two encode PEPC) support the hypothesis that *C. nodosa* could be a C4 species [78], similar to what has been observed in *P. acutifolius* (Extended Data Table 9). Notably, none of the studied seagrass species possesses the Serine-residue characteristic of C4 Phosphoenolpyruvate carboxylase (PEPC), thus likely ruling out that a terrestrial-like C4-based (biochemical) CCM system is operating in seagrasses. This would suggest the presence of some kind of C3-C4 intermediate metabolism. Alternatively, homologs to C4 genes could have a role in the resistance of seagrasses to a variety of abiotic stresses, including salt stress [79].

Consistent with an augmented need for light capture, seagrasses show an expansion of LHCB (encoding light-harvesting complex B) as compared to freshwater plants that occur close to the water surface (Supplementary Figure 5.6.2and Supplementary Note 5.6.2)**.** Only *C. nodosa* had a number of LHCB genes comparable to the freshwater *P. acutifolius* and *Spirodela* spp. Other components of the photosynthetic machinery, including Photosystems I and II, are similar in gene number to other species, either freshwater or terrestrial (Supplementary Figure 5.6.2). Seagrasses have conserved the full repertoire of orthologous genes encoding photosensory proteins and components of the light signaling systems (Supplementary Figure 5.6.4 and Supplementary Note 5.6.3) that evolved in the green lineages during the different stages of plant terrestrialization [80].

Species-specific adaptation to UV tolerance and downstream regulation, and its relation to light habitat features during the invasion of the marine environment, appear to have differed among seagrass lineages (Supplementary Note 5.6.3). Those living at lower latitudes with intense UV-B radiation throughout the year (*T. testudinum* and *C. nodosa*) have kept the typical *UVR8* of land plants along with their main regulatory proteins (encoded by *RUP1,2*). In contrast, *Z. marina,* as a higher latitude species, has lost the genes for both photoreceptors and their main negative regulatory proteins (Supplementary Figure 5.6.4), consistent with its lower exposure to UV-B radiation. In *P. oceanica*, a species restricted to the Mediterranean, the orthologous gene for *UVR8* lacks the sequence region C27 engaged in the regulation of UVR8 reversion state from the activated to the inactivated state. The species-specific adaptation in the UV-signaling and its negative feedback regulation (Supplementary Figure 5.6.5), further reinforce the idea that 'tweaking' and not massive change of key traits and their regulatory mechanisms facilitated the invasion of the marine environment.

Perception of surrounding light cues is also critical for the entrainment of the circadian clock system which in turn is essential for regulation of basic physiology and the life cycle, e.g., daily water and carbon availability, and hormone signaling pathways [81]. All seagrass species, except *T. testudinum* have lost the *TIMING OF CAB1* (encoded by *TOC1*) gene (Supplementary Figure 5.6.4). The general reduction of clock genes in aquatic species suggests that the "absence of drought", has led to a reduction of the regulatory daily-timing constraints for some metabolic and developmental plant processes. We find it interesting that all seagrasses have retained some genes related to the phytochromes light-signaling pathway. These include *PIF*s and *LAF1* (Supplementary Figure 5.6.4) following WGT and WGD events, as well as genes related to the circadian clock and photoperiodism such as GI and ZTL (Supplementary Figure 5.6.4).

No Apical Meristem (NAC) Transcription Factors (TF)

NAC transcription factors (TF) are among the largest plant-specific-transcription factor (TF) families involved in signaling crosstalk events. They mediate development and aging programs and environmental stress signals. While a comparable number of sequences are found in seagrasses as compared to land plants, freshwater and mongrove species, specific orthogroups were restricted to seagrasses. One of them is annotated as Transcription factor JUNGBRUNNEN 1 (encoded by *JUB1*), a central longevity regulator that is also involved in (salt) stress tolerance. A detailed screening of sequences annotated as JUB1 across other plant genomes reveals sequence similarities and functional reorganizations among JUB1 found in *C. nodosa* and *P. oceanica*. Besides the sequence similarity between the two species, only *C. nodosa* sequences are expressed (Supplementary Note 5.7 and Supplementary Figure 5.7). This difference in functional regulation could potentially be linked to the different ecological tolerance of the two species to environmental factors. Although the two species can coexist, *C. nodosa* can colonize enclosed and shallow environments, which have higher fluctuation range and speed of salinity, light and temperature.

Nitrogen Metabolism

Key genes linked to nitrogen uptake/transport and assimilation have been retained in all seagrasses examined, although nitrate transporters (encoded by NRTs) are strongly contracted (Extended Data Table 10 and Supplementary Note 5.8). This implies that seagrasses may have evolved alternative mechanisms for nitrogen uptake and utilization. Although our results are not particularly revealing in this regard, recent work on seagrass microbiomes has shown that nitrogen acquisition involves nitrogen-fixing bacteria in the roots [82] and that epiphytic micro-organisms on the leaves mineralize amino acids via their heterotrophic metabolism [83]. Gaining a more mechanistic understanding of the plant role in these interactions, is now possible for future investigations, given these new genomes.

Flower Development

Sexual reproduction in seagrasses occurs underwater (hydrophilous) by completely submerged male and female (unisexual) flowers. Their floral structures are simplified, often having reduced, or no, sepals and petals, which

may represent an adaption to hydrophilous, and mostly abiotic, pollination [84]. However, this striking morphological adaption is not reflected by a striking loss of genes defining the well-known ABC(D)E model for floral organ-specification [85,86] (Supplementary Table 5.9 and Figure 4a). In *Z. marina*, the B-function (encoding *PISTILATA*) homolog seems to be mainly expressed in the staminate ("male") flower, while two C-function homologs (*AGAMOUS*; *AGa* and *AGb*) were mainly expressed in the pistillate ("female") flower (Figure 4d), suggesting involvement of the B-function only in stamen development, and C-function in carpel development. In *P. oceanica*, the expression patterns differ from those in *Z. marina*, but largely agree with previously ascribed roles in floral organ patterning: B-function *PI* and C-function *AG* homologs are highly expressed in both staminate and pistillate flowers (Figure 4e). However, in both seagrasses, one A-function homolog, *AGL6*, is highly expressed in pistillate flowers, indicating the possibility of A-function neofunctionalization, transitioning from a role in sepals and petals to one being associated with pistillate flower development. The two *SEP* E-function homologs of three seagrasses are highly expressed in pistillate and staminate flowers, indicating an essential role of these flower-specific co-factors in organ specification. The discrepancy between the floral simplification and the presence of all types of floral organ identity genes in the seagrass genomes may reflect the instability of the floral ground plan between alismatid lineages [87], and is possibly affected by neofunctionalisation and shifts in expression domains of floral identity genes.

Hydrophilous pollination is extremely rare outside the seagrasses, leading to the proposal that it is one of the defining features of seagrasses [26]. The majority of seagrasses have flexible, filiform pollen in which a rigid exine layer is structurally reduced or absent [88], likely facilitating hydrophilous pollination. Consistent with the loss or severe reduction of the exine layer, many genes involved in the biosynthesis and secretion of the exine layer (Supplementary Note 5.9) are absent in *Z. marina* [15], while *C. nodosa,* P. *oceanica,* and *T. testudinum* show partial gene loss (Figure 4f). It will be of interest to also investigate the role of pollen-specific genes, such as an orthologs of RESTORER OF FERTILITY 1 *(*encoded by *RF-1)*, in the evolution of hydrophylous pollination. Supplementary Figure 5.9 shows flower and pollen development toolkit gene family expansion and contraction values for 96 species, including the 90 species-transcriptome data set of Chen et al. [16].

CONCLUSION

Seagrasses are now recognized as foundational species for invaluable ecosystems that provide multiple functions and services [9]. They prevent erosion and hence preserve coastal seascapes, serve as biodiversity hotspots for associated animals, algae and plants, and have recently been proposed as a nature-based solution for climate mitigation owing to their carbon storage capacity in belowground biomass [89]. Seagrasses also represent an extremely rare adaptation in the world of flowering plants, unlike (re-)adaptation to freshwater environments, which occurred at least 222 times in embryo-bearing plants [90]. As far as is known, in part due to an extremely poor fossil record, seagrasses have evolved only on three different occasions from freshwater ancestors to (a group of) species that lives continuously submerged in a highly saline environment, including subaqueous pollination (except in *Enhalus acoroides* [91]). Why only 84 species, spread across the three lineages, emerged in a time interval of 100 Mya, remains unresolved, but it may be related to high ocean connectivity on one hand [92], while within-species, ecological tolerance and phenotypic plasticity is high [93].

Comparative genome analysis has unveiled considerable convergence in seagrasses, but mainly for processes and pathways that have become redundant or even detrimental in a submerged marine environment. These include genes for stomata development, ethylene biosynthesis and signaling, pollen-coat formation, disease resistance, and heat shock transcription factors (HSFs). Jointly, these results illustrate that the invasion of the marine environment is associated with a significant loss of genes in multiple pathways that are no longer needed, a compelling example of "use it or lose it."

Clear evidence of convergent positive (or gain of function) adaptation among the different lineages of seagrasses is harder to establish. Rather than unveiling major biological innovations including the rewiring of biological networks, adaptation to the marine environment seems mainly to involve the fine-tuning of many different/supportive processes that likely all had to happen in parallel, possibly explaining why the transitioning to a marine lifestyle has been exceedingly rare. For instance, adaptation of seagrasses to a marine (saline) environment was not accompanied by massive changes to individual salt tolerance traits, but rather involved more subtle changes in gene copy number and regulatory mechanisms, along with structural adaptations of the cell walls. This gradual modulation of preexisting mechanisms is consistent with the presence of multiple less extreme halophytes within alismatid families [94]. The fine-tuning of many biological processes may also have facilitated the considerable phenotypic plasticity displayed by seagrass populations allowing their colonization from the tropics to the poles.

Many of the genes co-opted in different pathways in seagrasses seem to have been specifically retained following WGDs and WGTs that occurred long ago, suggesting important interdependencies of large-scale (or major) genome evolution events and evolutionary adaptation. Prime examples identified here are hypoxia-responsive genes, genes involved in salt tolerance, flavonoid metabolism, carbon acquisition, and C4-like photosynthesis. Therefore, the co-option of extra genes specifically retained following ancient whole genome duplications likely played a crucial role in facilitating survival in a marine environment.

We expect that the new, high-quality, seagrass genomes presented here will accelerate experimental and functional studies and contribute to transformative solutions in the management and conservation of seagrass ecosystems, which is an urgent concern in times of climate change and marine biodiversity crisis given the continuing worldwide loss of seagrass meadows.

METHODS

Sampling metadata, DNA and RNA preparation

Whole plants from each species were collected from the field, transported to the lab in a cool box, cleaned, frozen in LN_2 and then stored at -80°C. Collection and processing information are summarized in Supplementary Table 1.1. All samples were made with collection permits and followed the CBD-Nagoya Protocol. Care was taken to use tissue harvested from the basal area of young, clean leaves (10-cm pieces) to minimize epiphytic diatoms and bacteria If necessary. The seagrass tissues were then sent by overnight courier on dry ice to the Arizona Genomics Institute, Tucson, AZ, USA for extraction of nucleic acids (https://www.genome.arizona.edu). Quality controlled nucleic acid samples were then shipped on dry ice to the Joint Genome Institute (JGI) in Berkeley, CA, USA (https://jgi.doe.gov/) for further diagnostics and sequencing library preparation. For *P. acutifolius*, nucleic acids were extracted, QC'd and sequenced at the Max Planck-Genome-Centre Cologne, Germany (https://mpgc.mpipz.mpg.de /home/).

High Molecular Weight (HWM) DNA was extracted from young leaves of *T. testudinium*, *P. oceanica*, and *C. nodosa*, using the protocol of Doyle and Doyle (1987)[95] with minor modifications. Young leaves, that had been flash frozen in LN2 and kept frozen at -80C, were ground to a fine powder in a frozen pestle and mortar with LN2 followed by very gentle extraction in CTAB buffer (that included proteinase K, PVP-40 and β-mercaptoethanol) for 20 mins at 37°C, followed by 20 mins at 50°C. Following centrifugation, the supernatant was gently extracted twice with 24:1 chloroform: iso-amyl alcohol. The upper phase was adjusted to $1/10^{th}$ volume with 3M Sodium acetate (pH=5.2), gently mixed, and DNA precipitated with iso-propanol. DNA was collected by centrifugation, washed with 70% EtOH, air dried for few minutes and dissolved thoroughly in 1x TE at room temperature. Size was validated by pulsed field electrophoresis. HMW DNA for *P. acutifolius* was extracted from 2 g of young leaves with the NucleoBond HMW DNA kit (Macherey Nagel). Quality was assessed with a FEMTOpulse device (Agilent) and the quantity was measured by a Quantus fluorometer (Promega).

RNA was extracted from seagrass leaves, rhizomes, roots, and flowers (Supplementary Table 1.1) with the NucleoSpin RNA Plant and Fungi Kit (Macherey-Nagel, USA), and checked for integrity by capillary electrophoresis using an Agilent (Santa Clara, CA, USA) 2100 Bioanalyzer with the Agilent RNA 6000 Nano Kit following manufacturer's instructions. RNA was extracted from leaves and roots of *P. acutifolius* with the RNAeasys Plant Kit (Qiagen), including an on-column DNase I treatment. Quality was assessed with an Agilent Bioanalyser and the quantity was calculated by an RNA-specific kit from Quantus (Promega).

Genome Sequencing

The genomes of *T. testudinium*, *P. oceanica*, and *C. nodosa* were determined following a whole genome shotgun sequencing strategy and standard sequencing protocols. Sequencing reads were produced using the Illumina NovaSeq platform and the PacBio SEQUEL II platform at the Department of Energy (DOE) Joint Genome Institute (JGI) in Walnut Creek, California, and the Hudson Alpha Institute in Huntsville, Alabama. One 400bp insert 2x150 Illumina fragment library and one HiC library was sequenced for each organism. Technical sequencing statistics are summarized in Supplementary Table 2.1.1. Prior to assembly, Illumina fragment reads were screened for PhiX contamination and reads composed of >95% simple sequences were removed. Furthermore, Illumina reads <50bp, after trimming for adapter and checking for quality (q<20), were also removed. For the Illumina sequencing, the final combined read set consisted of 4,284,278,120 high-quality reads with 161x coverage for *T. testudinium*, 6,543,657,580 high-quality reads with 327x coverage for *P. oceanica*, and 693,903,610 high-quality reads with 208x coverage for C. nodosa. For the PacBio sequencing, a total of 18 PB chemistry 3.1 chips (30-hour movie time) were sequenced with a HiFi read yield of 231.8 Gb with 51.53x coverage, 238.3 Gb with 79.44x coverage and 39.6 Gb with 79.24x coverage for *T. testudinium*, *P. oceanica* and *C. nodosa*, respectively.

For *P. acutifolius*, all libraries (PacBio, RNA and Tell-seq) and PacBio HiFi sequencing were performed at the Max Planck-Genome-Centre Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). Short-read libraries and sequencing (RNA-seq and Tell-seq) were performed at Novogene Ltd (UK), using a NovaSeq 6000 S4 flowcell Illumina system. An Illumina-compatible was prepared with the NEBNext® Ultra™ II RNA Library Prep Kit for Illumina. PacBio-HiFi libraries were prepared according to the manual "Procedure & Checklist - Preparing HiFi SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0" with an initial DNA fragmentation by g-Tubes (Covaris) and final library size selection on BluePippin (Sage Science). Size distribution was again controlled by FEMTOpulse (Agilent). Size-selected libraries were sequenced on a Sequel II with Binding Kit 2.0 and Sequel II Sequencing Kit 2.0 for 30 h (Pacific Biosciences). The same genomic DNA was used for TELL-seq but without fragmentation. Library preparation was done as outlined in the manual "TELL-Seq™ WGS Library Prep User Guide" (ver. November 2020). Illumina "sequencing-by-synthesis" was performed on a HiSeq 2500, 2 x 250 bp with additional index sequencing cycles to read out the unique fragment barcodes. Sequences were analyzed as recommended by Universal Sequencing Technology (UST, Canton, U.S.A). The final combined read set consisted of 54,401,190 Illumina high-quality reads with 13.4 coverage and 1,900,000 PacBio HiFi reads with 43.5 coverage (Supplementary Table 2.1.1)

Genome assembly

For *T. testudinium*, *P. oceanica* and *C. nodosa*, the following assembly strategy was used: the PacBio HiFi data was assembled using HiFiAsm and subsequently polished using RACON (https://github.com/lbcb-sci/racon). Due to the high heterozygosity of our sequenced seagrasses, both haplotypes were nearly complete resulting in a genome assembly composed of a highly contiguous primary set of chromosomes and a more fragmented alternative set of chromosomes (Supplementary Figure 2.1.1). For *T. testudinium*, the initial primary assembly consisted of 1,987 contigs with a contig N50 of 483.4 Mb, and a total assembled size of 4,866.1 Mb. For *P. oceanica*, the initial primary assembly consisted of 3,470 contigs, with a contig N50 of 355.8 Mb, and a total assembled size of 3,192.0 Mb (Supplementary Table 2.1.2). For *C. nodosa*, we produced an initial primary assembly of 1,362 contigs, with a contig N50 of 18.5 Mb, and a total assembled size of 466.0 Mb (Supplementary Table 2.1.2). Misjoins in the assemblies were identified using HiC data as part of the JUICER/JuiceBox pipeline[96]

for each of the three seagrass genomes. After resolving the misjoins, the broken contigs were then oriented, ordered, and joined together with HiC data using the JUICER/JuiceBox pipeline. In *T. testudinum*, there were 5 misjoins identified in the polished primary assembly, and a total of 15 joins were applied to the primary assembly to form the final assembly consisting of 9 chromosomes. In both the *P. oceanica* and *C. nodosa* polished primary genomes, there were no misjoins identified. A total of 6 joins were applied to the primary assemblies of *P. oceanica* and *C. nodosa* to form the final assembly consisting of 10 chromosomes and 18 chromosomes, respectively. Each chromosome join is padded with 10,000 Ns. Significant telomeric sequence was identified using the $(TTTAGGG)_n$ repeat, and care was taken to make sure that contigs terminating in telomere were properly oriented in the production assembly. The remaining scaffolds were screened against bacterial proteins, organelle sequences, GenBank nr and removed if found to be a contaminant. Heterozygous SNP/indel phasing errors were corrected using the HiFi data (51.53x for *T. testudinum*, 79.44x for *P. oceanica* and 79.24x for *C. nodosa*). Finally, homozygous SNPs and indels were corrected in the releases using Illumina reads (2x150, 400bp insert). A total of 2,613 homozygous SNPs and 82,421 homozygous indels were corrected in *T. testudinum*. A total of 1,643 homozygous SNPs and 100,570 homozygous indels were corrected in *P. oceanica* and total of 1,426 homozygous SNPs and 12,492 homozygous indels were corrected in the *C. nodosa*. Due to the high heterozygosity of the three genomes, both haplotypes of each chromosome were well represented in the assemblies. The primary set of chromosomes were constructed from the primary assembly, while an alternative set of chromosomes were constructed from the alternate assembly. Chromosomes for the alternate haplotype were then oriented, ordered, and joined together using synteny from the primary chromosomes (Supplementary Table 2.1.3).

For *Potamogeton acutifolius*, we used HiFiAsm [97] to assemble a draft genome assembly of a total length of 611 Mb with N50 = 3.09 Mb and scaffolded it further with Tell-seq data (linked reads; bioRxiv 2019, 852947) using the ARCS software [98] and reaching final N50 = 4.45 Mb (6,705 scaffolds in total, the length of the largest scaffold = 31.2 Mb).

Genome annotation

Structural and functional annotation of genes

Our annotation pipeline integrated three independent approaches, the first one based on transcriptome data, the second one being an *ab initio* prediction and the third based on protein homology. Both RNA-seq and Iso-seq data from different tissues (Supplementary Table 3.2.1 – Supplementary Table 3.2.4) were used to aid the structural annotation and RNA-seq datasets were first mapped using Hisat2 (v2.1.0, arguments --dta) [99] and subsequently assembled into transcript sequences by Stringtie2 [100], whereas Iso-seq sequences were aligned to the seagrass genome using GMAP [101]. All transcripts from RNA-seq and Iso-seq were combined using Cuffcompare (v2.2.1) and subsequently merged with Stringtie2 (arguments --merge -m 150) to remove fragments and redundant structures [100]. Transdecoder v5.0.2 (github.com/TransDecoder) was then used to predict protein sequences with diamond v2.0.14 results (--evalue 1e-5 --max-target-seqs 1 -f 6). BARKER v2.1.2 [102] was used for *ab initio* gene prediction using model training based on RNA-seq data. Homology-based annotation was based on the protein sequences from related species (*Z. marina* v1.0, *Spirodela polyrhiza*, *Oryza sativa* and *Arabidopsis thaliana*) as query sequences to search the reference genome using TBLASTN with e-value ≤$1e^{-5}$, then regions mapped by these query sequences were subjected to Exonerate to generate putative transcripts. Additionally, an independent, homology-based gene annotation was performed using GeMoMa [103] using the same species with TBLASTN.

All structural gene annotations were joined with EvidenceModeller [104] v1.1.1, and BUSCO v4.0.4 (Benchmarking Universal Single-Copy Orthologs) [105] was used to assess the quality of the annotation results. Finally, we used GenomeView [106] to do the gene curations manually based on the RNA-seq and Iso-seq data. Putative gene functions were identified using InterProScan [107] with different databases, including PFAM, Gene3D, PANTHER, CDD, SUPERFAMILY, ProSite and GO. Meanwhile, functional annotation of these predicted genes was obtained by

aligning the protein sequences of these genes against the sequences in public protein databases and the UniProt database using BLASTP with the e-value ≤1e – 5.

Annotation of non-protein coding RNA families

Finished genome assemblies and annotations (genome.fasta and genome.gff files for *Z. marina*, *C. nodosa*, *P. oceanica, T. testudinum* and *P. acutifolius*) were uploaded to, and later downloaded from, JGI Phytozome [108]. Infernal v1.1.4 (Dec 2020) [109] was used to perform sequence similarity searches of each genome sequence versus the RFAM database (RNA families database, Dec2021) [110]. The output from Infernal was filtered, keeping only the hits with an E-value threshold E<0.01. A second filtering step was performed to remove redundant information, i.e., overlapping matches with similar hits. A third filtering step was performed by retaining all the hits matching with a coverage of at least 95% and removing all partial/fragmented matches with incomplete hits from the reference collection. rRNA, tRNA, snoRNA and miRNA regions were selected and annotated in the annotation.jff files for each species. An updated functional annotation including the identified loci in the genomes was performed by scanning the Uniprot database [111] with BLASTp [3]. Introns and the corresponding sequence regions were extracted by GenomeTools [112] and Bedtools [113] programs. The functional annotation of the long introns (>= 20kb) was performed by similarity searches in the NCBI nucleotide [114] database with the BLASTn tool [3].

Annotation of repeats and transposable elements (TEs)

Two complementary approaches were used to identify repetitive DNA sequences. First, a *de novo* repeat identification was carried out with RepeatModeler v2.0.1 (https://www.repeatmasker.org/RepeatModeler/) based on the default TE Rfam database, followed by RepeatMasker v4.1 (https://www.repeatmasker.org/) to discover and classify repeats based on the custom repeat libraries from RepeatModeler v2.0.1. Second, LTR_Finder [115] (v1.0.7), LTR_harvest [116] from genometools (v1.5.9) and LTR_retriever [117] (v2.9.0) were used to identify and trace the LTR elements, which were subsequently characterized at clade/lineage level by searching coding domains within the sequences, using the tool Domain based ANnotation of Transposable Elements (DANTE) (https://github.com/kavonrtep/dante). Transposable elements not classified by RepeatModeler were analyzed using DeepTE [118]. We merged the libraries from RepeatModeler, LTR_retriever and DeepTE using USEARCH [119] with 80% identity as the minimum threshold for combining similar sequences into the final non-redundant de novo repeat library. Finally, we used RepeatMasker v4.1.0 (-e rmblast -gff -xsmall -s -norna -no_is -lib) to identify and classify repeats in the genome assemblies of seagrasses and *Potamogeton*.

Dating bursts of repeats in seagrass genomes

The identification of high-quality intact LTR-RTs and the calculation of insertion age for intact LTR-RTs were carried out using LTR_retriever (v2.9.0), using the formula T=K/2r. The nucleotide substitution rate "r" was set to 1.3e-8 substitutions per site per year [120].

Identifying Whole Genome Duplications

K_S age distributions and gene tree-species tree reconciliation

K_s age distribution analysis was performed using the wgd package [121]. Anchor pairs (i.e., paralogous genes lying in collinear or syntenic regions of the genome) were obtained using i-ADHoRe [122]. Ks distribution analysis was also performed using the KSRATES software [123], which locates ancient polyploidization events with respect to speciation events within a phylogeny, comparing paralog and ortholog K_S distributions, while correcting for substitution rate differences across the involved lineages (see Supplementary Note 4.2.1).

OrthoFinder [124] was used to build orthologous gene families. For each orthogroup, a multiple sequence alignment (MSA) based on amino acid sequences was obtained using PRANK [125] and then used as input for Markov Chain Monte Carlo (MCMC) analysis in MrBayes [126]. A time-calibrated species tree was inferred by MCMCtree from the PAML package [127], using reference speciation times of 42–52 million years ago (MYA) for the divergence between *Oryzae sativa* and *Brachypodium distachyon*, 118-129 MYA for that between *Spirodela polyrhiza* and *Z. marina*,

and 130-140 for that between *Spirodela* and other terrestrial monocots [128]. A gene duplication-loss (DL)+WGD model, under critical and relaxed branch-specific rates, was implemented for the inference of the significance and corresponding retention rates of the assumed WGD events under Bayesian inference [25]. (see Supplementary Note 4.2.2)

Absolute dating of WGDs

Absolute dating of WGD events followed an approach previously described for *Zostera marina* [15]. Paralogous gene pairs located in duplicated segments (so-called anchors) and duplicated pairs lying under the WGD peak (so-called peak-based duplicates) were collected for phylogenetic dating. Anchors, which are assumed to correspond to the most recent WGD, were detected using i-ADHoRe 3.0 [122]. For each WGD paralogous pair, an orthogroup was created that included the two paralogues plus several orthologues from other plant species, as identified by InParanoid (v. 4.1) [129], using a broad taxonomic sampling. Gene duplicates were then dated using the BEAST v. 1.7 package [130] under an uncorrelated relaxed clock model with the LG+G (four rate categories) evolutionary model. A starting tree with branch lengths satisfying all fossil-prior-constraints was created according to the consensus APGIII phylogeny. Fossil calibrations were implemented using log-normal calibration priors (see Supplementary Note 4.2.3).

Time-calibrated tree construction

Protein sets were collected for 23 species (see Supplementary Note 4.3). These species were selected as representatives for monocots and eudicots, and representing different habitats from terrestrial, freshwater-floating, freshwater-submerged, to marine-submerged. Orthofinder v2.3 [131] was used to delineate gene families with mcl inflation factor 3.0. All-versus-all Diamond blast with an E-value cutoff of 1e−05 was performed and orthologous genes were clustered using OrthoFinder. Single-copy orthologous genes were extracted from the clustering results. MAFFT [132]) with default parameters was used to perform multiple sequence alignment of protein sequences for each set of single-copy orthologous genes, and to transform the protein sequence alignments into codon alignments after removing the poorly aligned or divergent regions using trimAl [133]. The resulting codon alignments from all single copy orthologs were then concatenated into one supergene for species phylogenetic analysis. A maximum-likelihood phylogenetic tree of single-copy protein alignments and codon alignments was constructed using IQ-TREE [134] with the GTR+G model and 1,000 bootstrap replicates. Divergence times between species were estimated using MCMCtree from the PAML package under the GTR+G model (see Supplementary Note 4.3).

Gene family comparisons

Gene families analyzed in the paper were searched in the output from Orthofinder and a master table was compiled to show the detailed information for each orthogroup, which is defined as the group of genes from multiple species descended from a single gene in the last common ancestor. For the superfamilies, we used the phylogenetic tree to further classify them into subfamilies. We adopted a custom criterion to assess the expansion and contraction of gene families. If the average gene number in seagrasses increased or reduced by >40% compared to non-seagrass species, we called it expansion or contraction. Syntenic analysis of genes are performed using MCScanX [135] and i-ADHoRe [122]. Lastly, circos plots were drawn using Circos [136].

Data availability

The DNA sequencing data for *C. nodosa* genome assembly has been deposited in the NCBI databases under the BioProject PRJNA1041560 via the link: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1041560 All assemblies and annotations for all seagrass species discussed in the current paper can be found at https://bioinformatics.psb.ugent.be/gdb/seagrasses/. Transcriptome data (including raw data and clean data) and sequencing QC Reports for *C. nodosa* can be found at https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Cymnodnscriptome_2; transcriptome data and sequencing QC Reports for *P. oceanica* can be found at

https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Posocenscriptome_2; transcriptome data and sequencing QC Reports for *T. testudinum* can be found at https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Thatesnscriptome_4; transcriptome data for *Z. marina* is from Jeanine et al. (2016). For the public databases, RFAM database v14.7 can be downloaded at https://ftp.ebi.ac.uk/pub/databases/Rfam/14.7/; UniProt database can be accessed from the web at http://www.uniprot.org and downloaded from http://www.uniprot.org/downloads; NCBI nucleotide database can be accessed via https://www.ncbi.nlm.nih.gov/

ACKNOWLEDGEMENTS

Y.VdP., J.L.O., T.B.H.R. and G.P. acknowledge funding from the US-Dept. of Energy, Joint Genome Institute, Berkeley, California, USA, under the Community Sequencing Program 2018, Project Number 504341 (Marine Angiosperm Genomes Initiative-MAGI). The CSP award also included support sequencing and plant bioinformatics from HudsonAlpha Institute for Biotechnology, Huntsville, AL; and DNA/RNA extraction and processing from the Arizona Genomics Institute, Tuscon, AZ. Y.VdP. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522) and from Ghent University (Methusalem funding, BOF.MET.2021.0005.01). P.N. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 497665889, 1606/3-1 for research on *Potamogeton*. M.K. acknowledge funding through the Helmholtz School for Marine Data Science (MarDATA), Grant No. HIDSS-0005. The work of G.P., E.D., J.P., and M.R. was partially supported by the project Marine Hazard, PON03PE_00203_1 (MUR, Italian Ministry of University and Research) and by the National Biodiversity Future Centre (NBFC) Program, Italian Ministry of University and Research, PNRR, Missione 4 Componente 2 Investimento 1.4 (Project: CN00000033). M.D.D., L.L.W., M.P.T. and Y.Y.S. acknowledge funding from Universiti Malaysia Terengganu (SRG Vot55317). The work (proposal: 10.46936/10.25585/60001196) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. The work of A.B. was performed within the Papanin Institute for Biology of Inland Waters RAS state assignment (theme 121051100099-5).

Author contributions table

X.M., V.S. and J.Ch. contributed equally to the work and are joint first authors. Y.VdP., J.L.O., T.B.H.R. and G.P. conceived the project, provided the overall evolutionary context, and wrote the proposal. Y.VdP., J.L.O., T.B.H.R. and G.P. and S.V. wrote and edited the main manuscript, and organized and further edited the individual contributions for the Supplementary Notes (as listed for the Supplementary Information below). All co-authors then provided specific feedback in forming the final version.

J.L.O., J.Ca., G.P., L.M.G., T.B.H.R., A.B., A.M., and P.N. contributed to sample tissue collection, preparation, and shipping for DNA extraction. S.Raj., L.B., G.H., J.W., M.Y performed the HMW DNA extractions and QC, as well as RNA extractions and QC for annotation assistance. J.S. and J.G. coordinated genome sequencing management steps for the seagrasses. A.M. and P.N. coordinated genome sequencing management steps for *Potamogeton*. K.B. was responsible for overall JGI technical coordination, liaison with principal investigators and project manager.

J.J., C.P., J.S., Y.VdP., S.R., A.S., J.vV. and T.B. performed analysis activities surrounding genome assembly (PacBio, HiC), supporting transcriptomics for annotation. J.J., Y.VdP., S.R., X.M. were responsible for deposition and maintenance of the species on the ORCAE site, and deposition of the new genomes to NCBI and Phytozome. J.Ch., X.M., S.M., J.Ch. were responsible for manuscript graphics.

Analysis of architectural features of genome evolution and annotation of specific gene families, including the written contributions to the main paper and Supplementary Information sections as follows: M.L.C., L.A. for the

Orthogroups Master Extended Data; M.L.C., A.S., X.M., JCh. for overview of gene families; H.C., X.M., J.Ch., Y.VdP. for Whole Genome Duplications/Triplications and dating; M.L.C., X.M. for Transposable Elements and repeat elements; M.K., T.B.H.R. for Organellar genomes; M.L.C., L.A. for Non-protein coding RNA families; S.V., X.M. for Stomata; S.V., X.M. for Volatile metabolites and signaling, ethylene; X.M., S.V. for Plant body development, lignification, vascular tissue; T.B. for Plant defense, R-genes; L.M.G. for Heat shock factors; S.V., X.M. for Flavenoids and phenolics; S.V. for Cellular salt tolerance; S.V., B.V. for Cell wall plasticity; S.V., X.M. for Hypoxia; G.P. for Light perception, photosynthesis, light harvesting, transcription factors; G.P., M.R. Carbon acquisition, CCMs; G.P. for UVB tolerance; G.P., E.D. for Clock genes; J.P. for NAC genes; D.M., L.W., M.P.T., Y.Y.S. for Nitrogen metabolism.

X.M., J.Ch., S.R. were responsible for data deposit on ORCAE platform. K.B., J.G. were in charge of data deposit on NCBI.

Competing Interests

The authors declare no competing interests

Figure legends

Figure 1. Distribution of the genomic features for the seagrass species *T. testudinum, P. oceanica, Z. marina* and *C. nodosa.* Tracks from the inner to outer side correspond to gene density (blue); LTR/Gypsy density (green); LTR/Copia (orange); DNA transposable elements (pink) and chromosomes (with length in Mb). Curved lines through the center denote synteny between different genomic regions. Grey lines in A, B and C reflect synteny involving the WGD, whereas the three colored lines represent synteny with WGTs. Colored lines in D represent synteny and strong intragenomic conservation and should not be compared with colors in A, B and C (see text for further details). The distribution of the genomic features for the longest scaffolds of *P. acutifolius*, can be found in Supplementary Figure 2.1.2.

Figure 2. Time-calibrated phylogeny and WGT/WGD events across flowering plants that have chromosome-level genome assemblies. The tree was inferred from 146 single-copy genes and show WGDs and WGTs based on inferences from the current study and previous analyses (Supplementary Table 4.2 and Supplementary Figure 4.2.8). For a more comprehensive tree showing the phylogenetic position of seagrasses within Alismatales, see Supplementary Figure 1.1. The dashed lines represent additional freshwater Alismatales species (phylogenetic position inferred using transcriptome data), mainly added for illustrative purposes to show non-monophyly of seagrass species. All branches have bootstrap support >98%. See text and Methods for details.

Figure 3. The loss, contraction, and expansion of gene families involved in the adaption to a marine environment. a) The normalized gene copy numbers for 4 seagrasses and 19 representative non-seagrass species. The normalization on the family dataset divides the gene count number of each species by the largest gene copy number within that family. The species order on the top of the heatmap is the same as that in Figure 1. The colors correspond to the different life-forms. The orange ones are terrestrial species; the green ones are emergent species (floating-leaved); the light blue ones are submerged species; the navy-blue ones are marine species (seagrasses) and the black ones are mangroves **b)** Salt stress signaling implies different ion channels. *HAK5* encodes HIGH-AFFINITY POTASSIUM TRANSPORTER 5; CNGC, CYCLIC NUCLEOTIDE GATE CATION CHANNELs; *AVP1* encodes Vacuolar H^+-PPases **c)** Stomata differentiation from meristemoid mother cells (MMC) to guard mother cell (GMC), to guard cells. **d)** Ethylene synthesis and signaling. **e)** The hypoxia-responsive signaling in which the direct (*ERF-VII*) and indirect responsive (*SnRK1*) pathways are expanded. The rate-limiting enzyme (encoded by *PFK4*) in the glycolysis pathway, along with Lactate dehydrogenase (encoded by *LDH*), a rate-limiting enzyme in fermentation, are also expanded. **F)** Simplified schematic of the lignin and flavonoid biosynthesis pathways. Only steps that have significantly changed are shown. *PAL* encodes phenylalanine ammonialyase, which is the gateway enzyme of the general phenylpropanoid pathway; *CHS* encodes chalcone synthase, which is the first enzyme of flavonoid biosynthesis that directs the metabolic flux to flavonoid biosynthesis; *GT1* encode flavonoid glycosyltransferases, which catalyze the final step of flavonoid biosynthesis to generate various flavonoid glycoside derivatives; *GH1* encode flavonoid beta-glucosidase & myrosinase, which are responsible for the recycling of carbohydrate-based flavonoids; *HCT* encode Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase, channels phenylpropanoids via the "esters" pathway to monolignols"; *LACCASEs* encode the final enzymes in the pathway that oxidize monolignols to facilitate their polymerization into lignin. Panels d) e) f) genes in red are expanded; blue means contracted; The dashed line in the pathway means multiple metabolic steps.

Figure 4. Flower development (like MADS-box genes) and pollen toolkit genes. a) Phylogenetic tree of type II MADS-box genes in seagrasses and *P. acutifolius*, including *Arabidopsis thaliana* (AT) and *Oryza sativa* (Os) for reference. b) Gene expression patterns for type II MADS-box genes from various organs of *Z. marina*. Expression values were scaled by log2(TPM+ 1). c) Gene expression patterns for type II MADS-box genes from various organs of *P. oceanica*. Expression values were scaled by log2(TPM+ 1). d) The flowering ABCE model in *Z. marina* specifying female and male organs as proposed based on gene expression values (bar heights) from b. e) The flowering ABCE model in *P. oceanica* specifying female and male organs as proposed based on the gene expression values (bar heights) from c. f) Normalized gene copy numbers for MADS-box and pollen toolkit genes for 4 seagrasses and 19 representative non-seagrass species. Normalization for each gene family was obtained by dividing the number of genes in that gene family for a particular species by the largest gene copy number within that family (considering all species). Genes in black are absent. Taxa are arranged phylogenetically and colored by life form.

References

1 Green, E. P. & Short, F. T. *World Atlas of Seagrasses. Prepared by the UNEP World Conservation Monitoring Centre.*, 48-58 (Univ. of California Press, Berkeley, USA, 2003).

2 Short, F., Carruthers, T., Dennison, W. & Waycott, M. Global seagrass distribution and diversity: A bioregional model. *Journal of Experimental Marine Biology and Ecology* **350**, 3-20 (2007). https://doi.org:https://doi.org/10.1016/j.jembe.2007.06.012

3 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). https://doi.org:10.1186/1471-2105-10-421

4 McKenzie, L. J. *et al.* The global distribution of seagrass meadows. *Environmental Research Letters* **15**, 074041 (2020). https://doi.org:10.1088/1748-9326/ab7d06

5 Duffy, J. E. *et al.* Toward a Coordinated Global Observing System for Seagrasses and Marine Macroalgae. *Frontiers in Marine Science* **6** (2019).

6 Gallagher, A. J. *et al.* Tiger sharks support the characterization of the world's largest seagrass ecosystem. *Nature Communications* **13**, 6328 (2022). https://doi.org:10.1038/s41467-022-33926-1

7 Bertelli, C. M. & Unsworth, R. K. F. Protecting the hand that feeds us: Seagrass (Zostera marina) serves as commercial juvenile fish habitat. *Marine Pollution Bulletin* **83**, 425-429 (2014). https://doi.org:https://doi.org/10.1016/j.marpolbul.2013.08.011

8 Nordlund, L., Koch, E., Barbier, E. & Creed, J. Seagrass Ecosystem Services and Their Variability across Genera and Geographical Regions. *PLOS ONE* **11**, e0163091 (2016). https://doi.org:10.1371/journal.pone.0163091

9 Unsworth, R. K. F., Cullen-Unsworth, L. C., Jones, B. L. H. & Lilley, R. J. The planetary role of seagrass conservation. *Science* **377**, 609-613 (2022). https://doi.org:10.1126/science.abq6923

10 Waycott, M. *et al.* Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc Natl Acad Sci U S A* **106**, 12377-12381 (2009). https://doi.org:10.1073/pnas.0905620106

11 Reusch, T. B. H. *et al.* Lower Vibrio spp. abundances in Zostera marina leaf canopies suggest a novel ecosystem function for temperate seagrass beds. *Marine Biology* **168**, 149 (2021). https://doi.org:10.1007/s00227-021-03963-3

12 Sievers, M. *et al.* The Role of Vegetated Coastal Wetlands for Marine Megafauna Conservation. *Trends in Ecology & Evolution* **34**, 807-817 (2019). https://doi.org:https://doi.org/10.1016/j.tree.2019.04.004

13 Duarte, C. M., Sintes, T. & Marbà, N. Assessing the CO2 capture potential of seagrass restoration projects. *Journal of Applied Ecology* **50**, 1341-1349 (2013). https://doi.org:https://doi.org/10.1111/1365-2664.12155

14 Macreadie, P. I. *et al.* Blue carbon as a natural climate solution. *Nature Reviews Earth & Environment* **2**, 826-839 (2021). https://doi.org:10.1038/s43017-021-00224-1

15 Olsen, J. L. *et al.* The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**, 331-335 (2016). https://doi.org:10.1038/nature16548

16 Chen, L.-Y. *et al.* Phylogenomic Analyses of Alismatales Shed Light into Adaptations to Aquatic Environments. *Molecular Biology and Evolution* **39**, msac079 (2022). https://doi.org:10.1093/molbev/msac079

17 Ma, X. *et al.* Improved chromosome-level genome assembly and annotation of the seagrass, Zostera marina (eelgrass). *F1000Res* **10**, 289 (2021). https://doi.org:10.12688/f1000research.38156.1

18 Yu, L. *et al.* Ocean current patterns drive the worldwide colonization of eelgrass (Zostera marina). *Nature Plants* **9**, 1207-1220 (2023). https://doi.org:10.1038/s41477-023-01464-3

19 Dubin, M. J., Mittelsten Scheid, O. & Becker, C. Transposons: a blessing curse. *Curr Opin Plant Biol* **42**, 23-29 (2018). https://doi.org:10.1016/j.pbi.2018.01.003

20 Vicient, C. M. & Casacuberta, J. M. Impact of transposable elements on polyploid plant genomes. *Annals of Botany* **120**, 195-207 (2017). https://doi.org:10.1093/aob/mcx078

21 Böse, M., Lüthgens, C., Lee, J. R. & Rose, J. Quaternary glaciations of northern Europe. *Quaternary Science Reviews* **44**, 1-25 (2012). https://doi.org:https://doi.org/10.1016/j.quascirev.2012.04.017

22 Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat Rev Genet* **18**, 411-424 (2017). https://doi.org:10.1038/nrg.2017.26

23 Murat, F., Armero, A., Pont, C., Klopp, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics* **49**, 490-496 (2017). https://doi.org:10.1038/ng.3813

24 Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* **38**, 530-532 (2022). https://doi.org:10.1093/bioinformatics/btab602

25 Zwaenepoel, A. & Van de Peer, Y. Inference of Ancient Whole-Genome Duplications and the Evolution of Gene Duplication and Loss Rates. *Mol Biol Evol* **36**, 1384-1404 (2019). https://doi.org:10.1093/molbev/msz088

26 Arber, A. *Water plants: a study of aquatic angiosperms*. (Cambridge University Press, 1920).

27 Den Hartog, C. *The seagrasses of the world*. (North Holland Publishing Co., 1970).

28 Harris, B. J., Harrison, C. J., Hetherington, A. M. & Williams, T. A. Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology* **30**, 2001-2012.e2002 (2020). https://doi.org:https://doi.org/10.1016/j.cub.2020.03.048

29 Shulaev, V., Silverman, P. & Raskin, I. Airborne signalling by methyl salicylate in plant pathogen resistance. *Nature* **385**, 718-721 (1997). https://doi.org:10.1038/385718a0

30 Golicz, A. A. *et al.* Genome-wide survey of the seagrass Zostera muelleri suggests modification of the ethylene signalling network. *J Exp Bot* **66**, 1489-1498 (2015). https://doi.org:10.1093/jxb/eru510

31 Sasidharan, R. & Voesenek, L. A. C. J. Ethylene-Mediated Acclimations to Flooding Stress. *Plant Physiology* **169**, 3-12 (2015). https://doi.org:10.1104/pp.15.00387

32 Hartman, S. *et al.* Ethylene-mediated nitric oxide depletion pre-adapts plants to hypoxia stress. *Nat Commun* **10**, 4020 (2019). https://doi.org:10.1038/s41467-019-12045-4

33 Van de Poel, B., Smet, D. & Van Der Straeten, D. Ethylene and Hormonal Cross Talk in Vegetative Growth and Development. *Plant Physiol* **169**, 61-72 (2015). https://doi.org:10.1104/pp.15.00724

34 Sogin, E. M. *et al.* Sugars dominate the seagrass rhizosphere. *Nat Ecol Evol* (2022). https://doi.org:10.1038/s41559-022-01740-z

35 Kuo, J., Cambridge, M. L. & Kirkman, H. in *Seagrasses of Australia: Structure, Ecology and Conservation* (eds Anthony W. D. Larkum, Gary A. Kendrick, & Peter J. Ralph) 93-125 (Springer International Publishing, 2018).

36 Barnabas, A. D. & Arnott, H. J. Zostera capensis Setchell: root structure in relation to function. . *Aquatic Botany* **27**, 309-322 (1987).

37 Taylor, A. R. A. Studies of the development of Zostera marina L.: II. Germination and seedling development. *Can J Botany* **35**, 477-499 (1957).

38 Zhuo, C. *et al.* Developmental changes in lignin composition are driven by both monolignol supply and laccase specificity. *Sci Adv* **8**, eabm8145 (2022). https://doi.org:10.1126/sciadv.abm8145

39 Zhao, Q. *et al.* Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. *Plant Cell* **25**, 3976-3987 (2013). https://doi.org:10.1105/tpc.113.117770

40 Barros, J. & Dixon, R. A. Plant Phenylalanine/Tyrosine Ammonia-lyases. *Trends Plant Sci* **25**, 66-79 (2020). https://doi.org:10.1016/j.tplants.2019.09.011

41 Wang, B. *et al.* Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol* **186**, 514-525 (2010). https://doi.org:10.1111/j.1469-8137.2009.03137.x

42 Strullu-Derrien, C., Selosse, M.-A., Kenrick, P. & Martin, F. M. The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist* **220**, 1012-1030 (2018). https://doi.org:https://doi.org/10.1111/nph.15076

43 Kohout, P. *et al.* Surprising spectra of root-associated fungi in submerged aquatic plants. *FEMS Microbiology Ecology* **80**, 216-235 (2012). https://doi.org:10.1111/j.1574-6941.2011.01291.x

44 Moora, M. *et al.* AM fungal communities inhabiting the roots of submerged aquatic plant Lobelia dortmanna are diverse and include a high proportion of novel taxa. *Mycorrhiza* **26**, 735-745 (2016). https://doi.org:10.1007/s00572-016-0709-0

45 Bohrer, K. E., Friese, C. F. & Amon, J. P. Seasonal dynamics of arbuscular mycorrhizal fungi in differing wetland habitats. *Mycorrhiza* **14**, 329-337 (2004). https://doi.org:10.1007/s00572-004-0292-7

46 Nielsen, S. L., Thingstrup, I. & Wigand, C. Apparent lack of vesicular–arbuscular mycorrhiza (VAM) in the seagrasses Zostera marina L. and Thalassia testudinum Banks ex König. *Aquatic Botany* **63**, 261-266 (1999). https://doi.org:https://doi.org/10.1016/S0304-3770(98)00123-5

47 Gomez-Roldan, V. *et al.* Strigolactone inhibition of shoot branching. *Nature* **455**, 189-194 (2008). https://doi.org:10.1038/nature07271

48 Chang, J. *et al.* The genome of the king protea, Protea cynaroides. *Plant J* **113**, 262-276 (2023). https://doi.org:10.1111/tpj.16044

49 Liu, Y. *et al.* An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Mol Plant* **14**, 2015-2031 (2021). https://doi.org:10.1016/j.molp.2021.08.001

50 Scharf, K. D., Berberich, T., Ebersberger, I. & Nover, L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochim Biophys Acta* **1819**, 104-119 (2012). https://doi.org:10.1016/j.bbagrm.2011.10.002

51 Papazian, S., Parrot, D., Burýskova, B., Weinberger, F. & Tasdemir, D. Surface chemical defence of the eelgrass Zostera marina against microbial foulers. *Sci Rep* **9**, 3323 (2019). https://doi.org:10.1038/s41598-019-39212-3

52 Lamb, J. B. *et al.* Seagrass ecosystems reduce exposure to bacterial pathogens of humans, fishes, and invertebrates. *Science* **355**, 731-733 (2017). https://doi.org:10.1126/science.aal1956

53 Teles, Y. C. F., Souza, M. S. R. & Souza, M. F. V. Sulphated Flavonoids: Biosynthesis, Structures, and Biological Activities. *Molecules* **23** (2018). https://doi.org:10.3390/molecules23020480

54 Grignon-Dubois, M. & Rezzonico, B. Phenolic chemistry of the seagrass Zostera noltei Hornem. Part 1: First evidence of three infraspecific flavonoid chemotypes in three distinctive geographical regions. *Phytochemistry* **146**, 91-101 (2018). https://doi.org:10.1016/j.phytochem.2017.12.006

55 Vilas-Boas, C., Sousa, E., Pinto, M. & Correia-da-Silva, M. An antifouling model from the sea: a review of 25 years of zosteric acid studies. *Biofouling* **33**, 927-942 (2017). https://doi.org:10.1080/08927014.2017.1391951

56 van Zelm, E., Zhang, Y. & Testerink, C. Salt Tolerance Mechanisms of Plants. *Annu Rev Plant Biol* **71**, 403-433 (2020). https://doi.org:10.1146/annurev-arplant-050718-100005

57 Gaxiola, R. A. *et al.* Drought- and salt-tolerant plants result from overexpression of the AVP1 H+-pump. *Proc Natl Acad Sci U S A* **98**, 11444-11449 (2001). https://doi.org:10.1073/pnas.191389398

58 Kumar, T., Uzma, Khan, M. R., Abbas, Z. & Ali, G. M. Genetic improvement of sugarcane for drought and salinity stress tolerance using Arabidopsis vacuolar pyrophosphatase (AVP1) gene. *Mol Biotechnol* **56**, 199-209 (2014). https://doi.org:10.1007/s12033-013-9695-z

59 Yang, Y. *et al.* Overexpression of a Populus trichocarpa H+-pyrophosphatase gene PtVP1.1 confers salt tolerance on transgenic poplar. *Tree Physiol* **35**, 663-677 (2015). https://doi.org:10.1093/treephys/tpv027

60 Duan, X. G., Yang, A. F., Gao, F., Zhang, S. L. & Zhang, J. R. Heterologous expression of vacuolar H(+)-PPase enhances the electrochemical gradient across the vacuolar membrane and improves tobacco cell salt tolerance. *Protoplasma* **232**, 87-95 (2007). https://doi.org:10.1007/s00709-007-0268-5

61 Nakamura, R. L. & Gaber, R. F. Ion selectivity of the Kat1 K+ channel pore. *Mol Membr Biol* **26**, 293-308 (2009). https://doi.org:10.1080/09687680903188332

62 Morris, E. R., Powell, D. A., Gidley, M. J. & Rees, D. A. Conformations and interactions of pectins. I. Polymorphism between gel and solid states of calcium polygalacturonate. *J Mol Biol* **155**, 507-516 (1982). https://doi.org:10.1016/0022-2836(82)90484-3

63 Gloaguen, V. *et al.* Structural characterization and cytotoxic properties of an apiose-rich pectic polysaccharide obtained from the cell wall of the marine phanerogam Zostera marina. *J Nat Prod* **73**, 1087-1092 (2010). https://doi.org:10.1021/np100092c

64 Byrt, C. S., Munns, R., Burton, R. A., Gilliham, M. & Wege, S. Root cell wall solutions for crop plants in saline soils. *Plant Science* **269**, 47-55 (2018). https://doi.org:https://doi.org/10.1016/j.plantsci.2017.12.012

65 Mølhøj, M., Verma, R. & Reiter, W. D. The biosynthesis of the branched-chain sugar d-apiose in plants: functional cloning and characterization of a UDP-d-apiose/UDP-d-xylose synthase from Arabidopsis. *Plant J* **35**, 693-703 (2003). https://doi.org:10.1046/j.1365-313x.2003.01841.x

66 Xu, S. *et al.* The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *National Science Review* **4**, 721-734 (2017). https://doi.org:10.1093/nsr/nwx065

67 Natarajan, P. *et al.* A reference-grade genome identifies salt-tolerance genes from the salt-secreting mangrove species Avicennia marina. *Communications Biology* **4**, 851 (2021). https://doi.org:10.1038/s42003-021-02384-8

68 Dolferus, R. *et al.* Functional analysis of lactate dehydrogenase during hypoxic stress in Arabidopsis. *Functional Plant Biology* **35**, 131-140 (2008).

69 Baena-González, E., Rolland, F., Thevelein, J. M. & Sheen, J. A central integrator of transcription networks in plant stress and energy signalling. *Nature* **448**, 938-942 (2007). https://doi.org:10.1038/nature06069

70 Cho, H.-Y., Lu, M.-Y. J. & Shih, M.-C. The SnRK1-eIFiso4G1 signaling relay regulates the translation of specific mRNAs in Arabidopsis under submergence. *New Phytologist* **222**, 366-381 (2019). https://doi.org:https://doi.org/10.1111/nph.15589

71 Monteiro, F. M., Pancost, R. D., Ridgwell, A. & Donnadieu, Y. Nutrients as the dominant control on the spread of anoxia and euxinia across the Cenomanian-Turonian oceanic anoxic event (OAE2): Model-data comparison. *Paleoceanography* **27** (2012). https://doi.org:https://doi.org/10.1029/2012PA002351
72 Selby, D., Mutterlose, J. & Condon, D. J. U–Pb and Re–Os geochronology of the Aptian/Albian and Cenomanian/Turonian stage boundaries: Implications for timescale calibration, osmium isotope seawater composition and Re–Os systematics in organic-rich sediments. *Chemical Geology* **265**, 394-409 (2009). https://doi.org:10.1016/j.chemgeo.2009.05.005
73 Kirk, J. Light and Photosynthesis in Aquatic Systems. *Light and Photosynthesis in Aquatic Ecosystems, third edition* **Vol. VI**, 1-651 (2010). https://doi.org:10.1017/CBO9781139168212
74 Campbell, J. E. & Fourqurean, J. W. Mechanisms of bicarbonate use influence the photosynthetic carbon dioxide sensitivity of tropical seagrasses. *Limnology and Oceanography* **58**, 839-848 (2013). https://doi.org:https://doi.org/10.4319/lo.2013.58.3.0839
75 Capó-Bauçà, S., Iñiguez, C., Aguiló-Nicolau, P. & Galmés, J. Correlative adaptation between Rubisco and CO2-concentrating mechanisms in seagrasses. *Nature Plants* **8**, 706-716 (2022). https://doi.org:10.1038/s41477-022-01171-5
76 Rubio, L. *et al.* Direct uptake of HCO3− in the marine angiosperm Posidonia oceanica (L.) Delile driven by a plasma membrane H+ economy. *Plant, Cell & Environment* **40**, 2820-2830 (2017). https://doi.org:https://doi.org/10.1111/pce.13057
77 Larkum, A. W. D., Davey, P. A., Kuo, J., Ralph, P. J. & Raven, J. A. Carbon-concentrating mechanisms in seagrasses. *Journal of Experimental Botany* **68**, 3773-3784 (2017). https://doi.org:10.1093/jxb/erx206
78 Koch, M., Bowes, G., Ross, C. & Zhang, X.-H. Climate change and ocean acidification effects on seagrasses and marine macroalgae. *Global Change Biology* **19**, 103-132 (2013). https://doi.org:https://doi.org/10.1111/j.1365-2486.2012.02791.x
79 Chen, S., Peng, W., Ansah, E. O., Xiong, F. & Wu, Y. Encoded C4 homologue enzymes genes function under abiotic stresses in C3 plant. *Plant Signal Behav* **17**, 2115634 (2022). https://doi.org:10.1080/15592324.2022.2115634
80 Han, X. *et al.* Origin and Evolution of Core Components Responsible for Monitoring Light Environment Changes during Plant Terrestrialization. *Mol Plant* **12**, 847-862 (2019). https://doi.org:10.1016/j.molp.2019.04.006
81 McClung, C. R. The Plant Circadian Oscillator. *Biology* **8** (2019).
82 Mohr, W. *et al.* Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine bacterium. *Nature* **600**, 105-109 (2021). https://doi.org:10.1038/s41586-021-04063-4
83 Tarquinio, F. *et al.* Microorganisms facilitate uptake of dissolved organic nitrogen by seagrass leaves. *ISME J* **12**, 2796-2800 (2018). https://doi.org:10.1038/s41396-018-0218-6
84 Kuo, J. & Hartog, C. d. in *SEAGRASSES: BIOLOGY, ECOLOGYAND CONSERVATION* (eds Anthony W. D. Larkum, Robert J. Orth, & Carlos M. Duarte) 51-87 (Springer Netherlands, 2006).
85 Krizek, B. A. & Fletcher, J. C. Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics* **6**, 688-698 (2005). https://doi.org:10.1038/nrg1675
86 Lohmann, J. U. & Weigel, D. Building beauty: the genetic control of floral patterning. *Dev Cell* **2**, 135-142 (2002). https://doi.org:10.1016/s1534-5807(02)00122-3
87 Remizowa, M. V., Sokoloff, D. D. & Rudall, P. J. EVOLUTIONARY HISTORY OF THE MONOCOT FLOWER. *Annals of the Missouri Botanical Garden* **97**, 617-645 (2010).
88 Ackerman, J. D. in *SEAGRASSES: BIOLOGY, ECOLOGYAND CONSERVATION* (eds Anthony W. D. Larkum, Robert J. Orth, & Carlos M. Duarte) 89-109 (Springer Netherlands, 2006).
89 Orth, R. J. *et al.* Restoration of seagrass habitat leads to rapid recovery of coastal ecosystem services. *Science Advances* **6**, eabc6434 https://doi.org:10.1126/sciadv.abc6434
90 Cook, C. D. K. The number and kinds of embryo-bearing plants which have become aquatic: a survey. *Perspectives in Plant Ecology, Evolution and Systematics* **2**, 79-102 (1999). https://doi.org:https://doi.org/10.1078/1433-8319-00066
91 Ackerman, J. D. in *Seagrasses: Biology, Ecology and Conservation* (eds W. D. Larkum, R. J. Orth, & C. M. Duarte) 89-109 (Springer, NL, 2006).
92 Waycott, M., Biffin, E. & Les, D. H. in *Seagrasses of Australia: Structure, Ecology and Conservation* (eds Anthony W. D. Larkum, Gary A. Kendrick, & Peter J. Ralph) 129-154 (Springer International Publishing, 2018).
93 Pazzaglia, J., Reusch, T. B. H., Terlizzi, A., Marín-Guirao, L. & Procaccini, G. Phenotypic plasticity under rapid global changes: The intrinsic force for future seagrasses survival. *Evolutionary Applications* **14**, 1181-1201 (2021). https://doi.org:https://doi.org/10.1111/eva.13212

94 Flowers, T. J., Galal, H. K. & Bromham, L. Evolution of halophytes: multiple origins of salt tolerance in land plants. *Functional Plant Biology* **37**, 604-612 (2010).

95 Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin (1987)

96 Dudchenko, O. *et al.* The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*, 254797 (2018). https://doi.org:10.1101/254797

97 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175 (2021). https://doi.org:10.1038/s41592-020-01056-5

98 Yeo, S., Coombe, L., Warren, R. L., Chu, J. & Birol, I. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* **34**, 725-731 (2018). https://doi.org:10.1093/bioinformatics/btx675

99 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-915 (2019). https://doi.org:10.1038/s41587-019-0201-4

100 Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**, 278 (2019). https://doi.org:10.1186/s13059-019-1910-1

101 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005). https://doi.org:10.1093/bioinformatics/bti310

102 Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108 (2021). https://doi.org:10.1093/nargab/lqaa108

103 Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. *Methods Mol Biol* **1962**, 161-177 (2019). https://doi.org:10.1007/978-1-4939-9173-0_9

104 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008). https://doi.org:10.1186/gb-2008-9-1-r7

105 Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* **1962**, 227-245 (2019). https://doi.org:10.1007/978-1-4939-9173-0_14

106 Abeel, T., Van Parys, T., Saeys, Y., Galagan, J. & Van de Peer, Y. GenomeView: a next-generation genome browser. *Nucleic Acids Research* **40**, e12-e12 (2012). https://doi.org:10.1093/nar/gkr995

107 Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-120 (2005). https://doi.org:10.1093/nar/gki442

108 Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* **42**, D26-31 (2014). https://doi.org:10.1093/nar/gkt1069

109 Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013). https://doi.org:10.1093/bioinformatics/btt509

110 Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res* **49**, D192-d200 (2021). https://doi.org:10.1093/nar/gkaa1047

111 Consortium, U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480-d489 (2021). https://doi.org:10.1093/nar/gkaa1100

112 Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10**, 645-656 (2013). https://doi.org:10.1109/TCBB.2013.68

113 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033

114 Benson, D. A. *et al.* GenBank. *Nucleic Acids Res* **41**, D36-42 (2013). https://doi.org:10.1093/nar/gks1195

115 Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268 (2007). https://doi.org:10.1093/nar/gkm286

116 Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008). https://doi.org:10.1186/1471-2105-9-18

117 Ou, S. & Jiang, N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons *Plant Physiology* **176**, 1410-1422 (2017). https://doi.org:10.1104/pp.17.01310

118 Yan, H., Bombarely, A. & Li, S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* **36**, 4269-4275 (2020). https://doi.org:10.1093/bioinformatics/btaa519

119 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010). https://doi.org:10.1093/bioinformatics/btq461

120 Ma, J. & Bennetzen, J. L. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci U S A* **101**, 12404-12410 (2004). https://doi.org:10.1073/pnas.0403715101

121 Zwaenepoel, A. & Van de Peer, Y. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **35**, 2153-2155 (2019). https://doi.org:10.1093/bioinformatics/bty915

122 Proost, S. *et al.* i-ADHoRe 3.0--fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res* **40**, e11 (2012). https://doi.org:10.1093/nar/gkr955

123 Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* (2021). https://doi.org:10.1093/bioinformatics/btab602

124 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019). https://doi.org:10.1186/s13059-019-1832-y

125 Löytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *P Natl Acad Sci USA* **102**, 10557-10562 (2005). https://doi.org:10.1073/pnas.0409137102

126 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001). https://doi.org:10.1093/bioinformatics/17.8.754

127 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007). https://doi.org:10.1093/molbev/msm088

128 An, D. *et al.* Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of Spirodela. *Proc Natl Acad Sci U S A* **116**, 18893-18899 (2019). https://doi.org:10.1073/pnas.1910401116

129 O'Brien, K. P., Remm, M. & Sonnhammer, E. L. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* **33**, D476-480 (2005). https://doi.org:10.1093/nar/gki107

130 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973 (2012). https://doi.org:10.1093/molbev/mss075

131 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015). https://doi.org:10.1186/s13059-015-0721-2

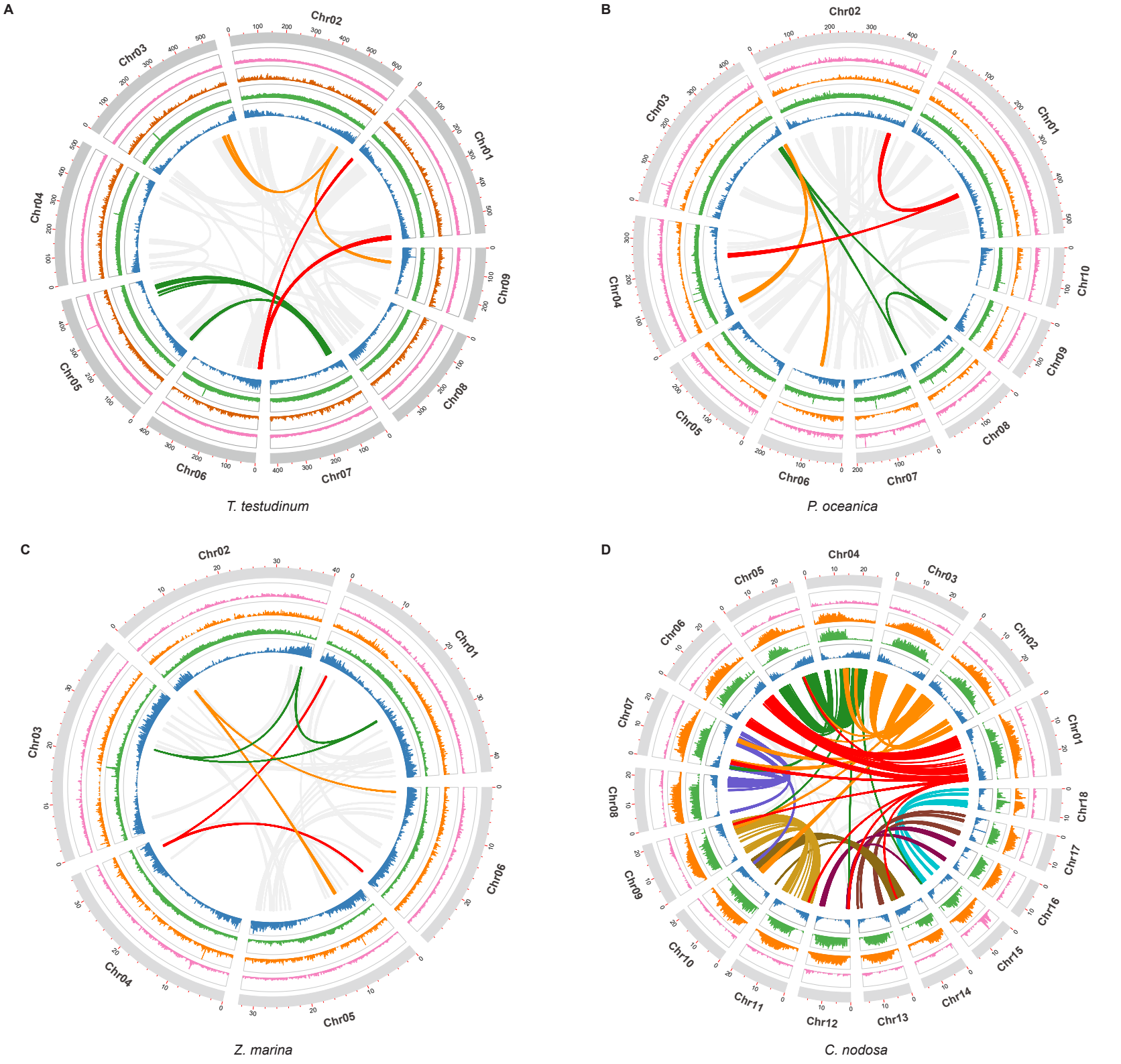
132 Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* **47**, W5-W10 (2019). https://doi.org:10.1093/nar/gkz342

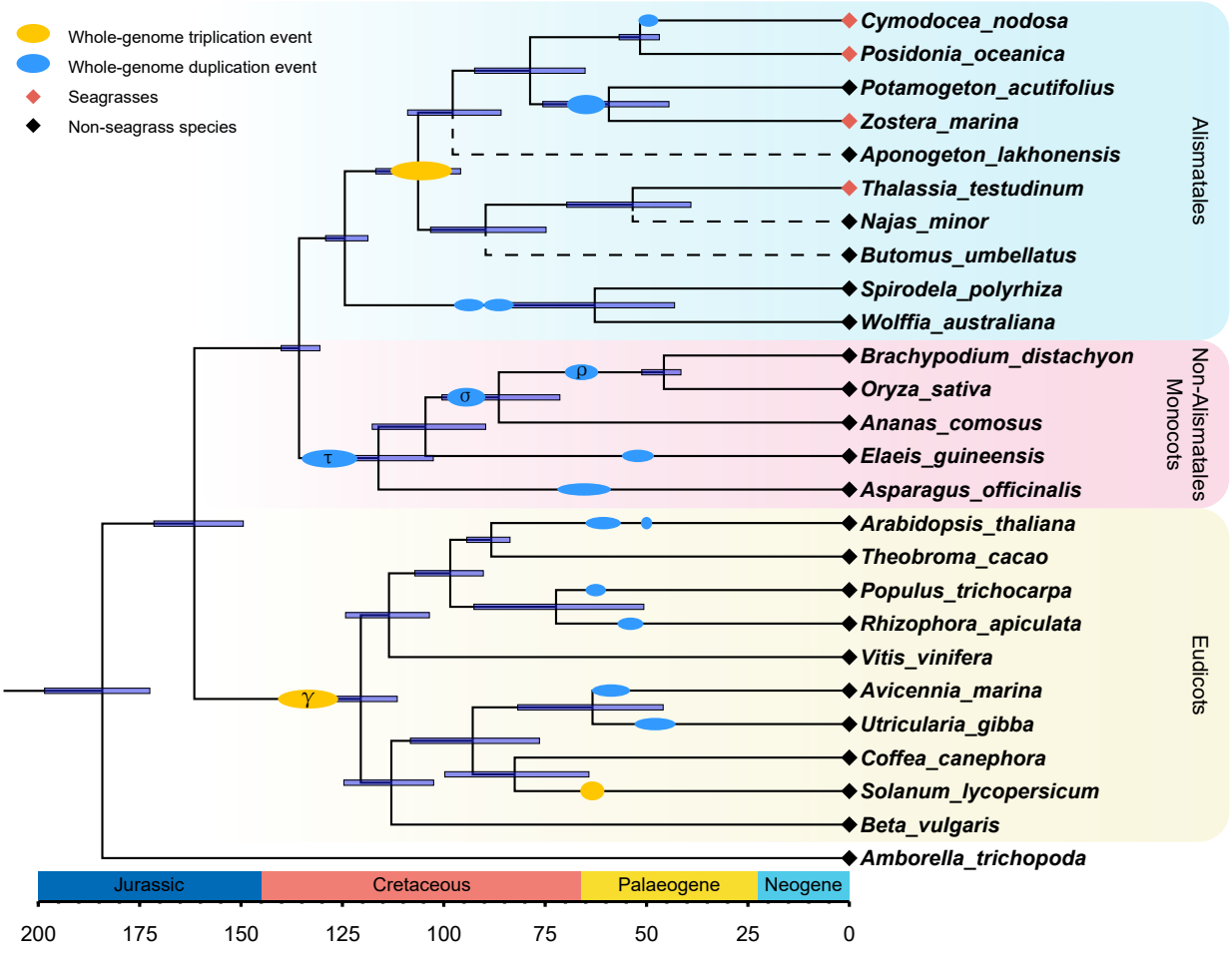
133 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009). https://doi.org:10.1093/bioinformatics/btp348

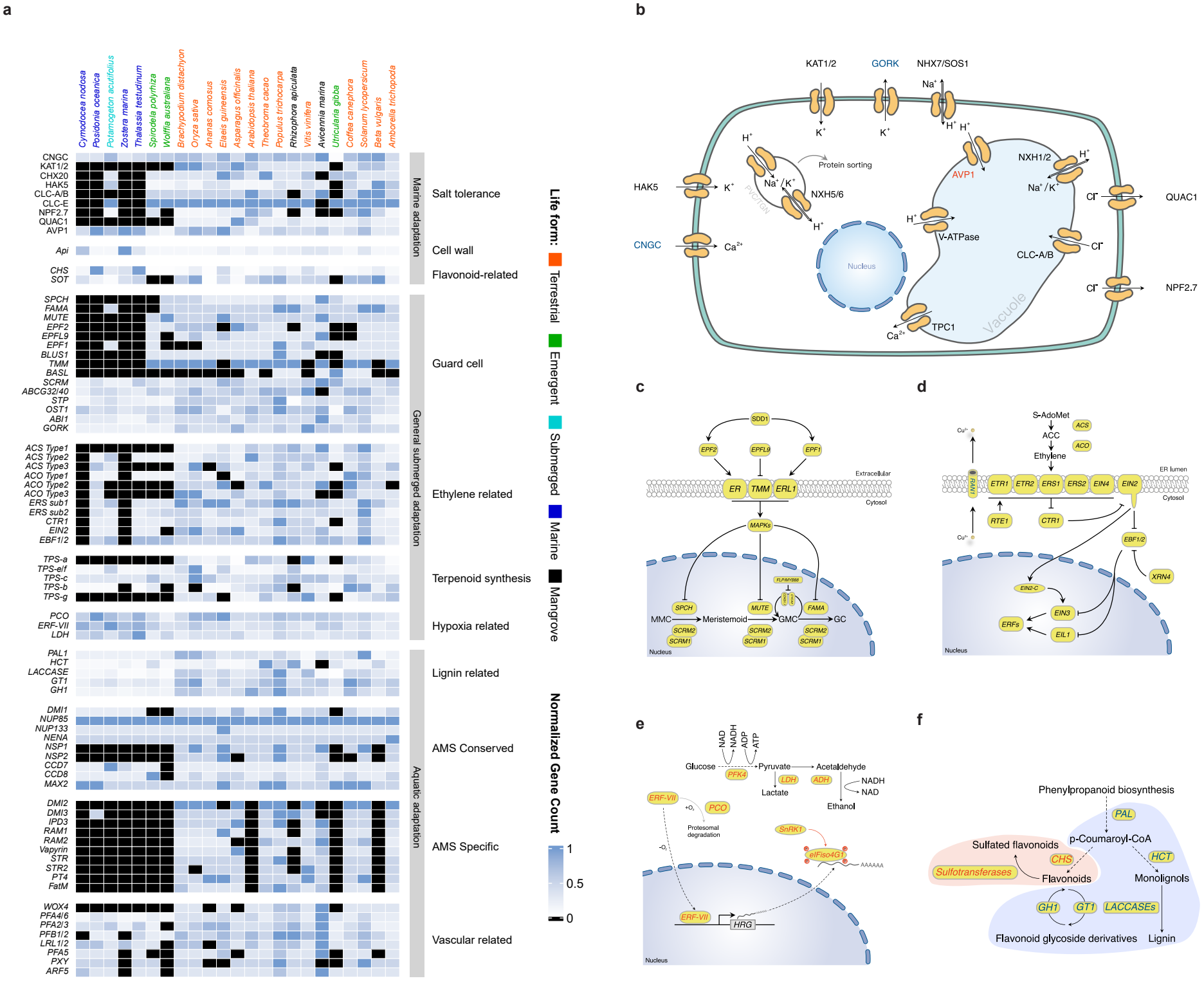
134 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020). https://doi.org:10.1093/molbev/msaa015

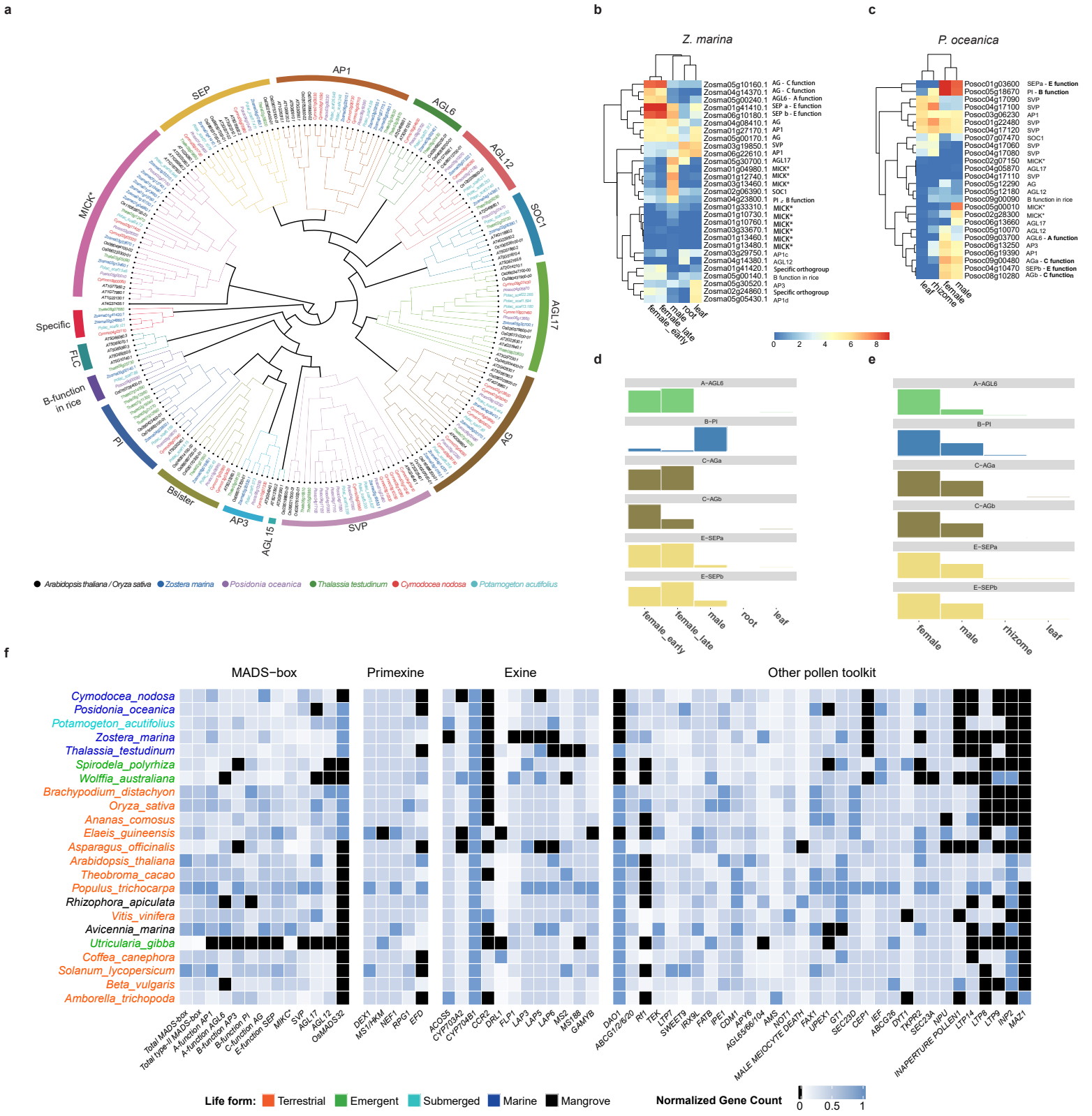
135 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012). https://doi.org:10.1093/nar/gkr1293

136 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009). https://doi.org:10.1101/gr.092759.109


A
Chr02
Chr01
Chr09
Chr08
Chr07
Chr06
Chr05
Chr04
Chr03
T. testudinum
B
Chr02
Chr01
Chr10
Chr09
Chr08
Chr07
Chr06
Chr05
Chr04
Chr03
P. oceanica
C
Chr02
Chr01
Chr06
Chr05
Chr04
Chr03
Z. marina
D
Chr04
Chr03
Chr02
Chr01
Chr18
Chr17
Chr16
Chr15
Chr14
Chr13
Chr12
Chr11
Chr10
Chr09
Chr08
Chr07
Chr06
Chr05
C. nodosa


Whole-genome triplication event
Whole-genome duplication event
Seagrasses
Non-seagrass species
Cymodocea_nodosa
Posidonia_oceanica
Potamogeton_acutifolius
Zostera_marina
Aponogeton_lakhonensis
Thalassia_testudinum
Najas_minor
Butomus_umbellatus
Spirodela_polyrhiza
Wolffia_australiana
Alismatales
Brachypodium_distachyon
Oryza_sativa
Ananas_comosus
Elaeis_guineensis
Asparagus_officinalis
Non-Alismatales Monocots
Arabidopsis_thaliana
Theobroma_cacao
Populus_trichocarpa
Rhizophora_apiculata
Vitis_vinifera
Avicennia_marina
Utricularia_gibba
Coffea_canephora
Solanum_lycopersicum
Beta_vulgaris
Eudicots
Amborella_trichopoda
ρ
σ
τ
γ
Jurassic
Cretaceous
Palaeogene
Neogene
200
175
150
125
100
75
50
25
0


a
Cymodocea nodosa
Posidonia oceanica
Potamogeton acutifolius
Zostera marina
Thalassia testudinum
Spirodela polyrhiza
Wolffia australiana
Brachypodium distachyon
Oryza sativa
Ananas comosus
Elaeis guineensis
Asparagus officinalis
Arabidopsis thaliana
Theobroma cacao
Populus trichocarpa
Rhizophora apiculata
Vitis vinifera
Avicennia marina
Utricularia gibba
Coffea canephora
Solanum lycopersicum
Beta vulgaris
Amborella trichopoda
CNGC
KAT1/2
CHX20
HAK5
CLC-A/B
CLC-E
NPF2.7
QUAC1
AVP1
Api
CHS
SOT
SPCH
FAMA
MUTE
EPF2
EPFL9
EPF1
BLUS1
TMM
BASL
SCRM
ABCG32/40
STP
OST1
ABI1
GORK
ACS Type1
ACS Type2
ACS Type3
ACO Type1
ACO Type2
ACO Type3
ERS sub1
ERS sub2
CTR1
EIN2
EBF1/2
TPS-a
TPS-e/f
TPS-c
TPS-b
TPS-g
PCO
ERF-VII
LDH
PAL1
HCT
LACCASE
GT1
GH1
DMI1
NUP85
NUP133
NENA
NSP1
NSP2
CCD7
CCD8
MAX2
DMI2
DMI3
IPD3
RAM1
RAM2
Vapyrin
STR
STR2
PT4
FatM
WOX4
PFA4/6
PFA2/3
PFB1/2
LRL1/2
PFA5
PXY
ARF5
Marine adaptation
Salt tolerance
Cell wall
Flavonoid-related
General submerged adaptation
Guard cell
Ethylene related
Terpenoid synthesis
Hypoxia related
Aquatic adaptation
Lignin related
AMS Conserved
AMS Specific
Vascular related
Life form:
Terrestrial
Emergent
Submerged
Marine
Mangrove
Normalized Gene Count
1
0.5
0
b
KAT1/2
GORK
NHX7/SOS1
HAK5
CNGC
NXH5/6
Protein sorting
PVC/TGN
AVP1
NXH1/2
V-ATPase
CLC-A/B
QUAC1
NPF2.7
Nucleus
Vacuole
TPC1
c
SDD1
EPF2
EPFL9
EPF1
ER
TMM
ERL1
Extracellular
Cytosol
MAPKs
SPCH
MUTE
FAMA
MMC
Meristemoid
GMC
GC
SCRM2
SCRM1
FLP/MYB88
Nucleus
d
S-AdoMet
ACC
Ethylene
ACS
ACO
ER lumen
RAN1
ETR1
ETR2
ERS1
ERS2
EIN4
EIN2
Cytosol
RTE1
CTR1
EBF1/2
XRN4
EIN2-C
EIN3
ERFs
EIL1
Nucleus
e
Glucose
PFK4
Pyruvate
Acetaldehyde
LDH
ADH
Lactate
Ethanol
NAD
NADH
ADP
ATP
ERF-VII
PCO
Proteasomal degradation
SnRK1
eIFiso4G1
HRG
Nucleus
f
Phenylpropanoid biosynthesis
PAL
p-Coumaroyl-CoA
Sulfated flavonoids
CHS
HCT
Sulfotransferases
Flavonoids
Monolignols
GH1
GT1
LACCASEs
Flavonoid glycoside derivatives
Lignin


a
AP1
AGL6
AGL12
SOC1
AGL17
AG
SVP
AGL15
AP3
Bsister
PI
B-function in rice
FLC
Specific
MIKC*
SEP
Arabidopsis thaliana / Oryza sativa
Zostera marina
Posidonia oceanica
Thalassia testudinum
Cymodocea nodosa
Potamogeton acutifolius
b
Z. marina
Zosma05g10160.1 AG - C function
Zosma04g14370.1 AG - C function
Zosma05g00240.1 AGL6 - A function
Zosma01g41410.1 SEP a - E function
Zosma06g10180.1 SEP b - E function
Zosma04g08410.1 AG
Zosma01g27170.1 AP1
Zosma05g00170.1 AG
Zosma03g19850.1 SVP
Zosma06g22610.1 AP1
Zosma05g30700.1 AGL17
Zosma01g04980.1 MIKC*
Zosma01g12740.1 MIKC*
Zosma03g13460.1 MIKC*
Zosma02g06390.1 SOC1
Zosma04g23800.1 PI + B function
Zosma01g33310.1 MIKC*
Zosma01g10730.1 MIKC*
Zosma01g10760.1 MIKC*
Zosma03g33670.1 MIKC*
Zosma01g13460.1 MIKC*
Zosma01g13480.1 MIKC*
Zosma03g29750.1 AP1c
Zosma04g14380.1 AGL12
Zosma01g41420.1 Specific orthogroup
Zosma05g00140.1 B function in rice
Zosma05g30520.1 AP3
Zosma02g24860.1 Specific orthogroup
Zosma05g05430.1 AP1d
female_early female_late male root leaf
0 2 4 6 8
c
P. oceanica
Posoc01g03600 SEPa - E function
Posoc05g18670 PI - B function
Posoc04g17090 SVP
Posoc04g17100 SVP
Posoc03g06230 AP1
Posoc01g22480 SVP
Posoc04g17120 SVP
Posoc07g07470 SOC1
Posoc04g17060 SVP
Posoc04g17080 SVP
Posoc02g07150 MIKC*
Posoc04g05870 AGL17
Posoc04g17110 SVP
Posoc05g12290 AG
Posoc05g12180 AGL12
Posoc09g00090 B function in rice
Posoc05g00010 MIKC*
Posoc02g28300 MIKC*
Posoc06g13660 AGL17
Posoc05g10070 AGL12
Posoc09g03700 AGLG - A function
Posoc06g13250 AP3
Posoc06g19390 AP1
Posoc09g00480 AGa - C function
Posoc04g10470 SEPb - E function
Posoc08g10280 AGb - C function
leaf rhizome female male
d
A−AGL6
B−PI
C−AGa
C−AGb
E−SEPa
E−SEPb
female_early female_late male root leaf
e
A−AGL6
B−PI
C−AGa
C−AGb
E−SEPa
E−SEPb
female male rhizome leaf
f
MADS−box
Primexine
Exine
Other pollen toolkit
Cymodocea_nodosa
Posidonia_oceanica
Potamogeton_acutifolius
Zostera_marina
Thalassia_testudinum
Spirodela_polyrhiza
Wolffia_australiana
Brachypodium_distachyon
Oryza_sativa
Ananas_comosus
Elaeis_guineensis
Asparagus_officinalis
Arabidopsis_thaliana
Theobroma_cacao
Populus_trichocarpa
Rhizophora_apiculata
Vitis_vinifera
Avicennia_marina
Utricularia_gibba
Coffea_canephora
Solanum_lycopersicum
Beta_vulgaris
Amborella_trichopoda
Total MADS-box
Total type-II MADS-box
A-function AP1
A-function AGL6
B-function AP3
B-function PI
C-function AG
E-function SEP
MIKCc
SVP
AGL17
AGL12
OsMADS32
DEX1
MS1/HKM
NEF1
RPG1
EFD
ACOS5
CYP703A2
CYP704B1
CCR2
DRL1
FLP1
LAP3
LAP5
LAP6
MS2
MS188
GAMYB
DAO1
ABCG1/2/6/20
RPT1
TEK
LTP7
SWEET9
IRX9L
FATB
IPE1
CDM1
APY6
AGL65/66/104
AMS
NOT1
MALE MEIOCYTE DEATH
FAX1
UPEX1
GT1
SEC23D
CEP1
IEF
ABCG26
DYT1
TKPR2
SEC23A
NPU
INAPERTURE POLLEN1
LTP14
LTP8
LTP9
INP2
MAZ1
Life form: Terrestrial Emergent Submerged Marine Mangrove
Normalized Gene Count
0 0.5 1