doi:10.1038/nature16548

Seagrass Genome Supplementary Notes

1 Seagrasses

1.1 Evolution and definition of seagrasses

Seagrasses have arisen independently at least three times within the Alismatales, a cosmopolitan, diverse and monophyletic clade of basal aquatic and marine monocots. The order comprises ~ 4500 species in 13 families (Angiosperm Phylogeny Group 2009) or alternatively 11 families (excluding Araceae and Tofieldiaceae) following Les & Tippery (2013). The seagrasses belong to four or five families depending on the taxonomic authority: the Hydrocharitaceae, Posidoniaceae, Cymodoceaceae/Ruppiaceae (separately or as one) and Zosteraceae. Hence, seagrasses constitute an ecological/functional grouping rather than a monophyletic clade as shown and reviewed in the molecular phylogenies (Les *et al.* 1997; Les & Tippery 2013).

A phylogenetic tree (**Supplementary Fig. S1.1)** based on the rbcL gene is still the most thorough molecular study of alismatids based on 69 species to date, covering all recognized core families and 83% of alismatid genera (Les & Tippery 2013). Additional studies (Peterson *et al.* 2006; Cuena *et al.* 2010) using mtDNA confirm the above phylogenetic relationships. Likewise the most recent phylogenetic analyses based on 17 plastid genes but a smaller set of taxa did not affect the basic relationships with respect to seagrasses (Iles *et al.* 2013; Iles *et al.* 2015).

Alismatales have a relatively extensive fossil record extending into the upper/late Cretaceous (100 to 66 Mya) including some putative seagrasses. The best confirmed seagrass fossils are, however, from the Ecocene (56-34 Mya, Den Hartog & Kuo 2006). Fossil-calibrated, molecular clock estimates suggest that the three major clades of seagrasses (Hydrocharitaceae, Posidoniaceae plus Cymodoceaceae/Ruppiaceae, and Zosteraceae) evolved at roughly the same time between 40 -78 Mya (**Supplementary Table S1.1**). See also **Supplementary Note 4.1** on whole genome duplication in *Z. marina*.

As angiosperms and with true root systems, seagrasses were able to occupy the previously empty niche in marine systems of shallow sedimentary shorelines. This may have been facilitated by the Cretaceous-Paleogene extinction event that took place roughly 70-65 Mya and coincides with a lineage-specific whole genome duplication in *Zostera*.

For a review of seagrass systematics, evolution, general biology, ecology, physiology and metabolism, including comparisons with their freshwater, alismatalean relatives, see respectively (Larkum *et al.* 2006b; Papenbrock 2012; Ackerman 2006). In general, seagrasses are characterized as:

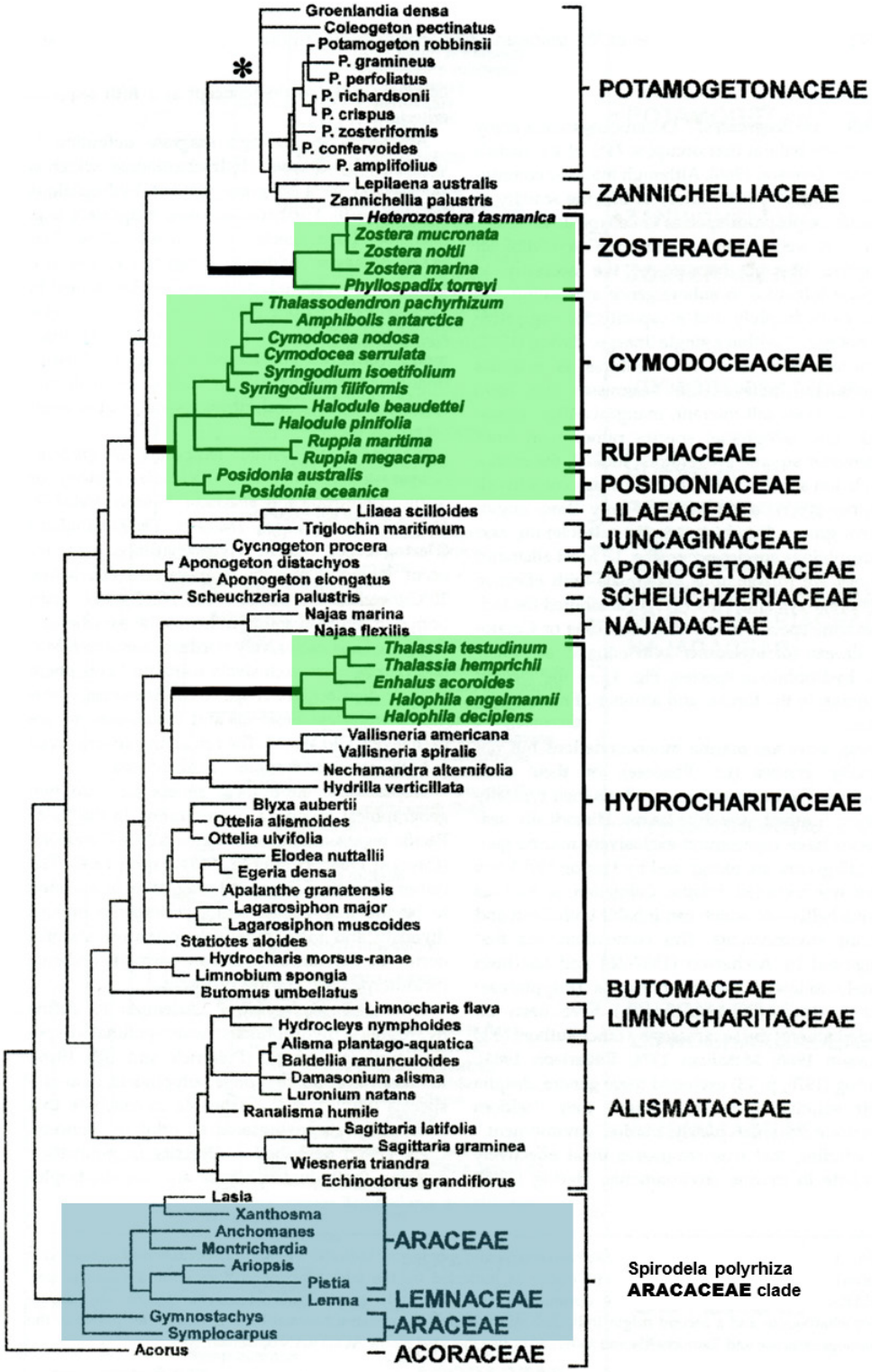
1) occurring in estuarine and fully marine environments and nowhere else. Although some freshwater species can withstand full marine salinities, e.g., *Potamogeton* and *Ruppia*, their reproductive systems and other morphological features are aquatic/freshwater.

2) possessing specialized leaves that lack stomata. Although a few aquatic plants developmentally lack stomata depending on the environment, loss of the genes for stomata (as reported here) is presently thought to be unique to seagrasses.

3) possessing cell walls that feature sulphated polysaccharides and other specialized polysaccharides reminiscent of marine seaweeds/macroalgae (Rhodophyta, Phaeophyta, Chlorophyta). No other angiosperms have these kinds of cell walls.

4) having a life history that is completed fully submerged underwater with submarine flowers (with the exception of *Enhalus acoroides*) and submarine pollination by specialized filiform pollen (with the exception of *Enhalus, Thalassia and Halophila*, all of which are in the Hydrocharitaceae). Note too that the relative importance of air, surface and underwater pollination to a particular species is not always clear-cut, but true hydrophily is known to be the dominant form of pollination in seagrasses as reviewed by (Ackerman 2006).

In the present genomic analyses we have compared genes and gene families across a range of monocots and dicots. Within the Alismatales there are only two genomes currently available: the fully marine *Zostera marina* (eelgrass) and the floating, aquatic/freshwater *Spirodela polyrhiza* (duckweed). Seagrasses share many features with their freshwater sister taxa but they also possess uncommon and unique traits, which constitute evolutionary innovation in their return to the sea, an extremely rare set of events. Freshwater angiosperms have arisen at least 100 times; seagrasses only three (Les *et al.* 1997; Les & Tippery 2013).



Supplementary Fig. S1.1 | Phylogenetic relationships within the Alismatales and the placement of seagrasses.

Maximum Parsimony tree based on plastid rbcL sequences from (Les *et al.* 1997, with permission). Seagrasses (green) are fully marine (here including *Ruppia*, which has both marine and brackish members). Sister taxa are aquatic/freshwater (e.g., Potamogetonaceae marked with *) thus confirming that seagrasses arose from freshwater ancestors.

Supplementary Table S 1.1 | Estimated molecular clock divergence times in the evolution of Alismatales and seagrass lineages.

Estimates by Iles (2013) are based on fossil calibration (four of 25 fossils relevant to the Alismatales) with constraints; and on sequence divergence in 13 plastid genes for (29 of 140 monocots; 6 seagrasses) using BEAST with random local clocks and exponential priors. The estimates of Coyer *et al.* (2013b) are restricted to the Zosteraceae and based only on the plastid rbcL gene (with constraints(as above using BEAST) but with complete species coverage including biogeographic sampling. The K-Pg extinction occurred ca. 65 Mya.

	Stem taxa Mya (range)	**Crown taxa Mya (range)**	**Reference**
Alismatales	134 (130-137	132 (124 - 137)	Iles 2013
Hydrocharitaceae	78 (73 - 82)	58 (56 - 59)	Iles 2013
Cymodoceaceae (including Ruppiaceae)	40 (35-45)	37 (34-41)	Iles 2013
Posidoniaceae	40 (35–45)	---	Iles 2013
Zosteraceae	39 (27–48)	---	Iles 2013
Zosteraceae/Potamogetonaceae-Posidoniaceae/Cymodoceaceae split	55* (45-65)	--	Iles 2013
Zosteraceae-*Ruppia/Posidonia* split	73* (66 - 86)	--	Coyer *et al*. 2013
Zosteraceae *(Zostera-Phylospadix* split)		23 (7 – 60)	Coyer *et al*. 2013

*not a stem, listed here for comparison

2 Genome sequencing and assembly

2.1 Nuclear genome

Supplementary Table S2.1 | Genomic libraries included in the *Zostera marina* genome assembly and their respective assembled sequence coverage levels in the final release; ver 2.1.

Library	Sequencing Platform	Average Insert Size	Read Number	Assembled Sequence Coverage (x)
MONE	Illumina piared-end fragment	763 ± 21	17,376,230	10.32
MTWO	Illumina paired-end fragment	775 ± 22	17,321,160	8.80
IWHB	Illumina mate pair	1,820 ± 190	37,108,670	8.44
IUSG	Illumina mate pair	3,520 ± 436	36,295,144	7.81
NUUS	Illumina mate pair	7,678 ± 828	45,575,590	5.20
NUUP	Illumina mate pair	8,133 ± 999	58,230,176	6.84
XTP	Sanger	35,450 ± 4,655	194,303	0.29
Total		N/A	212,101,273	47.70

Supplementary Table S2.2. | Summary statistics for *Z. marina* of the output of the whole genome shotgun assembly prior to screening, removal of organelles and contaminating scaffolds.

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

Minimum scaffold length	Number of scaffolds	Number of contigs	Scaffold size	Basepairs	% non-gap basepairs
2.5 Mb	2	143	5,248,631	5,072,686	96.65%
1 Mb	27	865	35,566,684	34,379,110	96.66%
500 Kb	121	3,114	100,845,174	97,053,084	96.24%
250 Kb	259	5,406	149,985,944	143,374,271	95.59%
100 Kb	462	7,747	183,481,400	173,935,305	94.80%
50 Kb	613	8,953	194,427,835	183,467,168	94.36%
25 Kb	783	9,965	200,453,753	188,165,649	93.87%
10 Kb	1,349	12,061	209,140,194	195,457,076	93.46%
5 Kb	2,225	13,606	215,196,637	201,118,403	93.46%
2.5 Kb	4,199	16,236	221,989,015	207,750,769	93.59%
1 Kb	13,875	27,657	235,864,335	221,416,006	93.87%
All	15,747	30,723	237,504,008	222,860,864	93.83%

Supplementary Table S2.3. | Summary of assembly statistics for *Z. marina* assembly V2.1.

V2.1 assembly	Scaffolds	Contigs
Sequences (#)	2,228	12,583
Total Length (b)	203,914,448	191,659,986
Max. Length (b)	2,654,544	642,312
Min. Length (b)	1,000	200
N50	124	623
L50 (b)	485,578	79,958

Supplementary Table S2.4. | Placement of the individual Fosmid clones and their contribution to the overall error rate in *Z. marina*.

Fosmid Clone ID	Length	Scaffold	Start	Stop	Discrepant Bases
16243	34,109	super_26	721,592	755,699	1
16250	35,814	super_88	463,243	498,762	5
16247	36,068	super_99	154,601	184,706	6
16245	35,858	super_19	714,416	750,280	9
16242	42,538	super_96	143,759	186,304	114
16248	32,137	super_236	162,103	190,033	246
16249	36,808	super_130	143,611	180,407	333
Total	253,332				714

2.2 Chloroplast genome

The chloroplast (cp) genome was manually assembled from two contigs into one circular sequence (**Supplementary Fig. S2.1**). The longest contig was 120,583 bp and started in the *ndhF* gene at the beginning of the short single copy region (SSC) and continued counterclockwise across one full inverted repeat (IR_B), the entire long single copy region (LSC) and ended 216 bases into the second inverted repeat region (IR_A). The second contig (3093 bp) overlapped with the *ndhF* gene and extended clockwise 881 bases into the IR_A region, thus bridging the junction between the IR_A and SSC regions. The overlaps between both contigs were found using BLASTn+ (Boratyn *et al.* 2013) and included only one mismatch that resided in the poly-A region.

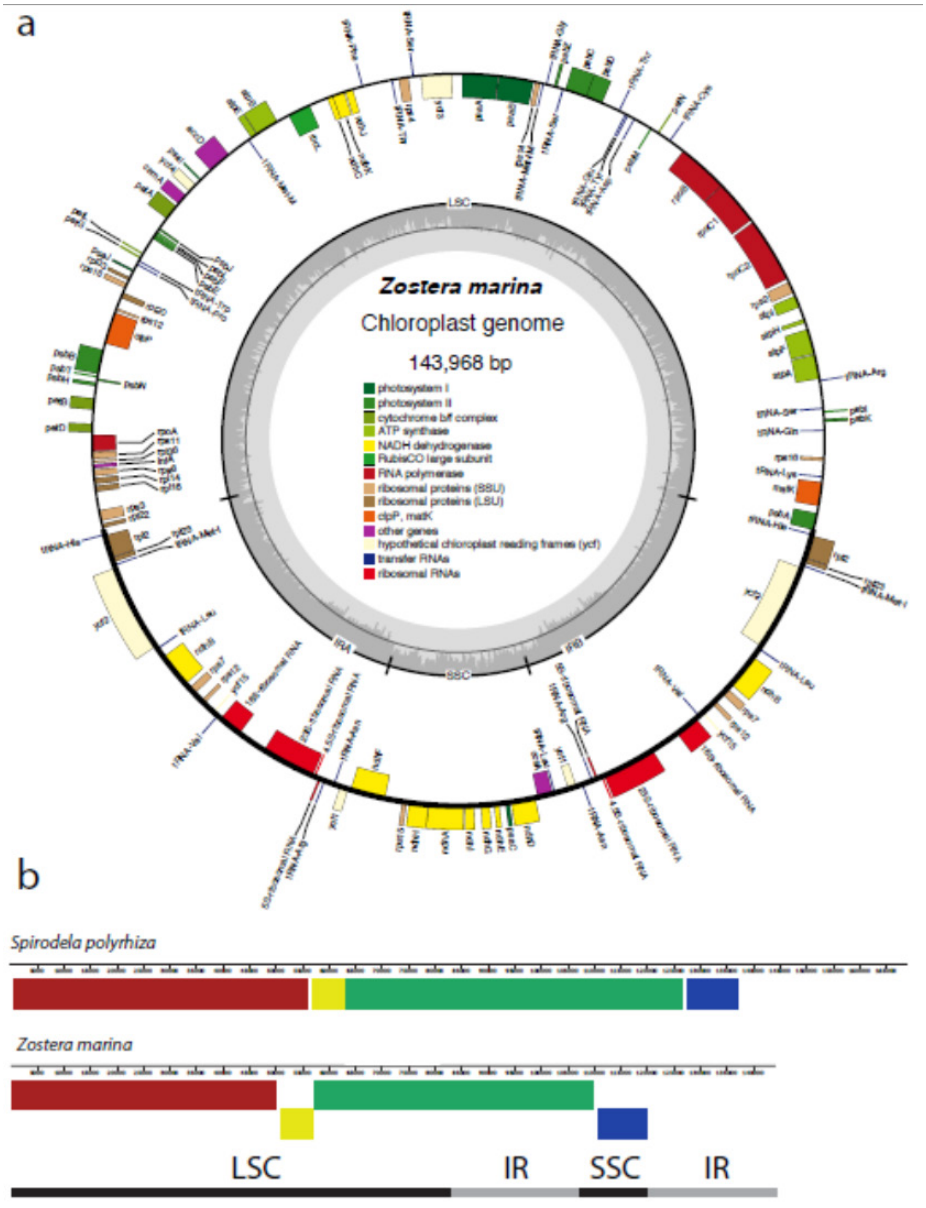
Paired-end reads were then mapped to the cp-genome sequence using BWA-MEM (http://bio-bwa.sourceforge.net/) and default settings in order to validate the junctions of the IR_A region. This was done twice, and the "circular" sequence was cut in different places each time. Coverage of reads varied between ~300-600X with no obvious region of lower coverage (which would have indicated an error in the manual assembly). The analysis further revealed a region of ~2,000 bp with an overrepresentation of polymorphic sites. This region corresponds to the junction between the SSC and the manually added IR_A region and is believed to be due to sequence variation in the two IR regions.

The gene content of the *Z. marina* chloroplast genome was annotated using the software CPGAVAS and DOGMA (Wyman et al. 2004). The analysis identified 79 protein-coding genes, four rRNAs (4.5S, 5S, 16S and 23S) and 25 tRNAs (only counting genes in the inverted repeats once). The two annotation analyses differed slightly in their result. For example, *psbM, rpl22, rpl32, rps19* and *ycf68* were not found by the CPGAVAS analysis but all except *rps19* where found with DOGMA. Ribosomal protein S19 (*rps19*) was found in the region between LSC and IR_A in the *Spirodela polyriza* cp-genome, which would correspond to the region between *rpl22* and *rpl2* in *Z. marina*. This is close to the region that was manually assembled, and was supported by a 216 bp overlap. Junctions of repetitive regions are difficult to correctly assemble and therefore the apparent absence of the *rps19* gene may be an assembly error. However, searching the rest of the assembly revealed no trace of the *rps19* gene, which suggests that it has genuinely been lost. In addition, *rps12, infA* and *rpl20,* which are not present in the *S. polyrhiza* chloroplast genome were found in *Z. marina*. The 79 amino acid sequences translated by CPGAVAS where BLASTed to the protein database at NCBI in order to identify obvious assembly errors (e.g. reads from bacteria assembled with *Z. marina* reads) or horizontally transferred genes, but no such indications were found.

The *Z. marina* chloroplast genome is 143,968 bp long, with an inverted repeat of 24,096 bp. The cp-genome of *Spirodela polyrhiza* is 168,788 bp long (NCBI accession JN160603; IR region 31,755 bp) (Wang & Messing 2011). *Lemna minor* (DQ400350; 165 955 bp; IR 31223 bp), *Wolffiella lingulata* (JN160604; 169,337 bp; IR 31,683 bp) and *Wolffia australiana* (JN160605; 168,704 bp; IR 31930 bp) also have cp-genomes and IR regions of approximately the size of *S. polyrhiza*, which is significantly larger than that of *Z. marina*. This difference is due to the fact that *Z. marina* only has very short *ycf1* genes (recently renamed

Translocon at the inner chloroplast envelope 214, Tic214 (Kikuchi *et al.* 2013a)) in the inverted repeat regions.

Synteny analysis between *Z. marina* and *S. polyrhiza* was preformed using the program MAUVE (Darling et al. 2004) to identify rearranged regions between the two species. This analysis shows that a part of LSC (including the regions *rbcL, atpB, atpE* and *trnM-CAT*; yellow box in **Supplementary Fig. S2.1**, as well as the main part of SSC in *Z. marina* has been reversed (blue box in **Supplementary Fig. S2.1).** The latter region is also reversed compared to the *Arabidopsis thaliana* cp-genome (data not shown) and is where the main part of the gene *ycf1* is found. There, the gene starts in the IR_B region (as in *Z. marina*) and then extends into the SSC region. One can speculate that the truncation of the *ycf1* gene happened at the same time as the SSC region was reversed.



Supplementary Fig. S2.1 | The *Zostera marina* chloroplast genome and comparison with *Spirodela polyrhiza*.

a, The size of the chloroplast genome differs significantly from that of other species in the Alismatales. A large part of this difference is due to the fact that *Z. marina* has a very short *ycf1* gene that codes for the protein Tic214, which is part of the protein translocation system in chloroplasts (Kikuchi *et al.* 2013b). The sequence was manually assembled from two contigs into one circular genome. Genes shown on the outside of the circle are transcribed counter-clockwise; genes on the inside are transcribed clockwise. The color-

coded legend in the center shows different classes of genes. Image generated using the http://ogdraw.mpimp-golm.mpg.de/index.shtml server.
b, The two rows of colored boxes indicate similar regions in the *S. polyrhiza* (top row) and *Z. marina* (bottom row) chloroplast genomes. Boxes displaced below the row indicate reversed regions**.** The yellow box indicates the reversed region around the genes *rbcL*, *atpB*, *atpE* and *tRNA-Met-M* in *Z. marina*, and the blue box represents the main part of SSC that is also reversed compared to *S. polyrhiza*. The thick black and grey line at the bottom of the figure indicates the location of the long single copy region (LSC), short single copy region (SSC) and the two inverted repeats (IR).

2.3 Mitochondrial genome

The assembly analysis resulted in two scaffolds for the mitochondrial (mt) genome: one of 188,248 bp (5,430 "N"characters) and one of 1,133 bp (100"N" characters) in length respectively. Annotation identified 104 genes.

Several different assemblies were attempted using the CLC assembly software. None were successful given many repetitive regions. The predicted number of genes and the estimated genome size in range with *S. spirodela* indicates that the two scaffolds probably include all or a significant part of the mitogenome. Further analyses are required to close the gaps within and between the two scaffolds in order to confirm the correct size of the mitogenome.

The estimated mitogenome size of *Z. marina* is significantly smaller than that for *Spirodela polyrhiza* (228 439 bp, 57 genes), which is the smallest known mitochondrial genome in the monocots (Wang *et al.* 2012). The predicted number of genes in *Z. marina* is almost twice the number in *S. polyrhiza,* and comparable to other monocot species, e.g., *Oryza sativa* [indica group] (JN861112; 100 genes, 454 894 bp genome). However, mt-genome sizes in seed plants vary greatly (Alverson *et al.* 2010) and the number of genes is not correlated with the mt-genome size, as seen in a comparison between *Sorghum bicolor* (NC_008360; 54 genes, 468,628 bp), *Phoenix dactylifera* (NC_016740; 44 genes, 715,001 bp) and *Zea mays* (AY506529; 218 genes, 569,630).

3 Genome characteristics and annotation

3.1 Genes and transposable elements

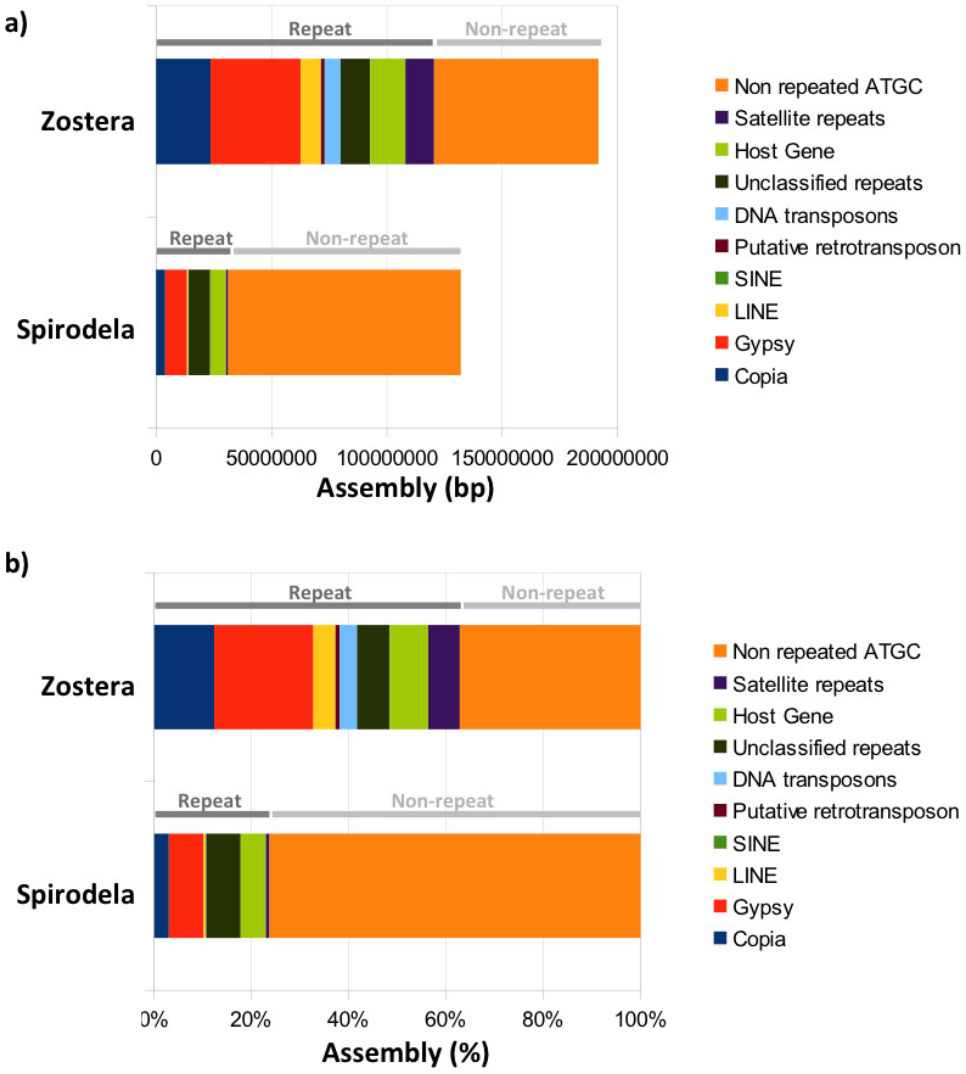
Supplementary Table S3.1 | Summary of genes and transposable elements in *Z. marina* and other plant genomes.

		A. thaliana	*O. sativa*	*B. distachyon*	*S. polyrhiza*	*Z. marina*
Genome assembly						
	Genome size (Mb)	119	374	271	128	203
	Assembly status	5 chromosomes	12 chromosomes	N50: 3 L50: 59.3 Mb	22 pseudomolecule	N50: 124 L50: 486 Kb
Genome annotation[1]						
	No. protein-coding genes	27,416	39,049	26,552	19,623	20,450
	No. multi-exon genes (%)	20,607 (75.2%)	28,071 (71.9%)	20,317 (76.5%)	15,817 (80.6%)	15,985 (78,2%)
	No. single-exon genes (%)	6,809 (24.8%)	10,978 (28.1%)	6,235 (23.5%)	3,806 (19.4%)	4,465 (21,8%)
	Avg. gene density (kb / gene)	4.4	9.6	10.2	6.5	9.9
	Avg. gene / CDS length (bp)	1,867/1,218	2,329/1,064	2,851/1,284	3,458 / 1,108	3,301 /1,177
	Avg. exon / intron length (bp)	237 / 152	258/401	256/384	213/ 559	227 /443
	Avg. exons per gene	5.1	4.1	5.0	5.2	5.2
	Avg. Intergenic (bp)	2,211	16,382	17,177	3,926	5,029
Transposable element						
	Overall TE content (%)	24[2]	35[3]	21.4[4]	13[5]	63
	Class 1 LTR Gypsy/Copia (%)	5.9/1.6	10.9/3.9	4.8 / 16.1 / 0.5	6.1 / 1.7	32/20
	Class 1LINE / unknown(%)	1/0.1	1.1 / 3.4	1.9 / 0.5	- / 5.3	7/1
	Class 2 DNA transposon (%)	12.1	12.9	4.8	-	6
	Unclassified (%)	0.2	1.8	-	-	11

1. Phytozome 9 annotation for *A. thaliana, O. sativa, B. distachyon and S. polyrhiza* gene models
2. A.thaliana(Maumus & Quesneville 2014)
3. O. sativa (Matsumoto *et al.* 2005)
4. B. distachyon (Vogel *et al.* 2010)
5. S. polyrhiza (Wang *et al.* 2014).

Supplementary Table S3.2 | Annotation results of transposable elements in the *Z. marina* genome

Type	Total bases
Copia	23,632,031
Gypsy	39,114,803
LINE	8,670,051
SINE	21,222
putative retrotransposon	1,613,655
DNA transposons	6,857,309
Unclassified	12,827,267
Host Gene	15,257,469
satellite repeats	12,406,906



Supplementary Fig. S3.1 | Repeat-driven genome size difference between *Zostera marina* and *Spirodela polyrhiza*.

Repeated and non-repeated fractions of the *Zostera* and *Spirodela* genome assemblies (in absolute base pairs (**a**) and in percent of assembly (**b**)) are indicated by grey bars above histograms for each species. The different classes of elements composing the repeated fractions are listed with their color code.

3.2 Transcriptome libraries, sequencing and assembly

Finnish clone library from leaf-meristem tissue. See Main Methods for tissue collection, handling and freezing in LN2 prior to RNA extraction. RNA extraction is described below.

Libraries from mesocosm-cultivated plants (flower, rhizome, root, leaf tissue). *Z. marina* plants (N=180, different genotypes) were collected from Falkenstein Beach area, Kiel Bight, Germany Lat. 54°.41 N, Long: 10°.19 E (on 27 Feb 2014), referred from here on as the Kiel samples. Within the same day, shoots were planted into flow-through seawater, aquaria (15 x 15 x 25 cm) within 600-L tanks. Six plants were planted in sandy sediment per aquarium. Water temperature was raised 1°C/day from ambient 5°C to 15°C. Light was provided by two halogen metal vapor lamps producing a light intensity of ~ 150 µmol photon sec^{-1} m^{-1} at the leaf surface. In order to mimic summer conditions and induce flowering, the light period was increased from 12 to 16 hours day light in increments of 30 min every third day. Several marine snails (*Littorina sp.*) were added to the aquaria to control epiphytic growth on *Z. marina* leaves. After 10 weeks, plants had developed flowers. Using a dissecting microscope, inflorescences were sampled and separated into male and female flowering tissue. Male flower tissue samples included the theca with pollen, whereas female tissue included pistils and styles. The dissecting procedure was performed as quickly as possible and tissues frozen in liquid nitrogen. Leaf tissue was also collected from each plant and cleaned with tissue paper. A 2-cm long piece was cut from the middle part of the leaf and placed in liquid nitrogen. Root samples were taken last. Sediment was cleaned from fine hairy roots by rinsing with seawater. Clean, 2-cm long root hairs were frozen in liquid nitrogen.

The same RNA extraction protocol was used for all samples using the InviTrap Spin Plant RNA Mini Kit (Stratek Molecular) following the manufacture's protocol. We used the provided RP buffer for lysis. RNA concentrations and purity were tested by Nanodrop® measurement (ND-1000, peQLab). RNA integrity was checked with an automated electrophoresis station Experion (Bio-RAD), using StdRNA chips and reagents (Bio-RAD). RNA concentrations ranged between 23-182 ng/µl, RQI values were >7.2.

For the mesocosm flower-leaf-root samples described above, 24 libraries were prepared (**Supplementary Table S3.3**) and sequenced at the University of Kiel Core Sequencing Centre (May 2014). Assembly stats can be found in **Supplementary Table S3.4**

For the Finnish clone genotype (leaf-sample described in the main Methods), a single library was created for high coverage sequencing at the University of Groningen Medical Centre, Genome Analysis Centre (March 2013) (**Supplementary Table S3.3**). Library preparation involved DNase 1 digestion of total RNA, mRNA isolation by use of oligo(dT) beads, mRNA fragmentation, first and second strand cDNA synthesis, end-repair, A-tailing, bar-coded adapter ligation and PCR amplification. Sequencing libraries were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) before sequencing. Paired-end (2 x 100 bp in both Groningen and Kiel) RNA Sequencing (RNASeq) data were generated using standard Illumina protocols and kits (TruSeq SBS KIT-HS v3, FC-401-3001; TruSeq PE Cluster Kit v3, PE-401-3001) and all sequencing was performed using the HiSeq 2000 (Groningen and Kiel) platforms.

All raw read data was transferred to the Vlaams Institute for Biotechnology, Gent and filtered by the TRIMMOMATIC (Bolger *et al.* 2014) (ver.0.30; parameters ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:10 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50) and only high quality reads with minimum read length longer than 50 bp were kept for further analysis.

Supplementary Table S3.3 | Summary of transcriptome sequencing in *Z. marina*.

Tissue	Library ID	Replicate	Number of raw read-pairs	Number of high quality reads (singletons)	% of aligned reads
Finnish clone genome (leaf/meristem)	BD1U11ACXX	1	165,574,710	148,535,158 (14,892,895)	88.19
Kiel Samples					
Female flower late	D1772	1	14,145,347	11,955,360 (1,464,507)	79.10
	D1773	2	16,776,929	14,261,336 (1,720,070)	71.70
	D1774*	3	5,103,708	313,424 (3,766,665)	0.40
Female flower early	D1775	1	14,580,265	12,371,137 (1,499,930)	79.90
	D1776	2	17,503,612	15,683,671 (993,116)	85.40
	D1777	3	15,621,403	13,302,315 (1,601,125)	78.00
	D1778	4	17,627,290	15,487,361 (1,236,579)	75.50
Male flower	D1779	1	15,141,096	13,194,752 (1,232,010)	81.90
	D1780	2	9,568,492	8177156 (795445)	78.40
	D1781	3	16,098,150	13,945,277 (1,338,284)	81.90
	D1782	4	17,121,565	14,920,337 (1,342,846)	81.60
	D1783	5	14,551,169	12,024,511 (1,797,798)	77.50
	D1784	6	10,492,845	8,937,642 (868,867)	76.30
	D1785	7	20,938,237	17,882,303 (1,790,373)	80.30
Leaf	D1786	1	11,832,954	9,998,704 (1,139,006)	61.60
	D1787	2	17,042,062	15,070,715 (1,087,460)	85.30
	D1788	3	15,838,023	13,985,218 (985,684)	85.10
	D1789	4	16,420,355	13,973,369 (1,559,967)	67.60
	D1790	5	16,031,062	13,946,115 (1,113,220)	87.20
	D1791	6	16,237,861	14,095,298 (1,108,672)	74.30
Root	D1792	1	17,474,087	15,173,223 (1,228,700)	83.70
	D1793	2	15,425,451	13,615,267 (955,448)	80.30

	D1794	3	13,432,406	11,476,883 (1,169,013)	78.00
	D1795	4	15,777,364	13,788,769 (1,021,423)	79.80
*Library failed during sequencing and was removed from further analysis					

Supplementary Table S.3.4 | Summary of transcriptome assembly following pooling of tissue-specific libraries in *Z. marina.*

	Number of assembled contigs	Average contig length after FrameDP correction (bp)	Contigs with expression support (FPKM >1 and IsoPct >1)	Collapsed by cd-hit
Finnish clone/genome (leaf/meristem)	52,623	Not available	Not available	Not available
Kiel Samples				
Female flower late	44,519	1,058.0	39,188	
Female flower early	55,977	1,131.3	47,482	
Male flower	53,753	1,081.2	40,617	
Leaf	90,421	871.5	72,043	
Root	63,804	1,021.5	53,336	
Total contigs				79,134

3.3 Differential gene expression analysis

Between 61% and 85% of the sequencing reads were aligned to the genome and the corresponding gene models. By comparing gene expressions profiles of four tissues (female flower early stage, female flower late stage, male flower and root) against the leaf tissue, the male flower sample showed the highest number of differentially expressed genes **(Supplementary Table S3.5, Extended Data Fig. 1, Supplementary Data 1-3)**. In particular one of the metallothionein genes (Zosma282g00060) was highly up-regulated in the male flower tissue (see **Supplementary Note 8**).

Supplementary Table S3.5 | Percentages of differentially expressed genes (log_2 >2) in comparison to leaf tissue in *Z. marina*.

	Differentially expressed compared to leaf	% of genes differentially expressed	Up regulated	Down regulated
Female flower early	3,538	15.5	1,249	2,289
Female flower late	2,148	9.4	571	1,577
Male flower	5,680	24.9	2,673	3,007
Root	2,478	10.9	895	1,583

3.4 miRNA methods and annotation

Using small RNA libraries from (Chavez Montes *et al.* 2014), we confirmed the presence of conserved miRNAs by the identification of their precursor loci. Thirty-six putative miRNA genes, distributed among 17 families, were identified with high confidence as described in Methods using criteria based on sequence homology, stem-loop structure and local small RNA expression profile (Meyers *et al.* 2008) (**Supplementary Table S3.6**). The mapping results confirmed the capability of the *Z. marina* genome to encode the miRNA families identified by the earlier genome-independent expression analysis (**Supplementary Table S3.7**). Putative targets of these conserved miRNAs were also investigated *in silico* using ~22,800 *Z. marina* functionally annotated coding sequences, confirming the conservation of several miRNA-regulated pathways (**Supplementary Table S3.8**).

Supplementary Table S3.6 | Precursors of known miRNAs identified in the genome of *Z. marina.*

miRNA family	Number of loci[1]
miR156	5
miR159	1
miR160	3
miR164	3
miR166	4
miR167	2
miR168	1
miR169	2
miR171	3
miR172	1
miR390	2
miR393	2
miR396	2
miR399	2
miR528	1
miR4414	1
miR5658	1
Total	**36**

[1]**see Supplementary Tables S2.7- 2.11.**

Supplementary Table S3.7 | miRNA families from the plant kingdom conserved in *Z. marina.*

Group[1]	miRNA families[2]
1 (ubiquitous)	**miR156**, **miR166**, **miR167**, **miR168**, **miR172**
2 (nearly ubiquitous)	miR158, **miR159**/miR319, **miR160**, miR162, **miR164**, **miR169**, **miR171**, **miR390**, **miR393**, miR394, **miR396**, miR397, miR529, miR535, **miR4414**
3 (absent in grasses)	miR391, miR472, miR482, miR858, miR1863, miR2111, miR2118, miR2950
4 (dicot specific)	miR403, miR473, miR477, miR530, miR1446, miR1515, miR3954, miR5234, miR5244, miR7122
5 (sparse)	miR395, miR398, miR399, miR408, miR479, miR827, ~~miR2911~~[3], miR5179, ~~miR5225~~[3]
6 (gymnosperm specific)	miR536, miR1083, miR1314
7 (monocot specific)	miR444, **miR528**, miR1120, miR1135, miR1318, miR1432, miR1436, miR1439, miR5048, miR5049, miR5067, miR5175, miR5281, miR5564, miR5565, miR5568, miR6201, miR6220, miR6225, miR6235
8 (Solanaceae specific)	miR1919, miR4376, miR5300, miR5301, miR6022, miR6023, miR6024, miR6025, miR6026, miR6027, miR6149
Other	**miR5658**

[1]: according to in Chávez-Montes et al. (2014)

[2]: bold = mapped on the genome; red = expressed over background noise

[3]: ~~strikethrough~~ = not a miRNA or not having characteristics typical of such in *Zostera marina*

Note: In contrast to small RNA expression data from Chávez-Montes et al. (Chavez Montes *et al.* 2014), we could not locate proper miRNA precursors for miR158, miR858, miR2911 and miR5225 in the genome. However, while the former two may be ascribed to incomplete genome assembly, it should be noted that the miR2911 family has been removed from the current miRBase release for its similarity with ribosomal RNA whereas miR5225 expression is in fact supported by extensive matching of its sequence to the genome but the local small RNA alignment profile exhibits properties typical of short-interfering RNA clusters rather than miRNAs. In addition, for two miRNA families, miR159 and miR168, we could not locate a *bona fide* precursor or a valid genome match for their most abundant representatives (5'-UUUGGAUUGAAGGGAGCUCUA-3' for miR159, 5'-UCGCUUGGUGCAGAUCGGGAC-3' for miR168). However, proper loci were identified for less abundant members of these two families.

Supplementary Table S3.8 | Predicted targets of conserved miRNAs in *Z. marina*.

miRNA family	Predicted targets	Conserved target
miR156	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein	Yes
miR159	MYB2 protein	Yes
miR160	Auxin response factor (ARF)	Yes
miR164	No Apical Meristem (member of NAC-domain transcription factors)	Yes
miR166	Homeobox-leucine zipper (HD-ZIP) protein	Yes
miR167	Nodal modulator 2-like (expected: ARF6, ARF8, IAR3)	No
miR168	AGO1	Yes
miR169	Glutamate synthase, putativ (expected: NF-YA)	No
miR171	SCARECROW-like proteins (SCL)	Yes
miR172	APETALA2	Yes
miR390	Not found (miR390 targets the non-coding *TAS3* locus)	Other
miR393	Not found (expected: TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX1 auxin receptor (TAAR))	Not found
miR396	Growth-regulating factor	Yes
miR399	Not found (expected: Pi transporter, DEAD box helicase, ubiquitin-conjugating E2 enzyme)	Not found
miR528	Not found (expected: CBP1 and other genes of unrelated function)	Not found
miR4414	Not found	Not found
miR5658	ABC transporter B family member, Mitogen-activated protein kinase, transcription factor bHLH123-like	No

Losses. Besides missing novel miRNA acquisitions belonging to the monocotyledon lineage, the data indicate *Z. marina* or its very close progenitors also experienced miRNA losses, especially within Group 2 of miRNAs described by (Chavez Montes *et al.* 2014), which includes broadly conserved but not ubiquitous miRNAs. Among these, miR162, miR394, miR397, miR529 and miR535 were not significantly expressed above background noise, nor encoded by the genome (**Supplementary Table S3.9**). A non-exhaustive list of known targets for these miRNAs in other plants is summarized in **Supplementary Table S3.8.**

Supplementary Table S3.9 | miRNAs lost in *Z. marina* and their targets.

miRNA family	Targets (non-exhaustive list)
miR162	dicer-1 (miRNA biogenesis pathway)
miR394	leaf curling responsiveness (leaf morphology)
miR397	copper protein plastocyanin (photosynthesis)
miR529	squamosa promoter binding protein-like (reproductive phase transition)
miR535	squamosa promoter binding protein-like (reproductive phase transition)

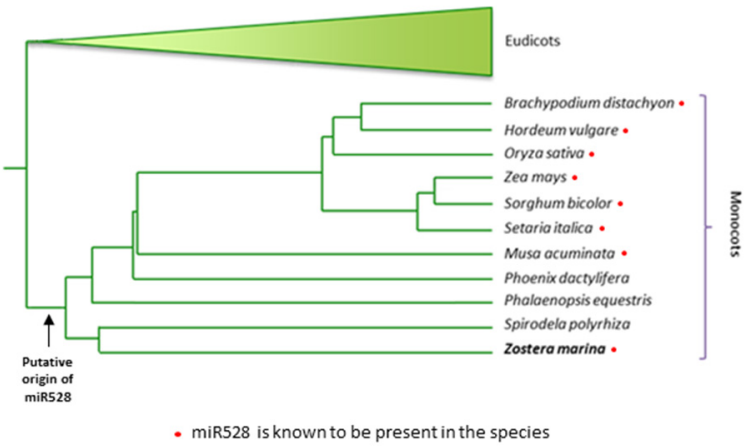
Expansions. The expansion of miRNA families as a result of tandem duplications, segmental duplications or repetitive elements, has been reported in various plant species (Sunkar *et al.* 2012). These and other less characterized mechanisms are thought to have elicited the expansion of miRNA loci, in particular those belonging to the miR156, miR159/319, miR166, miR169 and miR170 families, which are highly conserved and very abundant in several angiosperm genomes. Although some miRNA gene copies could have escaped our genomic screening due to stringent criteria, the overall results confirm the low copy number of most miRNA families and suggest that *Z. marina* did not experience or retain traces of prominent miRNA expansion. In addition, the genome of *Z. marina* lacks several miRNA families of more recent origin in monocots, most likely because the birth of these miRNAs occurred after the divergence of *Z. marina* from the other monocots investigated thus far. In particular, our previous results on the conservation and loss of miRNAs across *Z. marina*, banana and Poaceae (Chavez Montes *et al.* 2014) suggest that the majority of monocot-specific miRNAs (Group 7 miRNAs) originated and evolved more recently within the grass lineage rather than having been lost in *Z. marina.* As a result, *Z. marina* seems to reflect the ancestral repertoire of miRNA genes available to basal monocots prior to the divergence of the lineages that gave rise to *Z. marina,* banana and the grasses.

The lack of monocot-specific miRNAs also involved miR444, a MADs box-targeting miRNA widely conserved in monocots (Lu *et al.* 2008; Chavez Montes *et al.* 2014). Despite its old ancestry, miR444 was not found in the *Z. marina* and *S. polyrhyza* genomes, suggesting that its absence could be a property of the Alismatales and not specific to *Z. marina*. In contrast, the availability of the *Z. marina* genome sequence allowed us to identify a stem-loop precursor of miR528, another widely conserved monocot-specific miRNA, which had escaped the initial homology-based search due to a rare indel polymorphism (**Supplementary Fig. S3.2)**. The presence of a miR528 precursor and its expression as a mature miRNA in *Z. marina*, but not in *Spirodela polyrhiza* (Wang *et al.* 2014), rules out an indispensable role for this miRNA in the Alismatales but demonstrates that it is the only miRNA known so far that is basal to the entire monocotyledon lineage (Chavez Montes et al., 2014). Nearly all others conserved in monocots were found only in some clades, with the closest rival to mi528 for the most ancient being miR444.

a

```
other_monocots-miR528     TGGAAGGGGCATGCAGAGGAG-
zmr-miR528                TGGAAGGGGC-TGCAGAGGAGT
```

b



Supplementary Fig. S3.2 | Comparison of *Z. marina* vs other monocot miR528 sequences.

a, alignment of miR528; **b**, basal position of miR528 to the monocots

Unique. Novel miRNAs and their putative target sites were also discovered in *Z. marina*. Novel DCL1/AGO1-dependent miRNAs were predicted by selecting 5'-U small RNAs from *Z. marina* libraries (for which an abundance over background noise was present in at least in one small RNA library and that met the stem-loop requirements (see Methods for details). Excluding miRNA-like loci located in

transposable elements or other repeats, we identified nineteen stem-loop regions that were compatible with miRNA annotation criteria. Seven of these regions shared the same mature miRNA sequences with others bringing to fifteen the number of new putative miRNA families (provisionally named zmr-miR001-015; **Supplementary Tables S3.10** and **S3.11; Supplementary Data 4 and 5**). Putative target prediction identified a number of genes that may be under control of these miRNAs.

Supplementary Table S3.10 | Novel miRNA candidates in *Z. marina*.

miRNA family (temporary name)	Mature miRNA sequence[1]	Number of loci
zmr-miR001	UGUCCAAUCUAAUGUAUGGCC	2
zmr-miR002	UCUAUCAUCUGAGUCGAACGG	1
zmr-miR003	UUACAGCUCUGAGUAUGACAUUG	1
zmr-miR004	UUUAAUGUUGUCAUAAAGCUUA	1
zmr-miR005	UCUGAACUGCAGAUUCUCCUC	1
zmr-miR006	UCGUUAUAAAGAAGCUUCCUC	1
zmr-miR007	UUUCGAAACGCAUCGCAAUAA	1
zmr-miR008	UCAAGCAUGAAAUCUGACAAG	1
zmr-miR009	UCGGUCAACGAACUGGCAGAA	1
zmr-miR010	UUACGCAGGUGGAAGUUAACAU	1
zmr-miR011	UGGAUUGUAAAAGCAGGUCGAG	1
zmr-miR012	UUGUAACAAUCGGAGUCGUGAC	1
zmr-miR013	UAUAUAGAAGAGCUGAAAGAA	3
zmr-miR014	UUCUUGACCUUGUAAGACCCA	1
zmr-miR015	UUUGUAUGCUCAUCGACGGUU	2

[1]: see **Supplementary Data 5** - Novel miRNA loci and their predicted targets' for details on miRNA expression in the three tissues.

Supplementary Table S3.11 | Predicted target functions of novel miRNAs in *Z. marina*.

miRNA family (temporary name)	Predicted target function (only Allen score < 4)[1]
zmr-miR001	None
zmr-miR002	None
zmr-miR003	MET1-type DNA-methyltransferase
zmr-miR004	None
zmr-miR005	None
zmr-miR006	kinesin-4-like protein
zmr-miR007	None
zmr-miR008	None
zmr-miR009	argonaute 1-like protein
zmr-miR010	None
zmr-miR011	None
zmr-miR012	serine/threonine-protein kinase
zmr-miR013	KRTCAP2 homologous protein
zmr-miR014	auxin-responsive factor AUX/IAA-related
zmr-miR015	Methyl-CpG-binding domain protein

[1] see **Supplementary Data 5**. Novel miRNA loci and their predicted targets for a comprehensive list of potential targets.

Here we highlight three miRNA-target circuits. First, zmr-miR009 was predicted to target the *argonaute1* (*AGO1*) gene, encoding the effector protein in the active RNA-induced silencing (RISC) complex that is activated and guided by miRNAs (**Supplementary Fig. S3.3**). Moreover, *AGO1* (Zosma25g01490) was also confirmed as a target of miR168, which is already known to be involved in AGO1 homeostasis (Vaucheret *et al.* 2006), suggesting the intriguing possibility of a double miRNA control of AGO1in *Z. marina* (**Supplementary Fig. S3.3**). That zmr-miR009 expression appears to be restricted to the reproductive organs and its target transcripts are the highest in the male flower, may suggest a specific role in the female flower or a function in the fine tuning of target RNA levels. Future analysis of the expression of miR168, zmr-miR009, and their *AGO1* target in different *Z. marina* floral organs could provide further functional insights especially if combined with precise miRNA target cleavage now possible on a global scale (Jeong *et al.* 2011; Jeong *et al.* 2013). The second example involves two novel, unrelated miRNAs (zmr-miR003 and zmr-miR015) predicted to target genes associated with DNA methylation, i.e., a DNA methylase and a Methyl-CpG binding protein, respectively (**Supplementary Fig. S3.4**). If confirmed by experimental validation, our preliminary result suggests new means of miRNA-dependent control of DNA methylation that may mediate gene regulation. Third, zmr-miR014 putatively targets auxin response factor (ARF) genes (**Supplementary Fig. S3.5**). Considering that known targets for miR397 and miR393, also involved in the auxin response, failed to be predicted in the *Z. marina* genome, this observation seems to suggest a new and different mechanism for control of this pathway.

```
5' CUUCCGACUUGCACCAAGCAA 3' Transcript: Zosma25g01490:800-820 Slice Site:811
    |o|||||o||||||||||||
3' AAGGGCUGGACGUGGUUCGUU 5' Query: zmr-miR168

5' AUCUGCCAGUUCGUUGUCCGA 3' Transcript: Zosma25g01490:867-887 Slice Site:878
    |||||||||||||| ||||
3' AAGACGGUCAAGCAACUGGCU 5' Query: zmr-miR009
```

Supplementary Fig. S3.3 | Predicted target sites of miR168 and zmr-miR009 in Zosma25g01490 gene (putative AGO1).
The two sites are 67 nt apart.

```
5' CAAUAUCAUACUCAGAGCUGUAA 3' Transcript: Zosma14g00030:3558-3580 Slice Site:3571
   |||| ||||||||||||||||||
3' GUUACAGUAUGAGUCUCGACAUU 5' Query: zmr-miR003

5' UACU-UGGAUGAGCAUACAAA 3' Transcript: Zosma50g00420:601-620 Slice Site:611
    ||o | ||||||||||||||
3' UUGGCAGCUACUCGUAUGUUU 5' Query: zmr-miR015
```

Supplementary Fig. S3.4 | Predicted target sites of zmr-miR003 and zmr-miR015.

```
5' AAGGUCUUGCAAGGUCAAGAA 3' Transcript: Zosma24g01320:1209-1229 Slice Site:1220
     ||||||o||||||||||||
3' ACCCAGAAUGUUCCAGUUCUU 5' Query: zmr-miR014

5' AAGGUCUUGCAAGGUCAAGAA 3' Transcript: Zosma27g00270:1336-1356 Slice Site:1347
     ||||||o||||||||||||
3' ACCCAGAAUGUUCCAGUUCUU 5' Query: zmr-miR014
```

Supplementary Fig. S3.5 | Predicted target sites of zmr-miR014 in the putative ARF genes.

4 Comparative genome evolution

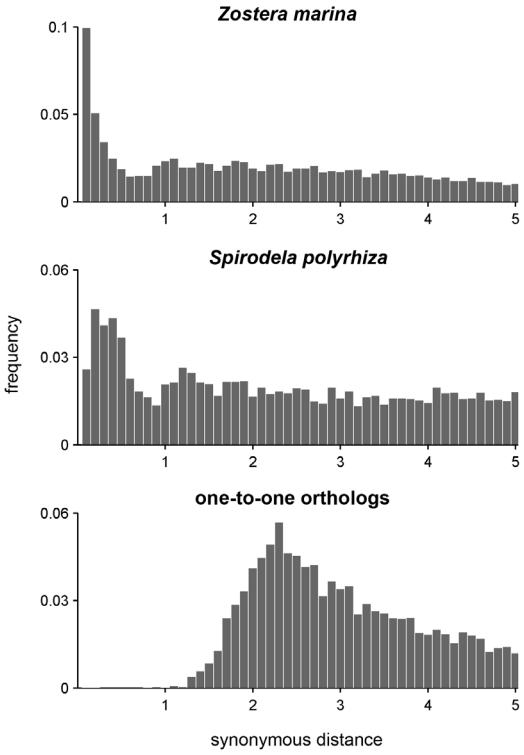
4.1 Whole genome duplication (WGD) analysis

Analysis of the number of synonymous substitutions per synonymous site (K_S) of the whole paranome indicates that a peak is present in the resulting K_S-based age distribution of *Z. marina* around a K_S of 1 (**Fig. 2a**). Because synonymous mutations accumulate at an approximately constant rate, they are putatively neutral and serve as a proxy for the time since duplication of paralogs (Kimura 1977). A peak therefore represents the signal of a burst of genes that were created contemporaneously during the evolutionary past of the species, most likely representing a whole genome duplication (WGD, Blanc & Wolfe 2004). Although this peak is located in a K_S range that falls well within the approximately linear response range of K_S estimates (Vanneste *et al.* 2013), it is elevated only slightly above the background. Therefore, we performed mixture modeling using EMMIX (McLachlan & Peel 1999) to investigate whether the peak corresponds to a *bona fide* WGD signature rather than stochastic variation of the background distribution of small-scale duplications (Schlueter *et al.* 2004; Cui *et al.* 2006) (**Fig. 2a**). Because model selection criteria used to identify the optimal number of components in the mixture model are prone to overfitting (Naik *et al.* 2007; Vekemans *et al.* 2012) we also used SiZER (Chaudhuri & Marron 1999) to distinguish components corresponding to WGD features from those that were fit to the background distribution (**Supplementary Table S4.1**). A component of the mixture model with its peak located at a K_S of 1.08 was confirmed by SiZER and corresponds to a WGD signature, indicating that *Z. marina* underwent a paleopolyploidy during its evolutionary history.

To investigate whether this WGD was shared or independent from the double WGD reported in *S. polyrhiza*, we also constructed the K_S age distributions of both the *S. polyrhiza* paranome and the one-to-one orthologs between *Z. marina* and *S. polyrhiza* (**Supplementary Fig. S4.1**). As reported previously (Wang *et al.* 2012), we found a clear WGD signature in the *S. polyrhiza* age distribution at a K_S of 1.23. The ortholog age distribution shows a clear peak at a K_S of 2.3, which corresponds to the speciation between these two lineages that has been estimated to be between 107 and 135 million years old (Bell *et al.* 2010). Although this divergence is old (so that these two species may be expected to display some rate heterogeneity (Shi *et al.* 2010)), the finding of an ortholog peak is located at such a large distance after the WGD duplication peaks in both *Z. marina* and *S. polyrhiza* indicates that these two species underwent separate WGDs.

Supplementary Table S4.1 | Mixture modeling and SiZer analysis of the *Z. marina* age distribution used in whole genome duplication analyses.

Number of duplicates	Number of components	BIC	mixture distribution means (K_s)	variance (K_s)	proportion	WGD signature*
3614	5	5456.90529	0.143	0.0007	0.082	no
			0.408	0.0473	0.296	no
			1.072	0.0628	0.317	yes
			1.531	0.0298	0.173	no
			1.841	0.0091	0.132	no
* Fitted components corresponding to WGD features are selected based on SiZer analysis.						

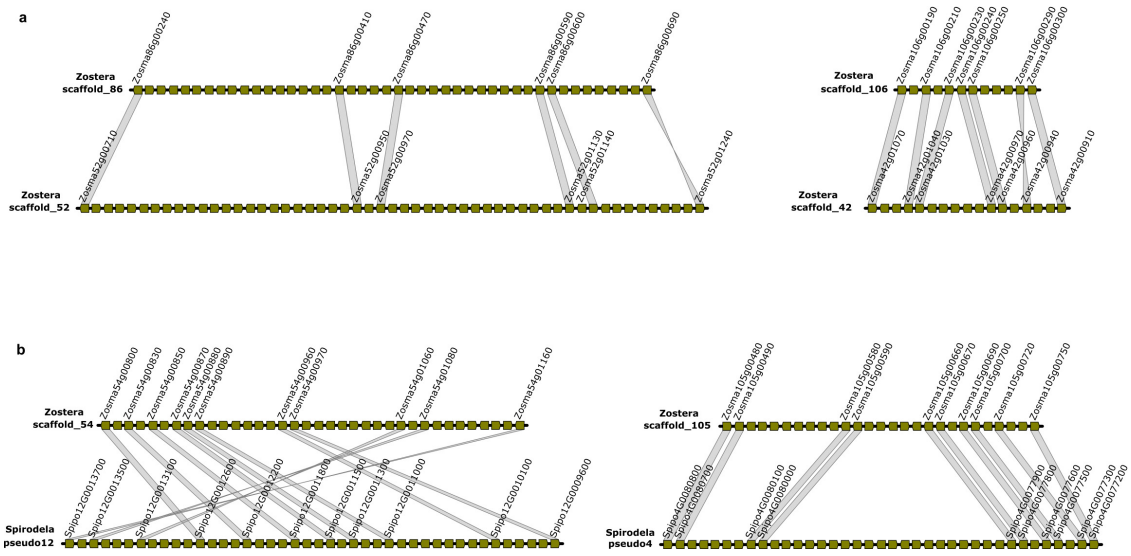


Supplementary Figure S4.1 | K_S-based age distributions for *Z. marina*, *Spirodela polyrhiza*, and their one-to-one orthologs.

The x-axis shows the synonymous distance (in bins corresponding to a K_S of 0.1), while the y-axis shows the frequency of retained duplicates per bin, for each distribution as indicated on top of the individual panels.

We also used collinearity-based methods to search for evidence of this paleopolyploidy in the *Z. marina* genome. While searching for evidence of *intra*genomic collinearity **(Supplementary Table S4.2)**, only one duplicated block was identified when only considering scaffolds larger than 1 Mb. Incorporation of all scaffolds indicated however that approximately 9% of the genome was in a duplicated state in support of the WGD discovered in the age distribution.

As a reminder more generally and as shown in **Extended Data Fig. 2**, the *Z. marina* genome is covered by regions of within-genome collinearity with about 29% of the *Z. marina* genome collinear with *Spirodela* as exemplified in **Supplementary Figure S4.2** below.



Supplementary Figure S4.2 | Examples of syntenic regions within *Z. marina* and between *Z. marina* and *S. polyrhiza*.

a, Examples of two syntenic regions within *Z. marina*. **b,** Examples of two syntenic regions between *Z. marina* and *Spirodela*.

Supplementary Table S4.2 | Detection of intragenomic collinearity within Z. marina using i-ADHoRe.

Multiplicity	**Number of multiplicons***
2	113
3	5
* Output is filtered for redundant level 2 multiplicons used in the profile search for higher-level multiplicons.	

While searching for evidence of *inter*genomic collinearity between *Z. marina* and *S. polyrhiza,* **(Supplementary Table S4.3)** and only considering scaffolds larger than 1 Mb, a complex duplication history was revealed with several *Z. marina* scaffolds displaying collinearity with multiple *S. polyrhiza* scaffolds and vice versa (not shown). This was anticipated because the complex duplication history of these two species (one independent WGD in *Z. marina* and two independent WGDs in *S. polyrhiza*) entails that any position in the *Z. marina* genome can display collinearity with up to four positions in the *S. polyrhiza* genome. Incorporation of all scaffolds confirmed this since many multiplicons (sets of homologous genomic segments) of higher level (the total number of included segments) were found.

Supplementary Table S4.3 | Detection of intergenomic collinearity between *Z. marina* and *S. polyrhiza* using i-ADHoRe.

Multiplicity	Number of multiplicons*
2	442
3	206
4	83
5	16
6	3

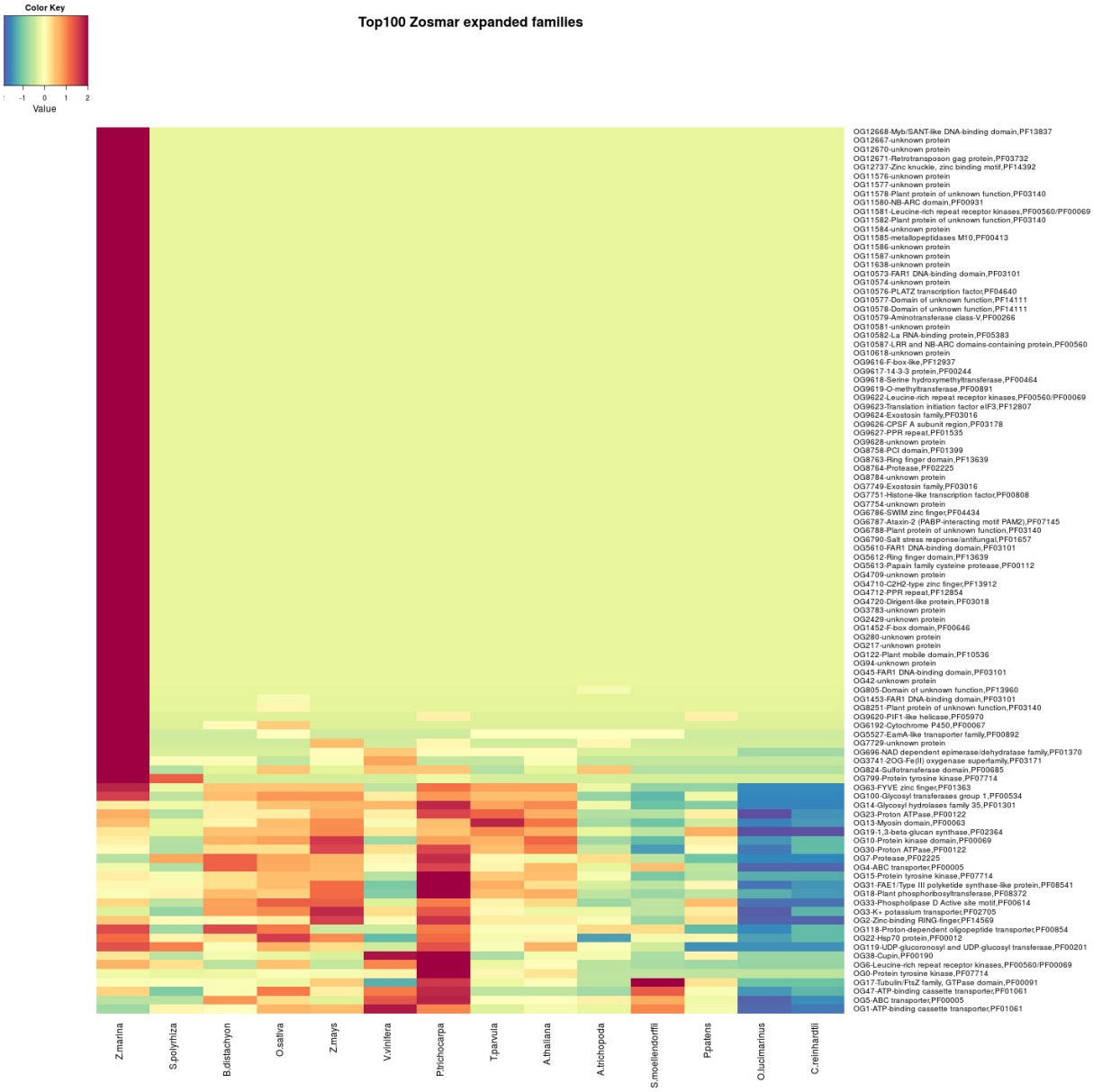
* Output is filtered for redundant level 2 multiplicons used in the profile search for higher-level multiplicons.

The absolute dating of genes present under the WGD peak between a K_S of 0.8 and 1.6 resulted in an age estimate for the WGD event of ~67 million years ago (**Fig. 2b**), with lower and higher 90% confidence interval limits of ~64 and ~72 million years ago, respectively (absolute dating using a peak with a narrower K_S range of 0.8–1.2 gave similar results, data not shown). This would place the paleopolyploidy event around the Cretaceous-Paleogene boundary, similar to many other angiosperms (Vanneste *et al.* 2014) (**Fig. 2c**) and before the origin of the Zosteraceae (Coyer *et al.* 2013a) at 23 Mya. Note, however, that the latter has a very broad 95% confidence interval (60 - 5 Mya). Our dating also coincides with the diversification within a diverse Alismatalean clade that includes (depending on the exact timing) at least three to four families in addition to the Zosteraceae: its sister family Potamogetonaceae, the two seagrass families Cymodoceaceae/Ruppiaceae and Posidoniaceae (as well as potentially the Juncaginaceae (**Supplementary Table S1.1, Supplementary Fig. S1.1**))(Janssen & Bremer 2004; Iles 2013; Iles *et al.* 2015). Determining whether this paleopolyploidy indeed occurred at or near the Cretaceous-Paleogene and which lineages share this WGD awaits the future availability of additional Alismatalean genomes.

Based on inferred synonymous substitution rates, Wang *et al.* (2014) dated two WGDs in *Spirodela* that occurred in close succession, at 95 Mya. We also dated the WGDs in *Spirodela* using the phylogenomic approach described above and in the Methods section of the main manuscript, and obtained a slightly younger estimate for the WGD event of ~79 Mya, with lower and higher 90% confidence interval limits of ~74 and ~83 Mya, respectively (**Fig. 2b**).

4.2 Gains and losses of gene families and Pfam domain analysis

The hundred largest *Z. marina* gene families, as compared to13 other genomes, including *S. polyrhriza*, were clustered (**Supplementary Fig. S4.3**). The top 20 GO terms enriched for gained and lost gene families are shown respectively in **Supplementary Table S4.4** and **S4.5**).



Supplementary Figure S4.3 | Hierarchical clustering of the 100 largest *Z. marina* gene families.

For each gene family in each species the z-score [(species gene-count – average family gene count)/standard deviation] was calculated to indicate the degree of family expansion/contraction and the resultant z-score matrix was clustered using Pearson correlation as a distance measure. Each row represents a gene family with the numerical gene family ID and corresponding functional description based on presence of a Pfam domain shown. The colour key indicates z-score with blue indicating contraction and red expansion.

Supplementary Table S4.4 | Top 20 GO terms enriched in gained gene families of *Z. marina* gained gene families.

MF: molecular function, BP: biological process, CC: cellular component.

Type	GO ID	Adjusted p-value	GO Term
MF	GO:0070011	4.58E-18	peptidase activity, acting on L-amino acid peptides
MF	GO:0008233	7.14E-17	peptidase activity
MF	GO:0008234	6.68E-16	cysteine-type peptidase activity
MF	GO:0008171	8.98E-14	O-methyltransferase activity
MF	GO:0008270	5.65E-13	zinc ion binding
MF	GO:0004175	3.96E-11	endopeptidase activity
MF	GO:0004650	6.11E-11	polygalacturonase activity
MF	GO:0046983	3.92E-10	protein dimerization activity
MF	GO:0046914	1.95E-06	transition metal ion binding
MF	GO:0005515	4.54E-06	protein binding
MF	GO:0043531	1.11E-05	ADP binding
MF	GO:0016787	1.44E-05	hydrolase activity
MF	GO:0051861	3.83E-05	glycolipid binding
MF	GO:0017089	3.83E-05	glycolipid transporter activity
MF	GO:0004222	4.53E-05	metalloendopeptidase activity
MF	GO:0042802	5.40E-05	identical protein binding
MF	GO:0003950	7.35E-05	NAD+ ADP-ribosyltransferase activity
MF	GO:0015298	0.000140177	solute:cation antiporter activity
MF	GO:0015299	0.000140177	solute:proton antiporter activity
MF	GO:0008237	0.000326357	metallopeptidase activity
BP	GO:0006508	2.62E-22	proteolysis
BP	GO:0044238	1.66E-09	primary metabolic process
BP	GO:0019538	6.52E-09	protein metabolic process
BP	GO:0071704	2.24E-08	organic substance metabolic process
BP	GO:0016032	3.07E-07	viral process
BP	GO:0043086	1.76E-06	negative regulation of catalytic activity
BP	GO:0044092	1.76E-06	negative regulation of molecular function
BP	GO:0043170	2.88E-06	macromolecule metabolic process
BP	GO:0005975	1.15E-05	carbohydrate metabolic process
BP	GO:1901264	1.23E-05	carbohydrate derivative transport
BP	GO:0046836	1.23E-05	glycolipid transport
BP	GO:0018149	1.81E-05	peptide cross-linking
BP	GO:0010876	3.03E-05	lipid localization
BP	GO:0006869	3.03E-05	lipid transport
BP	GO:0008152	4.84E-05	metabolic process
BP	GO:0060249	7.77E-05	anatomical structure homeostasis
BP	GO:0000723	7.77E-05	telomere maintenance
BP	GO:0032200	7.77E-05	telomere organization
BP	GO:0050790	0.000131602	regulation of catalytic activity
BP	GO:0044764	0.000145595	multi-organism cellular process

CC	GO:0031012	6.57E-15	extracellular matrix
CC	GO:0071944	0.000211565	cell periphery
CC	GO:0005839	0.000263543	proteasome core complex
CC	GO:0000502	0.000399430	proteasome complex
CC	GO:0005618	0.001984422	cell wall
CC	GO:0030312	0.001984422	external encapsulating structure
CC	GO:0000148	0.002305386	1,3-beta-D-glucan synthase complex
CC	GO:0005623	0.010472531	cell
CC	GO:0044464	0.010472531	cell part
CC	GO:0044459	0.017709891	plasma membrane part
CC	GO:0005886	0.024548454	plasma membrane
CC	GO:0005634	0.033141860	nucleus

Supplementary Table S4.5 | Top 20 GO terms enriched in analysis of lost gene families in *Z. marina*.

MF: molecular function, BP: biological process, CC: cellular component.

Type	GO ID	Adjusted p-value	Term
MF	GO:0001071	6.25E-27	nucleic acid binding transcription factor activity
MF	GO:0003700	6.25E-27	sequence-specific DNA binding transcription factor activity
MF	GO:0004672	1.84E-21	protein kinase activity
MF	GO:0004842	1.91E-21	ubiquitin-protein transferase activity
MF	GO:0042802	1.71E-20	identical protein binding
MF	GO:0017171	3.13E-18	serine hydrolase activity
MF	GO:0008236	3.13E-18	serine-type peptidase activity
MF	GO:0004568	1.93E-16	chitinase activity
MF	GO:0000155	3.32E-16	phosphorelay sensor kinase activity
MF	GO:0016775	3.32E-16	phosphotransferase activity, nitrogenous group as acceptor
MF	GO:0004673	3.32E-16	protein histidine kinase activity
MF	GO:0043565	3.91E-16	sequence-specific DNA binding
MF	GO:0030247	5.21E-15	polysaccharide binding
MF	GO:0001871	5.21E-15	pattern binding
MF	GO:0004474	7.83E-15	malate synthase activity
MF	GO:0008184	1.60E-14	glycogen phosphorylase activity
MF	GO:0004645	1.60E-14	phosphorylase activity
MF	GO:0016706	2.82E-14	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors
MF	GO:0004252	3.21E-13	serine-type endopeptidase activity
MF	GO:0016301	3.47E-13	kinase activity
BP	GO:0009856	4.95E-69	pollination
BP	GO:0009875	4.95E-69	pollen-pistil interaction
BP	GO:0008037	4.95E-69	cell recognition

BP	GO:0044706	4.95E-69	multi-multicellular organism process
BP	GO:0048544	4.95E-69	recognition of pollen
BP	GO:0044703	1.14E-68	multi-organism reproductive process
BP	GO:0022414	1.14E-68	reproductive process
BP	GO:0044707	2.51E-64	single-multicellular organism process
BP	GO:0032501	4.99E-64	multicellular organismal process
BP	GO:0051704	1.64E-61	multi-organism process
BP	GO:0019222	7.43E-41	regulation of metabolic process
BP	GO:2001141	2.76E-38	regulation of RNA biosynthetic process
BP	GO:0006355	2.76E-38	regulation of transcription, DNA-templated
BP	GO:0051252	2.76E-38	regulation of RNA metabolic process
BP	GO:0009889	7.10E-38	regulation of biosynthetic process
BP	GO:0031326	7.10E-38	regulation of cellular biosynthetic process
BP	GO:0010556	7.10E-38	regulation of macromolecule biosynthetic process
BP	GO:2000112	7.10E-38	regulation of cellular macromolecule biosynthetic process
BP	GO:0060255	1.99E-35	regulation of macromolecule metabolic process
BP	GO:0010468	5.89E-35	regulation of gene expression
CC	GO:0012511	1.15E-12	monolayer-surrounded lipid storage body
CC	GO:0005811	1.15E-12	lipid particle
CC	GO:0000151	1.25E-11	ubiquitin ligase complex
CC	GO:0098687	9.90E-11	chromosomal region
CC	GO:0000781	9.90E-11	chromosome, telomeric region
CC	GO:0000784	9.90E-11	nuclear chromosome, telomeric region
CC	GO:0016020	1.13E-09	membrane
CC	GO:0005849	6.06E-09	mRNA cleavage factor complex
CC	GO:1990234	1.09E-06	transferase complex
CC	GO:0030904	1.48E-05	retromer complex
CC	GO:0044454	2.72E-05	nuclear chromosome part
CC	GO:0000228	0.000128884	nuclear chromosome
CC	GO:0005875	0.000191211	microtubule associated complex
CC	GO:0071203	0.000200621	WASH complex
CC	GO:0031011	0.0002281	Ino80 complex
CC	GO:0033202	0.0002281	DNA helicase complex
CC	GO:0000790	0.0002281	nuclear chromatin
CC	GO:0070603	0.0002281	SWI/SNF superfamily-type complex
CC	GO:0097346	0.0002281	INO80-type complex
CC	GO:0070461	0.000580038	SAGA-type complex

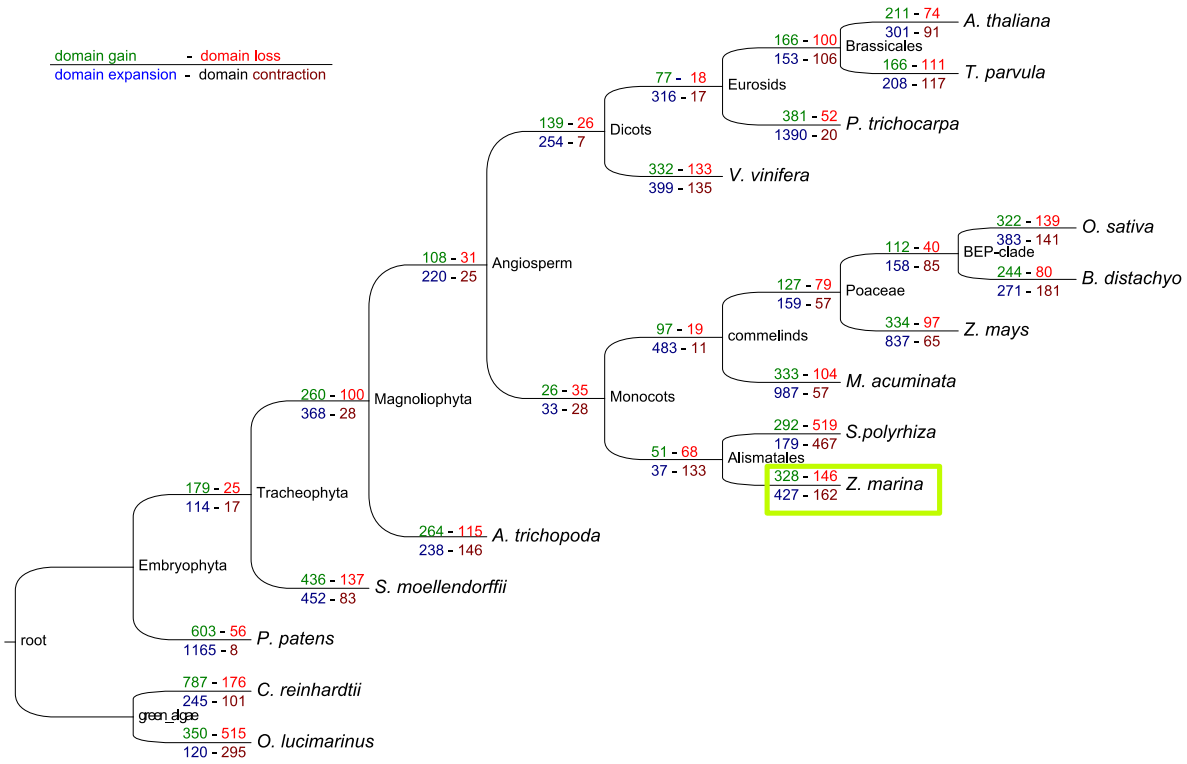
An additional approach to studying genome evolution utilizes the modular nature of proteins—their domains (Kersting *et al.* 2012). The fusion of genes and domain arrangements allows for jumps in protein evolution and may govern novel phenotypes. Modular evolution may be of particular importance for plants.

Domain annotations were conducted as follows. For genes with multiple splice variants, the longest splice variant was used. Genomes were analyzed with PFAM_SCAN.R (ftp://ftp.sanger.ac.uk/pub/databases/Pfam/Tools/), which uses HMMER3 (FINN *ET AL.* 2011) against the Pfam database (Finn *et al.* 2014). Pfam-A domains were assigned by the gathering thresholds provided by Pfam. Pfam-B domains were assigned with an e-value ≤0.001, domain length need to be < 0.3x of the hmm model length (Buljan *et al.* 2010). Pfam-A domains belonging to a clan were mapped on to it. Duplicated domains of the type 'Motif' or 'Repeat' were merged together within a gene. Overlapping domains were ordered by Pfam-A preferred over Pfam-B and then by retaining the domains with lower e-value.

Comparative domain gain, loss, expansion and contraction was performed with the ETE2-PACKAGE (Huerta-Cepas *et al.* 2010), which permits one to traverse the tree, both from tips to root and root to tips. Domain content of the ancestral nodes was calculated using maximum parsimony and a strict bifurcating tree (Kersting *et al.* 2012). Domain expansion and contraction were then calculated by count (Csuroes 2010) based on Wagner parsimony. Domains that were 1.5x more represented in the more derived node (as compared with the parent node) were designated as expanded; whereas domains that were 1.5x more represent in the parent node (as compared with the derived node) were designated as contracted.

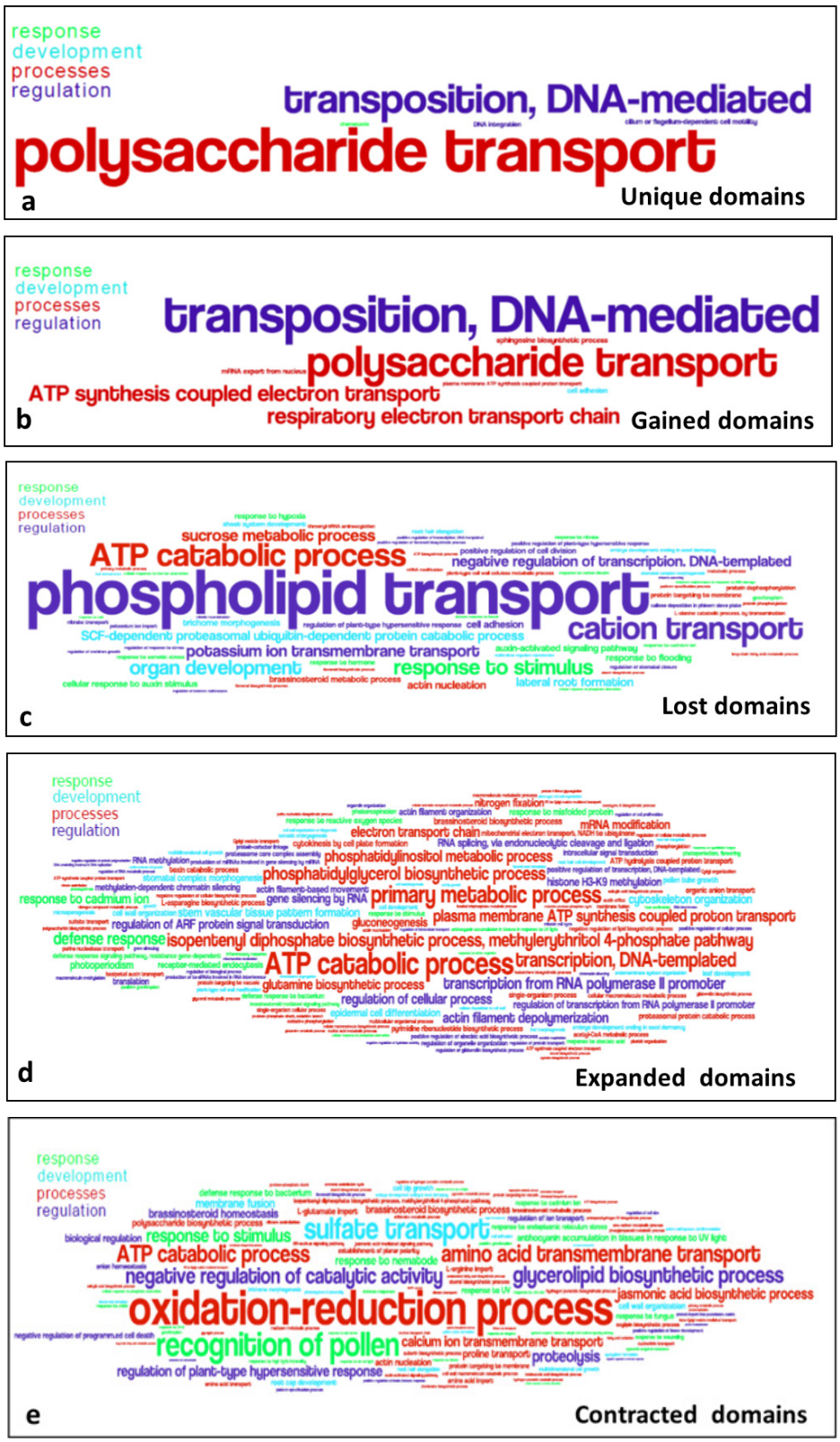
BLAST2GO (Conesa & Gotz 2008) was used to obtain the associated GO-terms (Ashburner & Lewis 2002) for all proteins. Overrepresentation of GO-terms for groups of proteins containing, e.g., expanded domains, was calculated with TOPGO in the Bioconductor package using FDR for multiple testing correction (Alexa *et al.* 2006). REVIGO (Supek *et al.* 2011)was used to visualize the overrepresentation.

Since the split from *Spirodela polyrhiza, Z. marina* has gained 328 domains, 83 of which appear to be unique; 146 domains have been lost, 427 domains have been expanded and 162 have contracted (**Supplementary Fig. S4.4).** An analysis of over-represented GO terms with unique, gained, lost, expanded or contracted domains revealed a number of results of interest for future investigation (**Supplementary Fig. S4.5, Supplementary Data 6**). Unique domains corresponded to polysaccharide transport, which correlated with the many unique carbohydrates of the cell wall in *Z. marina*. However, a check of proteins with unique domains against SwissProt was not fruitful. Lost domains generally corresponded to defense responses of insects, flooding, hypoxia, cold and CO_2—all of which are adaptations more applicable to a terrestrial life style than to *Z. marina*. Unfortunately, enrichment only captured the high-level receptor signaling protein activity rather than more specific proteins, e.g., ethylene. Expanded domains were predominantly ATP catabolic processes related to growth and photosynthesis. For example: Extensin_2, associated with cell wall strengthening (and not found in other monocots); sphingosine, which is involved in lipid signaling in diverse cellular processes; (Oxidored_q2, Cytochrom_B_C) and (Oxidored_q1, Cytochrom C1), which are involved in the electron transport and polysaccharide transport. Contracted domains were mainly related to responses to UV, UV-B and far red light and reduced defense response to nematodes and defense in general.



Supplementary Figure S4.4 | Pfam domain analysis of protein domain gain and loss in fourteen sequenced species.

Green is domain gain, red is domain loss, blue is domain expansion and brown is domain contraction.



Supplementary Figure S4.5 | Wordle diagram of Pfam domain, functional enrichment analysis.

a, unique; **b**, gained; or **c**, lost; **d**, expanded; **e,** contracted in *Z. marina*.

5 Stomata

5.1 Stomata genes

Stomata are microscopic structures formed by two specialized guard cells flanking a central pore on the surfaces of plant aerial organs/leaves that control gas exchange, water transpiration and osmoregulation between the plant interior and the aerial environment. Stomata have been found in vascular plant fossils (Edwards *et al.* 1992) and evolved their developmental program and regulatory gene network >400 MYA in mosses (Bowman 2011; Chater *et al.* 2013). Stomata allowed land plants to survive dry atmospheric conditions and exploit their new terrestrial niche, especially building upon CO_2 availability (Berry *et al.* 2010).

Although derived from land plants ~100 MYA, seagrasses do not possess stomata (Kuo & den Hartog 2006). The molecular basis driving the development of stomata has been investigated in detail in *Arabidopsis* (Pillitteri & Dong 2013). The main signaling pathways and individual components have been identified and appear to be conserved among different plant clades (Peterson *et al.* 2010) even if variations in the developmental program occur (Vaten & Bergmann 2012).

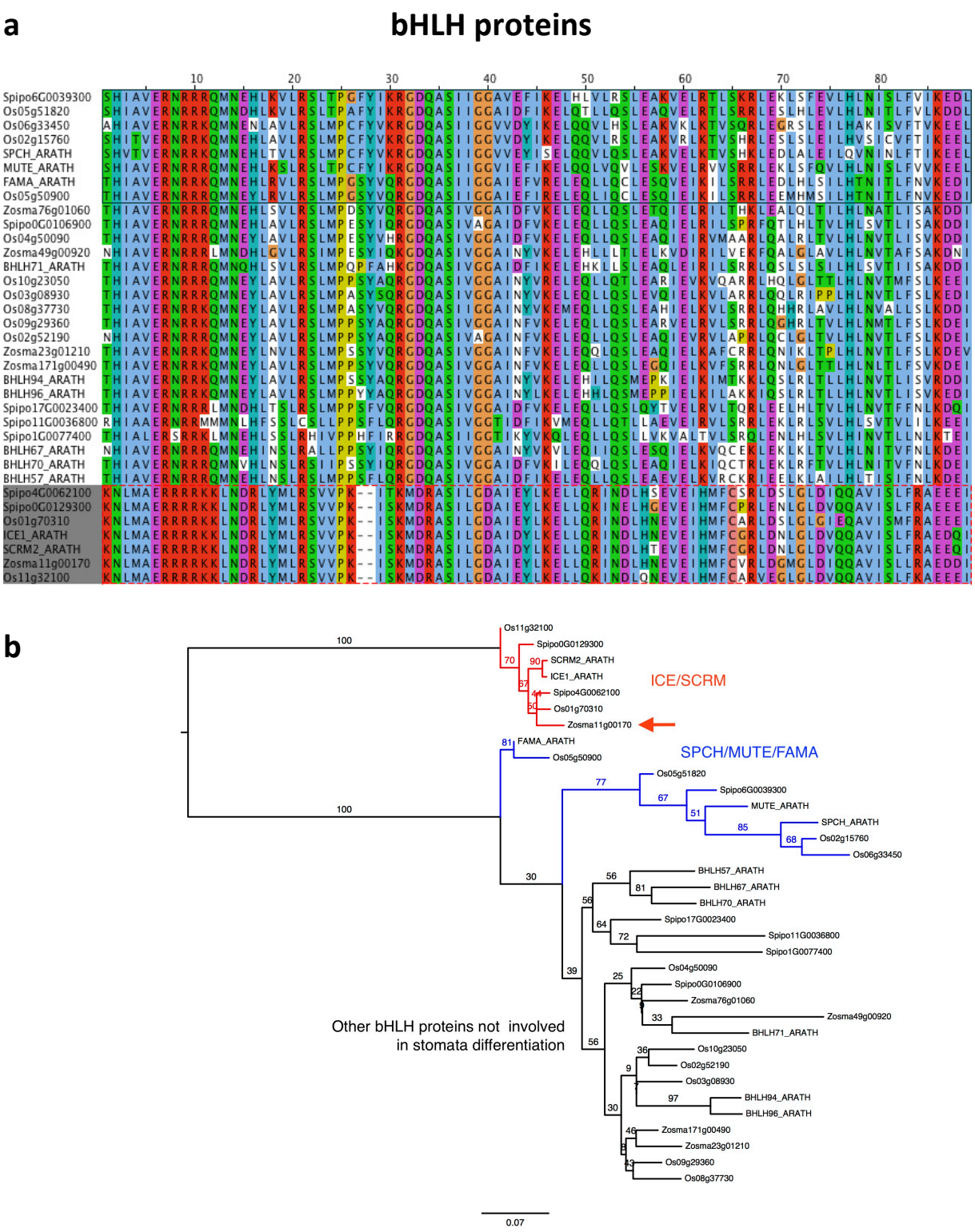
The gain/loss analysis provided a striking result: 11 genes which are specifically involved in stomatal development and patterning are missing in the genome of *Z. marina* (**Extended Data Table 1).** All bHLH proteins governing the three differentiation steps of epithelial cells into guard cells are missing, except SCRM (**Fig. 3a, Supplementary Figure S5.1**). The signaling EPF/EPFL involved in stomata patterning are also missing (**Fig. 3a, Supplementary Figure 5.2**), except for all members of the CHAL subfamily. TMM, the specific receptor-like kinase sensing this signal is missing as well, while its universal ER/ERL RLK partners are not. Finally, the proteinase-encoding SDD1 gene is also missing, an interesting observation suggesting that the only function of this gene is in stomatal development.

A closer look at the remaining genes shows that they are not dedicated to stomatal formation, having a broader role in other processes besides stomatal differentiation and patterning. This is clearly the case with the ER/ERL RLKs that have a pleiotropic role in plant development and environmental adaptation (Shpak 2013). Interestingly, ER has been shown to regulate inflorescence architecture through the binding of EPFL4 and EPFL6/CHAL, showing that genes in this EPFL/CHALf family play additional roles (Uchida *et al.* 2012). SCRM has in fact been first discovered for its role in freezing tolerance, hence its alternative name, ICE1 (Chinnusamy *et al.* 2003).

All of the above genes were mainly expressed in female and male flower tissues (**Supplementary Table S5.1)**. There are many occurrences of asymmetric divisions during plant development (Kajala *et al.* 2014) and it is likely that the genes involved in stomatal patterning are involved in other processes. For example, it has recently been shown that PAN1 and PAN2 were not dedicated to SMC polarization but were instead playing a broader role in plant development (Sutimantanapi *et al.* 2014). Our finding of POLAR and others in *Z. marina* suggests that this gene and others in the "remaining category" are involved in other processes such as female flower development (but not in male flowers or in roots).

As it has been shown in *Arabidopsis* that plants with a mutation in many of these genes are devoid of stomata (Pillitteri & Dong 2013), one might expect that the loss of a single gene in the pathway would result in the same loss-of-stomata phenotype. The finding that *Z. marina* has lost all of the genes that are known to be involved in stomatal differentiation (and the fact that the genes are not clustered in stomatal-bearing plants) indicates that this loss of stomata is an ancient and irreversible feature binding *Z. marina* to the sea.

Further we note that the freshwater Alismatid relative of *Z. marina*, the duckweed *Spirodela polyrhiza*, possesses stomata and all of the relevant stomatal genes.

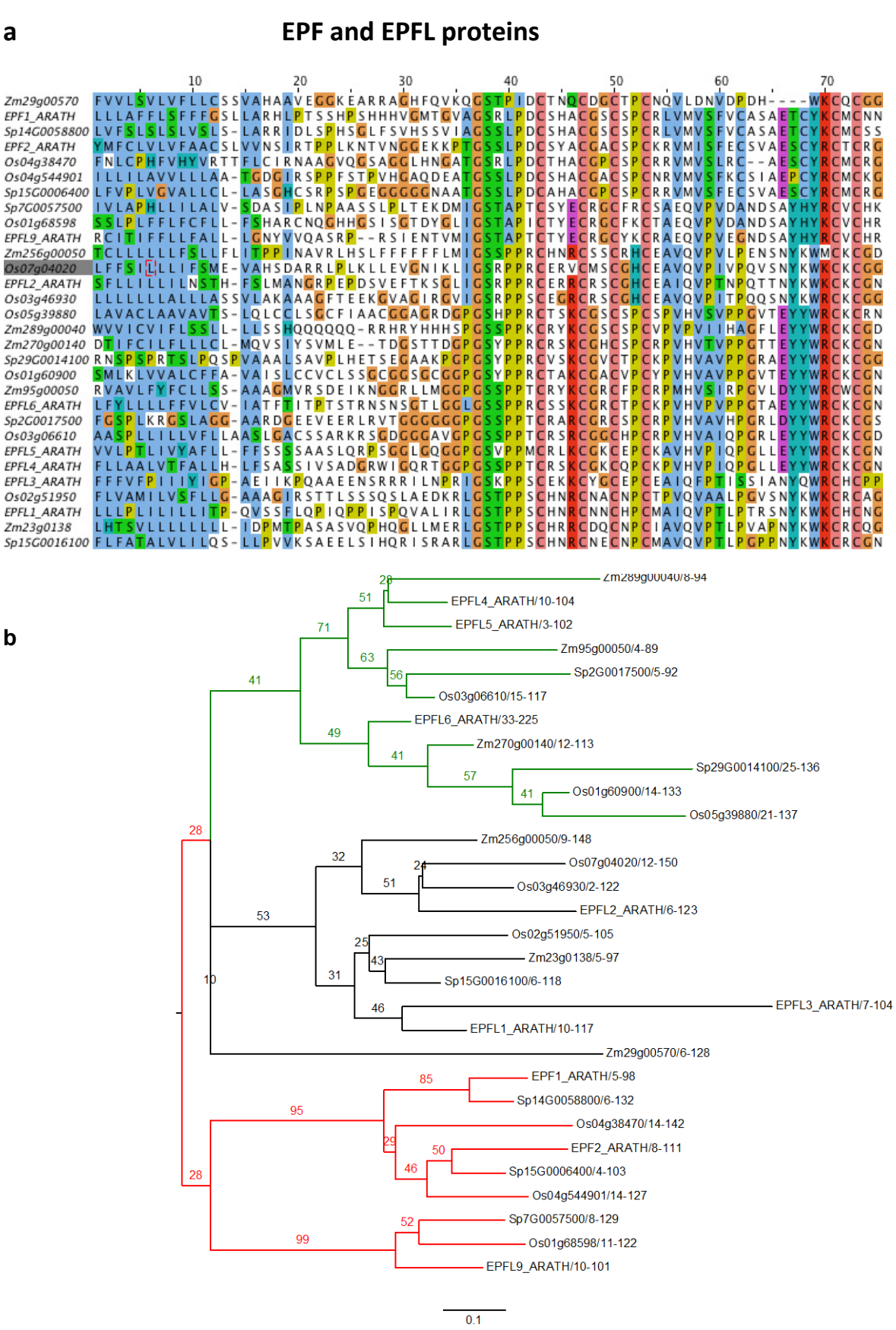


Supplementary Fig. S5.1 | Stomata development gene comparisons for basic-helix-loop-helix (bHLH) protein.

a, MUSCLE alignment of bHLH proteins, which are involved in stomata differentiation (SPCH, MUTE, FAMA, ICE/SCRM) or not (BHLHnn). Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes allowing 2 gaps in the same window.

b, Maximum Likelihood phylogenetic tree calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. ICE/SCRM, SPCH, MUTE,and FAMA are bHLH proteins shown to be involved in stomata differentiation in *Arabidopsis* (see text and **Extended Data Table**

1), while other bHLHs are not. The arrow indicates the ICE/SCRM ortholog from *Z. marina*, the only bHLH gene falling in the stomata cluster. Species abbreviations: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).



Supplementary Fig. 5.2 | Stomata development gene comparaisons for epidermal patterning factor (EPF) and epidermal patterning factor like (EPFL) proteins

a, The EPF family was aligned by MUSCLE using default settings. Alignment gaps and unconcerned regions were manually removed under Jalview. **b,** Maximum likelihood phylogenetic tree calculated in

RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. EPF1, EPF2, EPFL9 (stomatogen), EPFL4 (CHAL-like2), EPFL5 (CHAL-like1) & EPFL6 (Challah) are EPF and EPFL proteins shown to be involved in stomata differentiation in *Arabidopsis* (see text and **Extended Data Table 1**) while other EPFLs are not. This phylogeny shows the absence of EPF1, EPF2 and EPFL9 orthologs in *Zostera marina* (red clusters) while EPFL4, EPFL5 and EPFL6 do have orthologs in *Zostera* (green). Species abbreviations: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

Supplementary Table S5.1 | Normalized gene expression (FPKM) of stomata-related genes among *Z. marina* tissue types.

		Female flower early	Female flower late	Male flower	Root	Leaf
SCREAM / ICE1	zosma11g00170	25.19	58.20	1.45	3.88	1.61
ERECTA	zosma87g00130	52.98	22.15	2.73	5.58	5.78
ERECTA-LIKE1	zosma292g00090	12.70	7.95	0.13	0.10	4.68
ERECTA-LIKE2	zosma85g01030	25.25	13.95	1.96	1.07	2.58
CHALLAH/EPF-LIKE9	zosma270g00140	25.03	11.67	0.61	1.52	0.90
CHAL-LIKE1/EPF-LIKE5	zosma95g00050	6.40	5.28	0.68	2.11	0.58
CHAL-LIKE2/EPF-LIKE4	zosma289g00040	26.16	29.60	11.40	6.80	3.92
PANGLOSS1	zosma293g00080	22.43	21.28	0.41	7.00	7.27
PANGLOSS2	zosma30g00950	24.23	25.25	1.23	4.80	2.67
PANGLOSS2	zosma117g00680	18.31	21.48	2.24	4.49	3.26
POLAR	zosma16g01600	6.42	16.70	4.72	14.60	1.27
SCD1	zosma40g00290	30.19	21.41	26.65	22.47	26.56
SCD1	zosma40g00310	1.26	0.67	1.50	7.54	0.99

6 Plant Defense

Both basal secondary metabolites and immune system responses are reduced in *Z. marina*, a surprising result given that the marine environment harbors just as many (albeit different) herbivores and pathogens as on land.

6.1 Secondary metabolites and signaling

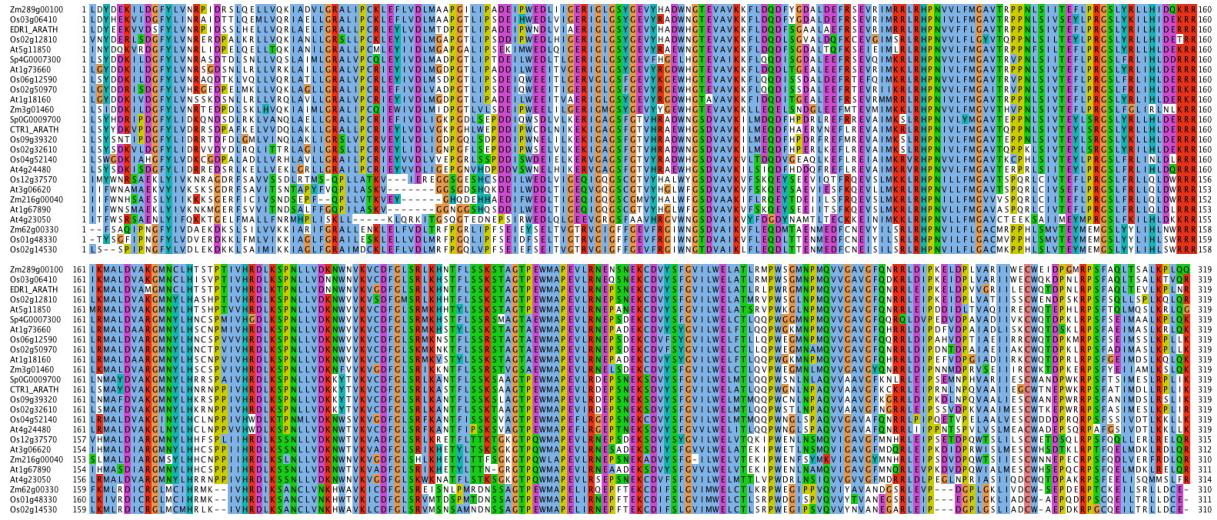
Flavonoids play various roles: pigments, UV protection, chemical messengers, and defense against pathogens (Lepiniec *et al.* 2006; Varin *et al.* 1997). Two large multigenic CAZy families are involved in the biosynthesis and recycling of carbohydrate-based flavonoids: the GT1 family (flavonoid glycosyltransferases) and the GH1 family (flavonoid beta-glucosidase & myrosinase) (Henrissat *et al.* 2001). *Z. marina* displays a reduced number of flavonoid-related genes with only 50 GT1 (122 GT1 in *Arabidopsis*, 202 in *Oryza*) and only 2 GH1 (48 GH1 in *Arabidopsis*, 35 in *Oryza)* **(Supplementary Table S7.4, Supplementary Data 8).**

Chitinases (GH18, GH19 and their associated chitin-binding modules [CBM18 and CBM50]) are produced to defend against terrestrial fungi (Stintzi *et al.* 1993). Chitinases are reduced in *Z. marina* with chitin catabolism involving only 4 GH18, 3 GH19, 1 CBM18 and 1 CBM50; as compared with *Oryza* (34 GH18, 17 GH19, 15 CBM18 and 7 CBM50). At present little is known about the role of fungi as seagrass pathogens **(Supplementary Data 8).**

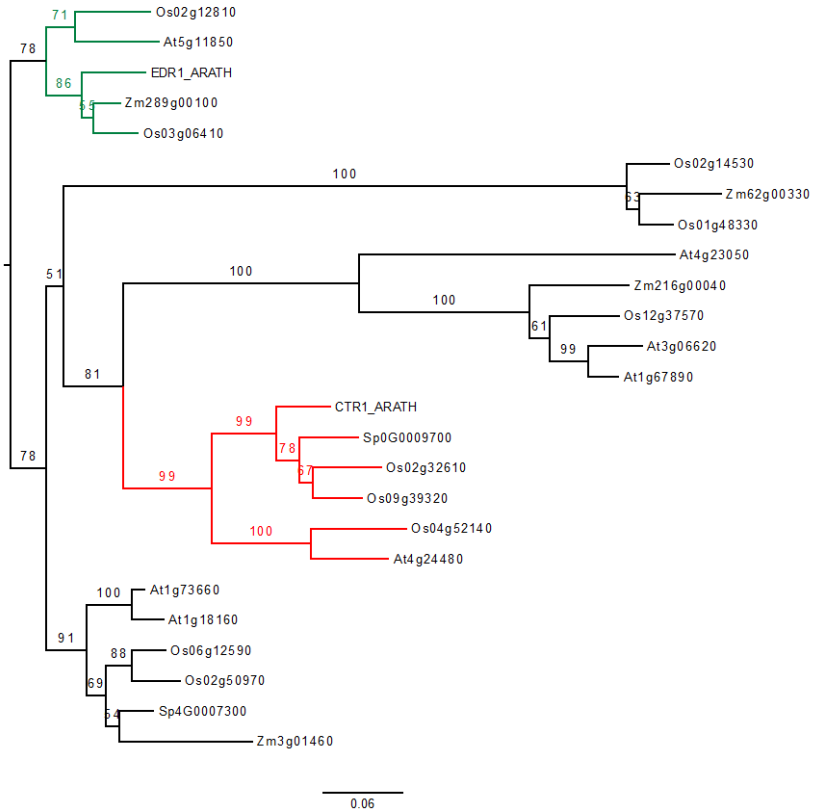
Ethylene is a key phytohormone involved in the regulation of virtually all developmental processes in plants through complex crosstalk and signaling, e.g., seedling emergence, leaf and flower senescence, formation of the vascular system, ripening of fruits, organ abscission. Ethylene also regulates responses to biotic and abiotic stresses, notably by the emission of volatile ethylene (through stomata) in response to herbivory by insects (which is further modulated by another phytohomromone, jasmonic acid).

Ethylene responsive genes (**Fig. 3b, Extended Data Table 2, Supplementary Fig. S6.1)** are entirely absent in *Z. marina*, which is consistent with the loss of stomata and the absence of insect herbivores in the sea (see paragraph at the end of **Supplementary Note 6.2**). The loss of ethylene genes was also analyzed in Golicz et al. (2015).

a **CTR1, EDR1 and related putative MAPK kinase kinases**



b



Supplementary Fig. S6.1| Ethylene response/receptor genes - constitutive triple response (CTR and CTR1), enhanced disease resistance (EDR1) and related putative mitogen-activated protein kinases (MAPK).

a, MUSCLE alignment of CTR1, EDR1 and related putative MAPK kinase kinases. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes allowing 2 gaps in the same window. **b,** Maximum likelihood phylogenetic tree calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. This phylogeny shows the orthologs of CTR1 in rice and *Spirodela* (red) but its absence from *Zostera marina*, while EDR1, a

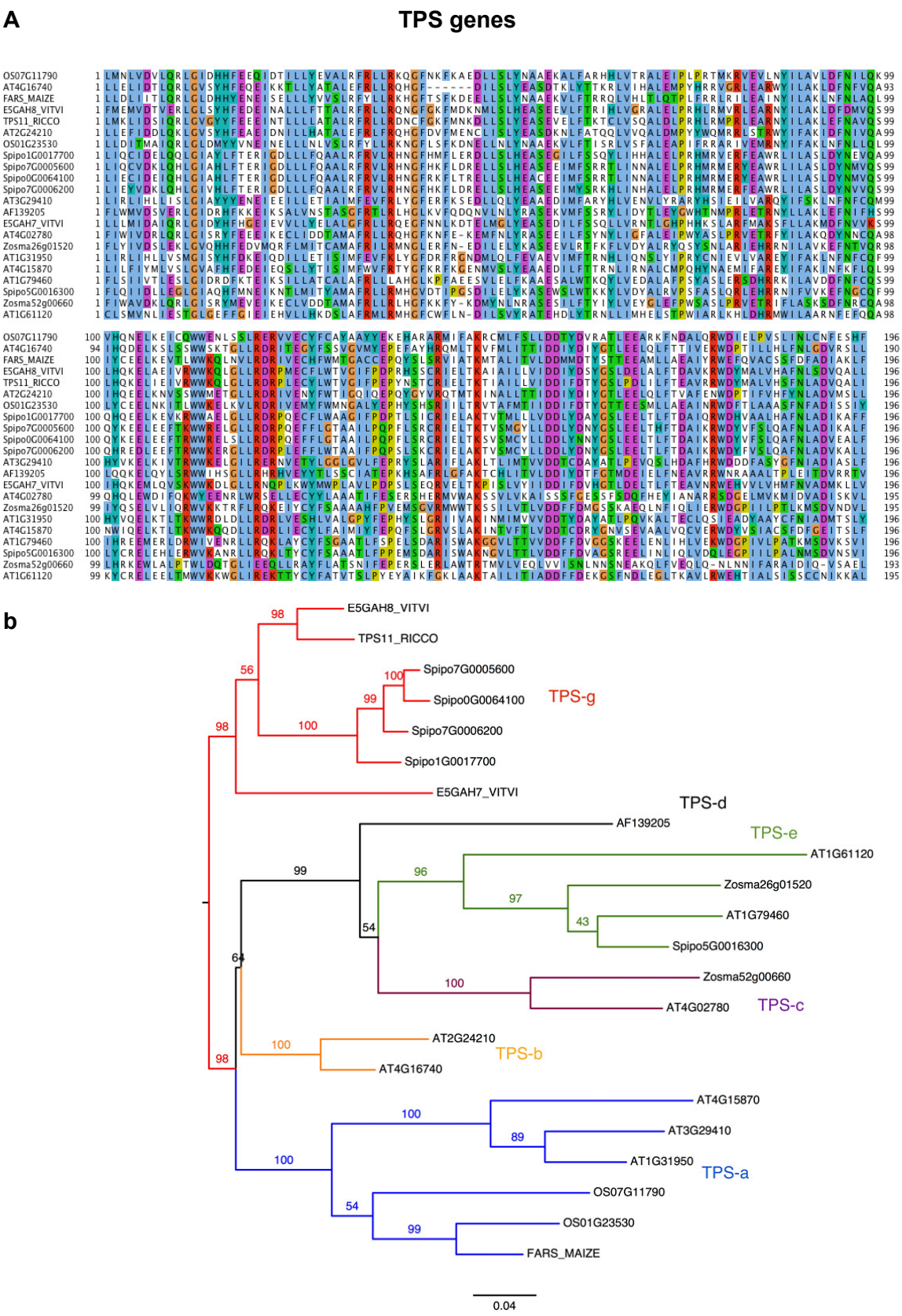
homologous MAPKKK that is involved in plant defense, is found in all species (in green). Species abbreviations: *Arabidopsis* (ARATH), rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

Terpenoids are all derived from isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP) (Chen *et al.* 2011). Although some are important players in plant primary metabolism (**Fig. 3c**), the vast majority are part of the largest and most diverse group of plant secondary metabolites, with the number of different structures having been estimated at >25,000. Most of these secondary terpenes are volatiles, which are expanded to serve specific functions within each plant taxon, e.g., plant reproduction and the attraction of animals, mainly insects, for pollination and seed dispersal. They are also main players in plant defense, acting to repel pests or as attractants towards their competitors or pathogens. The tremendous diversity of terpenoids in plants is produced through the action of a single family of Terpene Synthases (TPSs), each member being itself able to deal with a range of substrates and then to produce a variety of terpenes.

Zostera marina has only two Terpene Synthase (TPS) genes (Zosma52g00660 and Zosma26g01520). This is the lowest complement reported, putting it at the level of the bryophyte moss, *Physcomitrella patens*, as compared with 34 in *Arabidopsi,* 51 in *Oryza* (Chen *et al.* 2011) and a high of 113 in *Eucalyptus* (Myburg *et al.* 2014). Interestingly Zosma52g00660 and Zosma26g01520 encode an Ent-copalyl diphosphate synthase (CPS1/KSA) and an Ent-kaur-16-ene synthase (KS1/KSB), respectively. KSA and KSB are involved in the two consecutive steps to the gibberellin precursor ent-kaur-16-ene, then allowing the synthesis of the GA phytohormones (**Fig. 3c**).

Both Zosma52g00660 and Zosma26g01520 gene models are unambiguously supported by RNA-seq data, the first gene having low expression in male flowers and the second being strongly expressed in all organs. Their protein sequences nevertheless differ slightly from the KSA and KSB from other angiosperms; *Z. marina* KSA lacking a ca. 200aa C-terminal extension typical of plant KSAs and KSB being much shorter on its N-terminus.

A search for TPS in the *Spirodela* genome (following a proper check and re-annotation of the gene models) shows that besides the KSA and KSB genes (for which *Spirodela* has two copies of each) there are several other TPSs. At least four intact genes in the TPSg clade (**Supplementary Fig. S6.2**), as well as at least eight pseudogenes or partial genes are present. In contrast to *Spirodela*, there are no TPS genes in the *Zostera* genome other than those involved in GA synthesis. This means that *Zostera* cannot produce secondary volatile terpenes. As TPSg genes (which produce volatiles) are found in *Spirodela* (the only other sequenced alismatid), this lack of volatile terpenes is genuinely a *Zostera*-specific event.



Supplementary Fig. S6.2 | Terpene synthase (TPS) genes.

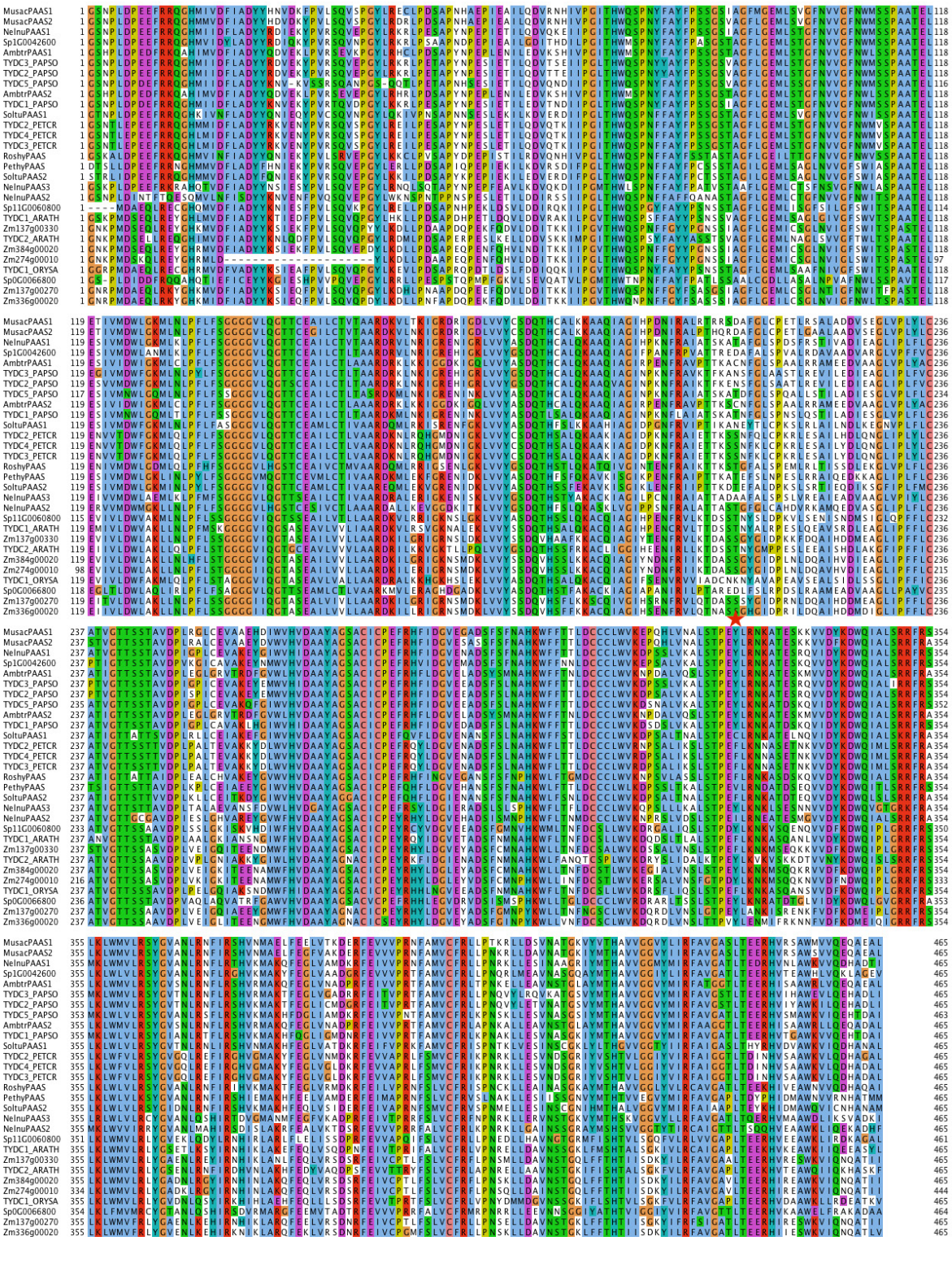
a, MUSCLE alignment. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10-amino acid window sizes allowing 2 gaps in the same window. **b,** Maximum Likelihood tree of TPS from *Z. marina, S. polyrhiza* and other plants calculated with RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. Sequences were chosen to capture a good representative set of TPS genes (see Chen *et al.* 2011): Arabidopsis (ARATH), Rice (Os), *Ricinus* (RICCO),

Maize, *Vitis* (VITVI) and *Picea abies* (AF139205) besides *Zostera marina* (Zm) and *Spirodela polyrhiza* (Sp). The two *Zostera* TPS with their KSA and KSB orthologs are shown in TPS-c and TPS-e clades, respectively. In addition to the KSA and KSB orthologs in these two clades, four *Spirodela* TPS were found in the TPS-g clade.

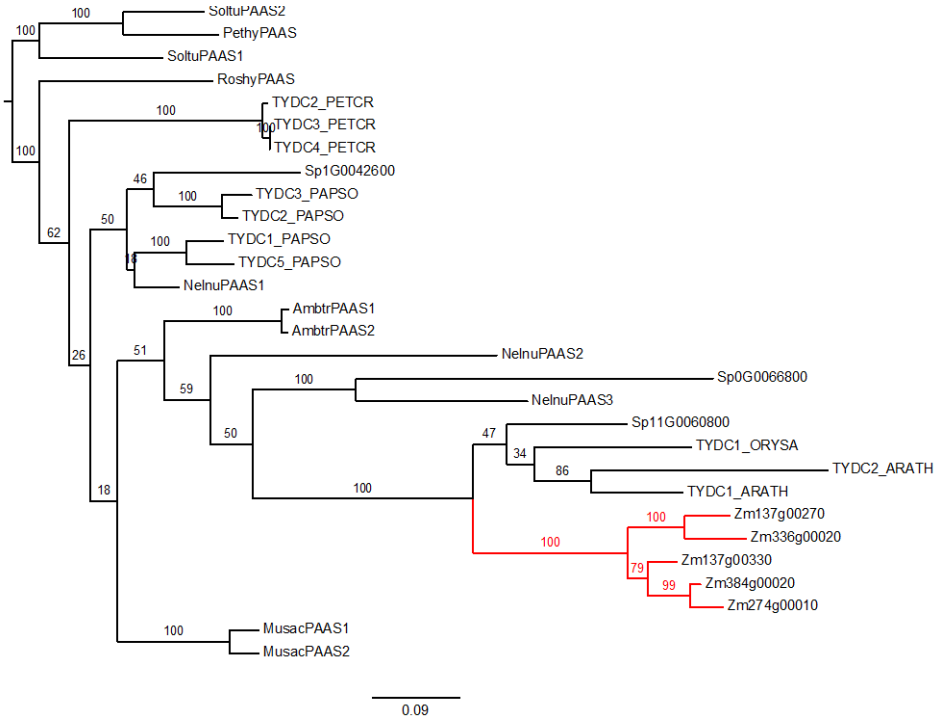
Aromatic amino acid decarboxylases (AAAD) genes, in contrast, are expanded in *Z. marina* with five genes; monocots typically have one and dicots two, although some plants have more, such as the opium poppy where they are involved in the biosynthesis of benzyloquinoline (**Supplementary Fig. S6.3**). The encoded PLP enzymes have been shown to be either tyrosine (aromatic amino acid) decarboxylases, i.e. producing tyramine (or another aromatic amine), or aromatic aldehyde synthases, i.e. producing phenylacetaldehyde (or another aromatic aldehyde). In *Arabidopsis* these genes are linked to defense against wounding herbivores and to flower scent production (Gutensohn *et al.* 2011).

Interestingly, AAADs from *Zostera* cluster together, far from most plant AAADs including *Spirodela,* but close to *Arabidopsis,* rice and *Spirodela* tyrosine/dihydroxyphenylalanine decarboxylase (TYDC). Whether or not *Z. marina* AAADs would serve as a defense mechanism against wounding is then an attractive hypothesis that remains to be tested.

a **AAAD proteins**



b



Supplementary Fig. S6.3 | Alignment and phylogenetic tree showing the expansion of arylalkylamine and aldehyde synthesizing aromatic amino acid decarboxylase (AAAD) proteins.

a, MUSCLE alignment of aromatic amino acid decarboxylase (AAAD) proteins. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes allowing 2 gaps in the same window. The F/Y active site has been highlighted (red star) as it corresponds to a shift in the enzymatic activity of the protein (Torrens-Spence *et al.* 2013), being either aldehyde synthase (F) or AAAD decarboxylase (Y). **b**, Maximum likelihood tree calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. Species abbreviations: *Arabidopsis* (ARATH), rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo)

6.2 Immune defense R-genes

Vascular plant defense is based on proteins containing the nucleotide binding site domain and a leucine rich repeat domain (NBS-LRR) with some possessing an additional coil-coil (CC) domain. These proteins recognize pathogen proteins either directly, conferring resistance (gene-for-gene model (Hammond-Kosack & Jones 1997), or indirectly by detecting pathogen induced changes in host proteins (guard hypothesis, Van der Biezen & Jones 1998). In both cases the LRR domain binds the target protein and the NBS domain invokes a signal by hydrolysis of ATP or GTP, which initiates the hypersensitive response. NBS-LRR genes are categorized into six classes, depending on the domain composition of the R-gene (**Supplementary Table S6.1**). In monocots the TIR domain class seems to be much less common than in dicots (Tarr & Alexander 2009). NBS genes show great diversity with the LRR domain evolving non-neutrally, consistent with diversifying selection (Ellis *et al.* 2000).

Supplementary Table 6.1 | Numbers of defense-associated nucleotide-binding site (NBS) genes by class in different plant species as compared with *Z. marina.*

CC, coiled coil; LRR, leucine-rich repeat; TIR, Toll interleukin receptor rich repeat;

	Arabidopsis	Musa	Oryza	Phoenix	Populus	Solanum	Sorghum	Zea	Spirodela	Zostera
CC-NBS	7	?	72	40	19	4	27	21	3	1
CC-NBS-LRR	43	27	42	19	119	103	40	10	30	11
NBS	0	177	260	69	27	41	93	79	11	10
NBS-LRR	11	89	118	16	71	213	176	26	41	19
TIR-NBS	15	?	1	0	32	22	0	0	4	3
TIR-NBS-LRR	83	0	0	0	91	55	0	0	0	0
Total	**159**	**293**	**493**	**144**	**359**	**438**	**336**	**136**	**89**	**44**

Arabidopsis thaliana (Guo *et al.* 2011),
Musa acuminata (D'Hont *et al.* 2012),
Oryza sativa (Mace *et al.* 2014),
Phoenix dactylifera (Al-Mssallem *et al.* 2013),
Populus trichocarpa (Kohler *et al.* 2008),
Solanum tuberosum (Jupe *et al.* 2012),
Sorghum bicolor (Mace *et al.* 2014),
Spirodela polyrhiza(Wang *et al.* 2014),
Zea mays(Mace *et al.* 2014).

R-genes in *Z. marina* and *S. polyrhizia* were initially identified using HMMER ver. 3.1B2 to find proteins containing the NB-ARC domain as defined by the Pfam entry (PF00931). Domains that were identified were cut out, aligned and used to produce a species-specific hidden markov model (hmm). This was used again to search for further genes with the NBS domain. The resulting list was manually curated to remove false positive genes with ATPase or GTPase domains. Genes were manually classified by searching for further domains (coiled coil, leucine rich repeat, TIR) using INTERPROSCAN. NBS domains were located in the protein sequences of all species by searching the Pfam NB-ARC domain using HMMSCAN, the sequence matching the domain was then cut out. All domain sequences were aligned using MAFFT ver. 7.215 and a tree was calculated using FASTTREE (http://microbesonline.org/fasttree/) with default settings.

Z. marina contains half (44) of the number of NBS resistance genes of *S. polyrhizia* (89) and only a small fraction of the numbers found in other plants (**Supplementary Table S6.1** and **S6.2**). The distribution among classes is typical for a monocot, with few TIR domain-containing proteins; but considering the low number of NBS genes in the first place, three such genes is comparatively many (7% of all R-genes) compared to rice (0.2%), maize (0%) and sorghum (0%). The length of the proteins is evenly distributed from 93 amino acids to 1277. The shortest variants are probably not functional as the NBS domain alone is usually ~280 amino acids long; they also have no or low expression (**Supplementary Table S6.3**). The total functional number of NBS r-genes could thus be even lower.

Supplementary Table 6.2 | List of 44 nucleotide-binding site (NBS) resistance genes in *Z. marina*.

CN, Coiled-coil-NBS. CNL, Coiled-coil-NBS-LRR. N, NBS. NL, NBS-LRR. TN, TIR-NBS.

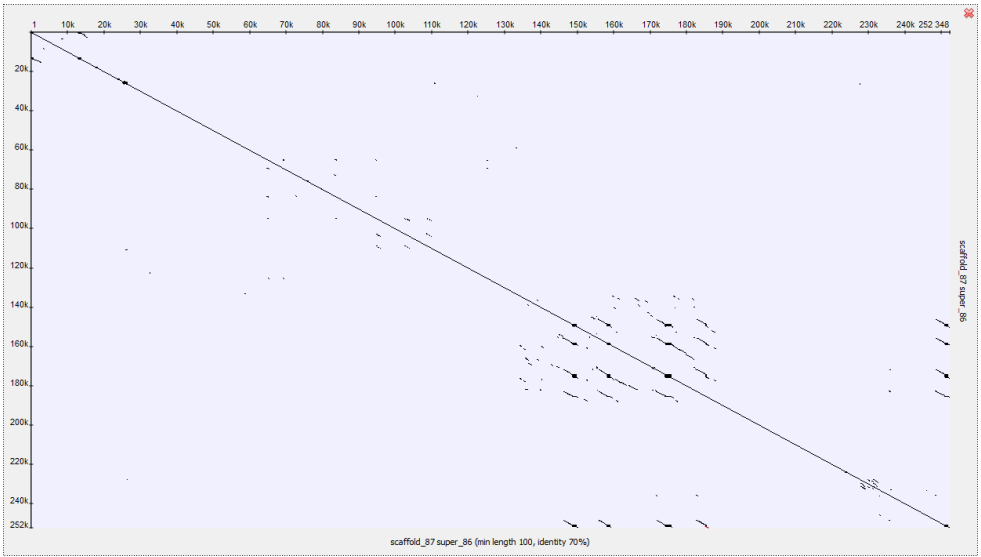
NBS gene	Class	NBS gene	Class
Zosma223g00100	CN	Zosma46g00470	NL
Zosma2195g00010	CNL	Zosma1597g00010	NL
Zosma113g00680	CNL	Zosma147g00160	NL
Zosma223g00060	CNL	Zosma87g00860	NL
Zosma20g01310	CNL	Zosma82g00030	NL
Zosma106g00700	CNL	Zosma87g00490	NL
Zosma176g00080	CNL	Zosma147g00170	NL
Zosma46g00640	CNL	Zosma147g00120	NL
Zosma90g00040	CNL	Zosma9g01570	NL
Zosma331g00140	CNL	Zosma87g00470	NL
Zosma115g00350	CNL	Zosma160g00400	NL
Zosma343g00010	CNL	Zosma58g01070	NL
Zosma73g00200	N	Zosma74g00260	NL
Zosma73g00190	N	Zosma91g00880	NL
Zosma48g00430	N	Zosma87g00840	NL
Zosma143g00210	N	Zosma264g00130	NL
Zosma326g00020	N	Zosma87g01050	NL
Zosma202g00090	N	Zosma77g00800	NL
Zosma513g00070	N	Zosma87g00830	NL
Zosma220g00080	N	Zosma31g00100	TN
Zosma408g00010	N	Zosma53g00050	TN
Zosma87g00900	N	Zosma15g01450	TN

Supplementary Table 6.3 | Normalized gene expression (FPKM) of nucleotide-binding site (NBS) R-genes in *Z. marina*.

Protein length and scaffold location are provided, colors in the scaffold column indicate genes located on the same scaffold.

			Expression					
NBS gene	Class	Length	Female flower early	Female flower late	Male flower	Root	Leaf	Scaffold
Zosma74g00260	NL	1038	3.26	4.73	159.01	11.74	5.19	scaffold_74
Zosma87g00830	NL	1277	3.49	6.10	3.29	5.18	83.94	scaffold_87
Zosma9g01570	NL	904	11.14	15.65	3.95	27.21	27.11	scaffold_9
Zosma58g01070	NL	1025	0.01	0.00	0.08	75.13	0.59	scaffold_58
Zosma53g00050	TN	1010	18.42	18.61	8.98	12.13	8.07	scaffold_53
Zosma87g00860	NL	720	0.03	0.11	0.13	61.52	1.54	scaffold_87
Zosma87g00470	NL	916	0.00	0.00	0.10	59.97	0.24	scaffold_87
Zosma91g00880	NL	1175	7.86	11.08	4.83	16.26	14.91	scaffold_91
Zosma15g01450	TN	1057	2.68	2.70	6.23	33.88	4.98	scaffold_15
Zosma77g00800	NL	1269	3.00	2.11	14.90	19.13	4.41	scaffold_77
Zosma264g00130	NL	1189	3.70	3.34	4.86	13.78	12.35	scaffold_264
Zosma87g00840	NL	1180	0.02	0.07	0.08	5.13	32.00	scaffold_87
Zosma31g00100	TN	985	14.80	11.74	2.12	5.96	2.64	scaffold_31
Zosma147g00170	NL	833	0.02	0.04	0.03	0.08	30.02	scaffold_147
Zosma87g00490	NL	801	0.00	0.00	0.10	29.70	0.20	scaffold_87
Zosma46g00640	CNL	949	11.53	6.92	3.09	4.19	1.28	scaffold_46
Zosma46g00470	NL	510	0.65	0.93	5.22	10.15	1.90	scaffold_46
Zosma160g00400	NL	976	0.00	0.00	0.19	12.38	0.00	scaffold_160
Zosma176g00080	CNL	864	0.08	0.06	0.07	4.96	6.94	scaffold_176
Zosma90g00040	CNL	977	0.08	0.11	0.07	8.47	0.11	scaffold_90
Zosma223g00060	CNL	663	0.01	0.00	0.07	0.01	7.59	scaffold_223
Zosma147g00120	NL	888	0.00	0.00	0.08	6.72	0.77	scaffold_147
Zosma87g00900	N	984	0.00	0.00	0.00	5.13	0.02	scaffold_87
Zosma343g00010	CNL	1219	0.00	0.00	0.01	3.82	0.01	scaffold_343
Zosma2195g00010	CNL	522	0.41	0.25	0.70	0.89	0.74	scaffold_2195
Zosma87g01050	NL	1195	0.00	0.00	0.02	2.86	0.01	scaffold_87
Zosma223g00100	CN	367	0.00	0.00	0.00	0.00	2.85	scaffold_223
Zosma115g00350	CNL	1026	0.12	0.30	0.43	0.91	0.83	scaffold_115
Zosma147g00160	NL	624	0.00	0.00	0.01	2.10	0.28	scaffold_147
Zosma82g00030	NL	794	0.00	0.00	0.21	1.06	0.00	scaffold_82
Zosma20g01310	CNL	746	0.03	0.02	0.02	0.78	0.00	scaffold_20
Zosma106g00700	CNL	794	0.00	0.00	0.03	0.42	0.00	scaffold_106
Zosma113g00680	CNL	626	0.00	0.00	0.00	0.42	0.00	scaffold_113
Zosma143g00210	N	223	0.00	0.00	0.36	0.00	0.00	scaffold_143
Zosma331g00140	CNL	1014	0.00	0.00	0.03	0.21	0.00	scaffold_331
Zosma220g00080	N	402	0.00	0.00	0.00	0.23	0.00	scaffold_220
Zosma408g00010	N	497	0.00	0.00	0.05	0.11	0.00	scaffold_408
Zosma48g00430	N	204	0.00	0.00	0.04	0.00	0.03	scaffold_48
Zosma202g00090	N	300	0.00	0.00	0.04	0.00	0.00	scaffold_202
Zosma513g00070	N	308	0.00	0.00	0.02	0.00	0.00	scaffold_513
Zosma73g00190	N	197	0.00	0.00	0.00	0.00	0.00	scaffold_73
Zosma73g00200	N	93	0.00	0.00	0.00	0.00	0.00	scaffold_73
Zosma326g00020	N	270	0.00	0.00	0.00	0.00	0.00	scaffold_326
Zosma1597g00010	NL	592	0.00	0.00	0.00	0.00	0.00	scaffold_1597

NBS R-genes in plants are often clustered on the genome suggesting a history of tandem duplications which in *Z. marina* seems only partly the case. There is one group of seven genes on scaffold 87, one group of three on scaffold 147 and the rest of the genes are only one or two per scaffold. The gene cluster on scaffold 87 consists of tandem repeats of genes, with two genes further away from the others (**Supplementary Fig. S6.4**).



Supplementary Fig. S6.4 | Dot plot of the scaffold_87 region containing r-genes.

(scaffold_87: 314040-566387, dot plot drawn with window size 100 bp, 70% min identity).

An alignment shows conservation over the NBS domain and further parts of the proteins, but little conservation in sequence or length in the C-terminal third **(Supplementary Fig. S6.5**). There is a small conserved domain of unknown function close to the C-terminus. Duplications are probably not recent due to the large number of differences.

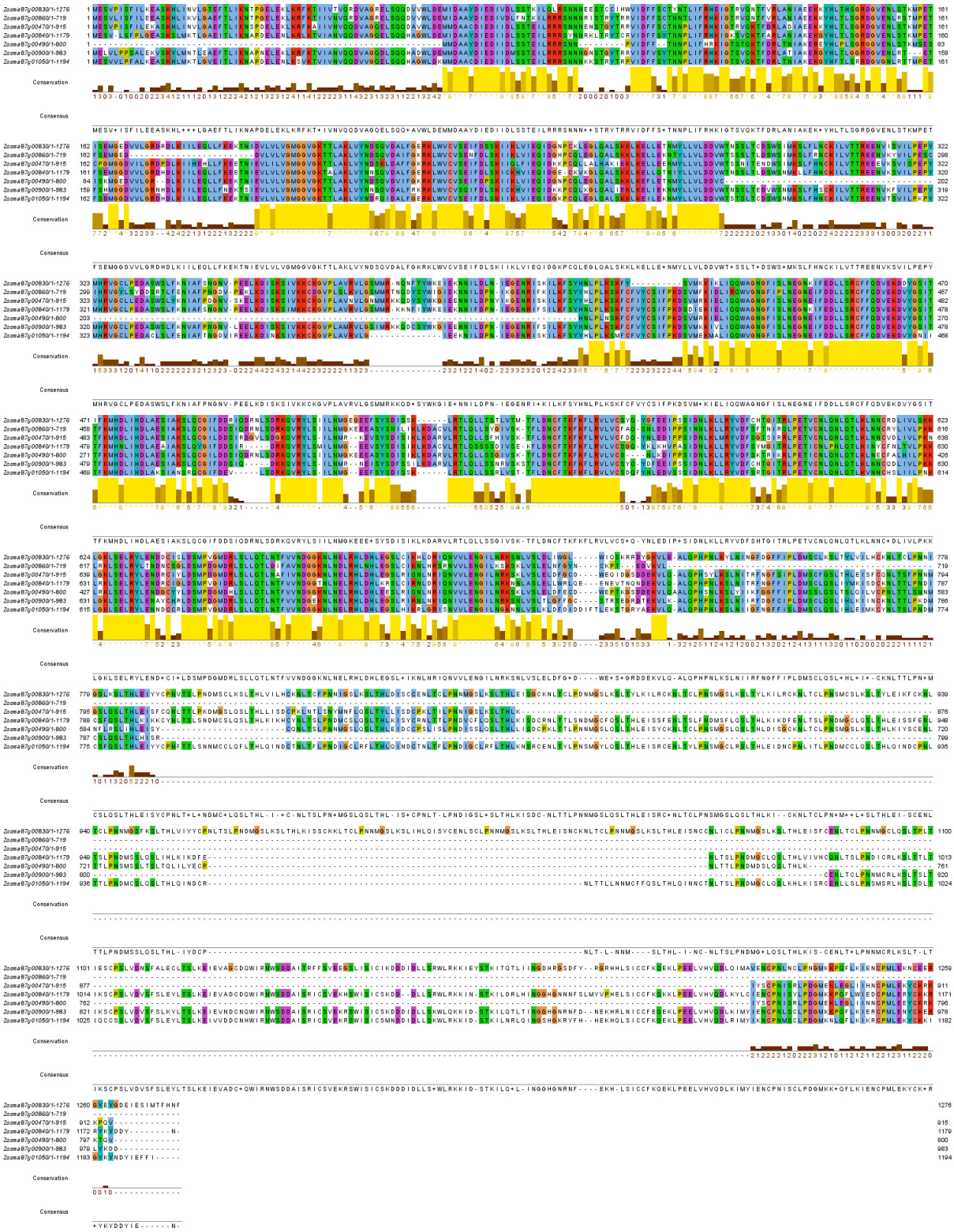
Though speculative, we propose three non-mutually exclusive hypotheses to account for the observed reduction in plant defense of this type.

First, the selective pressure on anti-herbivory defenses may be somewhat relaxed and/or involves other pathways than those associated with terrestrial plants. Direct consumption of *Z. marina* leaf tissue certainly occurs by small crustaceans (e.g., amphipods, isopods and crabs) and snails (who may trigger currently uncharacterized defense mechanisms), but in general, herbivore mesograzers preferentially graze on leaf-rich epiphytes rather than on the leaves themselves (Duffy 2006). In some geographic regions, where meadows occur in very shallow water (or that fall dry at low tide) they may be periodically subjected to geese grazing of the protein-rich rhizomes, but again, not the leaves.

Second, marine pathogens associated with *Z. marina* have scarcely been studied with the exception of the protist symbiont/mutalist/pathogen, *Labyrinthula zosterae*, which causes the "wasting disease" when present in large amounts; but which also modulates defense gene expression, possibly through bacteria (Brakel *et al.* 2014). In any case, the strongly divergent R-gene repertoire suggests that *Z. marina* is exposed to a very different set of plant pathogens as compared with terrestrial ancestors and also *Spirodela* (keeping in mind that *Spirodela* is a surface dweller).

Third, secondary metabolites in land plants are also involved in the attraction of pollinators, nitrogen-fixing bacteria and mycorrhizal fungi. In *Zostera,* however, submarine pollination occurs directly via water dispersal and pollinators are unknown (but see (Van Tussenbroek *et al.* 2012). Mycorrhizal fungi are also unknown even though Nitrogen-fixing cyanobacteria associated with the rhizosphere are well-documented (Welsh 2000). The precise mutualisitic association of the *Z. marina* microbiome is currently under investigation (http://seagrassmicrobiome.org/).

R-genes and NBS domain



Supplementary Fig. S6.5 | Alignment of *Z. marina* R-genes showing conservation over NBS domains.

Conserved regions are shown in brown and yellow below the alignment. Zoom for greater clarity.

7 Light

7.1 Photosynthesis/Photosensory/Carbon fixation

Seagrasses grow from the surface to as deep as 50 m in clear waters, including some species of *Zostera*. Irradiance decreases exponentially with depth and spectral quality is rapidly altered, shifting from the sunlight spectrum into a narrow band of blue-green light (Kirk 2011) within the first 1-2 m. Seawater strongly absorbs light in the red and infrared wavelengths. The absorption increases towards UV causing a dominance of blue-green (400–500 nm) spectral components in water as little as a meter deep. Nonetheless, despite the attenuation of the red and infrared light, these wavelengths are still present in the sea at low intensities.

Photosynthesis. In general, photosystem I and II and carbon-fixation genes of *Z. marina* are not significantly different from *S. polyrhiza* and other angiosperms (**Supplementary Table S7.1)**. Differential expression is discussed in **Supplementary Table S7.2** (see Carbon Fixation below).

Supplementary Table S7.1 | Photosynthesis/photosensory related genes *Z. marina*.

Spipo-prefixes are from Phytosome v.10.1, where a mix of nuclear and plastid genomes are included(Wang *et al.* 2014). Gi-prefixes are from GenBank, where the chloroplast genome is deposited (Wang & Messing 2011).

		Zostera marina		*Spirodela polyrhiza*	
Main process	**Protein name**	**Locus**	**Genome**	**Locus**	**Genome**
PSI	PsaO	Zosma2g03250	nuclear	Spipo9G0036800	n
	PsaG	Zosma112g00170	n	Spipo10G0029800	n
	PsaK	Zosma33g01310	n	Spipo8G0026800	n
	PasF	Zosma12g00570	n	Spipo13G0051600	n
		Zosma10g00340	n	Spipo8G0021700	n
	PsaH	Zosma5g00100	n	Spipo14G0018600	n
	PsaL	Zosma288g00080	n	Spipo4G0050200	n
	PsaN	Zosma3g00680	n	Spipo4G0095100	n
	PsaE	Zosma35g00710	n	Spipo11G0059900	n
				Spipo1G0013400	n
	PsaC	ZosmaCg00080	plastid	gi|341834121	p
	PsaD	Zosma35g00140	n	Spipo9G0011900	n
	PsaA	ZosmaCg00620	p	gi|341834071	p
				Spipo0G0152800	n
	PsaB	ZosmaCg00610	p	gi|341834070	p
				Spipo0G0152800	n
PSII	psbA	ZosmaCg00300	p	gi|341834051	p
				Spipo0G0140600	n
	psbB	ZosmaCg00940	p	gi|341834095	p
				Spipo0G0110400	n
	psbC	ZosmaCg00550	p	gi|341834067	p
				Spipo11G0005500	n
	psbD	ZosmaCg00540	p	gi|341834066	p

				Spipo0G0140600	n
	psbE	ZosmaCg00830	p	gi|341834088	p
				Spipo0G0132200	n
	psbF	ZosmaCg00820	p	gi|672374533	p
				Spipo0G0132300	n
	psbH	ZosmaCg00970	p	gi|341834098	p
				Spipo0G0186500	n
	psbI	ZosmaCg00360	p	gi|341834055	p
	psbJ	ZosmaCg00800	p	gi|672374532	p
	psbK	ZosmaCg00350	p	gi|341834054	p
	psbL	ZosmaCg00810	p	gi|341834086	p
	psbM	ZosmaCg00490	p	gi|341834065	p
	psbO	Zosma116g00700	n	Spipo3G0085000	n
	psbP	Zosma85g01220	n	Spipo8G0024400	n
		Zosma3g00940	n	Spipo4G0098600	n
		Zosma30g01090	n	-	
	psbT	ZosmaCg00950	p	gi|341834096	p
				Spipo0G0110500	n
	psbW	Zosma13g00760	n	Spipo9G0001000	n
				Spipo2G0058500	n
	psbY	Zosma72g00150	n	Spipo7G0062500	n
				Spipo15G0040400	n
	psbZ	ZosmaCg00570	p	gi|672374528	p
				Spipo0G0153000	n
	psbR	Zosma330g00140	n	Spipo32G0009500	n
	psbS	Zosma60g00630	n	Spipo13G0010400	n
				Spipo4G0115200	n
	pbsQ-llike	Zosma19g01210	n	Spipo7G0011400	n
	psbQ	Zosma162g00390	n	Spipo11G0048600	n
	psb27	Zosma43g00480	n	Spipo0G0061700	n
	psb28	Zosma37g00220	n	Spipo1G0013200	n
Cytochrome b6-f complex	petA	ZosmaCg00790	p	gi|341834084	p
				Spipo0G0109700	n
	petB	ZosmaCg00980	p	gi|341834099	p
	petD	ZosmaCg00990	p	gi|341834100	p
				Spipo0G0145200	n
	petG	ZosmaCg00850	p	gi|341834090	p
				Spipo0G0109900	n
	petL	ZosmaCg00840	p	gi|672374534	p
				Spipo0G0109800	n
	petM	Zosma33g00340	n	Spipo3G0112500	n
		-	-	Spipo2G0118900	n
	petN	ZosmaCg00480	p	gi|341834064	p
				Spipo0G0140500	n

	petP	-		-	
LHC superfamily	one-helix protein	Zosma13g01140	n		
		Zosma31g00600	n	Spipo21G0015500	n
				Spipo0G0048100	n
	ferrochelatase	Zosma172g00410	n	Spipo9G0038500	n
		Zosma87g00080	n	Spipo4G0076800	n
		Zosma34g00670	n		
	stress-enhanced protein	Zosma101g00320	n	Spipo12G0003300	n
		Zosma25g00660	n	Spipo2G0040100	n
	LIL3	Zosma106g00570	n	Spipo11G0014000	n
		Zosma361g00150	n		
	ELIP	Zosma152g00070	n	Spipo2G0018300	n
		Zosma48g00340	n		
	LHCA1	Zosma173g00090	n	Spipo22G0035800	n
	LHCA2	Zosma25g01660	n	Spipo4G0053500	n
	LHCA3	Zosma31g00760	n	Spipo10G0029400	n
	LHCA4	Zosma10g00190	n	Spipo2G0037000	n
	LHCA5	Zosma82g00850	n	Spipo2G0019400	n
	LHCA6	Zosma121g00630	n	Spipo0G0031300	n
	LHCB1	Zosma122g00170	n	Spipo3G0061200	n
		Zosma137g00310_20	n	Spipo19G0027400	n
		Zosma152g00200	n		
		Zosma122g00210	n		
		Zosma122g00230	n		
		Zosma59g00610	n		
		Zosma217g00280	n		
		Zosma152g00140	n		
		Zosma655g00010	n		
		Zosma122g00160	n		
	LHCB2	Zosma151g00070	n	Spipo3G0099600	n
	LHCB3	Zosma278g00220	n		
	LHCB4	Zosma137g00210	n	Spipo22G0045300	n
	LHCB5	Zosma83g00260	n	Spipo4G0080300	n
	LHCB6	Zosma89g01210	n	Spipo4G0013000	n
	LHCB7	Zosma186g00090	n	Spipo10G0017000	n
	PSBS	Zosma60g00630	n	Spipo13G0010400	n
				Spipo4G0115200	n
Chlorophyll synthesis	Chlorophyll a oxygenase (COA)	Zosma65g00150	n	Spipo9G0042000	n
				Spipo13G0010100	n
	Pheophorbide a oxygenase (PAO)	Zosma41g00880	n	Spipo1G0020000	n
				Spipo2G0064300	n
	Red chlorophyll catabolite reductase (RCCR)	Zosma4g00150	n	Spipo15G0020100	n

	Mg-chelatase subunit chlD	Zosma8g01570	n	Spipo3G0066700	n
				Spipo26G0021100	n
	Mg-protoporphyrin IX monomethyl ester cyclase	Zosma132g00650	n	Spipo3G0020600	n
	Chlorophyll a synthase (ChlG)	Zosma3g01180	n	Spipo23G0037100	n
		Zosma193g00090	n	Spipo23G0037100	n
	Mg-chelatase subunit chlH	Zosma64g00440	n	Spipo22G0000400	n
	Mg-chelatase subunit chlI	Zosma78g00070	n	Spipo26G0021100	n
	Mg- protoporphyrin IX methyltransferase	Zosma1g02580	n	Spipo11G0044700	n
	Glutamyl-tRNA synthetase	Zosma27g01290	n	Spipo10G0019500	n
	Glutamyl-tRNA reductase (HEMA)	Zosma40g00080	n	Spipo0G0023900	n
		Zosma203g00120	n	Spipo2G0126400	n
	Porphobilinogen synthase (HEMB)	Zosma78g00470	n	Spipo12G0014300	n
	Porphobilinogen deaminase (HEMC)	Zosma175g00020	n	Spipo5G0040400	n
	Urophorphyrin III synthase (HEMD)	Zosma25g01150	n	Spipo2G0046300	n
	Uroporphyrinogen decarboxylase (HEME)	Zosma2g00400	n	Spipo22G0044200	n
		Zosma84g00180	n	Spipo22G0027100	n
				Spipo29G0014200	n
	Coproporphyrinogen III oxidase (HEMF)	Zosma69g00410	n	Spipo5G0045800	n
	Protoporphyrinogen oxidase (HEMG)	Zosma29g01750	n	Spipo6G0014700	n
		Zosma166g00720	n	Spipo13G0021300	n
	Ferrochelatase (HEMH)	Zosma172g00410	n	Spipo9G0038500	n
		Zosma34g00670	n	Spipo4G0076800	n
		Zosma87g00080	n		
	Glutamate-semialdehyde aminotransferase (HEML)	Zosma89g00500_10	n	Spipo13G0023500	n
				Spipo4G0058400	n
	Heme oxygenase (HMOX)	Zosma219g00130	n	Spipo21G0001300	n
		Zosma25g01200	n	Spipo15G0018700	n
	NADPH-protochlorophyllide oxidoreductase¨(PORA)	Zosma15g01540	n	Spipo1G0039600	n
		Zosma53g00550	n	Spipo16G0033900	n
	Short-chain dehydrogenase/reductase (SDR)	Zosma408g00090	n	Spipo22G0006400	n
				Spipo22G0006000	n
				Spipo22G0006100	n
		Zosma21g00370	n	Spipo0G0000900	n
				Spipo0G0000400	n
				Spipo0G0000500	n
	protoheme IX farnesyltransferase	Zosma56g00600	n	Spipo0G0062200	n

	Uroporphyrin III methylase (HEMD)	Zosma126g00040	n	Spipo21G0042500	n
		Zosma209g00380	n		
		Zosma151g00230	n		
		Zosma1g01740	n		
Carotenoid biosynthesis	IPP isomerase (IPP)	Zosma29g00510	n	Spipo9G0066500	n
		Zosma74g00710	n		
	GGPP synthetase	Zosma86g00380	n		
	phytoene synthase (PSY)	Zosma16g00900	n	Spipo1G0051600	n
				Spipo8G0020300	n
	zeta-carotene desaturase	Zosma16g01160	n	Spipo2G0067200	n
	lycopene epsilon cyclase	Zosma70g00210	n	Spipo21G0023300	n
	lycopene beta cyclase	Zosma70g00580	n	Spipo12G0018200	n
	zeaxanthin epoxidase	Zosma82g00950	n	Spipo4G0008800	n
	violaxanthin deepoxidase	Zosma77g00860	n	Spipo7G0045700	n
	beta carotene ketolase	-			
	epoxycarotenoid cleavage enzyme	Zosma155g00260	n	Spipo5G0039400	n
		Zosma426g00050	n	Spipo8G0028800	n
		Zosma31g00080	n	Spipo8G0028900	n
		Zosma80g00490	n	Spipo13G0042400	n
		Zosma52g00590	n	Spipo11G0059000	n
				Spipo1G0017600	n
				Spipo30G0015100	n
				Spipo3G0097000	n
				Spipo15G0002100	n
				Spipo0G0117200	n
				Spipo0G0117500	n
Carbon fixation (See **Supplementary Table 7.2** for gene expression of carbon fixation genes)	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	ZosmaCg00710	p	gi|341834079	p
	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Zosma15g00370	n	Spipo0G0168300	n
		Zosma99g00830	n	Spipo14G0042700	n
		Zosma36g00370	n	Spipo10G0010300	n
	Phosphoenolpyruvate carboxylase	Zosma11g00980	n	Spipo19G0008900	n
		Zosma160g00340	n	Spipo19G0008900	n
		Zosma251g00350	n	Spipo19G0008900	n
		Zosma7g00770	n	Spipo19G0008900	n
		Zosma99g00660	n	Spipo18G0020200	n
	Carbonic anhydrase	Zosma171g00480	n	Spipo7G0035900	n

		Zosma216g00330	n	Spipo7G0035900	n
		Zosma366g00170	n	Spipo7G0035900	n
		Zosma64g00760	n	Spipo7G0035900	n
		Zosma1159g00010	n	Spipo7G0035900	n
		Zosma3g01540	n	Spipo7G0035900	n
		Zosma149g00090	n	Spipo4G0017000	n
		Zosma155g00330	n	Spipo14G0033900	n
		Zosma149g00090	n	Spipo4G0017000	n
	R/FR light photoreceptors				
Photoreceptors	Phytochrome A	Zosma79g00550	n	Spipo6G0014200	n
	Phytochrome B	Zosma216g00050	n	Spipo6G0031800	n
	Phytochrome C	-	-	Spipo2G0079900	n
	UV-A/blue photoreceptors				
	Cryptochrome 1	Zosma23g01480	n	Spipo15G0011900	n
	Cryptochrome 2	-	-	Spipo1G0003600	n
	Cryptochrome 5	-	-	Spipo21G0005400	n
	Cryptochrome DASH	-	-	Spipo12G0060800	n
	Phototropins			Spipo6G0048400	n
	phot1	Zosma114g00500	n	-	-
	phot2	Zosma14g01450	n	-	-
	LOV/F-box/Kelch- domain proteins				
	ZTL	Zosma25g00320	n	Spipo12G0055500	n
	Kelch- domain proteins	-	-	Spipo21G0020500	n
	UV-B				
	UV RESISTANCE LOCUS8	-	-	Spipo4G0097900	-
PHOTORECEPTORS RELATED PROTEIN FAMILY					
DNA Photolyase	Photolyase	Zosma22g01440	n	-	n
	DNA photolyase	Zosma66g00230	n	-	n
	Deoxyribodipyrimidine photolyase-like	Zosma158g00230	n	Spipo1G0090400	n
UV B RESPONSE	COP1	Zosma69g01020	n	Spipo31G0000500	n
		Zosma223g00310	n		
	HY5	Zosma4g01690	n	Spipo3G0080800	n
				Spipo6G0075600	n

Supplementary Table S7.2 | Summary of Carbon fixation genes including differential gene expression as normalized expression values (FPKM) in leaves, roots and flowers *Z. marina*.

	N° loci *Z. marina*	N° loci *Spiro-dela Poly-rhiza*	*Z. marina* locus	Leaf	Root	Femal e flow-er early	Femal e flow-er late	Male flowe r	Spirodela best match
RUBISCO Large subunit	1	1	**rbcL**						
			ZosmaC1g00620	7.67	-	9.15	11.08	1.91	AEK94350.1 (*)
RUBISCO Small subunit	3	4	**rbcS**						
			Zosma99g00830	23815	-	292.9	381	41.81	Spipo14G0042700
			Zosma36g00370	7.77	8.73	146.34	53.99	10.44	Spipo10G0010300
			Zosma15g00370	3289.4	-	104.9	87.20	very low	Spipo14G0042700
PHOSPHO ENOL PYRUVATE CARBOXYLASE	4	3	**PEPC**						
			Zosma11g00980	56.11	158.0	69.67	39.11	38.90	Spipo19G0008900
			Zosma160g00340	-	0.79	0.96	0.70	71.17	Spipo19G0008900
			Zosma251g00350	78.55	5.58	16.15	19.69	20.32	Spipo19G0008900
			Zosma7g00770	93.22	117.4	189.84	228.51	104.0	Spipo19G0008900
			Zosma99g00660	51.06	28.74	48.47	53.98	3.09	Spipo18G0020200
CARBONIC ANHYDRASE	9	3	**CA**						
			Zosma171g00480	94.71	74.83	3.02	5.64	16.77	Spipo7G0035900
			Zosma216g00330	-	-	137.23	954.72	12	Spipo7G0035900
			Zosma366g00170	851.05	-	-	-	-	Spipo7G0035900
			Zosma64g00760	-	-	0.44	-	142.3	Spipo7G0035900
			Zosma1159g00010	1.01	19.88	-	-	5.47	Spipo7G0035900
			Zosma3g01540	16.30	0.59	4.81	5.85	0.51	Spipo7G0035900
			Zosma149g00090	63.84	45.82	37.09	27.06	43.49	Spipo4G0017000
			Zosma155g00330	173.37	135.4	51.86	40.99	380.3	Spipo14G0033900
			Zosma149g00090	63.84	45.82	37.09	27.06	43.49	Spipo4G0017000
BORON TRANSPORTER	2	3	**HC03 Transport**						
			Zosma17g00880	66.17	68.14	13.95	15.61	9.36	Spipo27G0018500
			Zosma2g01810	14.40	90.33	1085.0	71.57	50.24	Spipo27G0011500

The *Z. marina* nuclear and chloroplast genomes encode the complete set of proteins necessary to constitute Photosystem (PSI), Photosystem II (PSII), and the Cytochrome b6-f complex **(Supplementary Table S7.1).** Genes encoding PSI (PsaA-PsaO) are mainly nuclear with the exception of PsaA-PsaC, while genes encoding PSII (PsbA-PsbZ) and the Cytochrome b6-f complex (PetA-PetN) are situated in the chloroplast genome, with the exception of PsbP, PsbR, PsbS, PsbY, PsbW and PetM. This situation is highly similar to that found in *S. polyrhiza* and terrestrial plants.

Photosystems I and II are complemented by a large set of antenna/light harvesting proteins which transfer light energy the chlorophyll molecules in the reaction centres, but may also be important for non-photochemical quenching (NPQ). The light-harvesting complex (LHCB) (Engelken *et al.* 2010) superfamily is expanded in *Z. marina* (see below) in accordance with high levels of light attenuation in coastal waters. Efficient regulation of light harvesting complexes and consequentially the expansion of related gene families have been proposed as an adaptation to the marine coastal environment (Cock *et al.* 2010). The origin of the LHC superfamily in eukaryotes is thought to be a cyanobacterial high light induced (HLIP)-like protein, which underwent several duplications, leading to a wide range of LHC superfamily proteins with different numbers of CB domains in extant eukaryotes.

Z. marina encodes two single-helix proteins (OHPs, Zosma13g01140 and Zosma31g00600), i.e. the eukaryote homolog of cyanobacterial HLIPs, and three ferrochelatases (Zosma172g00410, Zosma34g00670, and Zosma87g00080), two of which contain a CB domain. Like *A. thaliana*, *Z. marina* also possesses two stress-enhanced proteins (SEPs; Zosma101g00320 and Zosma25g00660) and two Light harvesting-like 3 proteins (LIL3s; Zosma106g00570 and Zosma361g00150), each containing two transmembrane helices. Two early light-induced proteins (ELIPs) and 22 LHC proteins were found, six likely to be associated with PSI (LHCA), and 16 likely to be associated with PSII (LHCB), each containing three transmembrane helices.

Ten of the 16 LHCB proteins belonged to the LHCB1 group, more than in most other plants (**Extended Data Figure 4, Supplementary Tables S7.1** and **S7.3;**). This is the result of an expansion of the LHCB1 sub-family that took place after the separation of the *Zosteraceae* and the *Araceae,* the latter represented by *S. polyrhiza.* In contrast, no members of the LHCF, LHCR, LHCX, or LHCZ families were found, as typical also for terrestrial plants. Lastly, the *Z. marina* genome encodes a single PSBS protein (Zosma60g00630) with four transmembrane helices. Some members of the LHCB1 family may be analogous to LHCX proteins in brown algae and LHCR proteins and OHPs in red algae and have acquired alternative functions, e.g., in photoprotection and NPQ. Although this hypotheses remains to be validated experimentally, at least one of the *Z. marina* LHCB1 proteins was found to be overexpressed in response to a heat wave treatment (Franssen *et al.* 2011) supporting a role of these proteins in stress response.

Supplementary Table S7.3 | Number of LHCA/B proteins encoded by different plant genomes and proteins within the LHCB1 family.

Species	LHCA/B	LHCB1
Zostera marina	22	10
Spirodela polyrhiza	14	2
Chamydomonas reinhardii	22 (+3 LHCR)	*7*
Amborella trichopoda	21	6
Vitis vinifera	18	5
Arabidopsis thaliana	21	5
Thellungiella halophila	25	5
Populus trichocarpa	32	5
Setaria italica	19	*4*
Oryza sativa	24	3
Setaria italica	19	*4*
Selaginella moellendorffi	12	1

Photosensory genes. Photoreceptors are proteins specialized for sensing and responding to the light environment. These proteins are able to perceive a broad range of light qualities and intensities. Plant photoreceptors include phytochromes, cryptochromes/photolyases, phototropins, and ultraviolet (UV)-B photoreceptors.

In flowering plants, up to twelve photoreceptors are present: five red/far-red light receptors (phytochromes) and seven blue light receptors (two cryptochromes (CRY1 and CRY2), two phototropins (phot1 and phot2) and three LOV/F-box/Kelch- domain proteins (ZTL, FKF, and LKP2)). In accordance with the drasically different light regime in marine waters, where both red/far red and UV are attenuated disporportionally strong, phytochromes and cryptochromes are reduced in *Z. marina* as compared with *S.polyrhiza* (**Supplementary Table S7.1**). In *Z. marina*, only two phytochromes genes are present: Zosma79g00550 encoding for PHY A, Zosma216g00050 encoding for PHY B, whereas it lacks of the PHY C gene. This may be significant in *Zostera* as PHYC is linked to photoperiodic control of flowering in most plants, whereas flowering seems to be more controlled by temperature in *Zostera*. Among cryptochromes, only one gene is present (CRY1), while CRY2 is missing.

The UV resistance locus 8 (UVR8) associated with UV photoprotection (Binkert *et al.* 2014) is entirely missing in *Z. marina* although associated genes COP1 (for which there are two putative homologues: Zosma69g01020 and Zosma223g00310) and the bZIP transcriptional factor HY5 (locus Zosma4g01690) are

present. An ortholog of the UVR8 gene is found in *Spirodela* (Spipo4G0097900). Therefore, the loss of the UVR8 cannot be seen as an Alismatales-specific feature but as unique to *Zostera* and may represent its adaptation to marine light, where UV wavelengths are attenuated.

Carbon fixation. Seagrass photosynthesis is of the C3 type (Larkum *et al.* 2006a). Carbonic anhydrases (CA) are part of the carbon concentrating mechanism that takes up bicarbonate from the dissociated CO_2 in seawater. *Z. marina* has nine CA as compared with three in *S. polyrhiza* (**Supplementary Table S7.2)**. Comparative gene expression data across tissues shows that this family has a strong stage/tissue-specific activation pattern **(Supplementary Table S7.2)**. Plants which employ the C3 pathway use ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) as the primary carboxylating enzyme (in chloroplasts), which usually consists of two types of protein subunits, the large chain (RBCL) encoded by a single gene in the chloroplast, and the small chain (rbcL) for which there are typically several related small-chain genes. In *Zostera*, the large-chain (rbcL) is encoded by the locus ZosmaCg00710 (**Supplementary Table S7.1**). The number of genes comprising the rbcS multigene family varies considerably among species. In *Zostera* the rbcS multigene family is composed by three members: Zosma15g00370, Zosma99g00830 and Zosma36g00370 (**Supplementary Table S7.1**). Zosma15g00370 and Zosma99g00830 are mainly expressed in the leaf tissue (**Supplementary Table S7.1**) and less so in flower tissues, while Zosma36g00370 is particularly expressed in female early flowers and, interestingly, it is also expressed in roots (**Supplementary Table S7.2**). Phosphoenolpyruvate carboxylase (PEPC, EC4.1.1.31) is an ubiquitous enzyme in plants that catalyzes the irreversible carboxylation of PEP to form oxaloacetate with Mg^{2+} or Mn^{2+} as essential cofactors. The form of inorganic carbon in this reaction is bicarbonate. There are four different genes encoding for PEP carboxylase in *Zostera* (**Supplementary Table S7.2**); as for rbcS, these genes showed stage specific expression patterns.

7.2 Carbohydrate storage

Z. marina possesses all the genes involved in the metabolism of **sucrose** (GT4 sucrose synthase(SuSy), and sucrose-phosphate synthases (SPS), GH36 raffinose synthase, GH32 and GH100 invertases (IVR)); of **trehalose** (GT20 trehalose synthase, GH37 trehalase); and of **starch** (GT5 starch synthase, GT35 starch phosphorylase, GH13 [branching enzyme, debranching enzyme, and alpha-amylases], GH14 beta-amylase, GH31 alpha-glucosidase, GH77 alpha-1,4-glucanotransferase, phosphoglucan, water dikinase). The number of sucrose- and trehalose-related genes is similar to that of *Arabidopsis* and *Oryza*; although sucrose transport proteins (SUT) genes are elevated, as would be expected. In contrast, the **starch**-related genes are reduced, especially in comparison with *Oryza*. *Z. marina* seems to have a minimal set of biosynthetic genes: 6 GT5, 2 branching enzymes (subfamily GH13_8), and 1 GH77 (versus 11 GT5, 5 GH13_8 and 2 GH77 in *Oryza*). This is also the case for starch recycling: only 1 GT35, 8 GH13, 6 GH14 (versus 2 GT35, 20 GH13 and 11 GH14 in *Oryza*). Given the reliance on sucrose, the most spectacular reduction is in the paucity of starch binding modules (1 CBM20, 2 CBM48, 1 CBM53, and even the absence of CBM45). In contrast, *Oryza* displays 6 CBM20, 4 CBM45, 16 CBM48, and 6 CBM53). See **Supplementary Data 8, Supplementary Table S7.4.**

Supplementary Table S7.4 | CAZyme expression comparisons between *Z. marina, A. thaliana* and *O. sativa.*

Glycoside Hydrolases (GH), Glycosyltransferases (GT), Polysaccharide Lyases (PL), Carbohydrate esterases (CE) and Carbohydrate Binding Molecules (CBM). Red indicates gene family expansion, Yellow indicates no change of gene family size and Blue indicates gene family size contraction.

	Total GH	Total GH families	GH1	GH2	GH3	GH5	GH9	GH10	GH13	GH14	GH16	GH17	GH18	GH19	GH20	GH27	GH28	GH29	GH31	GH32	GH33	GH35	GH36	GH37
Z. marina	237	33	4	2	9	7	23	3	8	6	15	40	4	3	3	2	53	2	6	5	0	13	3	1
A. thaliana	399	35	48	2	16	13	26	12	10	9	33	51	11	14	3	4	68	1	5	8	1	18	6	1
O. sativa	429	35	35	2	16	19	25	11	20	11	31	66	34	17	5	5	41	2	7	12	1	15	7	1

	GH38	GH43	GH47	GH51	GH63	GH77	GH79	GH81	GH85	GH89	GH95	GH100	GH116
Z. marina	2	2	4	1	1	1	1	1	1	0	1	7	3
A. thaliana	4	2	5	2	2	2	3	2	2	1	1	9	4
O. sativa	4	2	4	8	1	2	5	1	2	1	1	8	7

	Total GT	Total GT families	GT1	GT2	GT4	GT5	GT8	GT9	GT10	GT13	GT14	GT16	GT17	GT19	GT20	GT21	GT22	GT24	GT28	GT29	GT30	GT31	GT32	GT33
Z. marina	339	44	50	36	26	6	39	1	2	1	12	1	4	1	8	1	4	1	2	4	1	25	2	1
A. thaliana	468	44	122	42	24	6	42	1	3	1	11	1	7	1	11	1	3	1	4	3	1	33	6	1
O. sativa	577	45	202	47	25	11	39	1	2	1	12	1	4	1	11	1	4	1	4	5	1	40	2	2

	GT34	GT35	GT37	GT41	GT43	GT47	GT48	GT50	GT57	GT58	GT59	GT61	GT64	GT65	GT66	GT68	GT75	GT76	GT77	GT90	GT92	GT95	GT96
Z. marina	5	1	2	1	4	36	10	1	1	1	1	13	4	0	3	1	3	1	12	4	2	4	1
A. thaliana	8	2	10	2	4	39	13	1	2	1	1	8	3	0	2	3	5	1	19	9	5	3	2
O. sativa	6	2	18	3	10	35	11	1	2	1	1	25	3	1	2	1	3	1	16	7	5	4	2

	Total PL	Total PL families	PL1	PL4
Z. marina	16	1	16	0
A. thaliana	34	2	26	8
O. sativa	16	2	12	4

	Total CE	Total CE families	CE6	CE8	CE11	CE13
Z. marina	75	4	1	63	1	10
A. thaliana	85	4	2	67	4	12
O. sativa	53	4	4	38	1	10

	Total CBM	Total CBM families	CBM18	CBM20	CBM22	CBM43	CBM45	CBM48	CBM49	CBM50	CBM53
Z. marina	49	8	1	1	3	36	0	2	4	1	1
A. thaliana	122	9	10	4	17	62	6	16	3	1	3
O. sativa	122	9	15	6	12	52	4	16	4	7	6

Sucrose accounts for the largest portion (up to 90%) of carbon fixed photosynthetically by seagrass leaves, followed by glucose and fructose, reviewed in (Larkum *et al.* 2006a); starch can also represent considerable storage but is generally less (Burke *et al.* 1996; Vermaat & Verhagen 1996).

Sucrose can be synthesized in photosynthetic or storage cells by the sequential action of the two key enzymes sucrose phosphate synthase (SPS) and sucrose phosphate phosphatase (SPP). A third enzyme, sucrose synthase (SuSy) can reversibly interconvert sucrose and the cell wall precursor UDP-glucose, but generally acts in sucrose degradation (Wang *et al.* 2013). UDP-glucose is a nucleotide sugar central to polysaccharide biosynthesis pathways (starch, cellulose and monosaccharides for cell wall polymers). Invertases hydrolyze sucrose into hexose sugars and are thought to be important regulators of sucrose levels. Three main kinds exist: insoluble acidic cell-wall invertases, soluble acidic vacuolar invertases and neutral cytoplasmic invertases. Sucrose transport proteins (SUT) are one class of proteins responsible for loading sucrose from the apoplastic space into the phloem. The flux of sucrose from source cells, where it is produced into the apoplastic space seems to be mediated by a second family of SWEET sugar transporters. Only certain SWEET proteins transport sucrose, however, 7 of 17 in *A. thaliana* (AtSWEET10 to 15), and 5 of 21 in *O. sativa* (OsSWEET11 and 14)(Chen *et al.* 2012).

The *Z. marina* genome was searched for enzymes involved in sucrose synthesis and interconversion, invertases and sugar transporters using textual search and functional annotations. Protein sequences from the complete genomes of *Arabidopsis*, *Oryza*, *Vitis*, *Brachypodium, Musa* and *Spirodela* were obtained from Uniprot (Blast and text searches on annotation terms). Protein alignments were made using MUSCLE, and ML trees were built using PhyML (LG model, empirical amino-acid frequencies, invariable sites optimized, NNI tree-searching and aLRT branch support estimates, not shown). As anticipated, *Z. marina* has an expanded set of SuSy and SUT genes (**Supplementary Table S7.5**).

Supplementary Table S7.5 | Comparison of selected sucrose synthesis, interconversion and transportrelated genes with other plant genomes.

SPS=Sucrose phosphate synthase, SuSy=Sucrose synthase, SPP=Sucrose phosphate phosphatase, SUT=Sucrose transport protein, SWEET=sugar transporter, IVR=invertase (CW=insoluble cell wall, V=vacuolar, CP=soluble cytoplasmic forms, respectively).

Gene		**Number of homologs**					
	Zostera	*Spirodela*	*Brachypod.*	*Oryza*	*Musa*	*Vitas*	*Arab*
SPS	**6(5)***	**4**	**5**	**5**	**4**	**6**	**4**
SuSy	**9**	**3**	**7**	**7**	**7**	**4**	**6**
SPP	**1****	**1**	**3**	**3**	**2**	**2**	**4**
SUT	**9(7)#**	**2**	**4**	**5**	**3**	**8**	**9**
SWEET	**8**	**7**	**6**	**21**	**11**	**8**	**17**
IVR-CW	**3**	**?**	**7**	**6**	**4**	**5**	**6**
IVR-V	**2**	**?**	**4**	**4**	**3**	**4**	**2**
IVR-CP	**7(5)$**	**?**	**9**	**12**	**11**	**14**	**13**

* One truncated possible pseudogene with poor EST support
**Truncated gene (missing c-terminal region but gene model good including EST support)
#Two truncated genes
$Two truncated genes or possible pseudogene

8 Osmoregulation, ion homeostasis/sequestration, stress related traits

8.1 CAT, SOD, GPX, NHX, CHX, KEA, LEA, AHA/PM ATPases

$Na(^{+})/H(^{+})$ exchangers (NHX) family (**Supplementary Table S8.1**) are described as AtNHX1-8 in *Arabidopsis* and are involved in pH regulation and salt tolerance. There are five members of the NHX family in *Z. marina*, distributed among the three main clades: 1) the SOS (salt overly sensitive) family, and corresponding to *Arabidopsis* AtNHX7/8, 2) AtNHX5/6 localized to the Golgi/endosome, and 3) AtNHX1-4 localized to the tonoplast (Chanroj *et al.* 2012). *Z. marina* contains one SOS-like member that clusters with two *Cymodocea nodosum* homologues**.** These are described in *Arabidopsis* as plasma-membrane localized, widely distributed in the plant, and responsive to salt stress. A single NHX5/6 homologue is present. These are Golgi/endosome localized, and are also implicated in responses to salt stress in tomato (Rodriguez-Rosales *et al.* 2008) and *Arabidopsis*(Bassil *et al.* 2011). The vacuolar NHX1-4 family was most represented, with three genes. Based on entire protein sequences, *Z. marina* genes are consistently divergent from each other and basal on the tree (not shown).

Five $K(^{+})$ efflux antiporters (KEA) genes were identified in *Z. marina* with no evidence of expansion. Monocots typically have 16 cation/$H(^{+})$ exchangers (CHX) genes on average and *Z. marina* has 15. *Arabidopsis* has 28, most of which are expressed preferentially or specifically in pollen. There was no evidence for expansion and most of the genes are likely involved in K^{+} transport and homeostasis, apparently most associated with pollen **(Supplementary Table S8.1**).

Plasma membrane, H^{+}-ATPase (AHA) genes vary substantially between *Z. marina* and *Spirodela polyrhiza* as compared with other monocots and dicots (**Supplementary Table S8.2, Supplementary Fig. S8.1, Supplementary Data 7**). A cluster containing both dicot and monocot AHAs is missing in both *Spirodela* and *Zostera*. In contrast, a novel AHA cluster is found exclusively in *Zostera* and *Spriodela.* Within this cluster, two AHAs are interesting in that they have lost the typical auto-regulation of H^{+}-ATPases due to the missing C-terminal extension, which is responsible for the auto-regulation (i.e., their activity is now constitutive and no longer phosphorylation-regulated). In this cluster too, there is expansion of the *Zostera* AHAs in three members compared to one in *Spirodela* (green triangle in **Supplementary Fig. 8.1**). It is further noted that *Zostera* is missing an ortholog for AtAHA10, whereas *Spirodela* is not. Tissue expression is also noteworthy. The ZHA1 gene (Zosma294g00130) reported by Muramatsu (Muramatsu *et al.* 2002) as producing a salt-tolerant enzymatic activity, is indeed the most strongly expressed ZmAHA in vegetative tissue (leaf RPKM: 509). Further, three of these genes are highly and specifically expressed in male flower.

Because seagrasses have adapted to fully marine conditions, salinity *per se* cannot be considered a stress for them, yet many other stresses similar to those experienced by terrestrial plants clearly occur. Accordingly, at this stage in the exploration of seagrass osmoregulation and ion homeostasis, we can only report general observations with regard to reductions and expansions of the genes (and structures, e.g., cell wall composition, absence of stomata) associated with these enzymes.

As an environmental variable, salinity levels can alter general metabolic processes and enzymatic activities causing an increased production of reactive oxygen species (Modarresi *et al.* 2013). However, all plants must contend with a wide variety of other environmental stressors, e.g., desiccation, heat shock, heavy metal exposure, nutrient deprivation, mechanical stress and high/low light stress (Mittler 2002). Reactive oxygen intermediates such as superoxide anion, hydrogen peroxide and hydroxyl radicals are produced in all of these cases (Mittler 2002). In low quantities, these intermediates may function as a signal for the activation of stress-responses be rapidly converted to less reactive forms; in high quantities they can damage DNA, RNA and proteins.

Supplementary Table S8.1 | *Z. marina* cationic-H^+ antiporters.

Gene Family	A.t. localization	A. thaliana	O. sativa	S. polyrhiza	Z. marina	Z. marina expression
						Female flower early Female flower late Male flower, leaf, root
NHX / SOS						
Na+-K+ exchangers	vacuole	AtNHX1	Os07g47100	Sp6g0043500 Sp6g0003900	Zosma23g00150 Zosma306g00190	MF
		AtNHX2				FFE>FFL,R>L>MF
		AtNHX3	Os05g05590 Os11g42790	Sp6g0036300	Zosma426g00080	L>MF,R>FF
		AtNHX4	Os06g21360	Sp12g0030600	Not Found	
	Golgi endomembrane	AtNHX5	Os09g11450 Os09g30446*	Sp10g0028900	Zosma101g00730	L>R>MF>FF
		AtNHX6				
	plasma membrane	AtNHX7	Os12g44360	Sp20g0012600	Zosma28g00970	R,L>FFE>FFL,MF
		AtNHX8				
KEA						
K+ efflux antiporters	chloroplast	AtKEA1	Os04g58620	Sp11g0052800	Zosma8g01070	L>R>FF>MF
	chloroplast	AtKEA2				
	chloroplast	AtKEA3	Os12g42300	Sp2g0126600	Zosma49g00060	L>FF>R>MF
	SignalPeptide +	AtKEA5	Os03g03590*	Sp16g0005400	Zosma85g01070	R,MF>L>FF
	SignalPeptide +	AtKEA4	Os06g36590	Sp18g0001100 Sp21g0031300	Zosma5g02150 Zosma11g00530	FF,R,MF,L
	SignalPeptide +	AtKEA6				R>FFE>L>FFL>MF
CHX						
Cation / H+ antiporters	Pollen	AtCHX1	Os02g58660 Os09g37300 Os08g43690	Sp5g0065500	Zosma49g00790 Zosma234g00210	MF>>R,L MF>>R>FF>L
		AtCHX2				
	Pollen	AtCHX28	Os12g42300	Sp15g0011600	Zosma4g01020	
	Pollen / vegetative	AtCHX15	Os01g60140 Os05g40650 Os12g02840 Os11g03070	Sp0g0067200[s] Sp2g0057800 Sp0g0153100	Zosma87g01060[s] Zosma114g01000[s] Zosma2g01790 Zosma74g00350 Zosma134g00020	MF MF MF MF MF
	Pollen	AtCHX21				
	Pollen	AtCHX23				
	Pollen	AtCHX3	Os08g02450 Os11g01820 Os12g01820	Sp29g0015700	Zosma93g00780 Zosma49g00320 Zosma49g00330 Zosma106g00010	MF MF MF>>R MF
	Pollen	AtCHX4				
	Pollen	AtCHX5				
	Pollen	AtCHX6A				
	Pollen	AtCHX6B				
	Pollen	AtCHX7				
	Pollen	AtCHX8				
	Pollen	AtCHX9				
		AtCHX10				
		AtCHX11				

	Pollen	AtCHX12				
		AtCHX13				
		AtCHX14				
	Pollen	AtCHX26				
	Pollen	AtCHX27				
		AtCHX16	Os05g02240 Os12g42200 Os03g61290 Os05g19500	Sp3g0081700 Sp27g0004700	Zosma89g00780 Zosma15g01080[p] Zosma176g00140[p]	MF>L MF MF
	Roots / leaves	AtCHX17				
		AtCHX18				
		AtCHX19				
	Guard cells	AtCHX20				
	Pollen	AtCHX24	NF	NF	NF	
	Pollen	AtCHX25				

* : partial sequence

s : related to AtCH1/2 and AtCHX28 as well

p : pseudogene

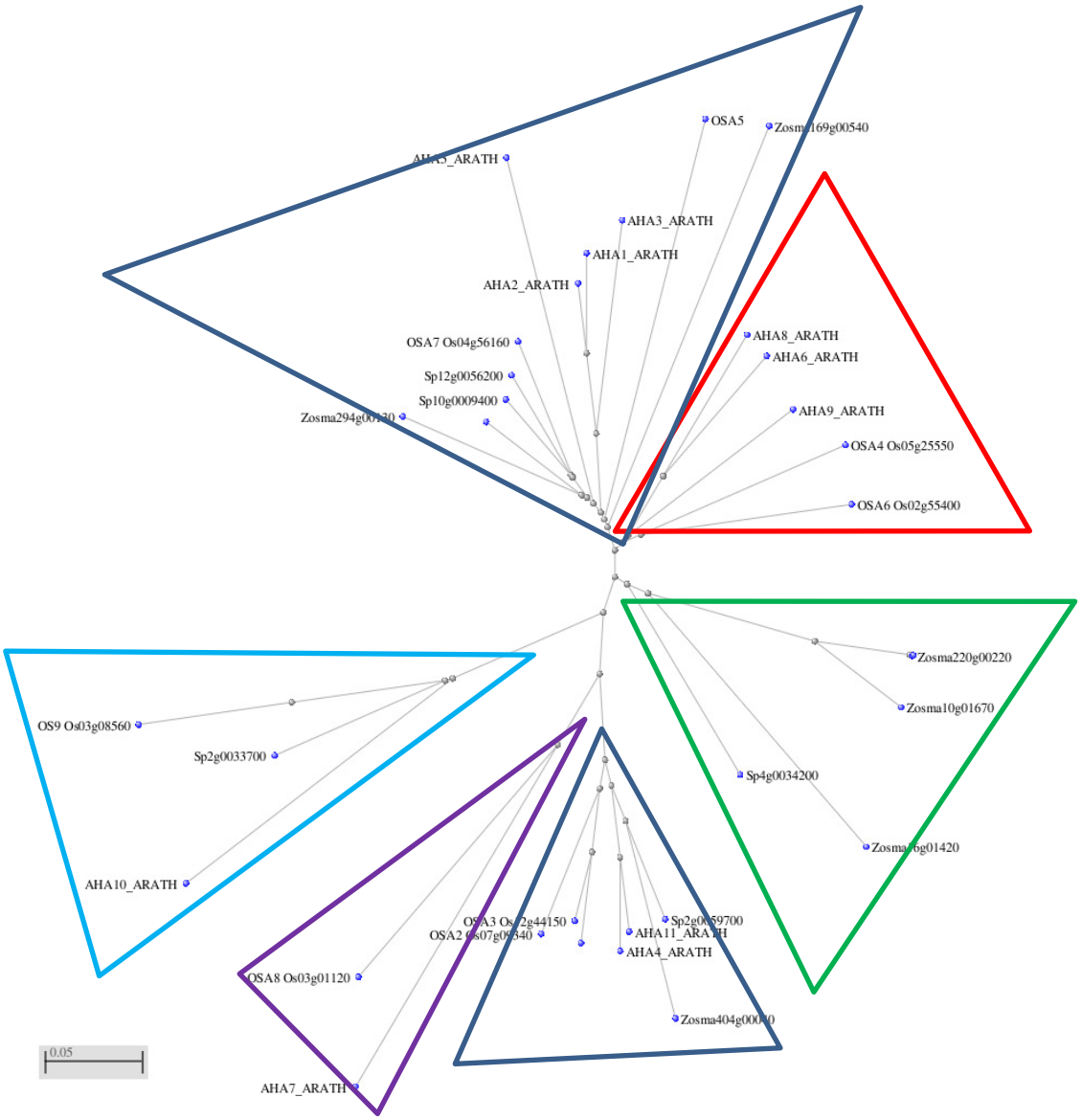
Supplementary Table 8.2 | Plasma membrane H^+-ATPase (AHA) genes and their expression in *Z. marina.*

Different tissues are female flower early, female flower late, male flower, root and leaf. Shaded/unshaded groups refer to phylogenetic clusters (**Supplementary Fig. S8.1)**

Gene Family	*A. thaliana*	*O. sativa*	*S. polyrhiza*	*Z. marina*	tissue expression in *Z. marina*				
AHA / PMA					FFE	FFL	MF	R	L
plasma membrane H+ -ATPase	AtAHA1	OsAHA5 OsAHA7	Sp12g0056200 Sp10g0009400	Zm169g00540 Zm294g00130[a]	0,4 35	1,1 27	1,4 2,0	1,2 2,7	0,5 **509**
	AtAHA2								
	AtAHA3								
	AtAHA5								
	AtAHA6	OsAHA4 OsAHA6	NF	NF					
	AtAHA8								
	AtAHA9								
	AtAHA7	OsAHA8	NF	NF					
	AtAHA10	OsAHA9	Sp2g0033700	NF					
	AtAHA4	OsAHA1 OsAHA2 OsAHA3	Sp2g0059700	Zm404g00040	3,4	4,3	0,7	2,1	1,1
	AtAHA11								
	NF	NF	Sp4g0034200[b]	Zm220g00220	0	0	**93**	0	0
				Zm10g01670	0	0	**45**	0	0
				Zm16g01420[b]	0,8	0	**225**	0	0

a : this is the salt-tolerant ZHA1 gene which is referred to in Muramatsu et al., 2002

b : these two H+-ATPases are lacking the C-terminal autoregulatory extension



Supplementary Fig. S8.1 | Plasma membrane H^+-ATPase (AHA) gene comparisons.

a, The alignment can be found in **Supplementary Data 9,** Neighbor Joining tree of the H^+-ATPase protein sequences from *Arabidopsis* (AtAHA), rice (OsAHA), *Spirodela polyrhiza* and *Zostera marina.* Clockwise from top: dark blue, present in all four taxa; red, missing in *Zostera* and *Spirodela*; green unique to *Zostera* and *Spirodela*; turquoise found in *Spirodela* but not in *Zostera*; purple missing in both *Zostera* and *Spirodela.*

Free radical detoxification enzymes such as catalase (CAT), superoxide dismutase (SOD) and glutathione peroxidase (GPX) are shown in **Supplementary Table S8.3**. There is only a single CAT gene in *Z. marina* that matches CAT2 in *Arabidopsis*. The absence of CAT1 is possibly related to the reduced xylem characteristic of submerged plants; the absence of CAT3 is not clear. Gene expression studies in *Z. marina* showed CAT down-regulation after *Labyrinthula zosterae* infection (Brakel *et al.* 2014). SOD expression was up-regulated under thermal stress in *Z. marina* (Winters *et al.* 2011) and up-regulated under salinity reduction in the seagrass *Thalassia hemprichii* (Jiang Zhijian *et al.* 2013). Differential tissue expression is shown in **Supplementary Table S8.4** and was highest in male flower.

LEA/dehydrin domains are clearly under-represented in both *Z. marina* and *Spirodela* genomes, relative to other genomes (**Supplementary Table S8.5**). Comparing with *Spirodela,* LEA III and IV are completely

missing in the *Z. marina* genome. Two S-adenosylmethionine synthase genes (Zosma467g00060 and Zosma17g00180) were highly expressed in all tissues and among the highest expressed in root and female flower. S-adenosylmethionine synthases have previously been found to be upregulated under salt stress conditions in tomato (Espartero *et al.* 1994; Sanchez-Aguayo *et al.* 2004), *Catharanthus roseus* (Schroder *et al.* 1997) and rice (Lee *et al.* 1997), and that transgenic overexpression promotes salt tolerance in tobacco (Qi *et al.* 2009). Further experimental investigation is merited.

Supplementary Table S8.3 | *Z. marina* reactive oxygen species (ROS) enzymes and comparisons with other species.

Gene Ids	Annotation	Top Blast	Zos-tera	Spiro-dela	Arabi-dopsis	Or-yza
Catalase			**1**	**2**	**3**	**3**
Zosma228g00030	Catalase	catalase [Ziziphus jujuba]				
Superoxide dismutase			**6**	**10**	**7**	**8**
Zosma72g00490	Copper chaperone SOD	PREDICTED: Cu/Zn-superoxide dismutase copper chaperone isoform X1 [Glycine max]				
Zosma16g01180	CuZn Superoxide dismutase	copper/zinc superoxide dismutase 2 [Musa acuminata AAA Group]				
Zosma5g01030	CuZn Superoxide dismutase	copper/zinc superoxide dismutase [Sandersonia aurantiaca]				
Zosma106g00140	Fe Superoxide dismutase	iron superoxide dismutase [Musa acuminata]				
Zosma16g00950	Fe Superoxide dismutase	hypothetical protein AMTR_s00044p00032580 [Amborella trichopoda]				
Zosma270g00070	Mn Superoxide dismutase	manganese superoxide dismutase [Elaeis guineensis]				
Glutathione peroxidase			**7**	**6**	**7**	**8**
Zosma180g00060	Glutathione peroxidase1	phospholipid hydroperoxide glutathione peroxidase 1 [Arabidopsis thaliana]				
Zosma269g00080	Putative Glutathione peroxidase7	glutathione peroxidase [Zantedeschia aethiopica]				
Zosma41g01010	Putative Glutathione peroxidase6	PREDICTED: probable phospholipid hydroperoxide glutathione peroxidase [Cucumis melo]				
Zosma44g01480	Glutathione peroxidase5	glutathione peroxidase 5 [Arabidopsis thaliana]				
Zosma117g00840	Glutathione peroxidase6	putative phospholipid hydroperoxide glutathione peroxidase 6 [Morus notabilis]				
Zosma96g00580	Putative Glutathione peroxidase3	hypothetical protein EUTSA_v10001645mg [Eutrema salsugineum]				
Zosma25g00150	Putative Glutathione peroxidase6	hypothetical protein PRUPE_ppa010771mg [Prunus persica]				

Supplementary Table S8.4 | Differential expression (FPKM) of the reactive oxygen species (ROS)-detoxification enzymes.

Catalase (CAT), Superoxide dismutase (SOD) and Glutathione peroxidase (GPx) in female early and late flowers, male flower, root and leaf tissues. The highest expression for each gene is reported in bold.

Gene Ids	Gene	Female flower early	Female flower late	Male flower	Root	Leaf
Zosma228g00030	CAT	90.82	99.43	**2714.72**	339.53	748.58
Zosma16g01180	CuZnSOD	12.49	14.91	**72.70**	**82.34**	12.26
Zosma5g01030	CuZnSOD	384.47	**461.64**	111.35	**446.72**	212.29
Zosma106g00140	FeSOD	**15.38**	**21.68**	3.00	2.97	5.31
Zosma16g00950	FeSOD	12.34	19.50	1.50	0.81	2.04
Zosma270g00070	MnSOD	**478.87**	**540.09**	245.70	129.85	290.60
Zosma180g00060	GPx1	12.08	32.160	85.27	1.17	**223.30**
Zosma96g00580	Putative GPx3	50.61	47.64	71.04	64.33	53.14
Zosma44g01480	GPx5	66.38	62.75	**233.21**	79.70	84.62
Zosma117g00840	GPx6	33.92	47.67	150.02	254.84	33.68
Zosma25g00150	Putative GPx6	95.27	96.30	**219.80**	108.35	**229.54**
Zosma269g00080	Putative GPx7	9.80	14.80	15.84	5.74	88.03
Zosma41g01010	Putative GPx6	77.02	141.79	62.68	**579.25**	**531.44**

Supplementary Table S8.5 | Number of late embryogenesis proteins (LEA) and dehydrins in the sequenced plant genomes based on the OrthoMCL gene family analysis.

LEA group	*Z. marina*	*S. polyrhiza*	*B. distachyon*	*O. sativa*	*Z. mays*	*V. vinifera*	*A. thaliana*
I	1	1	3	2	2	3	2
II	1	1	10	6	6	2	9
III	0	3	6	5	3	3	7
IV	0	1	6	4	3	4	3
Va	1	1	6	4	3	3	6
Vb	0	2	5	3	8	1	4
Vc	51	36	62	65	83	24	50
VI	0	0	1	0	1	0	3
VII	0	0	5	5	7	1	0

8.2 Plant Metallothioneins

Plant metallothioneins (MTs) are a family of small, multi-functional cytosolic proteins classically implicated in metal homeostasis. However, more recent studies show that they have a much broader role in, e.g., anti-oxidant activity, stress and developmental responses (Chaturvedi *et al.* 2012; Leszczyszyn *et al.* 2013). They have been most extensively characterized in *Arabidopsis thaliana* (Winter *et al.* 2007). Although considerable structural and physiological heterogeneity exists, there are four general types of plant MTs recognized (Leszczyszyn *et al.* 2013), albeit some authors (Zhou *et al.* 2006) split MT2 into to subclades: MT2 proper and MT4. MT4 differs from MT2 *sensu stricto* by the number of cysteines in the N-terminal cluster. It should be noted that these MT4 are different from (and should not be confused with) EC-MTs, which are often referred as MT4. We follow the aforementioned nomenclature in **Extended Data Figure 5**.

In most plants surveyed, plant-MT1 genes can account for up to 75% of an expression profile and in all tissues (none were found in *Z. marina*); plant-MT2/MT4 genes are mainly associated with leaves and roots, and in developing seed and early fruit (also the case in *Z. marina*); plant-MT3 genes are strongly expressed in the late ripening stage of fruit development (up to 50% of all ESTs expressed in ripening pineapple (Moyle *et al.* 2005) and also during senescence (Breeze *et al.* 2004) (only one was found in *Z. marina*); and EC genes, which are only found in plants, are typically embryo/seed specific but also found in both male and female reproductive organs, as well as some vegetative tissue (none were found in *Z. marina*). The combination of MT1, MT2 and MT3 is often up-regulated in ageing leaves or in leaves exposed to plant pathogens (Dauch & Jabaji-Hare 2006). *Z. marina* is characterized by high leaf replacement.

Ten MT genes were identified in *Z. marina* using documented MTs as queries **(Extended Data Figure 5, Supplementary Table 8.6**). The *Z. marina* repertoire is unusual in that: 1) the genes are not arranged in tandem, as is the case for *Arabidopsis* (7 MTS : 2 MT1, 2 MT2, 1 MT3, 2 EC) and rice (11 MTs: 2 MT1, 6 MT2s, 2 MT3, 1 EC); 2) no MT1 and EC have been found in either leaf, root or flower tissues; and 3) there are five "half metallothionein" (HMT) genes. Such HMTs have never been reported before. These genes encode very small and highly similar proteins (26 and 23 aa), the sequence of which contains the typical six Cys cluster observed at the C-terminus of the pMT1 and pMT3 subfamily. Because they may have been overlooked due to their very small size, we performed a more refined search in the *Arabidopsis* and rice genomes which yielded only the originally documented MTs in the MT1-MT3 subfamily.

Supplementary Table S8.6 | List of metallothioneins (MTs) and half-metallothioneins (HMTs) in *Z. marina* compared with *Arabidopsis, Rice* and *Spirodela*.

	Arabidopsis	*Zostera*	Rice	Spirodela
MT2	AT3G09390 AT5G02380	Zosma2g00100 Zosma2622g00020 Zosma282g00060 Zosma154g00400	Os01g05650	Spipo0G0112500
MT3	AT3G15353	Zosma148g00060	Os05g11320 Os01g10400	Spipo14G0028700
HMT	Not found	Zosma118g00003 Zosma118g00006 Zosma412g00015 Zosma379g00025 Zosma28g00595	Not found	Not found

MTs sequences have also been obtained from two other alismatids, the seagrass *Posidonia oceanica* and the freshwater duckweed *Spirodela polyrhiza*. A comparison of MT occurrence in this last species is especially interesting as it is the only other alismatid with a complete genome sequence (Wang *et al.* 2014) thus allowing for a fair comparison between MT repertoires. Five MT/MT-like genes were retrieved from the *S.*

polyrhiza genome, of which four were annotated (Spipo0G0112500: MT2, Spipo14G0028700: MT3, Spipo6G0071500: MT4L1 and Spipo0G0175800: EC) with MT2 and MT4L1 gene models revised to retrieve the proper first exon, and a novel one, MT4L2, found in the vicinity of MT4L1 (**Extended Data Figure 5)**.

Only four of the MTs of *Z. marina* have canonical features with other plant-MTs: MT2A, MT2B, MT2C and MT3. The very highly expressed MT2L differs from canonical MT2s in missing a segment of at least 18 residues at the N-terminus (six out of the eight cysteines in the Cys-cluster). Therefore, the metal-binding properties of MT2L must be at variance with the ones from canonical MT2s, including those from its highly expressed homologue from the seagrass *Posidonia oceanica*. We note that among the five *Spirodela* MTs, two MT4L are also divergent from canonical plant MTs. MT4L1 has N-terminal and C-terminal cysteine clusters reduced to four cysteines whereas the related MT4L2, with an unusual 4-Cys C-terminal cluster, has lost the entire N-terminus, including the Cys cluster. Since no expression data are currently publically available for the *Spirodela* genome, we do not know which MT/MT-like is expressed, when or where, and therefore cannot exclude the possibility that either or both MT4L would be pseudogenes.

The five HMTs are novel MT-like proteins. They do possess the conserved C-terminal cluster of six Cys-residues and, as such, very likely bind divalent metal ions. Interestingly, in HMT5 one Cys in the cluster is conservatively replaced by His, keeping intact its metal binding potential. With only a single Cys cluster, they are losing the features linked to interaction between the two Cys-domains (and those from the spacer as well). What their specific role is next to more or less canonical plant MTs and phytochelatins (the latter of which do exist in *Zostera*: Zosma147g00230 encodes a phytochelatin synthase) remains an open question.

MTs are often the most highly expressed genes in *Z. marina* under ambient/control conditions (Reusch *et al.* 2008; Wissler *et al.* 2009). This was also the case in a recent study conducted by Reusch and Brakel (unpublished; see below) with expression found in all tissues but especially leaves and late flowers for the conventional MT3 and MT2s (**Supplementary Table S8.7**) . Similarly high constitutive expression has also been found for MTs from *Posidonia oceanica*, an endemic seagrass to the Mediterranean (Wissler *et al.* 2009). There is also evidence of MT activity in heavy metal sequestration including mercury (Pergent-Martini 1998; Schlacher-Hoenlinger & Schlacher 1998; Giordani *et al.* 2000). Given the high volcanic activity in the Mediterranean, high expression of MTs in *P. oceanica* is not surprising. Similar evidence for *Z. marina* is lacking. Besides MTs, the novel HMTs have been shown to be very highly expressed as well, almost exclusively in roots, at least for some members (HMT1, HMT3 and/or HMT4) (**Supplementary Table S8.7**).

MT2B (Zosma282g00060) was not highly expressed in our tissue libraries but was 200th highest expressed in male flower and thus may be a tissue specific variant.

Differential expression of MT3 has been inconsistently found in heat stress experiments. MT3 accounted for 2.5-15% of all transcripts including a 6-fold down regulation under temperature stress in *Z. marina* (Reusch *et al.* 2008) and evidence for MT3 expression was also found using RT-PCR and an pMT-3 (Bergmann *et al.* 2010). However, high expression was not found in similar experiments by Franssen (Franssen *et al.* 2011), which focused on heat shock proteins, nor in the "ambient condition" transcriptome of the *Z. marina* Finnish clone used for the present genome sequencing project.

High MT expression has also been reported for *Z. marina* under salt stress (Kong *et al.* 2013) with inconclusive results. In our own experiment (J.B. & T.B.H.R.), using the ten MTs genes identified in *Z. marina* to measure gene expression based on qRT-PCR in a time series of exposures of *Z. marina* individuals to increasing and decreasing salinity, MT3 and MT2L were highly expressed under all conditions and time points. MT2B, MT2C, HMT1 and HMT2,3,4 were expressed at low levels in all conditions. HMT5 was not expressed. To confirm that the plants were definitely stressed by the salinity treatments, a set of housekeeping genes was also tested following earlier experiment in their lab (Bergmann *et al.* 2010; Winters *et al.* 2011). HSP70 was increased 3.5 fold under freshwater stress (15 to 10 PSU) at 24 hrs and APX gene expression increased 3.2 fold under salinity increase (15-27 PSU) after 24 hrs. Based on this experiment there was no indication of up or down regulation in any of the MT genes in response to salinity changes.

At present, the unique repertoire and high and variable expression of MTs and HMTs in *Z. marina* remains a mystery. Their role in metal sequestration and ion homeostasis are not in question but there are clearly other functions that have not yet been identified. For example, MTs are also implicated in development and senescence (Leszczyszyn *et al.* 2013). Leaf turnover in *Z. marina* can be extremely high. We speculate that the high constitutive expression of MT2s and MT3 may be involved in leaf senescence and male flower development, while the one of HMTs may have to do with sequestration of metals in roots.

Supplementary Table S8.7 | Differential tissue expression (FPKM) of metallothioneins (MTs) and half-metallothioneins (HMTs).

The five HMTs are highly similar; accordingly, the expression was computed for unambiguously assigned RNA-seq reads, which was possible for HMT1. HMT3 and HMT4 are almost identical; accordingly, the RNA-seq reads may mean that either one or the other or both genes are expressed as mentioned. For HMT2 and HMT5 no read perfectly matching their sequences were found.

FPKM	f flower early	f flower late	m flower	root	vegetative	type
Zosma2g00100	1304	4765	63.24	17.3	701	MT2A
Zosma282g00060	4.0	1.4	1067	0.00	0.1	MT2B
Zosma2622g00020	5.1	3.6	533	0.21	0.2	MT2C
Zosma154g00400	8.9	18.1	30.9	420	6007	MT2L
Zosma148g00060	2076	5562	790	1313	49537	MT3
Zosma28g00595	18.1	22.5	166	26574	43.3	HMT1
Zosma379g00025	0.00	0.00	0.00	0.00	0.00	HMT2
Zosma412g00015	0.00	0.00	4.8	21644	13.3	HMT3
Zosma118g00003	0.00	0.00	4.8	21644	13.3	HMT4
Zosma118g00006	0.00	0.00	0.00	0.00	0.00	HMT5

9 Fatty acid and lipid metabolism

9.1 Fatty acid and lipid metabolism

Fatty acids and lipids are structurally diverse molecules with important functions such as cell wall coatings, membrane barriers, storage of carbon and energy, protection against biotic and abiotic stress, and signaling.

There is limited bibliography on fatty acid and lipid composition in *Z. marina*, and more globally on other species of seagrass. Published studies on *Z. marina* have mainly described the presence of the glycerolipids such as phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), and sulfoquinovosyldiacylglycerol (SQDG), and their composition (Sanina *et al.* 2004; Sanina *et al.* 2008). This composition is quite similar to results observed for *Arabidopsis* tissues (Li-Beisson *et al.* 2013), in particular with the abundance of α-linolenic acid (18:3n-3, ALA) in MGDG and DGDG, and of 16:3n-3 (hexadecatrienoic acid, HTA) in MGDG, indicating a prevalence of n-3 PUFAs in lipids of photosynthetic membranes, and of n-6 PUFAs in extra-plastidial membranes. A striking difference is the detection of low amount of C20 polyunsaturated fatty acids (PUFAs), i.e. arachidonic acid (20:4n-6, ARA) and eicosapentaenoic acid (20:5n-3, EPA) (Khotimchenko 1993; Sanina *et al.* 2004; Sanina *et al.* 2008), in polar lipids of *Z. marina* and which have not been found in land plants except bryophytes. To our knowledge, no other categories of lipids have been analyzed in seagrass to determine their quantities and composition. This corresponds to storage lipids (triacylglycerols), sphingolipids (membrane components and bioactive molecules), cuticular waxes, suberin and cutin monomers.

A summary of fatty acid and lipid metabolism in *Z. marina* compared with *Arabidopsis* and *Spirodela* can be found in **Supplementary Table S9.1.** There is a considerable degree of homology in the fatty acid and lipid metabolism associated enzymatic steps and proteins between *Z. marina* and *Arabidopsis*, as well as with *Spirodela*. The degree to which this metabolism is important for the adaptation to seawater environment is probably not the presence or absence of different sets of genes, but the regulation of their expression.

Supplementary Table S9.1 | Summary of various categories (i.e., membranes, storage, defense, cell walls) of fatty acid and lipid metabolism in *Z. marina* compared with *Arabidopsis* and *Spirodela*.

Function (gene name)	*Arabidopsis* orthologue count	*Zostera* orthologue count	*Zostera* marine candidates genes	*Spirodela orthologue count*	*Spirodela candidate genes*
1) De novo plastid FA synthesis, elongation, desaturation and export					
Ketoacyl-ACP synthase III (KASIII)	1	1	Zosma11g00260	2	Spipo17G0041300 Spipo1G0028400
Ketoacyl-ACP reductase (KAR)	5	3	Zosma289g00290 Zosma181g00100 Zosma224g00060	3	Spipo7G0067500 Spipo12G0037200 Spipo10G0020500
Hydroxyacyl-ACP dehydrase (HAD)	2	1	Zosma25g01090	1	Spipo10G0034500
Enoyl-ACP reductase (ER)	1	1	Zosma4g00730	2	Spipo0G0076600 Spipo10G0020200
Ketoacyl-ACP synthase I (KASI)	1	1	Zosma11g00960	1	Spipo7G0061000
Malonyl-CoA : ACP malonyltransferase (MCMT)	1	1	Zosma141g00260	1	Spipo22G0028800
Holo-ACP synthase (HACPS)	2	1	Zosma168g00170	1	Spipo12G0055200
Lipoate synthase (LS)	1	1	Zosma46g00170	1	Spipo32G0005700
Lipoyltransferase (LT)	1	1	Zosma74g00060	1	Spipo27G0004100
Acyl carrier protein (ACP)	5	4	Zosma23g00140 Zosma3g01100 Zosma408g00070 Zosma408g00060	6	Spipo22G0010600 Spipo3G0042500 Spipo0G0087500 Spipo3G0052600 Spipo6G0053200 Spipo2G0080400
Acyl-ACP thioesterase B (FAT B) and A (FAT A)	3 (2A and 1B)	3	Zosma1g01370 Zosma178g00130 Zosma77g00830	4	Spipo21G0008900 Spipo27G0017100 Spipo0G0147200 Spipo4G0050900
Ketoacyl-ACP synthase II (KASII)	1	2	Zosma11g01300 Zosma176g00250	2	Spipo0G0033200 Spipo7G0013700
Stearoyl-ACP desaturase (SAD)	7	3	Zosma22g00970	7	Spipo4G0091300

			Zosma195g00240 Zosma166g00480		Spipo0G0009200 Spipo32G0010700 Spipo9G0033500 Spipo5G0077600 Spipo0G0123400 (partial) Spipo0G0123300 (partial)
Long-chain acyl-CoA synthetase (LACS)	9	6	Zosma9g01190 Zosma40g00520 Zosma166g00090 Zosma315g00010 Zosma35g00070 Zosma270g00010 (partial)	7	Spipo18G0011700 Spipo26G0016500 Spipo2G0116100 Spipo3G0077600 Spipo22G0000300 Spipo7G0043300 Spipo13G0009500
Long-chain-fatty-acid--[acyl-carrier-protein] ligase (AAE)	2	1	Zosma5g02540	1	Spipo4G0047900
ABC acyl transporter (ABCAT)	2	2	Zosma40g00440 Zosma78g00570	2	Spipo26G0017200 Spipo4G0100800
Acyl-CoA binding protein (ACBP)	6	6	Zosma74g01040 Zosma123g00500 Zosma277g00110 Zosma261g00210 Zosma48g00820 Zosma32g00460	7	Spipo11G0008200 Spipo5G0074300 Spipo1G0036000 Spipo22G0012000 Spipo6G0000800 Spipo1G0065900 Spipo17G0045000
2) Prokaryotic and eukaryotic galactolipid, sulfolipid, and phospholipid synthesis (plastid)					
NAD-dependent glycerol-3-phosphate dehydrogenase (GPDH)	4	3	Zosma56g01130 Zosma4g00800 Zosma87g00690	3	Spipo22G0024400 Spipo4G0036800 Spipo6G0033300
Glycerol-3-phosphate acyltransferase (GPAT)	10	10	Zosma29g01540 Zosma7g00950 Zosma379g00150 Zosma82g00900 Zosma28g00690 Zosma93g00640 Zosma112g00040 Zosma88g00100 (partial) Zosma129g00190 (partial) Zosma65g01010	12	Spipo30G0006700 Spipo23G0004300 Spipo11G0054200 Spipo23G0014600 Spipo32G0011400 Spipo6G0073200 Spipo3G0095200 Spipo0G0136700 Spipo6G0017900 Spipo8G0027600 Spipo6G0017600 Spipo7G0013300
1-acyl-sn-glycerol-3-phosphate acyltransferase 1 (LPAAT)	5	4	Zosma21g00930 Zosma6g00580 Zosma125g00200 Zosma9g01250	7	Spipo6G0030100 Spipo7G0018900 Spipo24G0004600 Spipo4G0068200 Spipo7G0051900 Spipo24G0001100 Spipo3G0111400
Phosphatidic acid phosphatase (PAP)	11	6	Zosma5g01560 Zosma307g00020 Zosma82g00780 Zosma55g00660 Zosma85g00740 Zosma75g00100	4	Spipo19G0026700 Spipo14G0041600 Spipo19G0026800 Spipo6G0068900
CDP-DAG synthase	6	2	Zosma1g01870 Zosma26g01260	2	Spipo2G0035400 Spipo16G0012300
Phosphatidylglycerol-phosphate synthase (PGPS)	2	2	Zosma302g00060 Zosma353g00080	2	Spipo22G0034100 Spipo14G0051000 (partial)
Phosphatidylglycerol-phosphate phosphatase (PGPP)	No sequence biochemically characterized in Arabidopsis	1 (id using yeast seq)	Zosma389g00130	1	Spipo7G0024500
Palmitate desaturase (FAD4)	3	1	Zosma7g00360	1	Spipo8G0007600
Delta-12 desaturase (FAD6)	1	1	Zosma90g00270	1	Spipo12G0044500
Delta-15 desaturase (FAD7 or FAD8)	1	2	Zosma121g00290 Zosma230g00390	2	Spipo23G0035900 Spipo23G0036000

Palmitoyl-monogalactosyldiacylglycerol delta-7 desaturase (FAD5)	9	1	Zosma234g00140	1	Spipo21G0000200
Monogalactosyldiacylglycerol synthase (MGDGS)	3	2	Zosma255g00090 Zosma50g00150	2	Spipo12G0049000 Spipo0G0074800
Digalactosyldiacylglycerol synthase (DGDGS)	2	2	Zosma146g00230 Zosma372g00090	2	Spipo4G0108300 Spipo5G0066700
UDP-sulfoquinovose:DAG sulfoquinovosyltransferase (SLS or SQD2)	1	1	Zosma239g00080	1	Spipo23G0004800
UDP-Sulfoquinovose Synthase (SQS or SQD1)	1	1	Zosma44g00360	1	Spipo3G0018400
UDP-Glucose Pyrophosphorylase (UGP)	1	2	Zosma190g00140 Zosma139g00370	2	Spipo23G0039100 Spipo8G0011100
Phospholipase D zeta	2	2	Zosma44g00840 Zosma96g00250	1	Spipo3G0029600
Phospholipase D delta	1	4	Zosma173g00170 Zosma313g00140 Zosma77g00370 Zosma77g00380	3	Spipo1G0071800 Spipo4G0071600 Spipo17G0027200
Phospholipase D beta and/or gamma	5	2	Zosma180g00160 Zosma46g00030	1	Spipo1G0095200
Phospholipase D alpha	1	3	Zosma4g01770 Zosma101g00130 Zosma132g00380	1	Spipo0G0089800
Phospholipase C	6	4	Zosma27g01110 Zosma401g00130 Zosma26g00960 Zosma225g00120	4	Spipo23G0002400 Spipo3G0061500 Spipo5G0063300 Spipo9G0060900
Phospholipase A2	9	8	Zosma53g00190 Zosma137g00170 Zosma111g00460 Zosma114g00160 Zosma84g00330 Zosma118g00390 Zosma218g00240 Zosma218g00230	4	Spipo16G0025400 Spipo15G0027600 Spipo22G0002900 Spipo22G0003200
Permease-like protein of inner chloroplast envelope (ABC transporter; TGD1)	1	1	Zosma85g00540	1	Spipo17G0022300
Phosphatidic acid-binding protein (TGD2)	1	1	Zosma38g01140	1	Spipo22G0037300
ATPase (TGD3)	1	3 (including one partial sequence)	Zosma2g00560 Zosma166g00680 Zosma230g00160 (partial)	1	Spipo4G0067700
Protein involved in lipid transport (TGD4)	1	1	Zosma8g00630	1	Spipo27G0008500
3) Eukaryotic phospholipid synthesis and editing (endomembrane)					
NAD-dependent glycerol-3-phosphate dehydrogenase (GPDH)	4	3	Zosma56g01130 Zosma4g00800 Zosma87g00690	3	Spipo22G0024400 Spipo4G0036800 Spipo6G0033300
Glycerol-3-phosphate acyltransferase (GPAT)	10	10	Zosma29g01540 Zosma7g00950 Zosma379g00150 Zosma82g00900 Zosma28g00690 Zosma93g00640 Zosma112g00040 Zosma88g00100 Zosma129g00190 Zosma65g01010	12	Spipo30G0006700 Spipo23G0004300 Spipo11G0054200 Spipo23G0014600 Spipo32G0011400 Spipo6G0073200 Spipo3G0095200 Spipo0G0136700 Spipo6G0017900 Spipo8G0027600 Spipo6G0017600 Spipo7G0013300
1-Acylglycerol-3-phosphate acyltransferase (LPAAT)	5	4	Zosma21g00930 Zosma6g00580 Zosma125g00200 Zosma9g01250	7	Spipo6G0030100 Spipo7G0018900 Spipo24G0004600 Spipo4G0068200 Spipo7G0051900 Spipo24G0001100 Spipo3G0111400
N-methylphospholipid methyltransferase (PLMT)	1	1	Zosma17g01200	1	Spipo28G0011600
CDP-DAG synthase (CDP-DAGS)	6	2	Zosma1g01870 Zosma26g01260	2	Spipo2G0035400 Spipo16G0012300
Phosphatidic acid phosphatase	11	6	Zosma5g01560	4	Spipo19G0026700

			Zosma307g00020 Zosma82g00780 Zosma55g00660 Zosma85g00740 Zosma75g00100		Spipo14G0041600 Spipo19G0026800 Spipo6G0068900
Diacylglycerol cholinephosphotransferase (AAPT)	2	1	Zosma62g00340	2	Spipo1G0032500 Spipo16G0036700
Phospholipase A2	9	8	Zosma53g00190 Zosma137g00170 Zosma111g00460 Zosma114g00160 Zosma84g00330 Zosma118g00390 Zosma218g00240 Zosma218g00230	4	Spipo16G0025400 Spipo15G0027600 Spipo22G0002900 Spipo22G0003200
1-acylglycerol-3-phosphocholine acyltransferase (LPCAT or LPLAT)	6	5	Zosma43g00130 Zosma18g00180 Zosma25g01630 Zosma361g00170 Zosma97g00530	4	Spipo17G0017700 Spipo24G0004600 Spipo1G0116800 Spipo1G0116900
Long chain acyl-coA synthetase (LACS)	9	6	Zosma9g01190 Zosma40g00520 Zosma166g00090 Zosma315g00010 Zosma35g00070 Zosma270g00010 (partial)	7	Spipo18G0011700 Spipo26G0016500 Spipo2G0116100 Spipo3G0077600 Spipo22G0000300 Spipo7G0043300 Spipo13G0009500
Choline-phosphate cytidylyltransferase (CCT)	2	1	Zosma1g01900	2	Spipo13G0054600 Spipo5G0070600
Choline kinase (CK)	4	2	Zosma97g00820 Zosma98g00370	2	Spipo21G0016000 Spipo21G0002900
Ethanolamine kinase (EK)	1	1	Zosma1g00630	1	Spipo21G0003000
Phosphoethanolamine N-methyltransferase (PEAMT)	3	2	Zosma452g00140 Zosma437g00020 (partial)	1	Spipo8G0069500
CTP:phosphorylethanolamine cytidyltransferase (PECT)	1	2	Zosma192g00290 Zosma18g00780	2	Spipo3G0010800 Spipo28G0024800
Phosphatidylserine decarboxylase (PSD)	3	2	Zosma31g01370 Zosma1g00140	2	Spipo11G0033700 Spipo4G0094200
Base-exchange-type phosphatidylserine synthase (BE-PSS)	1	1	Zosma124g00150	2	Spipo19G0023800 Spipo18G0028900
Phosphatidylinositol Synthase (PIS)	2	1	Zosma32g00860	2	Spipo25G0009900 Spipo7G0049200
***Myo*-inositol-3-phosphate synthase (MIPS)**	3	2	Zosma3g01680 Zosma89g01220	2	Spipo4G0013100 Spipo20G0028300
Phosphatidylglycerol-phosphate synthase (PGP or PGPS)	2	2	Zosma302g00060 Zosma353g00080	2	Spipo22G0034100
Cardiolipin synthase (CLS)	1	1	Zosma7g00230	1	Spipo14G0051000
Oleate delta12-desaturase (FAD2)	1	1	Zosma132g00220	1	Spipo0G0076400
Linoleate delta15-desaturase (FAD3)	1	2	Zosma193g00160 Zosma114g00740	2	Spipo23G0035900 Spipo23G0036000
4) Mitochondrial fatty acid synthesis					
Mitochondrial fatty acid and lipoic acid synthesis					
Malonyl-CoA synthase (MCS)	1	1	Zosma5g00290	1	Spipo14G0018000
Malonyl-CoA : ACP malonyltransferase (MCMT)	1	1	Zosma141g00260	1	Spipo22G0028800
Ketoacyl-ACP synthase (KAS II)	1	1	Zosma136g00210	4	Spipo5G0018100 Spipo7G0061000 Spipo0G0033200 Spipo7G0013700
Ketoacyl-ACP reductase (KAR)	1	2	Zosma181g00100 Zosma224g00060	2	Spipo12G0037200 Spipo10G0020500
Enoyl-ACP reductase (ER)	1	1	Zosma118g00310	1	Spipo23G0030600
Lipoate synthase (LS)	1	1	Zosma393g00050	3	Spipo8G0016400 Spipo32G0005700 Spipo8G0016100
Lipoyltransferase (LT)	1	1	Zosma68g00370	2	Spipo2G0116200 Spipo27G0004100
Acyl carrier protein (ACP)	3	2	Zosma408g00070 Zosma408g00060	6	Spipo22G0010600 Spipo3G0042500 Spipo0G0087500

					Spipo3G0052600 Spipo6G0053200 Spipo2G0080400
UDP-3-O-(3-hydroxymyristoyl)glucosamine N-acyltransferase (LpxD)	2	1	Zosma19g01230	1	Spipo1G0122500
UDP-N-acetylglucosamine acyltransferase (LpxA)	1	1	Zosma76g00820	1	Spipo21G0031100
UDP-3-O-acyl N-acetylglucosamine deacetylase (LpxC)	5	1	Zosma27g01380	1	Spipo2G0021000
Lipid-A-disaccharide synthase (LpxB)	1	1	Zosma114g01090	1	Spipo5G0027600
Tetraacyldisaccharide 4'-kinase (LpxK)	1	1	Zosma91g00170	1	Spipo7G0024300
3-deoxy-D-manno-octulosonic acid (Kdo) transferase (KtdA)	1	1	Zosma165g00380	1	Spipo23G0009400
Mitochondrial phospholipid synthesis					
Glycerol-3-phosphate acyltransferase (GPAT)	10	10	Zosma29g01540 Zosma7g00950 Zosma379g00150 Zosma82g00900 Zosma28g00690 Zosma93g00640 Zosma112g00040 Zosma88g00100 (partial) Zosma129g00190 (partial) Zosma65g01010	12	Spipo30G0006700 Spipo23G0004300 Spipo11G0054200 Spipo23G0014600 Spipo32G0011400 Spipo6G0073200 Spipo3G0095200 Spipo0G0136700 Spipo6G0017900 Spipo8G0027600 Spipo6G0017600 Spipo7G0013300
1-Acylglycerol-3-phosphate acyltransferase (LPAAT)	5	4	Zosma21g00930 Zosma6g00580 Zosma125g00200 Zosma9g01250	7	Spipo6G0030100 Spipo7G0018900 Spipo24G0004600 Spipo4G0068200 Spipo7G0051900 Spipo24G0001100 Spipo3G0111400
CDP-DAG synthase (CDP-DAGS)	6	2	Zosma1g01870 Zosma26g01260	2	Spipo2G0035400 Spipo16G0012300
Phosphatidylglycerol-phosphate synthase (PGPS)	2	2	Zosma302g00060 Zosma353g00080	1	Spipo22G0034100
Phosphatidylglycerol-phosphate phosphatase (PGPP) ***, identified by blast with sequence from yeast.**	?	1*	Zosma389g00130	1	Spipo7G0024500
Cardiolipin synthase (CLS)	1	1	Zosma7g00230	1	Spipo14G0051000
Cardiolipid-specific deacylase	1	1	Zosma72g00440	1	Spipo16G0022500
Phosphatidylserine decarboxylase (PSD)	3	1	Zosma31g01370 Zosma1g00140	2	Spipo11G0033700 Spipo4G0094200
5) Sphingolipid biosynthesis and fate of ceramide in the cells (also see defense genes)					
Long chain base (LCB)	3	4	Zosma17g00630 Zosma2g01090 Zosma37g00450 Zosma69g00760	4	Spipo17G0023700 Spipo0G0074000 Spipo0G0154400 Spipo0G0087700
3-ketodihydrosphinganine (KSR) reductase	2	3	Zosma67g00550 Zosma2g01050 Zosma123g00630	3	Spipo30G0013800 Spipo30G0013900 Spipo15G0000500
Sphingobase C4-hydroxylase (SBH)	2	4	Zosma20g01280 Zosma2g02590 Zosma248g00160 Zosma161g00220	11	Spipo2G0040900 Spipo15G0000600 Spipo4G0023300 Spipo0G0154100 Spipo15G0000100 Spipo0G0054300 Spipo15G0000200 Spipo14G0059500 Spipo0G0140900 Spipo0G0061400 Spipo21G0028100
Dihydrosphingosine delta-4 desaturase (DSD)	1	1	Zosma8g01130	1	Spipo1G0037700
Sphingosine transfer protein	2	2	Zosma30g00840 Zosma121g00660	1	Spipo8G0026500
Sphingosine kinase	3	2	Zosma415g00070 Zosma10g00890	2	Spipo8G0015200 Spipo1G0110100
Long chain base 1-phosphate phosphatase	2	2	Zosma46g00660 Zosma75g00110	2	Spipo2G0062800 Spipo23G0033800
Ceramide synthase (CS)	3	5	Zosma55g00060	5	Spipo12G0013700

			Zosma54g01160 Zosma4g00140 Zosma135g00230 Zosma25g01310		Spipo10G0040500 Spipo25G0017000 Spipo4G0009000 Spipo15G0019900
Sphingobase delta8-desaturase (SLD)	2	1	Zosma14g00770	1	Spipo26G0009900
Fatty acid 2-hydroxylase (FAH)	2	2	Zosma82g00620 Zosma228g00240	6 **Several duplicat-ions**	**Spipo13G0050300** **Spipo13G0050100** **Spipo13G0049800** **Spipo30G0013000** **Spipo0G0178000** **Spipo13G0049900**
Glucosylceramide synthase (UDP-glucose-dependent) (GCS)	1	1	Zosma8g00900	2	Spipo5G0072600 Spipo0G0016000
Inositolphosphorylceramide synthase (or phosphatidylinositol:ceramide inositolphosphotransferase)	3	3	Zosma7g01150 Zosma162g00240 Zosma99g00350	1	Spipo7G0008900
Glucosylceramide glucosidase (GCG) *** identified by blast with sequence from human.**	?	3*	Zosma24g01200 Zosma105g00720 Zosma161g00040	2	Spipo1G0007300 Spipo4G0077300
Ceramidase (or alkaline phytoceramidase) (CES)	1	2	Zosma182g00290 Zosma93g00540	1	Spipo14G0000400
Ceramide kinase (CERK)	1	1	Zosma203g00160	1	Spipo16G0001300
Dihydrosphingosine phosphate lyase (DPL)	1	1	Zosma145g00170	1	Spipo4G0054300
6) Synthesis of jasmonates and other oxylipins (also see defense genes)					
Lipoxygenase (LOX)	6	6	Zosma23g00670 Zosma20g00750 Zosma77g00290 Zosma303g00020 Zosma125g00680 Zosma48g00760 (partial)	11 **Several duplicat-ions**	**Spipo7G0050500** **Spipo0G0030100** **Spipo28G0005500** **Spipo15G0045200** **Spipo2G0068200** **Spipo4G0070100** **Spipo28G0005600** **Spipo2G0068500** **Spipo28G0005700** **Spipo28G0005300** **Spipo0G0094800**
PLAT/LH2 family proteins	3	1	Zosma91g00840	3	Spipo20G0025800 Spipo1G0084100 Spipo15G0010400
CYP74 family enzymes (family 74 of the cytochrome P450s), including Allene oxide synthase (AOS), hydroperoxyde lyase (HPL), and divinyl ether synthase (DES)	1	4	Zosma1g03370 Zosma9g01180 Zosma52g01010 Zosma52g01020	6	Spipo2G0116800 Spipo3G0112400 Spipo2G0116500 Spipo2G0116700 Spipo2G0116600 Spipo25G0002900
Allene oxide cyclase (AOC)	4	1	Zosma5g00590	2	Spipo26G0008700 Spipo14G0019700
12-oxophytodienoate reductase (12-OPR)	3	5	Zosma37g01460 Zosma123g00100 Zosma29g00550 Zosma153g00060 Zosma114g00650	7 **Massive gene duplicat-ions**	**Spipo4G0079600** **Spipo4G0079700** **Spipo4G0079300** **Spipo15G0046700** **Spipo4G0079500** **Spipo4G0079100** **Spipo4G0079200** **Spipo4G0079400**
4CLLs (4-coumarate--CoA ligase like), including OPC-8:CoA ligase	9	13	Zosma46g00040 Zosma153g00020 Zosma7g00380 Zosma204g00130 Zosma107g00380 Zosma89g00660 Zosma7g00110 Zosma123g00330 Zosma10g00620 Zosma182g00220 Zosma7g01830 Zosma506g00010 (partial) Zosma604g00010 (partial)	13	Spipo1G0075100 Spipo4G0066400 Spipo11G0055200 Spipo5G0067700 Spipo5G0006100 Spipo12G0064500 Spipo10G0010700 Spipo1G0017200 Spipo12G0053800 Spipo17G0024700 Spipo17G0024300 Spipo1G0050800 Spipo1G0032700
7) Triacylglycerols (TAGs) synthesis					

Diacylglycerol Cholinephosphotransferase	2	1	Zosma62g00340	2	Spipo1G0032500 Spipo16G0036700
Lysophospholipid acyltransferase (LPLAT)	5	5	Zosma43g00130 Zosma18g00180 Zosma25g01630 Zosma361g00170 Zosma97g00530	4	Spipo17G0017700 Spipo24G0004600 Spipo1G0116800 Spipo1G0116900
N-acylphosphatidylethanolamine synthase (NAPE synthase)	1	2	Zosma192g00110 Zosma2g03470	2	Spipo5G0035000 Spipo6G0046900
Phospholipase A2	5	8	Zosma53g00190 Zosma137g00170 Zosma111g00460 Zosma114g00160 Zosma84g00330 Zosma118g00390 Zosma218g00230 Zosma218g00240	4	Spipo16G0025400 Spipo15G0027600 Spipo22G0002900 Spipo22G0003200
Oleate delta12-desaturase (FAD2)	1	1	Zosma132g00220	1	Spipo0G0076400
Linoleate delta15-desaturase (FAD3)	1	2	Zosma193g00160 Zosma114g00740	2	Spipo23G0035900 Spipo23G0036000
Phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT)	1	2	Zosma68g00320 Zosma314g00020	1	Spipo6G0000100
Phospholipid:diacylglycerol acyltransferase (PDAT)	2	2	Zosma81g01180 Zosma49g00460	2	Spipo20G0011900 Spipo0G0156700 (partial)
Acyl-CoA:diacylglycerol acyltransferase (DGAT type 1)	1	1	Zosma110g00130	1	Spipo21G0027500
Acyl-CoA:diacylglycerol acyltransferase (DGAT type 2) OR monoacylglycerol acyltransferase (MAGAT)	1	1	Zosma462g00110	2	Spipo28G0006400 Spipo5G0061000
Choline kinase	3	2	Zosma97g00820 Zosma98g00370	1	Spipo21G0016000 Spipo21G0002900
Choline-phosphate cytidylyltransferase (CCT)	2	1	Zosma1g01900	2	Spipo13G0054600 Spipo5G0070600
Steroleosin-like proteins	8	1	Zosma42g00440	2	Spipo4G0087400 Spipo4G0087800
Oil-body oleosin	8	1	Zosma176g00380	4	Spipo18G0028200 Spipo18G0024200 Spipo21G0032000 Spipo17G0007900
Caleosin	8	2	Zosma215g00300 Zosma2g02520	6 **duplication**	**Spipo16G0047000** **Spipo15G0035100** **Spipo15G0035000** **Spipo15G0034900** **Spipo15G0034700** **Spipo15G0034800**
8) **Long chain fatty acid elongation and cuticular wax biosynthesis (also see cell walls)**					
ATP Citrate lyase A subunit (ATP-CL)	3	1	Zosma467g00110	1	Spipo6G0019800
UDP-N-acetylglucosamine acyltransferase	1	1	Zosma76g00820	1	Spipo21G0031100
Acetyl-CoA carboxylase (ACC)	2	1	Zosma224g00120	1	Spipo15G0009000
Long-chain acyl-CoA synthetase (LACS)	9	6	Zosma9g01190 Zosma40g00520 Zosma166g00090 Zosma315g00010 Zosma35g00070 Zosma270g00010 (partial)	7	Spipo18G0011700 Spipo26G0016500 Spipo2G0116100 Spipo3G0077600 Spipo22G0000300 Spipo7G0043300 Spipo13G0009500

3-ketoacyl-CoA synthase (KCS)	21	19	Zosma25g00560 Zosma573g00080 Zosma178g00240 Zosma10g00710 Zosma53g01070 Zosma83g00160 Zosma1g02610 Zosma81g01030 Zosma91g00330 Zosma10g00770 Zosma91g00320 Zosma319g00050 Zosma58g00510 Zosma1g01290 Zosma213g00110 Zosma34g00210 Zosma34g00080 Zosma206g00090 Zosma103g00240 Zosma330g00060 Zosma114g00150	20	Spipo11G0044800 Spipo21G0006400 Spipo14G0001700 Spipo2G0058800 Spipo2G0058600 Spipo8G0072000 Spipo25G0006100 Spipo0G0125400 Spipo9G0002800 Spipo2G0030600 Spipo3G0034400 Spipo0G0109400 Spipo0G0047000 Spipo10G0000700 Spipo24G0001000 Spipo2G0114100 Spipo4G0008300 Spipo15G0022200 Spipo1G0007500 Spipo1G0007600
Ketoacyl-CoA reductase (KCR)	2	4	Zosma194g00130 Zosma72g00140 Zosma17g01010 Zosma17g01030	11 **Massive gene duplicat-ions**	**Spipo4G0035000** **Spipo0G0147600** **Spipo7G0063800** **Spipo7G0063000** **Spipo7G0063300** **Spipo7G0062700** **Spipo7G0063200** **Spipo7G0063500** **Spipo7G0062800** **Spipo7G0063400** **Spipo7G0063100**
Hydroxyacyl-CoA dehydratase (HACD)	2	2	Zosma1g03480 Zosma42g00040	2	Spipo9G0029200 Spipo22G0035400
(trans)enoyl-CoA reductase (ECR)	1	1	Zosma2g00080	3	Spipo23G0019700 Spipo0G0125000 Spipo23G0019500
Lipid-A-disaccharide synthase	1	1	Zosma114g01090	1	Spipo5G0027600
Bifunctional wax ester synthase / diacylglycerol acyltransferase	11	8	Zosma26g00060 Zosma2g01620 Zosma313g00030 Zosma230g00210 Zosma322g00010 Zosma18g00750 Zosma76g00050 Zosma76g00040	7	Spipo0G0145600 Spipo31G0001800 Spipo6G0029300 Spipo6G0029000 Spipo6G0028800 Spipo19G0000500 Spipo0G0072200 (partial)
Midchain alkane hydroxylase (MAH; cytochrome P450s CYP96A)	13	3	Zosma122g00510 Zosma122g00500 Zosma122g00550	4	Spipo21G0016800 Spipo21G0016900 Spipo21G0016600 Spipo21G0017000
ABC transporter G family members	29	At least 9	Zosma4g01300 Zosma19g01370 Zosma291g00200 Zosma55g00700 Zosma6g02070 Zosma6g02080 Zosma129g00120 Zosma33g00040 Zosma291g00190	15	Spipo1G0040500 Spipo5G0073600 Spipo22G0001000 Spipo22G0000900 Spipo5G0073300 Spipo17G0040700 Spipo5G0073400 Spipo6G0068700 Spipo19G0025400 Spipo23G0028300 Spipo0G0088900 Spipo0G0012600 Spipo26G0017900 Spipo14G0023400 Spipo3G0062500
9) Pathways for the synthesis of the cutin and suberin (see cell walls)					
Long-chain acyl-CoA synthetase (LACS)	9	6	Zosma9g01190 Zosma40g00520 Zosma166g00090 Zosma315g00010 Zosma35g00070 Zosma270g00010 (partial)	7	Spipo18G0011700 Spipo26G0016500 Spipo2G0116100 Spipo3G0077600 Spipo22G0000300 Spipo7G0043300 Spipo13G0009500
Fatty acyl omega-hydroxylase (FAH;	6	5	Zosma26g00330	3	Spipo11G0054300

cytochromes P450s CYP86A and CYP86B)			Zosma16g00400 Zosma28g00680 Zosma57g00510 Zosma58g00840		Spipo0G0080900 Spipo13G0037200
Fatty acyl in-chain hydroxylase (FAIH; cytochrome P450s CYP77A)	1	1	Zosma2g00860	1	Spipo9G0033700
Omega-hydroxy fatty acyl dehydrogenase (HFADH)	1	4	Zosma76g00800 Zosma1g02060 Zosma37g00840 Zosma133g00060	3	Spipo21G0026600 Spipo0G0174400 Spipo5G0006500
Glycerol-3-phosphate acyltransferase (GPAT)	10	10	Zosma29g01540 Zosma7g00950 Zosma379g00150 Zosma82g00900 Zosma28g00690 Zosma93g00640 Zosma112g00040 Zosma88g00100 (partial) Zosma129g00190 (partial) Zosma65g01010	12	Spipo30G0006700 Spipo23G0004300 Spipo11G0054200 Spipo23G0014600 Spipo32G0011400 Spipo6G0073200 Spipo3G0095200 Spipo0G0136700 Spipo6G0017900 Spipo8G0027600 Spipo6G0017600 Spipo7G0013300
Hydrolase-like protein	1	2	Zosma4g01080 Zosma208g00060	1	Spipo1G0059400
Feruloyl Transferase (FT)	1	2	Zosma27g00520 Zosma225g00300	2	Spipo4G0058100 Spipo9G0059300
Fatty acyl-CoA reductase (FAR)	8	7	Zosma100g00270 Zosma166g00430 Zosma2g00840 Zosma27g01030 Zosma52g00750 Zosma487g00050 Zosma1g02800	3	Spipo0G0006600 Spipo4G0046600 Spipo0G0006700
Lipid transfer proteins (LTP)	70	At least 24	Zosma1g01320 Zosma46g00710 Zosma27g00660 Zosma93g01030 Zosma177g00070 Zosma36g00160 (partial) Zosma27g00650 Zosma111g00340 Zosma48g00740 Zosma155g00040 Zosma121g00270 Zosma16g01580 Zosma19g00250 Zosma121g00250 Zosma121g00240 Zosma166g00070 Zosma81g00400 Zosma121g00380 Zosma224g00100 Zosma9g00580 (partial) Zosma27g01220 Zosma49g00180 Zosma74g00610 Zosma15g00830	A least 10	Spipo9G0060100 Spipo2G0061600 Spipo9G0001300 Spipo15G0009100 Spipo1G0111700 Spipo9G0004600 Spipo9G0004800 Spipo9G0004400 Spipo5G0071300 Spipo3G0016300
Polyester synthase (putative cutin synthase)	2	2	Zosma15g01170 Zosma44g01120	2	Spipo0G0133400 Spipo16G0045200
Phenylalanine ammonia-lyase (PAL)	4	4	Zosma445g00020 Zosma69g00670 Zosma115g00180 Zosma49g00480	3	Spipo11G0025500 Spipo1G0003500 Spipo15G0044700
Cinnamic acid 4-hydroxylase (C4H)	1	3	Zosma48g00790 Zosma144g00200 Zosma56g01120	3	Spipo6G0058400 Spipo29G0013100 Spipo12G0063200
4-coumaric acid:CoA ligase (4CL)	At least 4	At least 4 (see also candidates for OPC-8:CoA ligase)	Zosma46g00040 Zosma153g00020 Zosma7g00380 Zosma204g00130 Zosma107g00380 Zosma89g00660 Zosma7g00110 Zosma123g00330 Zosma10g00620 Zosma182g00220	At least 8	Spipo4G0066400 Spipo11G0055200 Spipo5G0067700 Spipo5G0006100 Spipo12G0064500 Spipo10G0010700 Spipo1G0017200 Spipo12G0053800

			Zosma7g01830 Zosma506g00010 (partial) Zosma604g00010 (partial)		
Hydroxycinnamoyl-CoA shikimate hydroxycinnamoyl transferase (HCT or HST)	1	2	Zosma269g00050 Zosma50g00170		Spipo12G0048000 Spipo12G0047600
p-coumaroyl shikimate 3-hydroxylase (C3H, CYP98A3)	1	1	Zosma20g01450	1	Spipo6G0031900
Caffeoyl CoA 3-O-methyltransferase (CCoAOMT-like)	7	2	Zosma18g00720 Zosma251g00290	1	Spipo17G0012000
Cinnamoyl-CoA reductase (CCR)	2	2	Zosma45g00330 Zosma16g01350	11 proteins annotated as CRR **Massive gene duplication**	Spipo0G0185100 Spipo11G0026200 Spipo6G0037000 Spipo23G0040600 **Spipo28G0006800** **Spipo28G0002300** **Spipo28G0007200** **Spipo28G0006600** **Spipo28G0007100** **Spipo28G0007000** **Spipo28G0007500**
Ferulate 5-hydroxylase (F5H)	1	2	Zosma77g00240 Zosma427g00120	1	Spipo16G0010200
Caffeic acid/5-hydroxyferulic acid O-methyltransferase (COMT)	1	1	Zosma4g02110	2	Spipo1G0078000 Spipo1G0078200
Cinnamyl alcohol dehydrogenase (CAD-like)	9	4	Zosma195g00200 Zosma70g00830 Zosma203g00140 Zosma2g03110	4	Spipo12G0004300 Spipo1G0069500 Spipo17G0012300 Spipo2G0124600
Hydroxycinnamaldehyde dehydrogenase (HCALDH)	1	2	Zosma84g00060 Zosma65g00420	1	Spipo21G0015000
Aliphatic Suberin Feruloyl Transferase (ASFT1)	1	2	Zosma225g00300 Zosma27g00520	1	Spipo4G0058100
10) Fatty acid catabolism					
Triacylglycerol lipases (TAGL)	15	11	Zosma211g00130 Zosma93g00570 Zosma42g00750 Zosma42g00730 Zosma733g00030 Zosma112g00080 Zosma306g00130 Zosma156g00360 Zosma219g00070 Zosma155g00220 Zosma81g00830	24 Several partial sequences	Spipo3G0044300 Spipo3G0053300 Spipo3G0053200 Spipo25G0006800 Spipo13G0027500 Spipo7G0046100 Spipo7G0057100 Spipo19G0031600 Spipo19G0031500 Spipo3G0040800 Spipo32G0012500 Spipo0G0106600 Spipo0G0106700 Spipo5G0049300 Spipo1G0011500 Spipo26G0019000 Spipo4G0067900 Spipo0G0056100 Spipo3G0058900 Spipo11G0021400 Spipo3G0059100 Spipo32G0012100 Spipo8G0027000 Spipo1G0011500
Monoacylglycerol lipase (MAGL)	16	13	Zosma183g00030 Zosma56g01030 Zosma99g00820 Zosma356g00080 Zosma79g00660 Zosma194g00230 Zosma111g00070 Zosma30g00920 Zosma17g00940 Zosma78g00040 Zosma183g00340 Zosma12g00850 (p Zosma12g00840 (p	12	Spipo6G0048200 Spipo29G0020700 Spipo23G0027000 Spipo23G0026900 Spipo19G0035600 Spipo19G0035400 Spipo7G0000400 Spipo1G0001500 Spipo21G0007600 Spipo2G0018800 Spipo21G0035700 Spipo0G0170300
Fatty acid/acyl-CoA transporter	2	2	Zosma40g00440 Zosma78g00570	2	Spipo26G0017200 Spipo4G0100800
Acyl-CoA synthetase (LACS)	9	6	Zosma9g01190	7	Spipo18G0011700

			Zosma40g00520 Zosma166g00090 Zosma315g00010 Zosma35g00070 Zosma270g00010 (partial)		Spipo26G0016500 Spipo2G0116100 Spipo3G0077600 Spipo22G0000300 Spipo7G0043300 Spipo13G0009500
Acyl-CoA thioesterase (ACT)	4	2	Zosma16g00530 Zosma4g01190	3	Spipo11G0053600 Spipo11G0053500 Spipo4G0085300
Acyl-CoA oxidase (ACX)	6	5	Zosma42g01110 Zosma42g01120 Zosma89g01280 Zosma76g01030 Zosma19g00040 (partial)	3	Spipo16G0000800 Spipo20G0028000 Spipo14G0024600
Acyl-CoA dehydrogenase	1	1	Zosma56g00880	1	Spipo3G0031300
Multifunctional protein (MFP)	2	3	Zosma38g00040 Zosma5g00220 Zosma200g00110	2	Spipo14G0018200 Spipo3G0033000
Enoyl-CoA hydratase	3	3	Zosma56g00210 Zosma3g00180 Zosma66g00420	3	Spipo0G0135900 Spipo6G0016200 Spipo1G0073500
Dienoyl-CoA reductase	1	1	Zosma69g00380	0 No candidate found in S.p. genome. All <29%	-
Ketoacyl-CoA thiolase (KAT)	3	2	Zosma50g00960 Zosma221g00270	3	Spipo1G0017800 Spipo25G0018800 Spipo5G0021400
Enoyl CoA isomerase (ECI)	3	1	Zosma84g00400	1	Spipo18G0019700
Isovaleryl-CoA dehydrogenase	1	1	Zosma193g00150	1	Spipo29G0018500
Enoyl-coA hydratase / 3-hydroxyisobutryl-CoA hydrolase	4	6	Zosma101g00230 Zosma91g00210 Zosma85g00320 Zosma208g00160 Zosma2g02300 Zosma2g02290	4	Spipo1G0086200 Spipo25G0001800 Spipo11G0009000 Spipo21G0013100
α-dioxygenase-peroxidase	2	1	Zosma69g00230 Zosma69g00240 (partial)	2	Spipo0G0180000 (partial) Spipo5G0040500 (partial)
Fatty acid alcohol oxidase	4	3	Zosma2g01470 Zosma122g00890 Zosma206g00050	5	Spipo11G0001600 Spipo30G0014900 Spipo30G0015200 Spipo1G0005800 Spipo15G0002200 (partial)

10 Cell wall related traits

Polysaccharides constitute the major part of the cell wall and are classified into cellulose, hemicelluloses and pectin. The interaction among these three groups ensures a strong yet flexible set of structural properties. The interplay between cutin-cuticular waxes and suberin-lignin with (sulfated) polysaccharides in the cell wall involves a combination of compounds unique to seagrasses. In particular, the ecophysiological role of sulfated polysaccharides is hypothesized to provide mechanical flexibility in relation to wave action and osmoregulation through the creation of a polyanionic matrix (Kloareg & Quatrano 1988). Here we focus on the leaf tissue but it will be important to consider possible spatial differences in the occurrence of all of the cell wall components in the cell walls of the rhizome and roots.

10.1 Cutin, cuticular waxes, suberin

Most aerial organs of land plants (e.g. leaves, stems, floral organs, fruits) are coated with a protective cuticle that comprises primarily cutin and waxes. The fatty acid pathways for synthesis of cutin, cuticular waxes and suberin are of particular interest in *Z. marina* because of their involvement in the cell wall and the adaptation

to seawater. Cutin forms an insoluble polyester matrix overlaying the cell wall and serves as the main structural component of the cuticle. Intracuticular waxes are embedded in the cutin matrix, and epicuticular waxes, often in the form of crystals, cover the outer surface. The cutin-based cuticle, one of the largest biological interfaces in nature, is a major barrier against water loss and the first physical barrier encountered by most phytopathogens (Jetter *et al.* 2006). Suberin is a complex hydrophobic polymer and, unlike cutin, is deposited close to the plasma membrane in the cell wall of specific tissues. It functions as a barrier against uncontrolled water loss, ion movement and microbial aggression (Ranathunge *et al.* 2011). Suberization is locally induced upon wounding and in response to environmental stressors (Kolattukudy 2001).

In *Z. marina*, long chain fatty acid elongation and cuticular wax biosynthesis **(Supplementary Table S9.1 section 8)** involves the production of fatty acids and acyl-CoAs longer than 20 carbon atoms, saturated or mono-unsaturated. The fatty acid elongation and cuticular wax biosynthesis pathway is complete in *Zostera*, with number of genes for each enzymatic step similar to the situation in *Arabidopsis*, except for the mid-chain alkane hydroxylase (MAH), where a smaller number of genes has been identified in *Zostera.* Some differences are found in lipid transfer proteins (LPT), Cinnamoyl-CoA reductase (CCR) and Triacylglycerol lipases (TAGL) between *Zostera* and *Spirodela*.

Still, the full inventory of genes encoding enzymes involved in the synthesis of the cutin polyester and suberin aliphatic polyesters has yet to be identified in plants such as the omega-hydroxy fatty acyl epoxygenase (HFAE), the omega-hydroxy fatty acyl dehydrogenase (OFADH), and the epoxide hydrolase (EH).
Pathways and genes for the synthesis of these cutin/suberin genes in *Z. marina* are listed in **Supplementary Table S9.1 (section 9).** For suberin synthesis, there are two pathways to consider: the synthesis of the suberin aliphatic polyester fraction and the aromatic suberin monomers for the aliphatic domains. A further search revealed a full set of genes in *Z. marina* for the phenylpropanoid pathway implicating the second pathway.

10.2 Cellulose synthase superfamily

Enzymes of the cellulose synthase (CESA) family and CesA-like (CSL) families are responsible for the synthesis of celluloses and some hemicelluloses in plant cell walls. All of these proteins belong to Family 2 glycosyltransferases (GT2). Phylogenetic analyses have divided the CSL proteins into nine subfamilies: from CSLA to CSLI (Yin *et al.* 2014). Some CSL subfamilies have been functionally characterized: CSLA (mannan synthases) (Dhugga *et al.* 2004), CSLC (xyloglucan synthases) (Cocuron *et al.* 2007), CSLF (mixed-linkage glucan synthases) (Burton *et al.* 2006), and CSLH (mixed-linkage glucan synthases) (Doblin *et al.* 2009).

Zostera marina displays a large GT2 family (36 genes), fairly similar to that of *Arabidopsis* (42) and *Oryza* (47). Based on BLASTP, *Z. marina* possesses 12 CesA, 6 CSLA, 3 CSLC, 5 CSLD, 3 CSLE and 4 CSLG (**Supplementary Table S7.4, Supplementary Data 8**). CSLA, CSLC, CSLD are conserved in bryophytes, lycophytes and angiosperms, while CSLE are conserved in all angiosperms. The absence of the eudicot-specific CSLB is to be expected but the presence of CSLG in *Z. marina* was a surprise since these genes are usually thought to be restricted to eudicots. CSLF and CSLH, the mixed linkage glucan (MLG) synthases that are usually described as monocot-specific (Vega-Sanchez *et al.* 2013), are absent in *Z. marina*, which suggests the absence of MLG (β-1,3-1,4-glucans). Comparison with the genome of *Spirodela polyrhiza* (Wang *et al.* 2014) confirms the absence of CSLB, CSLF and CSLH. However, CLSG genes are present in *Z. marina* but not in *Spirodela,* which might be related to the adaptation to the marine environment.

10.3 Hemicelluloses

Hemicelluloses encompass several cell wall polysaccharides interacting with cellulose microfibrils and pectins(Carpita, N.C. and McCann, M. 2000), (Popper *et al.* 2011). *Z. marina* genome contains all the genes needed for the metabolism of the main hemicelluloses: arabinoxylans (GT8, GT14, GT47, GT61, GH3, GH10, GH51 and CE6), xyloglucans (GT2 CLSC, GT34, GT37 and GH16), gluco- and galacto-mannans (GT2 CSLA, GT34, GH5_7), callose (β-1,3-glucans; GT48, GH17, CBM43). Nevertheless, there are

significant variations in comparison to *Arabidopsis* and *Oryza* (Henrissat *et al.* 2001; Lombard *et al.* 2014)(**Supplementary Table S7.4**). In *Z. marina*, the genes involved in the recycling and modification of (arabino) xylans are reduced: GH10 xylanases, GH3 xylosidases, GH51 arabinofuranosidases). This is also the case for xyloglucans with half the GH16 xyloglucan endotransglucosylase/hydrolases. A major difference is the presence of only 2 GT37 xyloglucan fucosyltransferases, instead of 10 and 18 GT37 genes in *Arabidopsis* and *Oryza*, respectively. This suggests that the fine structure of *Zostera* xyloglucans differs. Although the number of callose synthases (GT48) is similar, *Z. marina* displays only half the GH17 β-1,3-glucanases (and β-1,3-glucan specific CBM43). See also **Supplementary Data 8** of CAZymes in *Z. marina* that provides detailed Gene ID, gene name, definition, family, EC number, best characterized hit, organism, GenBank ID and identity percentage.

10.4 Pectins

Pectin is an important cell wall polysaccharide allowing primary cell wall extension and plant growth. Along with the hemicelluloses they comprise the matrix in which cellulose microfibrils are embedded. In contrast to other cell wall metabolites, the number of pectin-related genes in Z. *marina* is not reduced as compared to land plants (**Supplementary Data 8**).

The number of biosynthetic genes (GT8 Galacturonosyltransferases (GAUTs), GT47 Rhamnogalacturonan II xylosyltransferases) are equivalent to *Arabidopsis* and *Oryza*. The catabolic genes are even expanded in comparison to the reference monocot *Oryza*: *Zostera* displays more GH28 polygalacturonases and PL1 pectate lyases (53 GH28 and 16 PL1 versus 41 GH28 and 12 PL1 in *Oryza*). Moreover *Zostera* has about twice the number of pectin methylesterases in comparison to *Oryza* (63 CE8 versus 38 CE8 in *Oryza*). This difference in pectin methylesterases is consistent with the fact that the *Z. marina* pectin (referred to as zosterin) has a very low degree of methylation in comparison to land plant pectin (Shibaeva *et al.* 1970). Therefore, the metabolism of pectin seems particularly crucial in *Z. marina*. The expansion of the CE8 family is very likely a key adaptation to marine environment because the low degree of methylation in pectin increases the polyanionic character of this matrix polymer. The presence of polyanionic polysaccharides (sulfated and uronic polysaccharides) is known to be important for ion homeostasis in marine algae and marine angisoperms (Kloareg & Mabeau 1987; Kloareg & Quatrano 1988; Aquino *et al.* 2005; Michel *et al.* 2010; Aquino *et al.* 2011; Popper *et al*. 2011).

10.5 Sulfated polysaccharides

As cell wall components in macroalgae or exopolysaccharides in microalgae, sulfated polysaccharides are found in all marine algae (green, red and brown macroalgae/seaweeds and microalgae/ phytoplankton). In contrast land plants do not produce sulfated polysaccharides. Fascinatedly, seagrasses do contain sulfated polysaccharides in their cell walls, whereas the aquatic/freshwater *Spirodela polyrhiza* does not.

Seagrass sulfated polysaccharides are reminiscent of the sulfated galactans found in red algae (Maeda *et al.* 1966; Aquino *et al.* 2005; Aquino *et al.* 2011). This suggests that the sulfation of cell wall polysaccharides was a needed step for angiosperms to readapt to the marine environment. Sulfate ions are abundant in seawater, but rare in terrestrial environment. Thus terrestrial plants have most likely lost these genes as a key adaptation to terrestrial constraints(Michel *et al.* 2010; Collen *et al.* 2013).

No homologues of carbohydrate sulfotransferases or sulfatases were found in the *Z. marina* genome. Instead, a unique multigenic family of aryl sulfotransferases (12 genes, see **Supplementary Table S10.1**) was identified. These genes are homologous to aryl sulfotransferases from terrestrial flowering plants, which are involved in the biosynthesis of sulfated flavonoids. Although these enzymes add sulfate groups to phenolic compounds they display almost no sequence similarity with carbohydrate sulfotransferases (Chapman *et al.* 2004). These aryl sulfotransferases display between 34 and 40% sequence identity with the characterized cytosolic sulfotransferase 12 from *Arabidopsis* (uniprot: P52839). There are no sulfotransferases in *Spirodela polyrhiza*.

Z. marina genome analysis confirms that plants have definitively lost classical carbohydrate sulfotransferases and sulfatases. There is no evidence of horizontal transfer of carbohydrate sulfotransferase and of sulfatase genes from marine algae. Nevertheless, *Z. marina* produces sulfated polysaccharides. Although theoretically, *Z. marina* could have evolved a completely new family of sulfotransferases, our analysis suggests that one or several aryl sulfotransferases have undergone functional change so as to be able to sulfate polysaccharides, since carbohydrate sulfatases are already known to be active on artificial aryl compounds such as methylumbelliferyl-sulfate (Hanson *et al.* 2004).

Supplementary Table S10.1 | Unique sulfotransferases in *Zostera marina*

Gene ID	Definition	Best characterized hit	organism	Uniprot ID	Identity (%)
Zosma104g00150	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	40
Zosma378g00070	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	38
Zosma637g00010	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	39
Zosma104g00160	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	38
Zosma378g00030	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	37
Zosma137g00410	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	37
Zosma167g00150	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	36
Zosma167g00070	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	37
Zosma167g00110	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	36
Zosma104g00170	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	37
Zosma141g00290	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	42
Zosma21g00400	Sulfotransferase	Cytosolic sulfotransferase 12	*A. thaliana*	P52839	34

11 Flowers and pollen

11.1 Pollen and pollination

Aside from the modifications of the female flower (no corolla, bifurcate stigma), hydrophilous pollination requires two critical adaptations that are shared among almost all seagrasses: the loss of the exine coat, which appears to be a critical factor in the transition from aerial to submerged pollination systems; and the evolution of filiform pollen that winds around the stigma for successful abiotic pollination under water. The hydrodynamics and biomechanics of filiform pollen and the success it confers on underwater pollination have been studied empirically in *Z. marina* by (Cox 1983) and reviewed in (Ackerman 2006).

The exine coat is a key innovation that led to the desiccation resistant pollen of terrestrial plants. Accordingly, amphibious freshwater taxa may have thinner exines. The absence of an exine is apparently extremely rare in aquatic alismatids (Furness 2013). *Spirodela* pollen has a well-defined exine. Exineless pollen is a feature of *Z. marina*, and almost all of the other seagrasses: *Phyllospadix, Amphibolis, Posidonia, Halophila, Cymodocea* and *Syringodium*. The only exception is *Enhalus acoroides* in the Hydrocharitaceae.

Filiform pollen is also extremely rare in freshwater taxa, most of which are surface pollinated (Furness 2013). *Spirodela* features a globose, ulcerated pollen grain. Filiform pollen has independently evolved within each of the seagrass lineages (Kuo & den Hartog 2006). It is found in all seagrasses with, again, the exception of *Enhalus acoroides* that is not hydrophilous (surface pollination). An intermediate, ellipsoidal form is found in *Halophila* spp. and a functionally filamentous "long-chain of spherical forms" is characteristic of *Thalassia* spp. (Ackerman 2006).

Pollination in terrestrial and aquatic/freshwater plants is typically insect- or wind-driven. In seagrasses pollination occurs entirely underwater/submarine (again with the exception of *Enhalus acoroides*). See discussion in (Les *et al.* 1997), as well as (Ackerman 2006). Contrary to some literature (reviewed in

(Ackerman 2006), surface pollination in seagrasses (e.g., those that periodically fall dry in the intertidal) is of limited importance relative to submarine pollination; likewise, underwater pollination in some aquatic relatives (i.e., Potamogetonaceae, Najadaceae and Zannichelliceae) is of limited importance relative to surface pollination. Although there are no insects in the sea and hence no insect pollinators, the possibility of animal-assisted pollination has been recently hypothesized. Micro-crustaceans and polychaete worms have been observed foraging on pollen in the tropical seagrass, *Thalassia testinudium*, with pollen attached to their bodies (Van Tussenbroek *et al.* 2012). Whether this contributes to pollination success remains to be tested.

Until now, nothing was known about the developmental program of this specialized, filiform pollen. We mined the *Z. marina* genome to determine which genes and pathways would eventually be involved in the synthesis of sporopollein (a still poorly chemically defined mixture of long chain aliphatic and aromatic polymers), which is the major constituent of pollen exine. We discovered that ten genes are missing among the many documented to be involved in male flower development in land plants (Ariizumi & Toriyama 2011), most often in *Arabidopsis* but also in rice and *Spirodela* (**Extended Data Table 3** and **Supplementary Table 11.1)**. Of these ten, four have recently been shown to act as a sporopollenin metabolon, with their products co-localized together in the Endoplasmic Reticulum (ER) (Lallemand *et al.* 2013). Most other genes also fit within the putative pathway allowing for biosynthesis of sporopollenin in the anther tapetum during male flower development (**Fig 3d**) (Ariizumi & Toriyama 2011). Two small gene families worth mentioning encode specific oleosins/glycine-rich proteins (McNeil & Smith 2010) and type III Lipid Transfer Proteins (LTP). Both are suggested to bind the sporopollenin precursors in *Arabidopsis*, after their secretion into the anther space from the tapetum, allowing their translocation to the pollen wall to achieve the build-up of the exine. While the existence or lack of an homologous oleosin gene in monocots cannot be anticipated due to the fast divergence of this family, a type III LTP is clearly seen in *Spirodela*, while absent from *Zostera*.

Supplementary Table 11.1 | Genes involved in pollen development of *Z. marina* compared to other angiosperms. The five genes encoding proteins associated on the ER-located sporopollenin metabolon in *Arabidopsis*[88] are highlighted in grey. The genes documented to be involved in pollen development in *Arabidopsis* or in rice were used as queries to find orthologs. Grey indicates genes associated on the ER-located sporopollenin metabolon. MF, male flowers; FFE, female flowers early; FFL, female flowers late; R, roots; L, leaves; NF-1, not found, supported by reciprocal Blast and phylogeny; NF-2, not found, single copy gene; amb, ambiguous because homologs are too similar to point to a specific ortholog. This is an expansion of **Extended Data Table 3**.

Gene Name	Symbol	*A. thaliana*	*O. sativa*	*S. polyrhiza*	*Z. marina*	Tissue expression in *Z. marina* (FPKM)				
						FFE	FFL	MF	R	L
ACYL-COA SYNTHETASE 5	ACOS5	At1g62940	Os04g24530	Sp12g0064500	NF-1					
POLYKETIDE SYNTHASE A	PKSA (LAP6)	At1g02050[1]	Os10g34360	Sp16g0013800	NF-1					
POLYKETIDE SYNTHASE B	PKSB (LAP5)	At4g34850	Os07g22850	Sp1g0062300	NF-1					
LESS ADHERENT POLLEN 3	LAP3	At3g59530	Os03g15710	Sp16g0030000	NF-1					
TETRAKETIDE a-PYRONE REDUCTASE 1	TKPR1 (DRL1)	At4g35420	Os08g40440	Sp10g0016700	NF-1					
TETRAKETIDE a-PYRONE REDUCTASE 2	TKPR2 (CCRL6)	At1g68540	Os01g03670	NF	NF-1					
CYTOCHROME P450 704B1/2	CYP704 B1[3] (CYP70 4B23)	At1g69500	Os03g07250	Sp2g0036600	Zm149g00275	NA	NA	NA	NA	NA
CYTOCHROME P450 703A2	CYP703 A2 (DEX2)	At1g01280	Os08g03682	Sp8g0027200	Zm121g00820	0.0	0.0	0.7	0.0	0.0
TYPE III LIPID TRANSFER PROTEINS	LTP3	At5g62080 At5g07230 At5g52160	Os08g43290 Os09g35700	Sp16g0007800	NF-1					

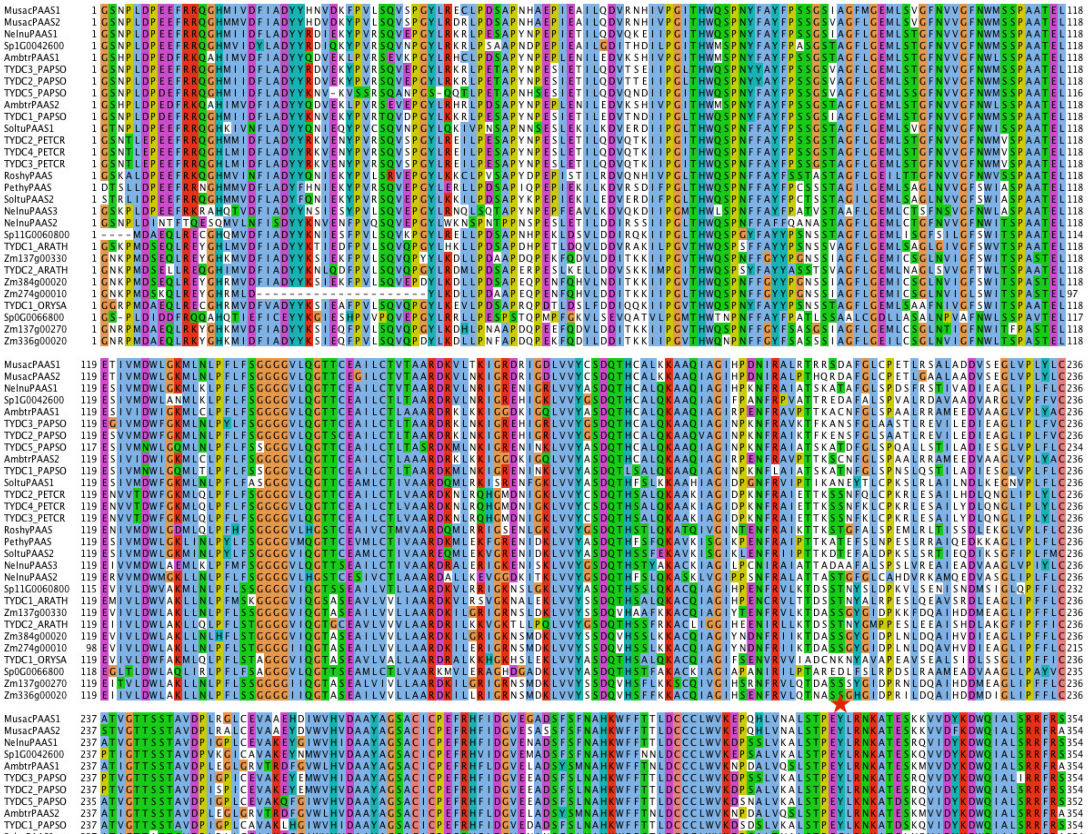
Type VII LIPID TRANSFERT PROTEIN	OsC6	NF	Os11g37280	Sp1g0111700	Zm350g00030	3.3	4.4	2.1	0.3	1.0
MALE STERILITY 1, PERSISTENT TAPETAL CELL1	MS1 (PTC1, HKM)	At5g22260	Os09g27620	Sp15g0041000[2]	Zm2g02440	0.2	0.2	0.2	0.2	0.2
MALE STERILITY 2	MS2 (FAR2, DPW)	At3g11980	Os03g07140	Sp4g0046600[2]	Zm100g00270	1.2	1.6	1.3	0.2	0.4
GA-regulated Myb-like Transcription Factor	GAMYB (MYB65, MYB33)	At3g11440 At5g06100	Os01g59660	Sp22g0020200	Zm6g00090	8.8	6.6	26.9	3.1	4.3
MALE MEIOCYTE DEATH 1	MMD1 (DUET)	At1g66170	Os03g50780	Sp20g0014200	Zm159g00130	0.0	0.2	1.6	0.0	0.0
ABC transporter G family member 15	ABCG15	At3g21090	Os05g13520	Sp22g0001000[4]	Zm291g00200	4.7	5.9	0.8	0.9	1.2
ABC transporter G family member 26	ABCG26 (WBC27)	At3g13220	Os06g40550	Sp25g0019600	Zm8g01840	0.0	0.3	0.3	1.7	0.2
FACELESS POLLEN-1, ECERIFERUM 3	FLP1 (ERC3, WAX2)	At5g57800	Os09g25850 Os02g08230 Os06g44300	Sp5g0009000[2]	NF-1					
ECERIFERUM 1, WAX DEFICIENT ANTHER 1	CER1 (WDA1, ERC1)	At1g02205	Os02g40784 Os10g33250	Sp0g0120500[2]	Zm37g00140 Zm215g00360	0 / 99.8	0 / 64.9	0.4 / 8.9	0.3 / 0.1	0 / 46.1
INAPERTURATE POLLEN 1	INP1	At4g22600	Os02g44250	Sp13g0036900	NF-2					
GLYCOSYLTRANSFERASE 1	GT1	At1g19710 At1g75420	Os01g15780	Sp14G0031700	Zm69g00440	10.1	8.1	41.7	224.1	62.4
CYSTEINE ENDOPEPTIDASE 1	CEP1[5]	At5g50260	Os08g44270 Os11g14900	Sp4g0036900 Sp11g0013300	NF-1					
NPK1-ACTIVATING KINESIN 1	NACK1 (HINKEL)	At1g18370	Os01g33040	Sp8g0065000	Zm111g00200	5.4	8.2	1.2	2.8	0.6
NPK1-ACTIVATING KINESIN 2	NACK2 (STUD, TETRASP)	At3g43210			Zm15g00280	0.0	0.0	0.1	0.0	0.0
DEFECTIVE IN EXINE FORMATION 1	DEX1	At3g09090	Os03g61050	Sp23g0006500	Zm7g01020	56.5	51.0	45.4	51.3	36.6
TAPETUM DEGENERATION RETARDATION / ABORTED MICROSPORES	TDR (AMS)	At2g16910	Os02g02820	Sp1g0068700 Sp1g0068800	Zm8g01850	0.0	0.0	0.4	4.2	0.6
ETERNAL TAPETUM 1, DELAYED TAPETUM DEGENERATION	EAT1 (DTD)	amb	Os04g51070	Sp6g0020000[2]	Zm123g00440	0.3	0.1	2.0	0.4	0.9
TDR INTERACTING PROTEIN 2	TIP2 (AtbHLH089, AtbHLH091)	At2g31210 At1g06170	Os01g18870	amb	amb					
TAPETAL DETERMINANT1	TPD1	At4g24972	Os12g28750 Os10g14020	Sp4g0058300 Sp7g0048400	Zm69g00720	0.0	0.0	0.9	0.0	0.0
EXCESS MICROSPOROCYTES 1, MULTIPLE SPOROCYTE1	EMS1 (EXS, MSP1)	At5g07280	Os01g68870	Sp4g0050100	Zm42g00660	0.4	0.7	0.8	0.4	0.4
DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION 1	TDF1 (MYB35)	At3g28470	amb	amb	amb					
DYSFUNCTIONAL TAPETUM 1	DYT1 (UDT1)	At4g21330	Os07g36460	amb	amb					
MALE STERILE 188	MS188 (MYB80)	At5g56110	Os04g39470	Sp4g0087200	Zm262g00100	5.9	4.6	42.5	9.4	2.8
MYB26/MALE STERILE35	MYB26 (MS35)	At3g13890	amb	amb	amb					

BARELY ANY MERISTEM 1 BARELY ANY MERISTEM 2	BAM1 BAM2	At5g65700 At3g49670	Os03g56270 Os07g04190	Sp2g0107200 Sp3g0108500	Zm161g00370 Zm9g01440	4.1 / 46.9	2.7 / 41.7	0.3 / 2.8	1 / 8.1	1.1 / 4
REDUCED MALE FERTILITY	RMF	At3g61730	Os01g11990	Sp8g0046600	Zm65g01040	0.0	0.0	0.7	0.0	0.0
β-KETOACYL REDUCTASE	KAR (KCR1, KCR2)	amb	amb	amb	amb					
RUPTURED POLLEN GRAIN 1	RPG1 (SWEET 8)	amb	amb	amb	amb					
KAONASHI 2	KNS2 (SPS2)	At5g11110	amb	amb	amb					
Fatty Acyl Thioesterase B	FATB	At1g08510	Os06g05130	Sp21g0008900[2]	Zm1g01370	115.1	104.8	213.7	129.3	124.3
Glycosyl transferase family GT31	B3GALT 7	At1g77810	Os02g35870	Sp4g0008000	Zm155g00180	4.4	1.4	354.3	10.5	2.2
SPOROCYTELESS/NOZZLE	SPL (NZZ)	At4g27330	NF	NF	NF					
NO EXINE FORMATION	NEF1	At5g13390	Os11g32470	Sp9g0054300[6]	Zm5g01490	28.4	23.6	14.7	25.3	18.8

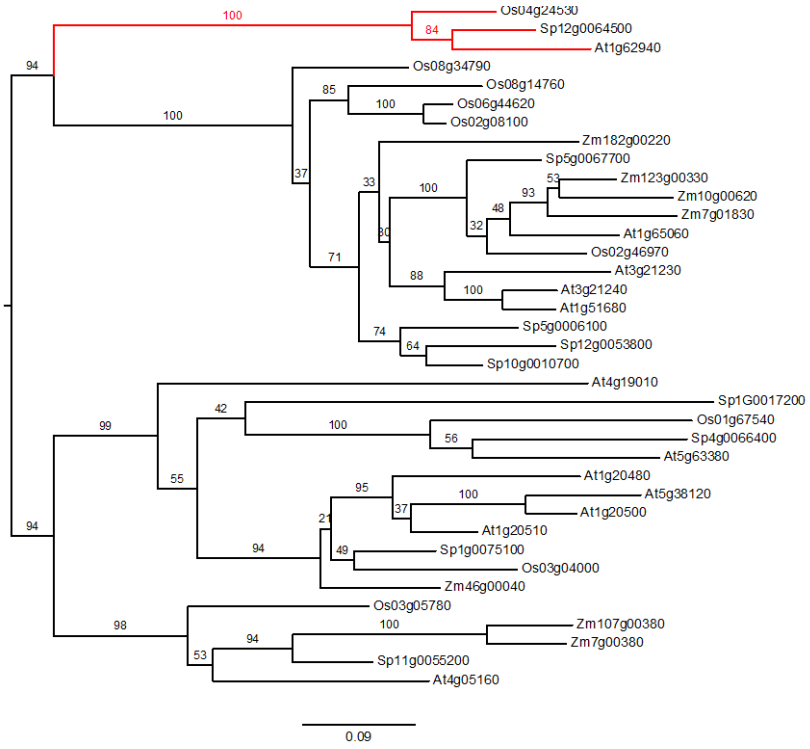
1. there is an additional, not documented paralog: At4g00040
2. partial sequence
3. CYP704B1 in Arabidopsis, CYP704B2 in rice
4. gene model corrected
5. this protein (and its orthologs) has a KDEL C-terminus
6. split gene. Merged with Sp9g00543200

In support of our findings, multiple sequence alignments and trees for Acyl-CoA Synthetase (ACOS5), KDEL-tailed cysteine endopeptidase (CEP1), faceless pollen1 (FLP1)/Eceriferum3 (CER3), Less adherent pollen3 (LAP3), lipid transfer proteins (LPTs) and tetraketide alpha-pyrone reductase (TKPRs) are presented in **Supplementary Fig. S11.1** (multi-part below).

I. Acyl-CoA Synthetase 5 (ACOS5)

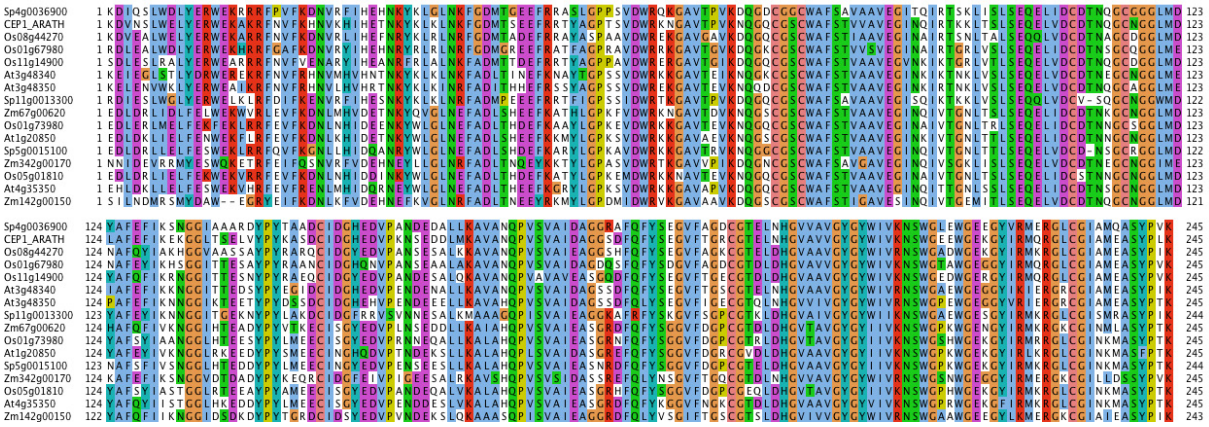


a, MUSCLE alignment of Acyl-CoA Synthetase 5 (ACOS5) from *Arabidopsis* (At1g62940) with paralogs and homologs from the rice, *Spirodela* and *Zostera* genomes. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.

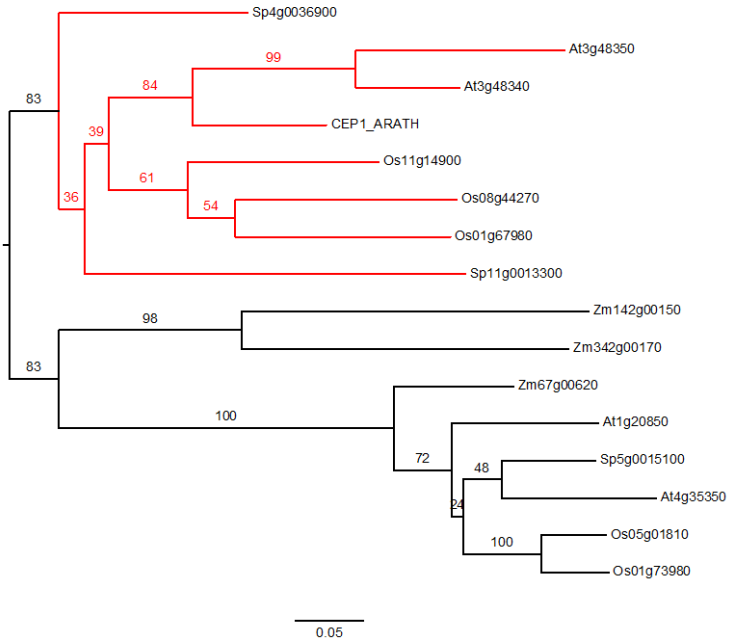


b, Maximum likelihood phylogenetic tree of Acyl-CoA Synthetase 5(ACOS5) calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. This phylogeny shows the orthologs of ACOS5 in rice and *Spirodela* (in red) but its absence in *Zostera marina*. Species abbreviations: *Arabidopsis* (ARATH), rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

II. KDEL-tailed cysteine endopeptidase CEP1

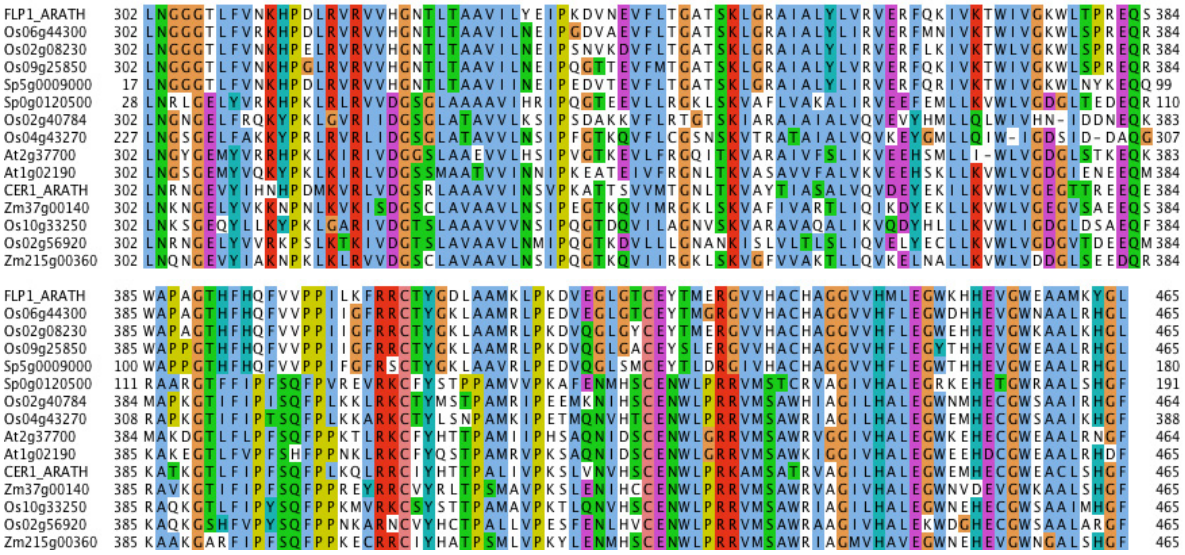


a, MUSCLE alignment of KDEL-tailed cysteine endopeptidase CEP1 from *Arabidopsis* and homologs from the rice, *Spirodela* and *Zostera* genomes. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.

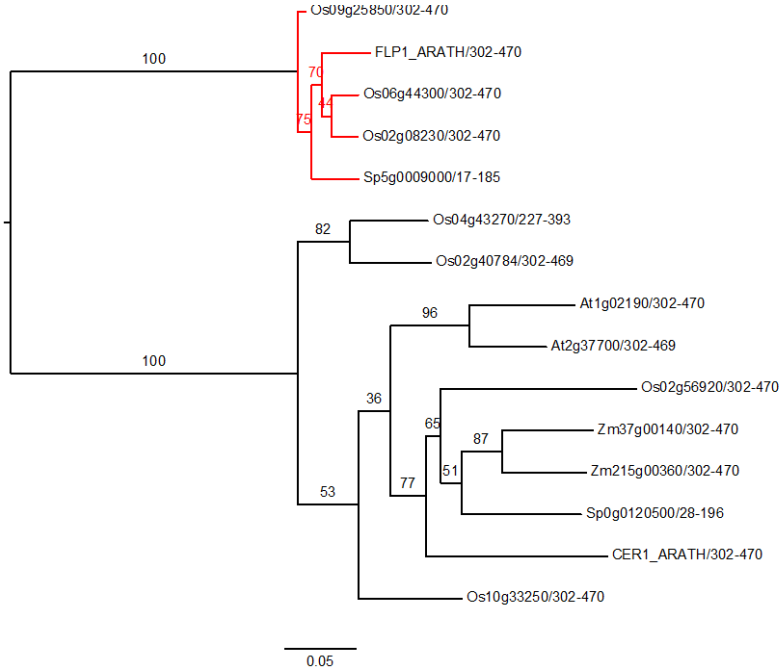


b, Maximum Likelihood phylogenetic tree of KDEL-tailed cysteine endopeptidase CEP1 calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000**.** Homologs of the KDEL-tailed cysteine endopeptidases CEP1 do occur in the rice and *Spirodela* genome (in red) but are not found in the *Zostera marina* genome. The C-terminal tetrapeptide is displayed next to each gene-ID showing a separate clustering of the proteases with and without a KDEL tail. Species abbreviations: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

III. FACELESS POLLEN 1 (FLP1) / ECERIFERUM 3 (CER3)

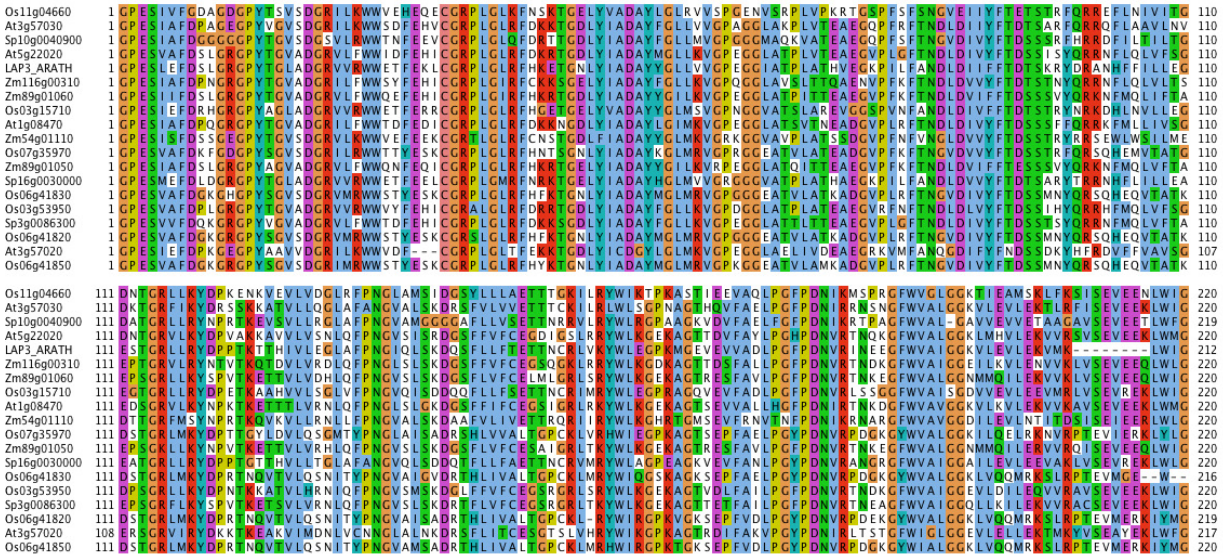


a, MUSCLE alignment of proteins encoded by the FACELESS POLLEN 1 (FLP1) / ECERIFERUM 3 (CER3) gene from *Arabidopsis thaliana* and homologs from the rice, *Spirodela polyrhiza* and *Zostera marina* genomes. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.

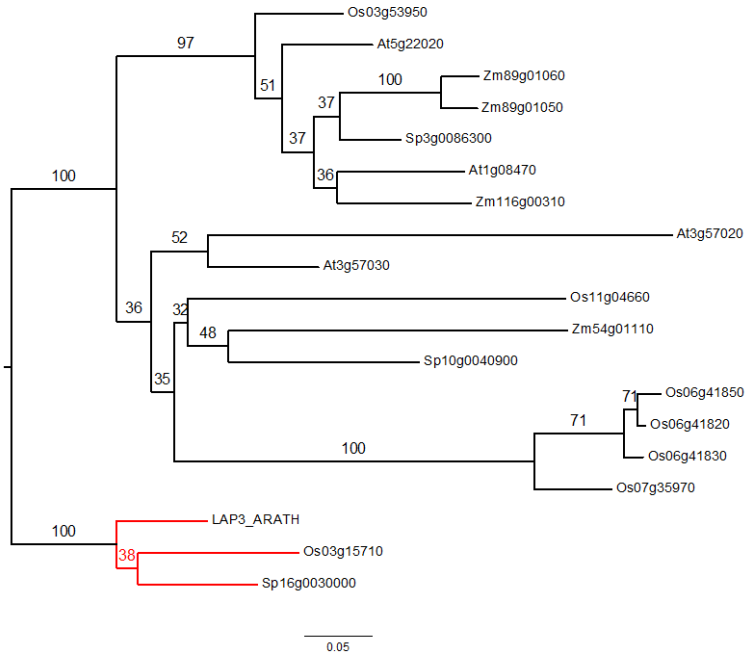


b, Maximum Likelihood phylogenetic tree of FLP1 calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. This phylogeny shows homologs of FLP1 in the rice and Spirodela genomes (in red) and its absence from *Z. marina.* Species abbreviations: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

IV. LESS ADHERENT POLLEN 3 (LAP3)



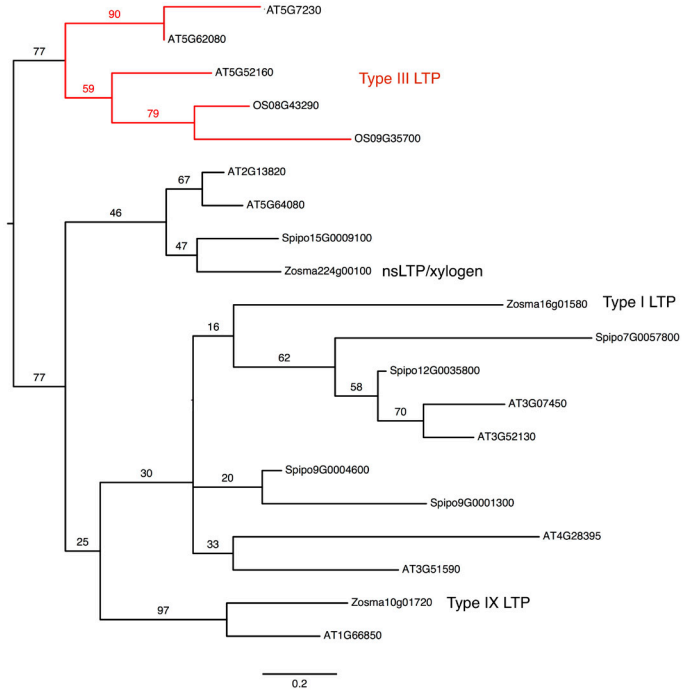
a, MUSCLE alignment of the Strictosidine Synthase-like 13 protein encoded by the LESS ADHERENT POLLEN 3 (LAP3) gene in *Arabidopsis thaliana* with homologs from the rice, *Spirodela polyrhiza* and *Zostera marina* genomes. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.



b, Maximum Likelihood phylogenetic tree of the LAP3 protein calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. This phylogeny shows homologs of LAP3 in rice and in *Spirodela* (in red) and its absence from *Z. marina.* Species abbreviations: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

V. LTPs

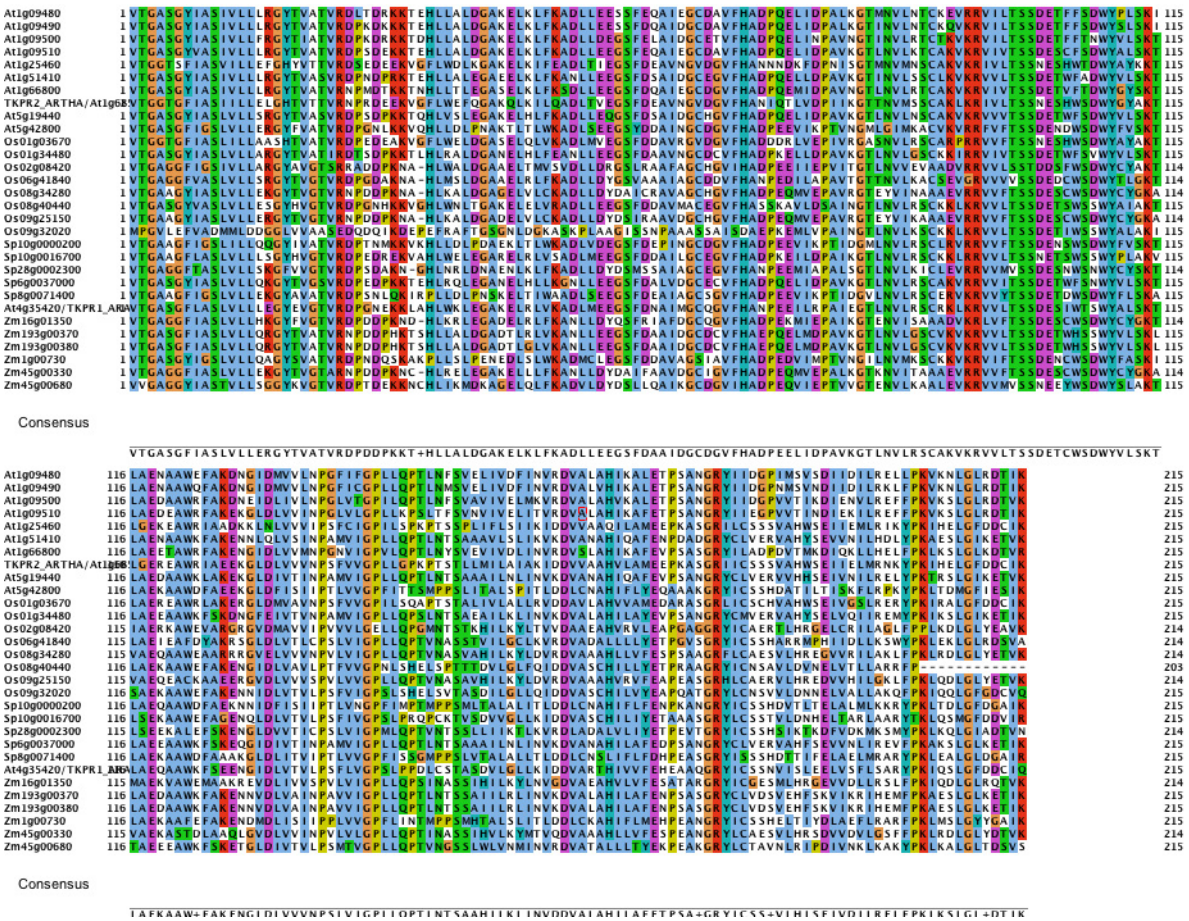
a, MUSCLE alignment of type III Lipid Transfer Proteins (LTPs) and other LTPs. Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.



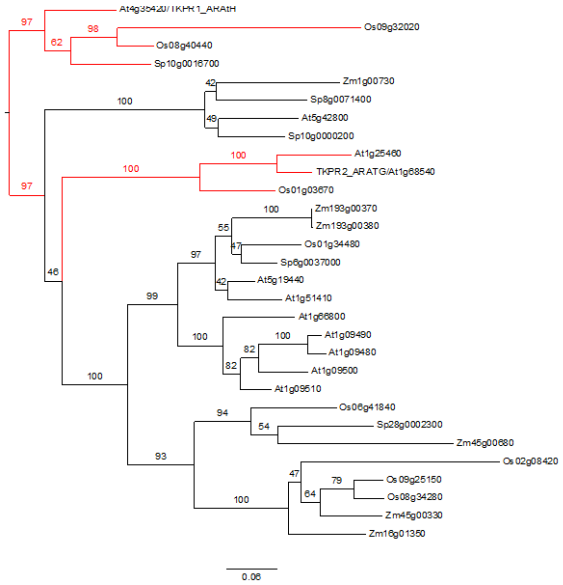
b, Maximum Likelihood phylogenetic tree of Type III LTPs and related LTPs calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000.

This phylogeny shows orthologs in the rice and *Spirodela* genomes for type III LTPs from *Arabidopsis* and documented to be exclusively expressed in anther tapetum (Huang & Huang 2013) (shown in red) and its absence from *Zostera marina*. The closest homologs from *Zostera* cluster with LTP proteins from other types (Zosma16g01580: Type I LTP, Zosma10g01720: Type IX LTP and Zosma224g00100: nsLTP/xylogen). Species name: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

VI. TetraKetide alpha-Pyrone Reductases (TKPRs)



a, MUSCLE alignment of *Arabidopsis* TetraKetide alpha-Pyrone Reductases (TKPRs). Non-conserved positions were removed based on BLOSUM62 scoring matrix in 10 amino acid window sizes.



b, Maximum Likelihood phylogenetic tree of TKPRs calculated in RAxML v8.1.3 with LG4X substitution models + GAMMA model of rate heterogeneity, bootstraps 1000. This phylogeny is shows orthologs of TKPR1 andTKPR2 in rice and TKPR1 in duckweed (in red) but the absence of both TKPR1 & TKPR2 from

Z. marina. Species name: *Arabidopsis* (ARATH), Rice (Os), *Zostera marina* (Zosma) and *Spirodela polyrhiza* (Spipo).

[Above] Supplementary Fig S11.1 | Multiple sequence alignments and phylogenetic trees to support pollen development gene comparisons in support of Extended Data Table 3.

11.2 MADS box genes

MADS-box genes are important developmental regulators in flowering plants, where they occur with high diversity (50-200 genes for most plants but as low as 20 in a few cases, e.g., *Amborella trichocpoda*) (Gramzow & Theissen 2013). Most MADS-box transcription factors regulate the development of flowers, pollen and seeds.

In order to identify MADS-box genes in *Z. marina* the conserved MADS-box Pfam domain (SRF-TF, PF00319) was searched for among all of the annotated *Z. marina* genes directly and then using the "MADS-short" hidden Markov model (hmm) profile (Gramzow & Theissen 2013), which did not yield additional hits. In addition, a *Z. marina* specific hmm profile was generated by aligning the previously identified MADS-box domains using HMMR (http://hmmer.janelia.org) to generate a new profile that did not result in any additional matches. To find possible genes missed by the annotation, the whole *Z. marina* genome was six-frame translated and searched with the hmm profiles. Hits were evaluated by checking for expression through the mapped RNASeq reads.

For generating alignments, *Z. marina* MADS-box genes were separated into types I and II by blasting against annotated genes from other species and checking for the presence of a K-box domain. For making alignments, only the MADS-box domain was extracted from the amino acid sequences.
All sequences were aligned using MAFFT (ver 7; http://mafft.cbrc.jp/alignment/software/) with the L-IN-SI settings. Alignments were manually cropped to length. Maximum likelihood trees were calculated with RAxML (ver. 8.0) (http://sco.h-its.org/exelixis/web/software/raxml/index.html)using the PROTGAMMALG model and the fast bootstrap option. Final trees were calculated using the CIPRES Science Gateway (v. 3.3) (http://www.phylo.org/)

The *Z. marina* genome contains 50 MADS-box genes. Of these, 28 can be classified to type I and 22 contain a K-box domain of type II (**Supplementary Table S11.1**). *Z. marina* (and also *Spirodela*) has fewer MADS-box genes than other sequenced plants that average around 100 (**Supplementary Table S11.2).** The distribution among classes roughly agrees with other species, most genes of type I are in class Mα (**Supplementary Fig. S11.2)**. Type II genes are spread across all classes with few exceptions (**Supplementary Fig. S11.3**). The low number of MADS genes fits with the simplified structure that is characterized by a much reduced perianth (corona and corolla) in both *Z. marina* and *Spirodela* flowers compared to most flowering plants(Kuo & den Hartog 2006).

Surprisingly, most of the genes were not expressed in flower tissue (**Supplementary Table S11.3)** with the exception of Zosma173g00150, Zosma311g00030 and Zosma136g00360, which are specifically expressed in male flowers. They are all close homologs of the agamous-like-30 (AGL30) gene from *A. thaliana* that is implicated in pollen development (see below) (Verelst *et al.* 2007; Adamczyk & Fernandez 2009).

Supplementary Table S11.1 | MADS-box genes in *Z. marina* by Type and Class.

Class assignment marked with ? is tentative.

Type I		Type II	
Gene ID	**Class**	**Gene Id**	**Class**
Zosma114g00480	Mα	Zosma21g00680	AG
Zosma120g00700	Mα	Zosma76g00140	AG
Zosma136g00360	Mα	Zosma2g02220	AGL12
Zosma1406g00020	Mα	Zosma83g00290	AGL12
Zosma147g00340	Mα	Zosma203g00090	AGL17
Zosma173g00150	Mα	Zosma127g00090	AGL2
Zosma180g00230	Mα	Zosma76g00210	AGL6
Zosma20g00030	Mα	Zosma7g01530	AGL9
Zosma25g00050	Mα	Zosma203g00280	DEF
Zosma296g00130	Mα	Zosma146g00340	GLO
Zosma2g02180	Mα	Zosma89g01160	MIKC*
Zosma2g02980	Mα	Zosma76g00080	OsMADS32
Zosma2g03630	Mα	Zosma12g00350	SQUA
Zosma311g00030	Mα	Zosma54g00160	SQUA
Zosma353g00060	Mα	Zosma55g00480	SQUA
Zosma36g00050	Mα	Zosma6g00550	SQUA
Zosma54g00230	Ma	Zosma105g00590	STK
Zosma54g00260	Mα	Zosma83g00280	STK ?
Zosma1g02510	Mβ	Zosma68g00830	StMADS11
Zosma1g02530	Mβ	Zosma106g00500	TM3
Zosma209g00220	Mβ	Zosma7g01540	TM8
Zosma34g00390	Mβ	Zosma89g00440	TM8
Zosma16g00320	Mγ		
Zosma22g01150	Mγ		
Zosma34g00560	Mγ		
Zosma37g00440	Mγ		
Zosma3972g00010	Mγ		
Zosma4g00620	Mγ		

Supplementary Table S11.2 | Comparison of MADS-box genes between *Z. marina* and other sequenced plant species.

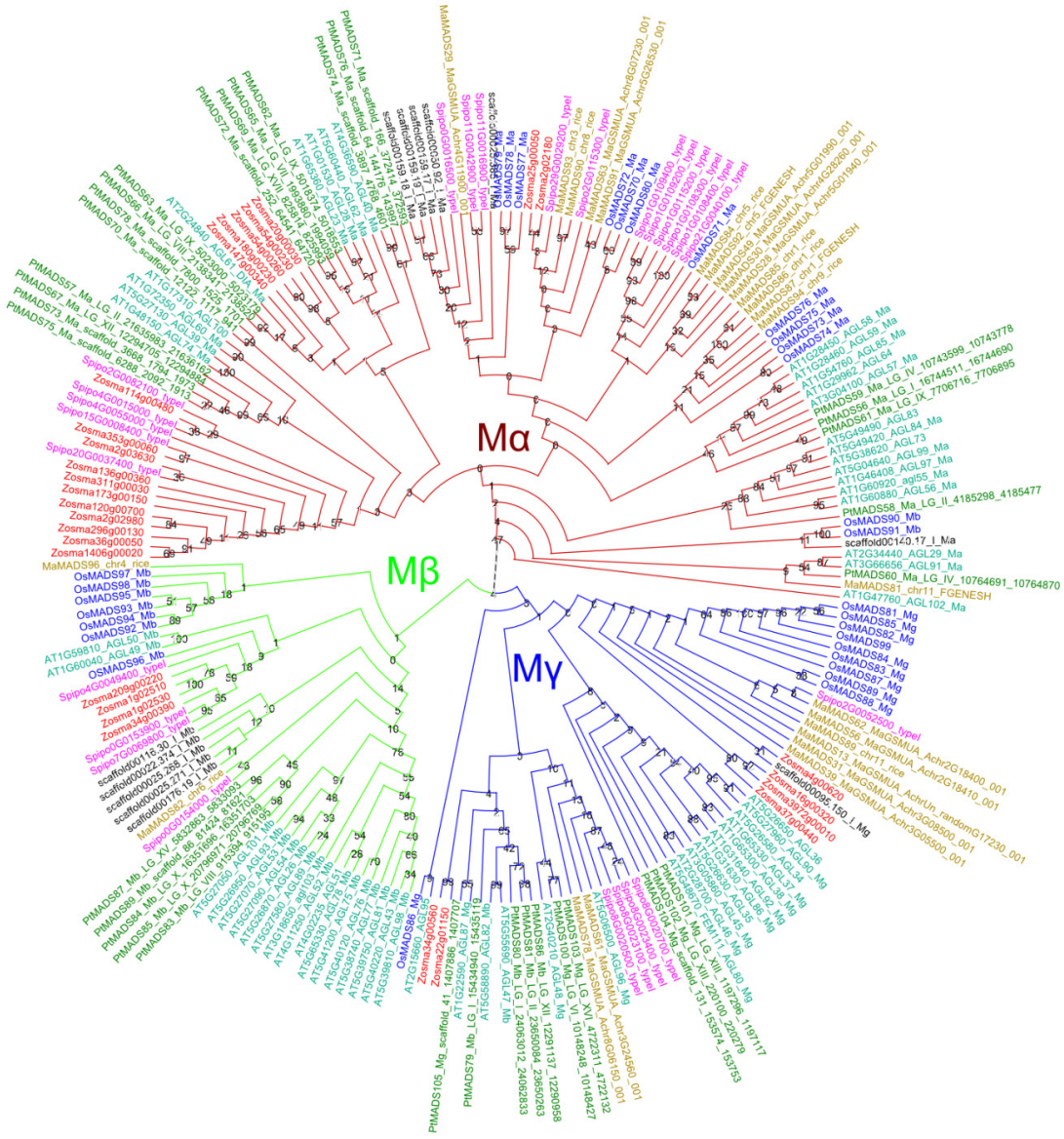
	Amborella trichocpoda	*Arabidopsis thaliana*	*Musa acuminata*	*Oryza sativa*	*Populus trichocarpa*	*Spirodela polyrhiza*	*Zostera marina*
Type I							
Ma	6	24	15	12	20	16	18
Mb	0	18	2	7	5	4	4
Mg	1	17	8	10	10	5	6
Total Type I	**7**	**59**	**25**	**29**	**35**	**25**	**28**
Type II							
AG	1	3	4	2	2	1	2
AGL12	1	1	2	2	2	0	2
AGL15	2	2	2	0	2	1	0
AGL17	1	4	6	6	4	1	1
AGL2	1	3	5	3	1	2	1
AGL6	1	2	5	2	3	1	1
AGL9	2	1	2	2	2	0	1
DEF	2	1	2	1	1	1	1
GGM13	2	2	2	3	3	1	0
GLO	2	1	2	2	2	2	1
MIKC*	1	7	5	3	8	0	1
OsMADS32	1	0	1	1	0	0	1
SQUA	1	4	9	4	4	2	4
STK	1	1	3	2	2	1	2
StMADS11	1	2	3	3	8	8	1
TM3	2	6	11	2	5	3	1
TM8	0	0	0	0	2	1	2
FLC	0	6	5	2	6	0	0
Total type II	**22**	**46**	**69**	**40**	**57**	**25**	**22**
Total	**29**	**105**	**94**	**69**	**92**	**50**	**50**

Supplementary Table S11.3 | Normalized gene expression (FPKM) of MADS box genes across tisues in *Z. marina*.

***** implicated in pollen development. See also **Supplementary Data 1, 2** and **3.**

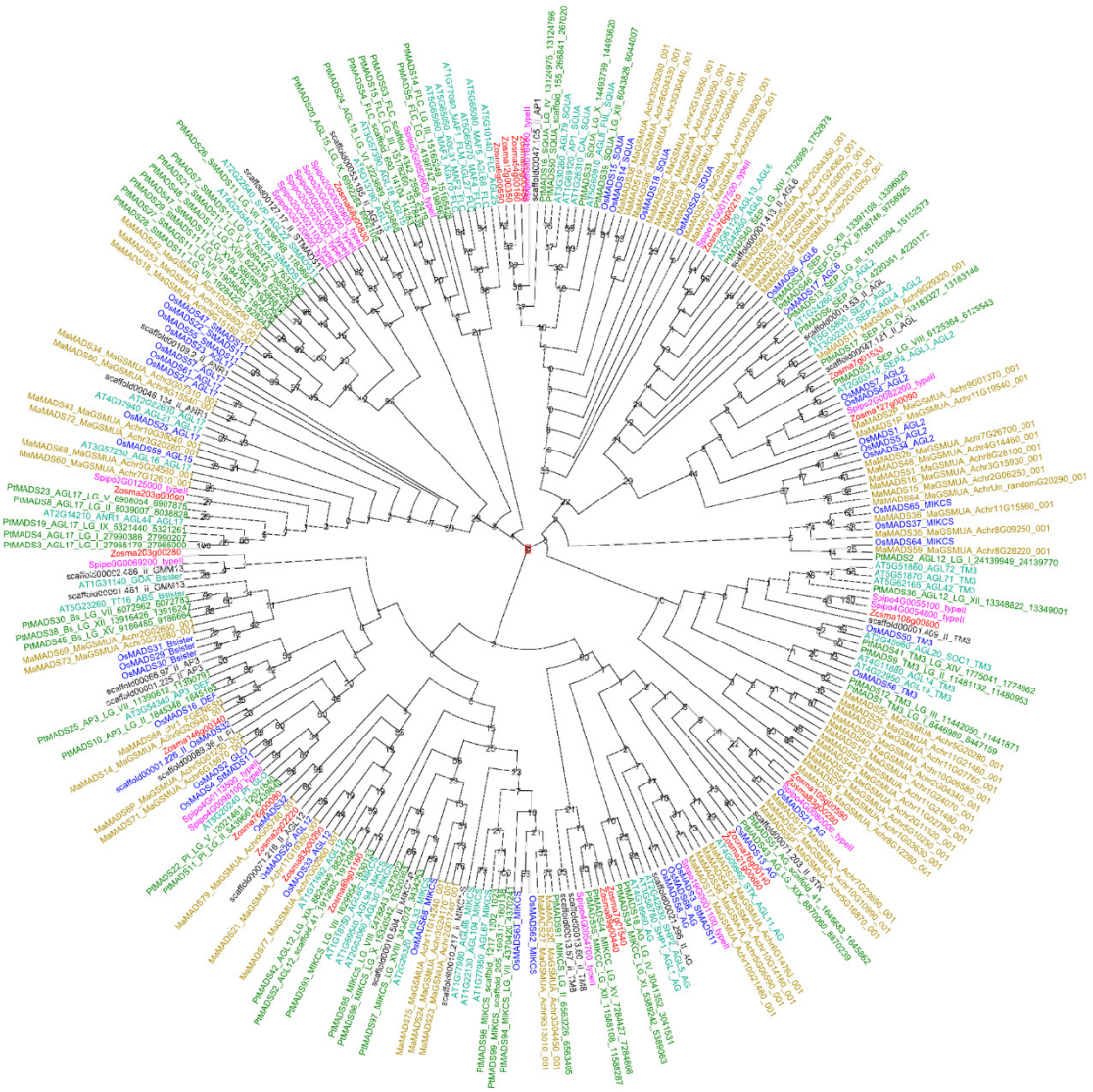
		FPKM				
Gene ID	Gene class	Female flower early	Female flower late	Male flower	Root	Leaf
Zosma7g01530	AGL9	473.46	528.13	166.18	1.51	18.86
Zosma173g00150*	Ma	2.10	0.19	499.95	0.06	0.10
Zosma127g00090	AGL2	262.50	341.17	206.20	1.28	6.88
Zosma106g00500	TM3	2.44	1.95	340.44	2.80	2.77
Zosma311g00030	Ma	1.04	0.10	293.50	0.13	0.28
Zosma105g00590	STK	212.35	258.06	4.17	0.52	0.75
Zosma203g00090	AGL17	0.85	0.06	251.01	19.38	0.21
Zosma353g00060	Ma	0.32	0.07	149.28	0.05	0.08
Zosma21g00680	AG	20.95	18.78	90.09	6.98	1.79
Zosma76g00140	AG	29.02	28.06	88.85	4.97	44.64
Zosma136g00360	Ma	0.22	0.10	86.76	0.10	0.03
Zosma55g00480	SQUA	3.33	2.61	14.68	49.59	84.74
Zosma68g00830	StMADS11	3.45	3.20	7.40	81.62	73.05
Zosma114g00480	Ma	42.39	50.86	0.00	8.08	0.00
Zosma12g00350	SQUA	27.49	30.38	32.16	4.41	47.88
Zosma76g00210	AGL6	37.40	43.64	0.80	0.08	1.20
Zosma54g00160	SQUA	5.43	6.35	1.16	0.00	37.80
Zosma203g00280	DEF	2.82	2.46	14.86	2.93	36.82
Zosma2g02220	AGL12	0.12	0.10	1.21	34.38	0.04
Zosma83g00280	STK ?	32.09	14.34	0.32	0.00	0.00
Zosma89g00440	TM8	10.41	7.33	0.71	0.03	22.46
Zosma83g00290	AGL12	0.00	0.00	0.06	15.65	0.04
Zosma147g00340	Ma	0.04	0.09	15.49	0.09	0.01
Zosma7g01540	TM8	8.50	14.28	0.11	0.00	0.64
Zosma54g00260	Ma	2.49	12.91	0.93	0.32	7.10
Zosma146g00340	GLO	1.79	1.30	11.30	0.20	0.25
Zosma76g00080	OsMADS32	8.54	10.35	0.86	3.84	2.78
Zosma6g00550	SQUA	0.57	1.87	0.12	6.47	3.88
Zosma34g00390	Mb	0.02	0.04	3.62	0.00	0.00
Zosma2g03630	Ma	0.09	0.63	3.13	1.12	0.28
Zosma2g02180	Ma	1.93	2.85	2.85	0.24	1.35
Zosma25g00050	Ma	0.00	0.05	2.77	0.03	0.04
Zosma20g00030	Ma	2.43	2.71	0.00	1.26	0.00
Zosma34g00560	Mg	0.32	1.09	2.52	1.08	2.43
Zosma22g01150	Mg	0.06	2.14	0.42	0.00	0.00
Zosma209g00220	Mb	0.00	0.00	1.82	0.00	0.00
Zosma120g00700	Ma	0.01	0.00	1.49	0.00	0.00
Zosma37g00440	Mg	0.18	0.11	0.95	0.19	0.09
Zosma1g02530	Mb	0.19	0.42	0.54	0.20	0.02

Zosma89g01160	MIKC*	0.36	0.52	0.00	0.00	0.09
Zosma4g00620	Mg	0.02	0.45	0.00	0.00	0.00
Zosma16g00320	Mg	0.24	0.22	0.29	0.02	0.05
Zosma180g00230	Ma	0.02	0.09	0.19	0.02	0.24
Zosma2g02980	Ma	0.00	0.00	0.21	0.00	0.00
Zosma54g00230	Ma	0.00	0.00	0.20	0.00	0.00
Zosma1g02510	Mb	0.06	0.07	0.18	0.00	0.00
Zosma36g00050	Ma	0.00	0.00	0.05	0.00	0.00
Zosma296g00130	Ma	0.00	0.00	0.04	0.00	0.00
Zosma1406g00020	Ma	0.00	0.00	0.03	0.00	0.00
Zosma3972g00010	Mg	0.00	0.00	0.00	0.00	0.00



Supplementary Fig. S11.2 | Maximum likelihood tree of the MADS domain of type I MADS box genes, cladogram view.

Branch colors show the gene classes (red: Mα, green: Mβ, blue Mγ), label colors designate species (Red: *Z. marina*. Purple: *S. polyrhiza*. Dark green: *P. trichocarpa*. Turquoise: *A. thaliana*. Blue: *O. sativa*. Yellow/gold: *M. acuminate*. *B*lack *as scaffold*: *A. trichopoda*). Node labels are bootstrap values (1000 bootstraps).



Supplementary Fig. S11.3 | Maximum likelihood tree of the MADS domain of type II MADS box genes, cladogram view.

Label colors designate species (Red: *Z. marina*. Purple: *S. polyrhiza*. Dark green: *P. trichocarpa*. Turquoise: *A. thaliana*. Blue: *O. sativa*. Yellow/gold: *M. acuminate*. Black *as scaffold*: *A. trichopoda*). Node labels are bootstrap values (1000 bootstraps).

References

Ackerman JD (2006) Sexual reproduction of seagrasses: pollination in the marine context. In: *Seagrasses: Biology, Ecology and Conservation* (eds. Larkum AWD, Orth RJ, Duarte CM) pp. 89-109 Springer, Dordrecht, Netherlands.

Adamczyk BJ, Fernandez DE (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiology,* **149**, 1713-1723.

Alexa A, Rahnenfuehrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics,* **22**, 1600-1607.

Al-Mssallem IS, Hu S, Zhang X, et al (2013) Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications,* **4**, 2274.

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution,* **27**, 1436-1448.

Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders of flowering plants. *Botanical Journal of the Linnean Society,* **161**, 105.

Aquino RS, Landeira-Fernandez AM, Valente AP, Andrade LR, Mourao PAS (2005) Occurrence of sulfated galactans in marine angiosperms: evolutionary implications. *Glycobiology,* **15**, 11-20.

Aquino RS, Grativol C, Mourao PAS (2011) Rising from the Sea: Correlations between sulfated polysaccharides and salinity in plants. *Plos ONE,* **6**, e18862.

Ariizumi T, Toriyama K (2011) Genetic regulation of sporopollenin synthesis and pollen exine development. *Annual Review Plant Biology,* **62**, 437-460.

Ashburner M, Lewis S (2002) On ontologies for biologists: the Gene Ontology - untangling the web. In: *In Silico Simulation of Biological Processes.* (eds. Bock G, Goode JA) pp. 63-83 Novartis Foundation, London, UK.

Bassil E, Ohto M, Esumi T, et al (2011) The *Arabidopsis* Intracellular Na+/H+ antiporters NHX5 and NHX6 are endosome associated and necessary for plant growth and development. *Plant Cell,* **23**, 224-239.

Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-visited. *American Journal of Botany,* **97**, 1296-1303.

Bergmann N, Winters G, Rauch G, et al (2010) Population-specificity of heat stress gene induction in northern and southern eelgrass *Zostera marina* populations under simulated global warming. *Molecular Ecology,* **19**, 2870-2883.

Berry JA, Beerling DJ, Franks PJ (2010) Stomata: key players in the earth system, past and present. *Current Opinion in Plant Biology,* **13**, 233-240.

Binkert M, Kozma-Bognar L, Terecskei K, De Veylder L, Nagy F, Ulm R (2014) UV-B-Responsive association of the *Arabidopsis* bZIP Transcription Factor ELONGATED HYPOCOTYL5 with target genes, including its own promoter. *Plant Cell,* **26**, 4200-4213.

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell,* **16**, 1667-1678.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England),* **30**, 2114-2120.

Boratyn GM, Camacho C, Cooper PS, et al (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Research,* **41**, W29-33.

Bowman JL (2011) Stomata: Active portals for flourishing on land. *Current Biology,* **21**, R540-R541.

Brakel J, Werner FJ, Tams V, Reusch TBH, Bockelmann A (2014) Current European *Labyrinthula zosterae* are not virulent and modulate seagrass (*Zostera marina*) defense gene expression. *Plos ONE,* **9**, e92448.

Breeze E, Wagstaff C, Harrison E, et al (2004) Gene expression patterns to define stages of post-harvest senescence in *Alstroemeria* petals. *Plant Biotechnology Journal,* **2**, 155-168.

Buljan M, Frankish A, Bateman A (2010) Quantifying the mechanisms of domain gain in animal proteins. *Genome Biology,* **11**, R74-R74.

Burke MK, Dennison WC, Moore KA (1996) Non-structural carbohydrate reserves of eelgrass *Zostera marina*. *Marine Ecology Progress Series,* **137**, 195-201.

Burton RA, Wilson SM, Hrmova M, et al (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science,* **311**, 1940-1942.

Carpita, N.C. and McCann, M. (2000) The cell wall. In: *Biochemistry and Molecular Biology of Plants* (eds. Buchanan B, Gruissem W, Jones R) pp. 52-108 American Society of Plant Physiologists, Rockville, MD, USA.

Chanroj S, Wang G, Venema K, Zhang MW, Delwiche CF, Sze H (2012) Conserved and diversified gene families of monovalent cation/H+ antiporters from algae to flowering plants. *Frontiers in Plant Science,* **3**, 25.

Chapman E, Best MD, Hanson SR, Wong CH (2004) Sulfotransferases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie-International Edition,* **43**, 3526-3548.

Chater C, Gray JE, Beerling DJ (2013) Early evolutionary acquisition of stomatal control and development gene signalling networks. *Current Opinion in Plant Biology,* **16**, 638-646.

Chaturvedi AK, Mishra A, Tiwari V, Jha B (2012) Cloning and transcript analysis of type 2 metallothionein gene (SbMT-2) from extreme halophyte *Salicornia brachiata* and its heterologous expression in *E. coli*. *Gene,* **499**, 280-287.

Chaudhuri P, Marron JS (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association,* **94**, 807-823.

Chavez Montes RA, de Fatima Rosas-Cardenas F, De Paoli E, et al (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nature Communications,* **5**, 3722.

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant Journal,* **66**, 212-229.

Chen LQ, Qu XQ, Hou BH, Osorio S, Fernie AR, Frommer WB (2012) Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science,* **335**, 207-2011.

Chinnusamy V, Ohta M, Kanrar S, et al (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes and Development,* **17**, 1043-1054.

Cock JM, Sterck L, Rouze P, et al (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature,* **465**, 617-621.

Cocuron J, Lerouxel O, Drakakaki G, et al (2007) A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proceedings of the National Academy of Sciences of the United States of America,* **104**, 8550-8555.

Collen J, Porcel B, Carre W, et al (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences of the United States of America,* **110**, 5247-5252.

Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics,* **2008**, 619832-619832.

Coyer JA, Hoarau G, Kuo J, Tronholm A, Veldsink J, Olsen JL (2013) Phylogeny and temporal divergence of the seagrass family Zosteraceae using one nuclear and three chloroplast loci. *Systematics and Biodiversity,* , 1-14.

Csuroes M (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics,* **26**, 1910-1912.

Cuena A, Petersen G, Seberg O, Davis JI, Stevenson DW (2010) Are substitution rates and RNA editing correlated? *BMC Evolutionary Biology,* **10**, 349.

Cui L, Wall PK, Leebens-Mack JH, et al (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Research,* **16**, 738-749.

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research,* **14**, 1394-1403.

Dauch AL, Jabaji-Hare SH (2006) Metallothionein and bZIP transcription factor genes from velvetleaf and their differential expression following *Colletotrichum coccodes* infection. *Phytopathology,* **96**, 1116-1123.

Den Hartog C, Kuo J (2006) Taxonomy and Biogeography of Seagrasses. In: *Seagrasses: Biology, Ecology and Conservation* (eds. Larkum AWD, Orth RJ, Duarte CM) pp. 1-23 Springer, Dordrecht, Netherlands.

D'Hont A, Denoeud F, Aury J, et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature,* **488**, 213-+.

Dhugga KS, Barreiro R, Whitten B, et al (2004) Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. *Science,* **303**, 363-366.

Doblin MS, Pettolino FA, Wilson SM, et al (2009) A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America,* **106**, 5996-6001.

Duffy JE (2006) Biodiversity and the functioning of seagrass ecosystems. *Marine Ecology Progress Series,* **311**, 233-250.

Edwards D, Davies KL, Axe L (1992) A vascular conducting strand in the early land plant *Cooksonia*. *Nature,* **357**, 683-685.

Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. *Current Opinion in Plant Biology,* **3**, 278-284.

Engelken J, Brinkmann H, Adamska I (2010) Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. *BMC Evolutionary Biology,* **10**, 233.

Espartero J, Pintor-Toro J, Pardo JM (1994) Differential accumulation of S-adenosylmethionine synthetase transcripts in response to salt stress. *Plant Molecular Biology,* **25**, 217-227.

Finn RD, Bateman A, Clements J, et al (2014) Pfam: the protein families database. *Nucleic Acids Research,* **42**, D222-30.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research,* **39**, W37.

Franssen SU, Gu J, Bergmann N, et al (2011) Transcriptomic resilience to global warming in the seagrass *Zostera marina*, a marine foundation species. *Proceedings of the National Academy of Sciences of the United States of America,* **108**, 19276-19281.

Furness CA (2013) Diversification of pollen and tapetum in early-divergent monocots. In: *Early Events in Monocot Evolution* (eds. Wilkin P, Mayo SJ) , Systematics Association Special Volume Series edn, pp. 1-22 Cambridge University Press, Cambridge, UK.

Giordani T, Natali L, Maserti BE, Taddei S, Cavallini A (2000) Characterization and expression of DNA sequences encoding putative type-II metallothioneins in the seagrass *Posidonia oceanica*. *Plant Physiology,* **123**, 1571-1581.

Gramzow L, Theissen G (2013) Phylogenomics of MADS-Box genes in Plants - Two opposing life styles in one gene family. *Biology,* **2**, 1150-64.

Guo Y, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D (2011) Genome-wide comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-encoding genes in *Arabidopsis*. *Plant Physiology,* **157**, 757-769.

Gutensohn M, Klempien A, Kaminaga Y, et al (2011) Role of aromatic aldehyde synthase in wounding/herbivory response and flower scent production in different *Arabidopsis* ecotypes. *Plant Journal,* **66**, 591-602.

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology,* **48**, 575-607.

Hanson SR, Best MD, Wong CH (2004) Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie-International Edition,* **43**, 5736-5763.

Henrissat B, Coutinho PM, Davies GJ (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Molecular Biology,* **47**, 55-72.

Huang MD, Huang AHC (2013) Abundant Type III Lipid Transfer Proteins in *Arabidopsis* tapetum are secreted to the locule and become a constituent of the pollen exine. *Plant Physiology,* **153**, 1218-1229.

Huerta-Cepas J, Dopazo J, Gabaldon T (2010) ETE: a Python environment for tree exploration. *BMC Bioinformatics,* **11**, 24-24.

Iles WJD, Smith SY, Graham SW (2013) Refining our understanding of the phylogenetic backbone of Alismatales. In: *Early Events in Monocot Evolution* (eds. Wilkin P, Mayo SJ) , Systematics Association Special Volume edn, pp. 1-28 Cambridge University Press, Cambridge, UK.

Iles WJD (2013) The Phylogeny and Evolution of Two Ancient Lineages of Aquatic Plants: the Hydatellaceae and Alismatales. *Ph.D Thesis, University of British Columbia, Vancouver, Canada,* , 1-237.

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

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

Jeong D, Park S, Zhai J, et al (2011) Massive Analysis of Rice Small RNAs: Mechanistic Implications of Regulated MicroRNAs and Variants for Differential Target RNA Cleavage. *Plant Cell,* **23**, 4185-4207.

Jeong D, Schmidt SA, Rymarquis LA, et al (2013) Parallel analysis of RNA ends enhances global investigation of microRNAs and target RNAs of *Brachypodium distachyon*. *Genome Biology,* **14**, R145.

Jetter R, Kunst L, Samuels AL (2006) Composition of plant cuticular waxes. In: *Biology of the Plant Cuticle* (eds. Riederer M, Muller C) pp. 145-175 Blackwell, Oxford, UK.

Jiang Zhijian, Huang Xiaoping, Zhang Jingping (2013) Dynamics of nonstructural carbohydrates in seagrass *Thalassia hemprichii* and its response to shading. *Acta Oceanologica Sinica,* **32**, 61-67.

Jupe F, Pritchard L, Etherington GJ, et al (2012) Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics,* **13**, 75-78.

Khotimchenko SV (1993) Fatty-acids and polar lipids of seagrasses from the Sea of Japan. *Phytochemistry,* **33**, 369-372.

Kikuchi S, Bedard J, Hirano M, et al (2013) Uncovering the Protein Translocon at the Chloroplast Inner Envelope Membrane. *Science,* **339**, 571-574.

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature,* **267**, 275-276.

Kirk JTO (2011) *Light and Photosynthesis in Aquatic Ecosystems* Cambridge University Press, Cambridge, UK.

Kloareg B, Quatrano RS (1988) Structure of the dell-walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology,* **26**, 259-315.

Kohler A, Rinaldi C, Duplessis S, et al (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Molecular Biology,* **66**, 619-636.

Kolattukudy PE (2001) Suberin from plants. In: *Biopolymers, Volume 3a, Polyesters I - Biological Systems and Biotechnological Production* (eds. Doi Y, Steinbüchel A) pp. 41-68 Wiley-Blackwell, Oxford, UK.

Kong F, Zhou Y, Sun P, Liu L, Mao Y (2013) Generation and analysis of expressed sequence tags from the salt-tolerant eelgrass species, *Zostera marina*. *Acta Oceanologica Sinica,* **32**, 68-78.

Kuo J, den Hartog C (2006) Seagrass morphology, anatomy, and ultrastructure. In: *Seagrasses: Biology, Ecology and Conservation* (eds. Larkum AWD, Orth RJ, Duarte CM) pp. 51-87 Springer, Dordrecht, Netherlands.

Lallemand B, Erhardt M, Heitz T, Legrand M (2013) Sporopollenin biosynthetic enzymes interact and constitute a metabolon localized to the endoplasmic reticulum of tapetum cells[W]. *Plant Physiology,* **162**, 616-625.

Larkum AWD, Drew EA, Ralph PJ (2006) Photosynthesis and metabolism in seagrasses at the cellular level. In: *Seagrasses: Biology, Ecology and Conservation.* (eds. Larkum AWD, Orth RJ, Duarte CM) pp. 323-345 Springer, Dordrecht, Netherlands.

Lee J, Chae HS, Lee J, et al (1997) Structure and expression of two cDNAs encoding S-adenosyl-l-methionine synthetase of rice (*Oryza sativa* L.)1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* **1354**, 13-18.

Lepiniec L, Debeaujon I, Routaboul J, et al (2006) Genetics and biochemistry of seed flavonoids. *Annual Review of Plant Biology,* **57**, 405-430.

Les DH, Tippery NP (2013) In time and with water…the systematics of alsimatid monocotyledons. In: *Early Events in Monocot Evolution* (eds. Wilkin P, Mayo SJ) pp. 118-164 Cambridge University Press, Cambridge, UK.

Les DH, Cleland MA, Waycott M (1997) Phylogenetic studies in Alismatidae, II: Evolution of marine angiosperms (seagrasses) and hydrophily. *Systematic Botany,* **22**, 443-463.

Leszczyszyn OI, Imam HT, Blindauer CA (2013) Diversity and distribution of plant metallothioneins: a review of structure, properties and functions. *Metallomics,* **5**, 1146-1169.

Li-Beisson Y, Shorrosh B, Beisson F, et al (2013) Acyl-lipid metabolism. *The Arabidopsis book / American Society of Plant Biologists,* **11**, e0161-e0161.

Liu C, Shi L, Zhu Y, et al (2012) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics,* **13**, 715.

Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research,* **42**, D490-D495.

Lu C, Jeong D, Kulkarni K, et al (2008) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 4951-4956.

Mace E, Tai S, Innes D, et al (2014) The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biology,* **14**, 253-253.

Maeda M, Koshikaw M, Nisizawa K, Takano K (1966) Cell wall constituents especially pectic substance of a marine phanerogam *Zostera marina*. *Botanical Magazine-Tokyo,* **79**, 422-431.

Matsumoto T, Wu JZ, Kanamori H, et al (2005) The map-based sequence of the rice genome. *Nature,* **436**, 793-800.

Maumus F, Quesneville H (2014) Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. *Nature Communications,* **5**, 4104-4104.

McLachlan G, Peel D (1999) The EMMIX algorithm for the fitting of normal and t-components. *Journal of Statistical Software,* **4**, 1-14.

McNeil KJ, Smith AG (2010) A glycine-rich protein that facilitates exine formation during tomato pollen development. *Planta,* **231**, 793-808.

Meyers BC, Axtell MJ, Bartel B, et al (2008) Criteria for annotation of plant microRNAs. *Plant Cell,* **20**, 3186-3190.

Michel G, Tonon T, Scornet D, Cock JM, Kloareg B (2010) The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus:* Insights into the evolution of extracellular matrix polysaccharides in eukaryotes. *New Phytologist,* **188**, 82-97.

Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science,* **7**, 405-410.

Modarresi M, Nematzadeh GA, Moradian F (2013) Salinity response pattern and isolation of catalase gene from halophyte plant *Aeluropus littoralis*. *Photosynthetica,* **51**, 621-629.

Moyle R, Fairbairn DJ, Ripi J, Crowe M, Botella JR (2005) Developing pineapple fruit has a small transcriptome dominated by metallothionein. *Journal of Experimental Botany,* **56**, 101-112.

Muramatsu Y, Harada A, Ohwaki Y, Kasahara Y, Takagi S, Fukuhara T (2002) Salt-tolerant ATPase activity in the plasma membrane of the marine angiosperm *Zostera marina* L. *Plant and Cell Physiology,* **43**, 1137-1145.

Myburg AA, Grattapaglia D, Tuskan GA, et al (2014) The genome of *Eucalyptus grandis*. *Nature,* **510**, 356-+.

Naik PA, Shi P, Tsai C (2007) Extending the Akaike Information Criterion to mixture regression models. *Journal of the American Statistical Association,* **102**, 244-254.

Papenbrock J (2012) Highlights in seagrasses' phylogeny, physiology, and metabolism: What makes them special? *International Scholarly Research Network Botany,* **2012**, 1-15.

Pergent-Martini C (1998) *Posidonia oceanica*: a biological indicator of past and present mercury contamination in the Mediterranean sea. *Marine Environmental Research,* **45**, 101-111.

Peterson G, Seberg O, Davis JI, Stevenson DW (2006) RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. *Taxon,* **55**, 871.

Peterson KM, Rychel AL, Torii KU (2010) Out of the mouths of plants: The molecular basis of the evolution anddDiversity of stomatal development. *Plant Cell,* **22**, 296-306.

Pillitteri LJ, Dong J (2013) Stomatal development in *Arabidopsis*. *The Arabidopsis book / American Society of Plant Biologists,* **11**, e0162-e0162.

Popper ZA, Michel G, Herve C, et al (2011) Evolution and diversity of plant cell walls: From algae to flowering plants. *Annual Review of Plant Biology,* **62**, 567-588.

Qi Y, Wang F, Zhang H, Liu W (2009) Overexpression of suadea salsa S-adenosylmethionine synthetase gene promotes salt tolerance in transgenic tobacco. *Acta Physiologiae Plantarum,* **32**, 263-269.

Ranathunge K, Schreiber L, Franke R (2011) Suberin research in the genomics era: New interest for an old polymer. *Plant Science,* **180**, 399-413.

Reusch TBH, Veron AS, Preuss C, et al (2008) Comparative analysis of expressed sequence tag (EST) libraries in the seagrass *Zostera marina* subjected to temperature stress. *Marine Biotechnology,* **10**, 297-309.

Rodriguez-Rosales PM, Xingyu J, Galvez JF, Aranda N, Cubero B, Venema K (2008) Overexpression of the tomato K+/H+ antiporter LeNHX2 confers salt tolerance by improving potassium compartmentalization. *New Phytologist,* **179**, 366-377.

Sanchez-Aguayo I, Rodraguez-Galain JM, Garcia R, Torreblanca J, Pardo JM (2004) Salt stress enhances xylem development and expression of S-adenosyl-l-methionine synthase in lignifying tissues of tomato plants. *Planta,* **220**, 278-285.

Sanina NM, Goncharova SN, Kostetsky EY (2004) Fatty acid composition of individual polar lipid classes from marine macrophytes. *Phytochemistry,* **65**, 721-730.

Sanina NM, Goncharova SN, Kostetsky EY (2008) Seasonal changes of fatty acid composition and thermotropic behavior of polar lipids from marine macrophytes. *Phytochemistry,* **69**, 1517-1527.

Schlacher-Hoenlinger MA, Schlacher TA (1998) Accumulation, contamination, and seasonal variability of trace metals in the coastal zone - patterns in a seagrass meadow from the Mediterranean. *Marine Biology,* **131**, 401-410.

Schlueter JA, Dixon P, Granger C, et al (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome Biology,* **47**, 868-876.

Schroder G, Eichel J, Breinig S, Schroder J (1997) Three differentially expressed S-adenosylmethionine synthetases from *Catharanthus roseus*: molecular and functional characterization. *Plant Molecular Biology,* **33**, 211-222.

Shi T, Huang H, Barker MS (2010) Ancient genome duplications during the evolution of kiwifruit (Actinidia) and related Ericales. *Annals of Botany,* **106**, 504.

Shpak ED (2013) Diverse roles of ERECTA family genes in plant development. *Journal of Integrative Plant Biology,* **55**, 1238-1250.

Stintzi A, Heitz T, Prasad V, et al (1993) Plant pathogenesis-related proteins and their role in defense against pathogens. *Biochimie,* **75**, 687-706.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and fisualizes long lists of gene ontology terms. *Plos ONE,* **6**, e21800-e21800.

Tarr DEK, Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Research Notes,* **2**, 197-197.

Torrens-Spence MP, Liu P, Ding H, Harich K, Gillaspy G, Li J (2013) Biochemical evaluation of the decarboxylation and decarboxylation-deamination activities of plant aromatic amino acid decarboxylases. *Journal Biological Chemistry,* , 2376-87.

Uchida N, Lee JS, Horst RJ, et al (2012) Regulation of inflorescence architecture by intertissue layer ligand-receptor communication between endodermis and phloem. *Proceedings of the National Academy of Sciences of the United States of America,* **109**, 6337-6342.

Van der Biezen EA, Jones JDG (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences,* **23**, 454-456.

Van Tussenbroek B, Montory-Velazquez L, Solis-Weiss V (2012) Meso-fauna foraging on seagrass pollen may serve in marine zoophilous pollination. *Marine Ecology Progress Series,* **469**, 1-6.

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

Vanneste K, Van de Peer Y, Maere S (2013) Inference of genome duplications from age distributions revisited. *Molecular Biology and Evolution,* **30**, 177-190.

Varin L, Marsolais F, Richard M, Rouleau M (1997) Biochemistry and molecular biology of plant sulfotransferases. *FASEB Journal,* **11**, 517-525.

Vaten A, Bergmann DC (2012) Mechanisms of stomatal development: an evolutionary view. *Evodevo,* **3**, 11.

Vaucheret H, Mallory AC, Bartel DP (2006) AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Molecular Cell,* **22**, 129-136.

Vega-Sanchez ME, Verhertbruggen Y, Scheller HV, Ronald PC (2013) Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signaling and Behavior,* **8**, e23143-e23143.

Vekemans D, Proost S, Vanneste K, et al (2012) Gamma paleohexaploidy in the stem lineage of core eudicots: Significance for MADS-box gene and species diversification. *Molecular Biology and Evolution,* **29**, 3779-3806.

Verelst W, Twell D, de Folter S, Immink R, Saedler H, Muenster T (2007) MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology,* **8**, R249.

Vermaat JE, Verhagen FCA (1996) Seasonal variation in the intertidal seagrass *Zostera noltii* Hornem: Coupling demographic and physiological patterns. *Aquatic Botany,* **52**, 259-281.

Vogel JP, Garvin DF, Mockler TC, et al (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature,* **463**, 763-768.

Wang L, Cui K, Zhang KY, Fan HY, Li TL (2013) Research advance of sucrose phosphate synthase (SPS) in high plants. *International Journal of Agriculture and Biology,* **15**, 1221-1226.

Wang W, Haberer G, Gundlach H, et al (2014) The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nature Communications,* **5**, 1-13.

Wang W, Messing J (2011) High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *Plos ONE,* **6**, e24670.

Wang W, Wu Y, Messing J (2012) The mitochondrial genome of an aquatic plant, *Spirodela polyrhiza*. *PLoS ONE,* **7**, e46747.

Welsh DT (2000) Nitrogen fixation in seagrass meadows: Regulation, plant-bacteria interactions and significance to primary productivity. *Ecology Letters,* **3**, 58-71.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *Plos ONE,* **2**, e718.

Winters G, Nelle P, Fricke B, Rauch G, Reusch TBH (2011) Effects of a simulated heat wave on photophysiology and gene expression of high- and low-latitude populations of *Zostera marina*. *Marine Ecology Progress Series,* **435**, 83-95.

Wissler L, Dattolo E, Moore AD, et al (2009) Dr. Zompo: an online data repository for *Zostera marina* and *Posidonia oceanica* ESTs. *Database-the Journal of Biological Databases and Curation,* **2009**, bap009.

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics,* **20**, 3252-3255.

Yin Y, Johns MA, Cao H, Rupani M (2014) A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. *BMC Genomics,* **15**, 260.