The Genome of a Southern Hemisphere Seagrass Species (*Zostera muelleri*)[1][OPEN]

HueyTyng Lee, Agnieszka A. Golicz, Philipp E. Bayer, Yuannian Jiao, Haibao Tang, Andrew H. Paterson, Gaurav Sablok, Rahul R. Krishnaraj, Chon-Kit Kenneth Chan, Jacqueline Batley, Gary A. Kendrick, Anthony W.D. Larkum, Peter J. Ralph, and David Edwards*

School of Agriculture and Food Sciences, University of Queensland, St Lucia, Queensland 4072, Australia (H.T.L., A.A.G., R.R.K., J.B., D.E.); School of Plant Biology, University of Western Australia, Crawley, Western Australia 6009, Australia (H.T.L., P.E.B., C.-K.K.C., J.B., G.A.K., D.E.); State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China (Y.J.); Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602 (Y.J., A.H.P.); Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou 350002, China (H.T.); and Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, New South Wales 2007, Australia (G.S., A.W.D.L., P.J.R.)

ORCID IDs: 0000-0001-8561-3161 (H.T.L.); 0000-0001-8530-3067 (P.E.B.); 0000-0002-8987-2782 (Y.J.); 0000-0002-3460-8570 (H.T.); 0000-0002-0276-6064 (G.A.K.); 0000-0002-0420-134X (A.W.D.L.); 0000-0001-7599-6760 (D.E.).

Seagrasses are marine angiosperms that evolved from land plants but returned to the sea around 140 million years ago during the early evolution of monocotyledonous plants. They successfully adapted to abiotic stresses associated with growth in the marine environment, and today, seagrasses are distributed in coastal waters worldwide. Seagrass meadows are an important oceanic carbon sink and provide food and breeding grounds for diverse marine species. Here, we report the assembly and characterization of the *Zostera muelleri* genome, a southern hemisphere temperate species. Multiple genes were lost or modified in *Z. muelleri* compared with terrestrial or floating aquatic plants that are associated with their adaptation to life in the ocean. These include genes for hormone biosynthesis and signaling and cell wall catabolism. There is evidence of whole-genome duplication in *Z. muelleri*; however, an ancient pan-commelinid duplication event is absent, highlighting the early divergence of this species from the main monocot lineages.

Seagrasses are a group of flowering plants that originated in the Cretaceous period and have evolved to live submerged in the marine environment (Larkum et al., 2006). Around 72 species of seagrass form three families, Zosteraceae, Hydrocharitaceae, and Cymodoceaceae complex, within the order Alismatales (Short et al., 2011) and represent less than 1% of all flowering plant species (Les et al., 1997). Morphologically, most seagrass species have long, strap-shaped leaves and grow in monospecific meadows resembling terrestrial grasses. Parallel evolution from a freshwater ancestor of terrestrial origin is believed to have occurred at least three times to form these seagrass lineages (Les et al., 1997).

The survival of seagrasses in the marine environment is the result of successful adaptation to a range of abiotic stresses, mostly distinct from their terrestrial cousins. Seagrasses have horizontal rhizomes for effective anchorage under tidal currents, aerenchyma in roots and rhizomes to support growth in an anaerobic environment, and leaves with reduced cuticles and no stomata. Seagrasses also are adapted to the slow diffusion rate of CO_2 in seawater; some have evolved to utilize HCO_3^- as a carbon source (Beer et al., 2002; Borum et al., 2016), and they also have adapted to variable and low levels of light that attenuates quickly in seawater into blue or green wavelengths of the spectrum (Larkum et al., 2006). Studies of gene (Dattolo et al., 2013, 2014; Salo et al., 2015) and protein (Mazzuca et al., 2009) expression, together with genome methylation signatures (Greco et al., 2013) of seagrasses in response to differing light conditions, revealed the differential expression of

[1] This work was supported by the Australian Research Council (project nos. LP0882095 and LP0989200), the Australian Genome Research Facility, the Queensland Cyber Infrastructure Foundation, the Pawsey Supercomputing Centre, with funding from the Australian Government and the Government of Western Australia, and the National Computational Infrastructure, which is supported by the Australian Government.

* Address correspondence to dave.edwards@uwa.edu.au.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: David Edwards (dave.edwards@uwa.edu.au).

H.-T.L. performed the analysis together with A.A.G., P.E.B., R.R.K., and G.S.; Y.J., H.T., and A.H.P. analyzed the whole-genome duplication events; C.-K.K.C. provided computational support; J.B. and D.E. supervised the research; G.A.K., A.W.D.L., and P.J.R. provided insights into seagrass biology and evolution; A.W.D.L. and P.J.R. provided sequence data; H.-T.L. wrote the article with assistance from D.E.; all authors read and contributed to the writing.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.00868

genes for protein turnover and enzymes related to cellular stress and photosynthesis. Salt tolerance mechanisms in seagrass are of particular interest due to their potential to be applied to improve salinity tolerance in crop plants, such as the use of a *Zostera marina* salt tolerance gene in transgenic *Oryza sativa* research (Zhao et al., 2013). Other studies aim to understand the effect of climate change on seagrass, following major declines reported worldwide (Orth et al., 2006). Seagrass meadows are valuable ecosystems functioning as carbon sinks for long-term carbon storage (Fourqurean et al., 2012). Seagrass growth is predicted to change in an elevated CO_2 environment (Alexandre et al., 2012; Ow et al., 2015), and gene expression (Winters et al., 2011) and transcriptomic responses have been assessed under simulated heatwaves (Franssen et al., 2014) and growth under light limitation (Ralph et al., 2007) in an attempt to predict the response of seagrass to climate change.

The genetic mechanisms underlying the marine adaptation of seagrass lineages remain a fundamental question. EST libraries of two seagrass species, *Posidonia oceanica* and *Z. marina*, revealed three functional groups of positively selected genes, including those encoding glycolysis enzymes, ribosomal genes, and photosynthesis genes (Wissler et al., 2011). In a separate study, ortholog groups of the *Z. marina* transcriptome were found to be more similar to the land plant *O. sativa* than the seaweed *Porphyra yezoensis* (Kong et al., 2014).

A comparative genomics approach was applied to characterize gene conservation and loss in seagrass in comparison with four terrestrial plants and one aquatic plant using unassembled whole-genome shotgun sequence data of *Zostera muelleri* (Golicz et al., 2015). Genes associated with ethylene synthesis and signaling pathways were found to be lost, suggesting that seagrasses have evolved to survive and grow without this important plant hormone, and this loss was confirmed recently in the genome assembly of *Z. marina* (Olsen et al., 2016). When compared with *Spirodela polyrhiza*, the giant duckweed, genes for stomatal differentiation, terpenoid synthesis, and UV light resistance also were lost in both *Z. muelleri* and *Z. marina*, while genes for salinity tolerance and stress resistance were conserved.

Here, we describe the assembly and characterization of the *Z. muelleri* genome, a southern hemisphere temperate to subtropical seagrass species, focusing on seagrass evolution and the mechanisms of adaptation to the marine environment. Specifically, genome comparisons of *Z. muelleri* with terrestrial plants and aquatic floating plants revealed the loss or modification of genes involved in multiple biological processes, including the hormone response and pectin catabolism. Whole-genome duplication (WGD) analysis of *Z. muelleri* has refined the phylogenetic timing of an ancient pan-commelinid WGD event, Tau, and suggested at least one species-specific genome duplication that is not shared by another *Zostera* spp. Together, these results shed light on the response of seagrass to multiple stressors during the evolutionary process.

RESULTS AND DISCUSSION

A whole-genome shotgun sequencing strategy was applied to the southern hemisphere temperate seagrass species *Z. muelleri* ssp. *capricorni* to produce a 632-Mb draft assembly represented by 56,823 scaffolds (Table I).

The genome size of *Z. muelleri* has not been measured previously. However, micrographs showed a chromosome number of $2n = 24$ (Kuo, 2001). A frequency plot of 17-mers from *Z. muelleri* sequence reads estimates the genome size to be 889 Mb (Fig. 1), implying that 71% of the genome was assembled. A minor peak at 75 represents duplicated sequences in the genome, possibly gained through a WGD event, while a small peak at 18 suggests relatively low heterozygosity.

Over 80% of the assembly is represented by 12,904 scaffolds. The scaffold N50 length is 36,732 bp, and the length of the largest scaffold is 303,880 bp. The quality of the draft assembly was evaluated using the CEGMA (Core Eukaryotic Genes Mapping Approach; Parra et al., 2007) and BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al., 2015) data sets (Table II). Eighty-one percent of core eukaryotic genes and 77% of plant BUSCOs searched were detected in the *Z. muelleri* scaffolds. While this metric measures the completeness of genic regions, it also is affected by the degree of conservation of the orthologous data sets, particularly in organisms that are distantly related from model species and the assembly level (contigs/scaffolds/chromosomes) of the draft genome. To account for these factors and also the lack of published BUSCO results for plant species, the assemblies of six other species were searched using the BUSCO data set (Fig. 2).

The assemblies of *Musa acuminata* (3% BUSCOs missing), *Brassica napus* (3% BUSCOs missing), and Arabidopsis (2% BUSCOs missing) at the chromosome

Table I. *Statistics of the Z. muelleri genome assembly*

Parameter	Contigs		Scaffolds		Final Scaffolds	
	No.	Length	No.	Length	No.	Length
		bp		*bp*		*bp*
Total (% of Ns)	237,933	657,815,437 (2.94)	45,719	632,821,539 (3.38)	56,823	632,070,940 (3.49)
Total 10 kb or greater	20,456	425,803,897	14,657	571,260,147	16,674	552,311,609
N90	54,064	1,911	14,488	10,250	18,533	7,926
N50	12,560	14,953	4,043	47,119	5,114	36,732
Longest	–	135,048	–	327,369	–	303,880
Mean length	–	2,765	–		–	11,124

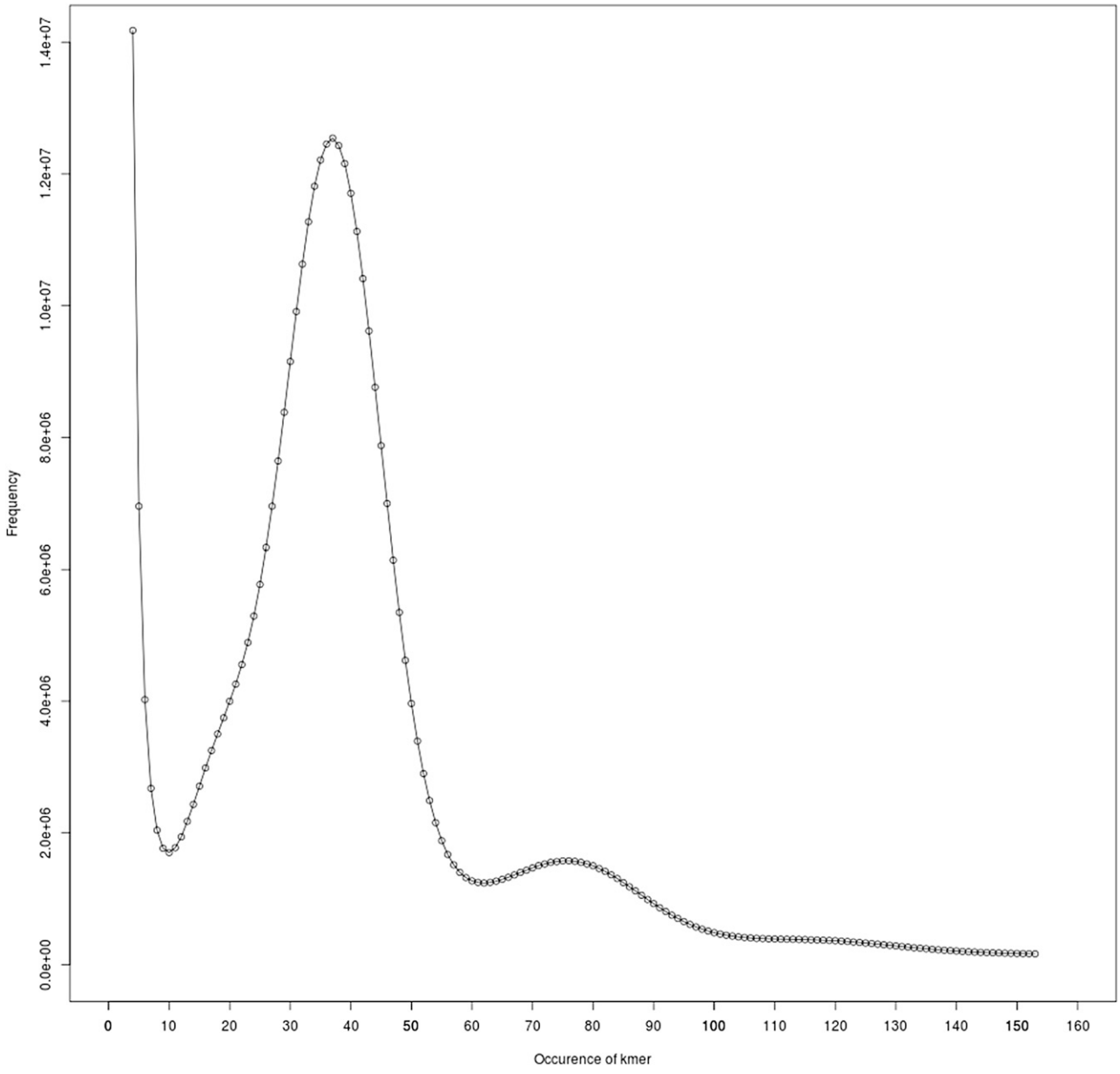

Figure 1. Distribution of 17-mer coverage of *Z. muelleri* read libraries.

level served as well-assembled standards. In order to address the 17% missing BUSCOs in the *Z. muelleri* assembly, *Z. marina* and the contigs of *Amborella trichopoda*, instead of the scaffolds, were analyzed, to control for the degree of conservation of BUSCO orthologs and the comparability of assembly statistics. Since the *A. trichopoda* contigs have 76% complete BUSCOs, whereas the scaffolds have 93% (data not shown), it is highly likely that the amount of missing BUSCOs in *Z. muelleri* is due to the level of assembly.

Interestingly, the number of duplicated BUSCOs detected in *Z. muelleri* is more than double that in *Z. marina*. A total of 63.7% of CEGMA genes also were duplicated in *Z. muelleri*. To elucidate whether this indicates a highly erroneous assembly (Simão et al., 2015) or the occurrence of genome duplications, two species that recently underwent WGD were included for comparison. The *B. napus* genome is highly multiplied due to allopolyploidy (Chalhoub et al., 2014) and triplication around 13 to 17 million years ago (Mya; Yang et al., 1999; Town et al., 2006). As expected, 90% of BUSCOs are duplicated in *B. napus*. Similarly, in *Linum usitatissimum*, which was duplicated 5 to 9 Mya (Wang et al., 2012), 69% of BUSCOs are duplicated. In contrast, the *A. trichopoda* genome, which only retains evidence from the ancient epsilon WGD (Amborella Genome Project, 2013), has 8% duplicated BUSCOs. The presence of three lineage-specific WGDs in *M. acuminata* dated at

Table II. *Assessment results of Z. muelleri scaffolds using CEGMA and BUSCO orthologous groups*

Parameter	CEGMA (Percentage of Total Orthologs Searched)	BUSCO (Percentage of Total Orthologs Searched)
Complete orthologs	201 (81.0)	737 (77.1)
Single-copy	43 (17.3)	365 (38.2)
Duplicated	158 (63.7)	372 (38.9)
Fragmented orthologs	31 (12.5)	52 (5.4)
Missing orthologs	16 (6.5)	167 (17.5)
Total orthologs	248	956

around 65 to 100 Mya (D'Hont et al., 2012), however, is not reflected in the percentage of duplicated BUSCOs, possibly because these WGDs are relatively ancient.

From these results, two conclusions are drawn: (1) the single-copy orthology status of the BUSCO plant data set is valid in genomes with no recent duplications or chromosomal doubling; and (2) *Z. muelleri* possibly underwent at least one recent large-scale genome duplication or is the result of recent hybridization between *Zostera* spp., causing the duplication of gene content.

The Tau WGD, an ancient pan-commelinid WGD (Jiao et al., 2014), identified in most of the sequenced genomes in the monocotyledon lineage including grasses, *M. acuminata*, and date palm (*Phoenix dactylifera*; International Brachypodium Initiative, 2010; D'Hont et al., 2012; Al-Mssallem et al., 2013), appears to be absent from *Z. muelleri*. Tau potentially echoes the pan-core eudicot Gamma event, but the exact timing of these duplications needs refinement (Jiao et al., 2014). Tau was estimated to occur 150 Mya, coinciding with the origin of monocots (Bremer et al., 2009), whereas the split of Alismatales from monocots occurred 130 Mya, based on analysis of the chloroplast gene *rbcL* (Janssen and Bremer, 2004). Phylogenetic analysis shows the absence of Tau in *Z. muelleri* and another member of Alismatales, *S. polyrhiza*, the giant duckweed (Fig. 3; Supplemental Figs. S1 and S2), consistent with recent observations (Ming et al., 2015). These results support that Tau occurred after the divergence of the Alismatales from the main monocot lineage, consistent with a previous study, and that the existing age estimation needs to be investigated further. The presence of the Tau event in *Z. marina* has been suggested (Fig. 2; Olsen et al., 2016), but supporting evidence was not provided. Multiple *Zostera* spp. independent WGD events also were detected (Fig. 3; Supplemental Figs. S1 and S2). However, since the *Z. marina* genome was not available at the time of this analysis, the *Z. muelleri*-specific duplication as observed in BUSCO results is not demonstrated in the phylogenetic trees.

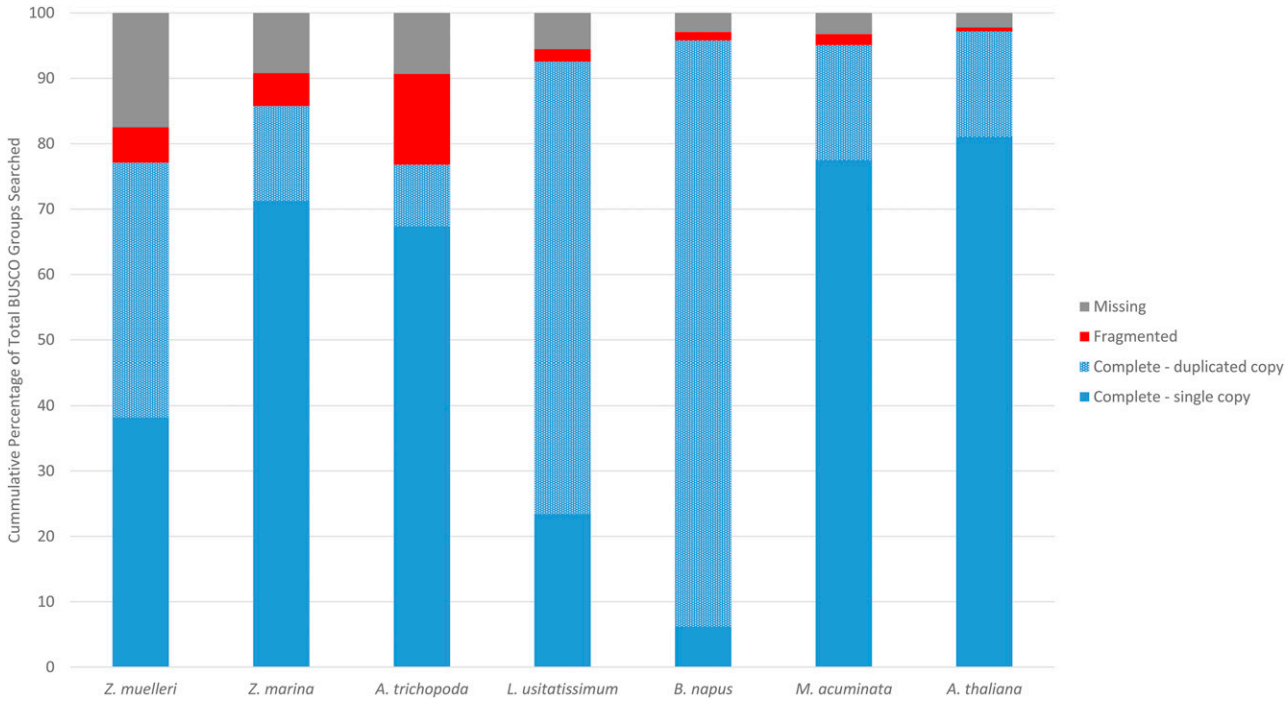

Figure 2. Stack plots of the categories (complete single copy, complete duplicated copy, fragmented, and missing) of BUSCO groups searched in draft assemblies of *Z. muelleri* and six other plant species (*Z. marina*, *Amborella trichopoda*, *Linum usitatissimum*, *Brassica napus*, *Musa acuminata*, and Arabidopsis [*Arabidopsis thaliana*]).

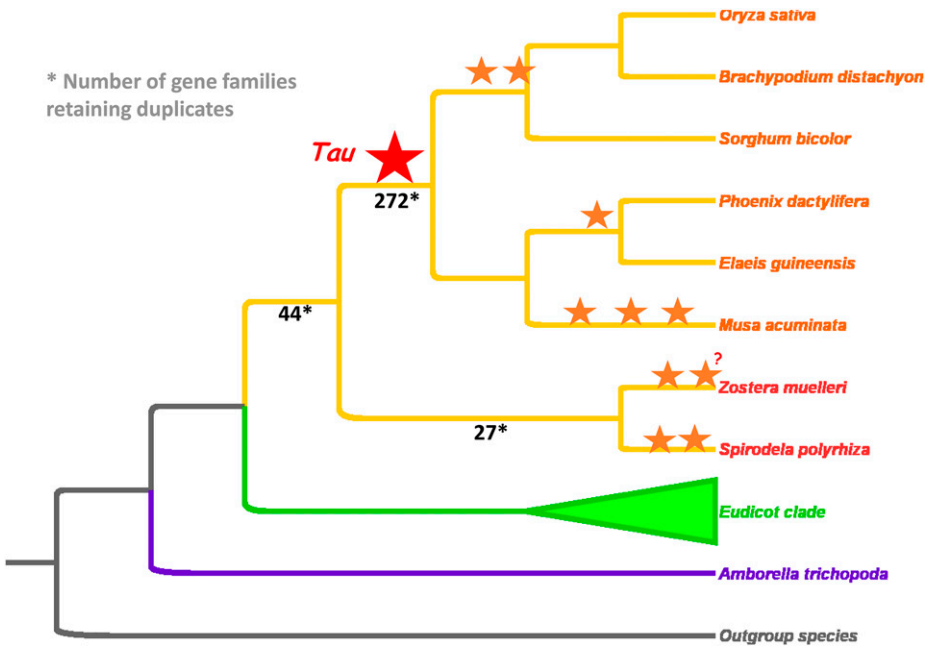

Figure 3. Phylogenetic timing of inferred gene duplications. Values given are numbers of orthogroups with at least one *Z. muelleri* gene showing duplications at the specified branches on the tree. Stars represent WGD events (red, Tau; and orange, clade- or species-specific duplications).

De novo and homology-based repeat prediction using RepBase (version 18.06) identified 339 Mb (56%) of repetitive sequence in the assembly (Supplemental Table S1). Consistent with other sequenced plant species, long terminal repeat (LTR) retrotransposons (22%) are more abundant than non-LTR retrotransposons (3%) and DNA transposons (7%). The frequency ratio of the LTR retrotransposon superfamilies Gypsy/Copia is 2.1, comparable to the seagrasses *P. oceanica* and *Z. marina*, which were recently reported as 1.8 (Barghini et al., 2015) and 1.7 (Olsen et al., 2016).

A total of 35,875 protein-coding genes were predicted, with average gene length of 3,154 bp, coding sequence length of 984 bp, and 5.7 exons (Table III). A cumulative distribution function curve of the annotation edit distance (AED) scores of all *Z. muelleri* genes shows that more than 85% of annotations are supported by homology-based or transcript evidence (AED < 1; Supplemental Fig. S3). AED measures the fit of each annotation to its supporting evidence in the MAKER pipeline (Yandell and Ence, 2012). The sharp increase of annotations from AED 0.9 to 1 observed in the cumulative distribution function curve represents 4,781 ab initio predictions with no supporting evidence. These de novo predictions are retained in the gene set, since the amount of seagrass data (transcripts and proteins) provided is relatively low compared with well-studied plant families and, therefore, might not sufficiently support all true *Z. muelleri* genes. The comparison between *Z. marina* and *Z. muelleri* annotations supports this hypothesis, where 975 of the ab initio *Z. muelleri* genes are orthologous with *Z. marina* genes. A total of 30,135 (95%) genes had identity to sequences in the Swiss-Prot, TrEMBL, and InterProScan databases, while 5,740 remain unannotated (Table IV). A total of 81% (4,655) of the unannotated genes are supported by transcript evidence, with high-confidence AED scores (AED ≤ 0.5).

Based on previous phylogeny studies, *Z. marina* and *Z. muelleri* belong to different clades that diverged around 14 Mya (Coyer et al., 2013). With a genome size that is 70% smaller, *Z. marina* has fewer genes (20,450; Olsen et al., 2016). They shared 11,623 gene families (Supplemental Fig. S4) comprising 20,103 *Z. muelleri* and 14,915 *Z. marina* genes, suggesting the presence of expanded gene families in *Z. muelleri*. Among the shared families are housekeeping genes, such as DNA replication repair genes (GO:0006281 and GO:000724), and genes identified as gained in *Z. marina* (Olsen et al., 2016), such as those involved in metal ion binding (GO:0046872 and GO:0008270) and peptidase activity (GO:0008233 and GO:0008236). Reciprocal BLAST methods show that around 52% of these shared *Z. marina* genes have two copies in *Z. muelleri* (Supplemental Fig. S5). K-mer frequency comparison of the coding sequences of 1,944 randomly selected orthologous pairs also revealed a copy number ratio of 1:1.6. The rate of substitutions per synonymous sites (*Ks*) of paralogous

Table III. *Summary of predicted genes in the Z. muelleri genome*

Total Genes Predicted	Average Gene Size	Average Coding Sequence Length	Average Exons per Gene	Average Exon Length
	bp	*bp*		*bp*
35,875	3154.8	986.8	5.7	225.6

Table IV. *Functional annotation of consensus-predicted genes in the Z. muelleri genome*

Database	No.	Percentage
Total	35,875	100
Annotated		
Swiss-Prot	9,857	27.5
TrEMBL	16,235	45.4
InterProScan	30,135	84.2
Gene Ontology (GO)	18,916	52.9
Unannotated	5,740	16.0

pairs in *Z. muelleri* and *Z. marina* further suggested a species-specific WGD event. Excluding small-scale background duplications, two significant components were demonstrated in *Z. muelleri* paralogous pairs (Supplemental Fig. S6), whereas three were found in *Z. marina* pairs (Supplemental Fig. S7). The peak at 0.17 (red in Supplemental Fig. S6), which was absent in *Z. marina*, was identified as a potential extra WGD event in *Z. muelleri* that occurred at 5.7 to 10.5 Mya, based on assumptions of synonymous mutation rate per base (Koch et al., 2000; Lynch and Conery, 2000). The peak shared between two species at 1.1 (gray in Supplemental Fig. S6 and blue in Supplemental Fig. S7) suggests that the WGD event at 64 to 72 Mya, as shown in *Z. marina* previously (Olsen et al., 2016), is pan lineage. Data from other seagrass species are needed to substantiate this hypothesis.

By comparison with the Arabidopsis, *M. acuminata*, *O. sativa*, *S. polyrhiza*, and *Z. muelleri* genome assemblies, we identified 7,404 gene families, comprising 78,683 genes, common to these five species (Fig. 4). The number of *Z. muelleri*-specific gene families is relatively low (2,233) and represents 16% of all *Z. muelleri* genes. Despite the unique habitat and morphology of seagrasses, these observations suggest that *Z. muelleri* maintains a similar gene content to the land plant species and that the early divergence and marine adaptation of seagrass did not involve a major increase in the abundance of novel gene families. The large number of *O. sativa*-specific gene families (2,880) is consistent with previous studies that demonstrate a high level of gene diversification in Poaceae (Goff et al., 2002). Similarly, a relatively low number of *S. polyrhiza*-specific gene families have been reported (Wang et al., 2014). The large number of gene families (1,131) shared between Arabidopsis, *M. acuminata*, *O. sativa*, and *S. polyrhiza* (AMOS) that are not found in *Z. muelleri* (Fig. 4) suggests that seagrass may have adapted through the loss of terrestrial plant genes rather than the generation of novel, seagrass-specific genes.

GO enrichment analysis identified 30 GO terms that were significantly overrepresented in the biological process category mapped by 196 of the AMOS families (Table V). In a separate analysis, 143 of the AMOS families mapped to 85 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table VI). These results suggest that genes related to at least five processes (pectin catabolism, hormone response, photosynthesis, heat response, and ribosome) were either lost or modified in seagrass. The complete list of enriched and mapped AMOS genes is included in Supplemental Tables S1 and S2.

The GO identifier that represents ethylene signaling (GO:0010105) was enriched significantly in AMOS families (Table V). Previous analyses demonstrated the loss of multiple ethylene biosynthesis and signaling

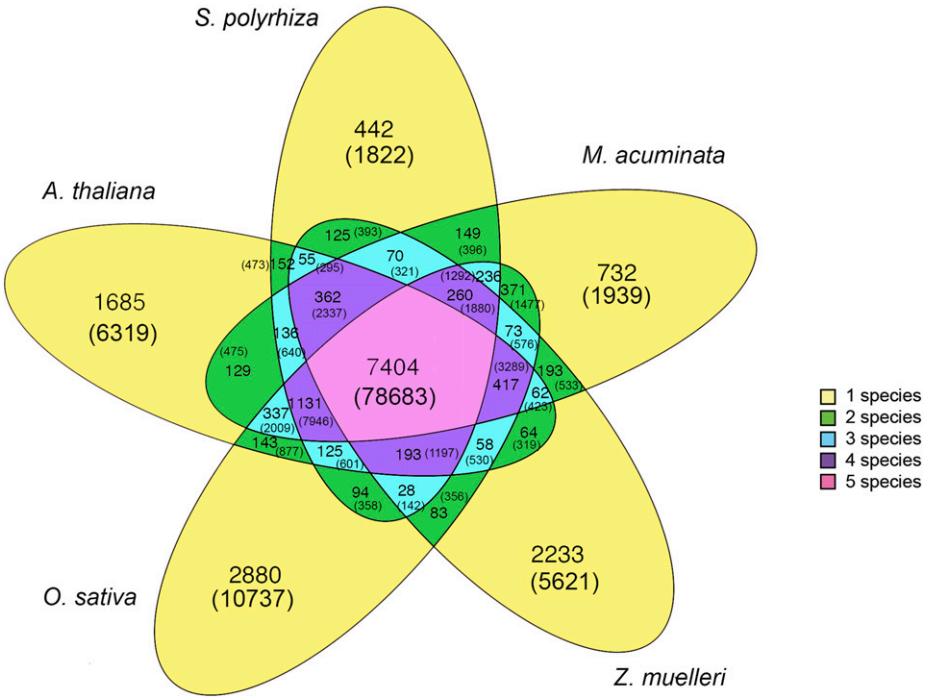

Figure 4. Venn diagram showing the distribution of shared gene families and genes (in parentheses) among five species (*M. acuminata, O. sativa, S. polyrhiza,* Arabidopsis, and *Z. muelleri*). Areas are color coded by the number of species sharing the gene families.

Table V. *Significantly enriched biological process GO terms in AMOS families*

GO Identifier	Term	*P*
GO:0042545	Cell wall modification	1.7e-07
GO:0045490	Pectin catabolic process	4.5e-06
GO:0010105	Negative regulation of ethylene-activated signaling pathway	5.9e-05
GO:0005983	Starch catabolic process	0.00017
GO:0042542	Response to hydrogen peroxide	0.00043
GO:0016567	Protein ubiquitination	0.00057
GO:0009408	Response to heat	0.00155
GO:0000023	Maltose metabolic process	0.00228
GO:0009737	Response to abscisic acid	0.00344
GO:0009965	Leaf morphogenesis	0.00387
GO:0010190	Cytochrome b_6f complex assembly	0.00481
GO:0016236	Macroautophagy	0.00620
GO:0043043	Peptide biosynthetic process	0.00620
GO:0009644	Response to high light intensity	0.00636
GO:0071281	Cellular response to iron ion	0.00696
GO:0002237	Response to molecules of bacterial origin	0.00699
GO:0045892	Negative regulation of transcription, DNA-templated	0.00777
GO:0009739	Response to GA	0.00781
GO:0019252	Starch biosynthetic process	0.00994
GO:0030307	Positive regulation of cell growth	0.01013
GO:0010029	Regulation of seed germination	0.01115
GO:0019375	Galactolipid biosynthetic process	0.01436
GO:0050665	Hydrogen peroxide biosynthetic process	0.01436
GO:0046685	Response to arsenic-containing substance	0.01439
GO:0080060	Integument development	0.01487
GO:0050994	Regulation of lipid catabolic process	0.01487
GO:0009755	Hormone-mediated signaling pathway	0.01487
GO:0010200	Response to chitin	0.01790
GO:0046777	Protein autophosphorylation	0.01860
GO:0008356	Asymmetric cell division	0.02005

genes in *Z. muelleri* and *Z. marina* (Golicz et al., 2015; Olsen et al., 2016). The overrepresentation of gene families encoding 1-aminocyclopropane-1-carboxylic acid synthase, 1-aminocyclopropane-1-carboxylic acid oxidase, ethylene receptors, and ethylene-activated transcription factors in AMOS, and the absence of these genes in seagrass-specific families, support these previous findings.

Genes involved in responses to abscisic acid (GO:0009737), GA (GO:0009739), and the hormone-mediated signaling pathway (GO:0009755) also are enriched in AMOS families. Four out of five *Z. muelleri* proteins with sequence similarity to DELLA proteins clustered in *Z. muelleri*-specific orthologous groups, indicating sequence diversification. DELLA proteins are repressors of the GA signaling pathway, so these changes may be associated with hormone-related cross talk (Davière and Achard, 2016). Although there is no overrepresentation of genes involved in the jasmonate pathway in AMOS, closer examination reveals the loss of genes encoding jasmonate methyltransferase (JMT) in the predicted gene sets of both *Z. muelleri* and *Z. marina*. JMT is an enzyme that converts jasmonic acid (JA) to one of its derivatives, methyl jasmonate (MeJA), which is released as a volatile compound. This loss is further confirmed by the absence of sequences with similarity to plant JMT in the RNA sequencing (RNA-seq) data of *Z. muelleri*. JMT also is absent in transcript and EST collections for other seagrass species, *P. oceania* and *Zostera noltii*. However, key genes for jasmonate biosynthesis (AOS1, AOC, LOXs, OPR3, and JAR1) and signaling (COI1, JAZs, MYCs, and NINJA), and also MeJA esterase, which hydrolyzes MeJA back to JA, are all present in the *Z. muelleri* and *Z. marina* gene sets. Further studies are required to understand the feedback regulation of JA synthesis in the absence of MeJA. Hormone signaling acts as a central regulator of plant responses to environmental stresses. For example, cross-talk between members of the ethylene and jasmonate pathways during the activation of plant defense responses had been shown in several studies (Lorenzo et al., 2003; Onkokesung et al., 2010; Song et al., 2014). With the loss of genes for ethylene production and perception in seagrasses and possible modification of GA and jasmonate signaling pathways, seagrasses provide an interesting model in which to test hypotheses about the roles and interactions of these hormone signaling pathways.

Seagrass pectin is known to contain a rare class of pectin homogalacturonan that is apiose substituted (AGA; Ovodov et al., 1971), and this AGA is also found in the floating plant duckweed (Hart and Kindel, 1970). Seagrass AGA, known as zosterin, has a low level of methyl esterification (Khotimchenko et al., 2012). Cell wall modification (GO:0042545) and pectin catabolic

Table VI. *KEGG pathways mapped by AMOS families*

KEGG Pathway Identifier	KEGG Pathway Name	No. of AMOS Genes Mapped
ath01100	Metabolic pathways	58
ath01110	Biosynthesis of secondary metabolites	35
ath04075	Plant hormone signal transduction	15
ath00040	Pentose and glucuronate interconversions	11
ath00500	Starch and sucrose metabolism	10
ath00940	Phenylpropanoid biosynthesis	8
ath03010	Ribosome	8
ath04626	Plant-pathogen interaction	8
ath01200	Carbon metabolism	8
ath04141	Protein processing in endoplasmic reticulum	7
ath03040	Spliceosome	7
ath00195	Photosynthesis	6
ath00270	Cys and Met metabolism	6
ath01230	Biosynthesis of amino acids	5
ath04144	Endocytosis	5
ath00620	Pyruvate metabolism	4
ath00630	Glyoxylate and dicarboxylate metabolism	4
ath00190	Oxidative phosphorylation	4
ath00130	Ubiquinone and other terpenoid-quinone biosynthesis	3
ath03013	RNA transport	3
ath00330	Arg and Pro metabolism	3
ath00030	Pentose phosphate pathway	3
ath00561	Glycerolipid metabolism	3
ath00010	Glycolysis/gluconeogenesis	3
ath00260	Gly, Ser, and Thr metabolism	2
ath04122	Sulfur relay system	2
ath00052	Gal metabolism	2
ath00790	Folate biosynthesis	2
ath04712	Circadian rhythm	2
ath00520	Amino sugar and nucleotide sugar metabolism	2
ath00410	β-Ala metabolism	2
ath00480	Glutathione metabolism	2
ath03410	Base excision repair	2
ath04145	Phagosome	2
ath03440	Homologous recombination	2
ath00592	α-Linolenic acid metabolism	2
ath00380	Trp metabolism	2
ath00860	Porphyrin and chlorophyll metabolism	2
ath00564	Glycerophospholipid metabolism	2
ath04120	Ubiquitin-mediated proteolysis	2
ath00670	One-carbon pool by folate	2
ath00603	Glycosphingolipid biosynthesis	2
ath03018	RNA degradation	2
ath00903	Limonene and pinene degradation	2
ath03015	mRNA surveillance pathway	2
ath00053	Ascorbate and aldarate metabolism	1
ath00350	Tyr metabolism	1
ath00904	Diterpenoid biosynthesis	1
ath00280	Val, Leu, and Ile degradation	1
ath00360	Phe metabolism	1
ath00511	Other glycan degradation	1
ath00920	Sulfur metabolism	1
ath00240	Pyrimidine metabolism	1
ath04140	Regulation of autophagy	1
ath00531	Glycosaminoglycan degradation	1
ath00230	Purine metabolism	1
ath04146	Peroxisome metabolism	1
ath00062	Fatty acid elongation	1
ath00590	Arachidonic acid metabolism	1
ath00196	Photosynthesis	1
ath00908	Zeatin biosynthesis	1

(*Table continues on following page.*)

Table VI. (*Continued from previous page.*)

KEGG Pathway Identifier	KEGG Pathway Name	No. of AMOS Genes Mapped
ath00071	Fatty acid degradation	1
ath00600	Sphingolipid metabolism	1
ath00604	Glycosphingolipid biosynthesis	1
ath00945	Stilbenoid, diarylheptanoid, and gingerol biosynthesis	1
ath00400	Phe, Tyr, and Trp biosynthesis	1
ath00710	Carbon fixation in photosynthetic organisms	1
ath00020	Citrate cycle	1
ath00460	Cyanoamino acid metabolism	1
ath04070	Phosphatidylinositol signaling system	1
ath03420	Nucleotide excision repair	1
ath03008	Ribosome biogenesis in eukaryotes	1
ath00310	Lys degradation	1
ath03050	Proteasome	1
ath00073	Cutin, suberin, and wax biosynthesis	1
ath00510	*N*-Glycan biosynthesis	1
ath00900	Terpenoid backbone biosynthesis	1
ath04130	SNARE interactions in vesicular transport	1
ath00906	Carotenoid biosynthesis	1
ath00910	Nitrogen metabolism	1
ath00051	Fru and Man metabolism	1
ath03430	Mismatch repair	1
ath03030	DNA replication	1
ath00340	His metabolism	1
ath00730	Thiamine metabolism	1

process (GO:0045490) are the top two enriched GO terms in AMOS. Sequence alignments and motif finding identified 135 predicted seagrass proteins containing pectin methylesterase (PME)-related domains. Nearly half of these remain in seagrass-specific families and do not cluster with PME genes from the four terrestrial plant species. This suggests that the methylesterification level of pectin is tightly controlled in seagrass, possibly to maintain marine osmoregulation. Pectin degradation is partly regulated by ethylene (De Paepe et al., 2004; Onkokesung et al., 2010), and it is unknown whether the modified pectin in seagrass is associated with the loss of ethylene signaling in seagrass (Golicz et al., 2015). This expansion of pectin catabolic genes also is observed in *Z. marina* (Olsen et al., 2016).

Genes involved in starch catabolic (GO:0005983) and biosynthetic (GO:0019252) processes, which were found previously to be lost or heavily modified in *Z. muelleri* reads (Golicz et al., 2015) and *Z. marina* (Olsen et al., 2016), also were significantly enriched in AMOS families (Table V). Genes responsible for the biosynthesis of secondary metabolites that were reduced in *Z. marina* (Olsen et al., 2016) also are found in AMOS (ath01110; Table VI). In addition, eight ribosome-related gene families were highly represented in AMOS (Table VI), and these also were found to be positively selected in *Z. marina* and *P. oceania* ESTs (Wissler et al., 2011).

CONCLUSION

This study of the *Z. muelleri* genome highlights modifications of multiple biological processes associated with adaptation to the sea. These findings provide a base for the further investigation of pathways and responsible genes in submerged marine plants: hormonal cross talk and cell wall modification for effective underwater osmoregulation. The absence of the Tau genome duplication in *Z. muelleri* has further confirmed its phylogenetic placement after the divergence of Alismatales from monocotyledons. The copy number of orthologs between *Z. muelleri* and another *Zostera* spp., together with K_s value analysis, also suggest a species-specific duplication event dated 5.7 to 10.5 Mya. Further study is needed to determine the age estimation of seagrass divergence in relation to Tau and the phylogenetic timing of duplications in the seagrass lineage. The *Z. muelleri* genome contains predicted genes that have no sequence similarity to genes from other species. Further analysis of these genes may assist in our understanding of the basal monocot genome and plant adaptation to marine life.

MATERIALS AND METHODS

Genome and RNA-seq Assembly

Whole-genome sequence data of 300 bp insert size are available in the Sequence Read Archive (SRA) (SRR1714574). Following clone removal using the custom Perl script remove_possible_clones.pl and quality-based trimming using Sickle (Joshi and Fass, 2011), a total of 170,178,513 (89%) paired-end reads were assembled using VelvetOptimizer (Zerbino, 2010) with the following parameters: -s 15 -e 99 -f\-shortPaired -fasta. A total of 79,843 contigs larger than 500 bp were then scaffolded with SSPACE version 2.0 (Boetzer et al., 2011) using 99,002,169 mate-paired reads of 2-kb insert size (SRR3398780). The genome size of *Zostera muelleri* is estimated as (total number of K-mers)/(peak depth). K-mer counting and histogram plotting were performed using Khmer (Pell et al., 2012).

Z. muelleri (DRA identifier ERP010473) and *Zostera marina* transcripts were generated by assembling two sources of publicly available RNA-seq data (SRA identifiers SRP035489 and SRP022957) using Trinity (Haas et al., 2013) with default parameters.

Assessment of Genome Assembly

CEGMA version 2.4 was run on the *Z. muelleri* genome assembly using default parameters. The early release of the plant data set and the necessary software of the program BUSCO were downloaded upon request to the authors. BUSCO was run on *Z. muelleri* and seven other genome assemblies (downloaded from Phytozome: *Z. marina* version 2.1, *Amborella trichopoda* contigs version 1.0, *A. trichopoda* scaffolds version 1.0, *Linum usitatissimum* version 1.0, *Brassica napus* version 1.4 Darmor-*bzh*, *Musa acuminata* version 1, and Arabidopsis [*Arabidopsis thaliana*] The Arabidopsis Information Resource 10) using default parameters.

WGD Analysis

Eighteen genomes were selected for WGD analysis: seven eudicots (Arabidopsis, *Populus trichocarpa*, *Theobroma cacao*, *Vitis vinifera*, *Solanum lycopersicum*, *Solanum tuberosum*, and *Nelumbo nucifera*), eight monocots (*Oryza sativa*, *Brachypodium distachyon*, *Sorghum bicolor*, *Elaeis guineensis*, *Phoenix dactylifera*, *M. acuminata*, *Spirodela polyrhiza*, and *Z. muelleri*), one basal angiosperm (*A. trichopoda*), one lycophyte (*Selaginella moellendorffii*), and one moss (*Physcomitrella patens*). Genome data of all taxa except for *Z. muelleri* were downloaded from Phytozome (version 6) or the respective project Web sites. Orthogroups classification and phylogenomic timing analysis were performed based on a published approach (Jiao et al., 2011, 2014).

Z. muelleri genes were used as queries to align against *Z. marina* genes using the following parameters: blastp -evalue 0.00001 -max_target_seqs 1. The number of query genes per target hit was calculated and plotted in a histogram. The 17-mer frequencies of 1,944 randomly selected orthologous pairs between the two species were analyzed based on a published approach (Vu et al., 2015).

K_s analysis was performed according to a previous approach (Cui et al., 2006). Paralogous pairs of sequences in *Z. muelleri* and *Z. marina* were identified from best reciprocal matches in all-by-all BLAST searches. The EMMIX software was used to fit a mixture model of multivariate normal components to a given data set. The mixed populations were modeled with one to four components.

Gene Prediction and Functional Annotation

A combination of ab initio, evidence-driven, and homology-based annotation methods was used in the MAKER (version 2.31.8) pipeline (Holt and Yandell, 2011). The transposable elements file within the MAKER package, repeats of organisms under the lineage Liliopsida in RepBase, and a *Z. muelleri*-specific repeat library were provided for repeat masking. RepeatModeler (Smit and Hubley, 2008), which employs two repeat-finding programs, RECON (Bao and Eddy, 2002) and RepeatScout (Price et al., 2005), was used to identify de novo repeat families and build a species-specific repeat library. Augustus (Stanke and Waack, 2003) and SNAP (Korf, 2004) were trained using a set of 232 core eukaryotic genes (core eukaryotic genes from CEGMA) and assembled *Z. muelleri* transcripts, respectively. Three sets of evidence were used to assist the annotation process: assembled *Z. muelleri* transcripts, assembled *Z. marina* transcripts from five libraries of published data (Franssen et al., 2014; Kong et al., 2014), and *Z. marina* ESTs from the publicly available database Dr. Zompo (version 2.0; Wissler et al., 2009). Predicted protein sequences for *B. distachyon* and *M. acuminata* were downloaded from Ensembl (release 26) for homology-based annotations. A two-pass MAKER run was performed where the predicted genes from the first MAKER run were used to retrain Augustus and SNAP. Results of the second MAKER run were obtained and evaluated with AED scores and functional annotation.

Predicted proteins were compared with TIGRFAM, ProDom, Panther, PfamA, PrositeProfile, and PrositePatterns using InterProScan (version 5.3; Jones et al., 2014) for motif and domain annotation. A total of 201 proteins containing transposase and reverse transcriptase-associated domain signatures were removed from the gene set. Gene functions were annotated from the best match alignments of predicted genes to Swiss-Prot and TrEMBL plant proteins (release 2015_05) using BLASTX (BLAST+ version 2.28; Camacho et al., 2009) with the parameters blastx -evalue 0.00001 -max_target_seqs 1 and further filtered with 40% or greater identical matches and 60% or greater subject sequence covered.

To identify gene family groups between *Z. muelleri* and terrestrial plants, we downloaded predicted protein sequences for Arabidopsis (Phytozome version 10: Athaliana_167_TAIR10), *O. sativa* (Phytozome version 10: Osativa_204_v7.0), *M. acuminata* (Ensembl Genomes: Musa_acuminata.MA1.22), and *S. polyrhiza* (SpirodelaBase: http://tinyurl.com/lsz3992) and compared them all against all using BLASTP with the parameters blastp -evalue 0.00001 and OrthoMCL (Li et al., 2003). The Venn diagram was plotted using the R package gplots (Warnes et al., 2012) and colored using Adobe Photoshop software.

The same method as above was used to identify gene family groups between *Z. muelleri* and *Z. marina* (Zosma_marina version 2.1).

GO Enrichment and KEGG Pathway Analysis

GO annotation and enrichment were performed on 1,131 Arabidopsis orthologs, each representing one AMOS gene family, based on previous approach (Golicz et al., 2015) with the Arabidopsis whole proteome (2010-09-TAIR10) as background. The Arabidopsis orthologs also were mapped to KEGG pathways (Kanehisa and Goto, 2000) using the KEGG Mapper Web interface (http://www.kegg.jp/kegg/tool/map_pathway2.html).

Accession Numbers

Accession numbers are as follows: SRA identifier SRR1714574, *Z. muelleri* whole-genome sequence paired-end library; SRA identifier SRR3398780, *Z. muelleri* whole-genome sequence mate-paired library; DRA identifier ERP010473, *Z. muelleri* RNA-seq data; SRA identifiers SRP035489 and SRP022957, *Z. marina* RNA-seq data. The assembled genome, annotation, and OrthoMCL groups can be downloaded at http://appliedbioinformatics.com.au/index.php/Seagrass_Zmu_Genome.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Exemplar orthogroup tree (orthoID_1773) showing duplication events in monocots.

Supplemental Figure S2. Exemplar orthogroup tree (orthoID_2030) showing duplication events in monocots.

Supplemental Figure S3. Cumulative distribution function curve of AED scores of all predicted *Z. muelleri* genes.

Supplemental Figure S4. Venn diagram of gene families shared between *Z. muelleri* and *Z. marina*.

Supplemental Figure S5. Histogram of orthologous gene ratio of *Z. muelleri* to *Z. marina*.

Supplemental Figure S6. K_s of paralogous gene pairs in *Z. muelleri* with three distinct components.

Supplemental Figure S7. K_s of paralogous gene pairs in *Z. marina* with four distinct components.

Supplemental Table S1. Repetitive sequences in *Z. muelleri* scaffolds.

Supplemental Table S2. List of genes significantly enriched in AMOS with corresponding GO annotation.

Supplemental Table S3. List of genes in AMOS corresponding to KEGG pathways.

Received June 28, 2016; accepted June 28, 2016; published July 3, 2016.

LITERATURE CITED

Alexandre A, Silva J, Buapet P, Björk M, Santos R (2012) Effects of CO_2 enrichment on photosynthesis, growth, and nitrogen metabolism of the seagrass *Zostera noltii*. Ecol Evol **2:** 2625–2635

Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, Liu J, Pan L, Zhang T, et al (2013) Genome sequence of the date palm *Phoenix dactylifera* L. Nat Commun **4:** 2274

Amborella Genome Project (2013) The Amborella genome and the evolution of flowering plants. Science **342:** 1241089

Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res **12:** 1269–1276

Barghini E, Mascagni F, Natali L, Giordani T, Cavallini A (2015) Analysis of the repetitive component and retrotransposon population in the genome of a marine angiosperm, *Posidonia oceanica* (L.) Delile. Mar Genomics **24:** 397–404

Beer S, Bjork M, Hellblom F, Axelsson L (2002) Inorganic carbon utilization in marine angiosperms (seagrasses). Funct Plant Biol **29:** 349–354

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:** 578–579

Borum J, Pedersen O, Kotula L, Fraser MW, Statton J, Colmer TD, Kendrick GA (2016) Photosynthetic response to globally increasing CO of co-occurring temperate seagrass species. Plant Cell Environ **39:** 1240–1250

Bremer B, Bremer K, Chase MW, Fay MF, Reveal JL, Soltis DE, Soltis PS, Stevens PF, Anderberg AA, Moore MJ, et al (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc **161:** 105–121

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics **10:** 421

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014) Plant genetics: early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345:** 950–953

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. Syst Biodivers **11:** 271–284

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res **16:** 738–749

Dattolo E, Gu J, Bayer PE, Mazzuca S, Serra IA, Spadafora A, Bernardo L, Natali L, Cavallini A, Procaccini G (2013) Acclimation to different depths by the marine angiosperm *Posidonia oceanica*: transcriptomic and proteomic profiles. Front Plant Sci **4:** 195

Dattolo E, Ruocco M, Brunet C, Lorenti M, Lauritano C, D'Esposito D, De Luca P, Sanges R, Mazzuca S, Procaccini G (2014) Response of the seagrass *Posidonia oceanica* to different light environments: insights from a combined molecular and photo-physiological study. Mar Environ Res **101:** 225–236

Davière JM, Achard P (2016) A pivotal role of DELLAs in regulating multiple hormone signals. Mol Plant **9:** 10–20

De Paepe A, Vuylsteke M, Van Hummelen P, Zabeau M, Van Der Straeten D (2004) Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in *Arabidopsis*. Plant J **39:** 537–559

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature **488:** 213–217

Fourqurean JW, Duarte CM, Kennedy H, Marba N, Holmer M, Mateo MA, Apostolaki ET, Kendrick GA, Krause-Jensen D, McGlathery KJ, et al (2012) Seagrass ecosystems as a globally significant carbon stock. Nat Geosci **5:** 505–509

Franssen SU, Gu J, Winters G, Huylmans AK, Wienpahl I, Sparwel M, Coyer JA, Olsen JL, Reusch TBH, Bornberg-Bauer E (2014) Genome-wide transcriptomic responses of the seagrasses *Zostera marina* and *Nanozostera noltii* under a simulated heatwave confirm functional types. Mar Genomics **15:** 65–73

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92–100

Golicz AA, Schliep M, Lee HT, Larkum AW, Dolferus R, Batley J, Chan CK, Sablok G, Ralph PJ, Edwards D (2015) Genome-wide survey of the seagrass *Zostera muelleri* suggests modification of the ethylene signalling network. J Exp Bot **66:** 1489–1498

Greco M, Chiappetta A, Bruno L, Bitonti MB (2013) Effects of light deficiency on genome methylation in *Posidonia oceanica*. Mar Ecol Prog Ser **473:** 103–114

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc **8:** 1494–1512

Hart DA, Kindel PK (1970) Isolation and partial characterization of apiogalacturonans from the cell wall of *Lemna minor*. Biochem J **116:** 569–579

Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12:** 491

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763–768

Janssen T, Bremer K (2004) The age of major monocot groups inferred from 800+ rbcL sequences. Bot J Linn Soc **146:** 385–398

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell **26:** 2792–2802

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature **473:** 97–100

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics **30:** 1236–1240

Joshi NA, Fass JN (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files, version 1.33. https://github.com/najoshi/sickle

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res **28:** 27–30

Khotimchenko Y, Khozhaenko E, Kovalev V, Khotimchenko M (2012) Cerium binding activity of pectins isolated from the seagrasses *Zostera marina* and *Phyllospadix iwatensis*. Mar Drugs **10:** 834–848

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol Biol Evol **17:** 1483–1498

Kong F, Li H, Sun P, Zhou Y, Mao Y (2014) De novo assembly and characterization of the transcriptome of seagrass *Zostera marina* using Illumina paired-end sequencing. PLoS ONE **9:** e112245

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics **5:** 59

Kuo J (2001) Chromosome numbers of the Australian Zosteraceae. Plant Syst Evol **226:** 155–163

Larkum AWD, Orth RJ, Duarte CM (2006) Seagrasses: Biology, Ecology and Conservation. Springer, Dordrecht, The Netherlands

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

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13:** 2178–2189

Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R (2003) ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. Plant Cell **15:** 165–178

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155

Mazzuca S, Spadafora A, Filadoro D, Vannini C, Marsoni M, Cozza R, Bracale M, Pangaro T, Innocenti AM (2009) Seagrass light acclimation: 2-DE protein analysis in *Posidonia* leaves grown in chronic low light conditions. J Exp Mar Biol Ecol **374:** 113–122

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

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

Onkokesung N, Baldwin IT, Gális I (2010) The role of jasmonic acid and ethylene crosstalk in direct defense of *Nicotiana attenuata* plants against chewing herbivores. Plant Signal Behav **5:** 1305–1307

Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, Hughes AR, Kendrick GA, Kenworthy WJ, Olyarnik S, et al (2006) A global crisis for seagrass ecosystems. Bioscience **56:** 987–996

Ovodov YS, Ovodova RG, Bondaren OD, Krasikov IN (1971) The pectic substances of Zosteraceae. Part IV. Pectinase digestion of zosterine. Carbohydr Res **18:** 311–318

Ow YX, Collier CJ, Uthicke S (2015) Responses of three tropical seagrass species to CO_2 enrichment. Mar Biol **162:** 1005–1017

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23:** 1061–1067

Pell J, Hintze A, Canino-Koning R, Howe A, Tiedje JM, Brown CT (2012) Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. Proc Natl Acad Sci USA **109:** 13272–13277

Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. Bioinformatics (Suppl 1) **21:** i351–i358

Ralph PJ, Durako MJ, Enriquez S, Collier CJ, Doblin MA (2007) Impact of light limitation on seagrasses. J Exp Mar Biol Ecol **350:** 176–193

Salo T, Reusch TBH, Bostrom C (2015) Genotype-specific responses to light stress in eelgrass *Zostera marina*, a marine foundation plant. Mar Ecol Prog Ser **519:** 129–140

Short FT, Polidoro B, Livingstone SR, Carpenter KE, Bandeira S, Bujang JS, Calumpong HP, Carruthers TJB, Coles RG, Dennison WC, et al (2011) Extinction risk assessment of the world's seagrass species. Biol Conserv **144:** 1961–1971

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

Smit AFA, Hubley R (2008) RepeatModeler Open-1.0. http://www.repeatmasker.org

Song S, Huang H, Gao H, Wang J, Wu D, Liu X, Yang S, Zhai Q, Li C, Qi T, et al (2014) Interaction between MYC2 and ETHYLENE INSENSITIVE3 modulates antagonism between jasmonate and ethylene signaling in *Arabidopsis*. Plant Cell **26:** 263–279

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics (Suppl 2) **19:** ii215–ii225

Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, et al (2006) Comparative genomics of Brassica oleracea and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell **18:** 1348–1359

Vu GTH, Schmutzer T, Bull F, Cao HX, Fuchs J, Tran TD, Jovtchev G, Pistrick K, Stein N, Pecinka A, et al (2015) Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. Plant Genome **8:** doi/10.3835/plantgenome2015.04.0021

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

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, et al (2012) The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. Plant J **72:** 461–473

Warnes GR, Bolker B, Lumley T (2012) gplots: various R programming tools for plotting data. R package version 2.6.0, https://cran.r-project.org/web/packages/gplots/index.html.

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*. Mar Ecol Prog Ser **435:** 83–95

Wissler L, Codoñer FM, Gu J, Reusch TB, Olsen JL, Procaccini G, Bornberg-Bauer E (2011) Back to the sea twice: identifying candidate plant genes for molecular evolution to marine life. BMC Evol Biol **11:** 8

Wissler L, Dattolo E, Moore AD, Reusch TBH, Olsen JL, Migliaccio M, Bornberg-Bauer E, Procaccini G (2009) Dr. Zompo: an online data repository for Zostera marina and Posidonia oceanica ESTs. Database (Oxford) **2009:** 10.1093/database/bap009

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet **13:** 329–342

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol **48:** 597–604

Zerbino DR (2010) Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics **Chapter 11:** Unit 11.15

Zhao F, Li Q, Weng M, Wang X, Guo B, Wang L, Wang W, Duan D, Wang B (2013) Cloning of TPS gene from eelgrass species *Zostera marina* and its functional identification by genetic transformation in rice. Gene **531:** 205–211