



Comparative Chloroplast Genomes of *Zosteraceae* Species Provide Adaptive Evolution Insights Into Seagrass

Jun Chen^{1†}, Yu Zang^{3†}, Shuai Shang^{1,4}, Shuo Liang¹, Meiling Zhu¹, Ying Wang^{1,2*} and Xuexi Tang^{1,2*}

¹ College of Marine Life Sciences, Ocean University of China, Qingdao, China, ² Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, ³ Key Laboratory of Marine Eco-Environmental Science and Technology, First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China, ⁴ College of Biological and Environmental Engineering, Binzhou University, Binzhou, China

OPEN ACCESS

Edited by:

Levi Yant,
University of Nottingham,
United Kingdom

Reviewed by:

Sunil Kumar Sahu,
Beijing Genomics Institute (BGI),
China

Dennis William Stevenson,
New York Botanical Garden,
United States

*Correspondence:

Ying Wang
ywang@ouc.edu.cn
Xuexi Tang
tangxx@ouc.edu.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 14 July 2021

Accepted: 23 August 2021

Published: 23 September 2021

Citation:

Chen J, Zang Y, Shang S,
Liang S, Zhu M, Wang Y and Tang X
(2021) Comparative Chloroplast
Genomes of *Zosteraceae* Species
Provide Adaptive Evolution Insights
Into Seagrass.
Front. Plant Sci. 12:741152.
doi: 10.3389/fpls.2021.741152

Seagrasses are marine flowering plants found in tropical and sub-tropical areas that live in coastal regions between the sea and land. All seagrass species evolved from terrestrial monocotyledons, providing the opportunity to study plant adaptation to sea environments. Here, we sequenced the chloroplast genomes (cpGenomes) of three *Zostera* species, then analyzed and compared their cpGenome structures and sequence variations. We also performed a phylogenetic analysis using published seagrass chloroplasts and calculated the selection pressure of 17 species within seagrasses and nine terrestrial monocotyledons, as well as estimated the number of shared genes of eight seagrasses. The cpGenomes of *Zosteraceae* species ranged in size from 143,877 bp (*Zostera marina*) to 152,726 bp (*Phyllospadix iwatensis*), which were conserved and displayed similar structures and gene orders. Additionally, we found 17 variable hotspot regions as candidate DNA barcodes for *Zosteraceae* species, which will be helpful for studying the phylogenetic relationships and interspecies differences between seagrass species. Interestingly, nine genes had positive selection sites, including two ATP subunit genes (*atpA* and *atpF*), two ribosome subunit genes (*rps4* and *rpl20*), two DNA-dependent RNA polymerase genes (*rpoC1* and *rpoC2*), as well as *accD*, *clpP*, and *ycf2*. These gene regions may have played key roles in the seagrass adaptation to diverse environments. The Branch model analysis showed that seagrasses had a higher rate of evolution than terrestrial monocotyledons, suggesting that seagrasses experienced greater environmental pressure. Moreover, a branch-site model identified positively selected sites (PSSs) in *ccsA*, suggesting their involvement in the adaptation to sea environments. These findings are valuable for further investigations on *Zosteraceae* cpGenomes and will serve as an excellent resource for future studies on seagrass adaptation to sea environments.

Keywords: *Zosteraceae*, seagrass, chloroplast genome, genome structure, adaptive evolution

INTRODUCTION

Seagrasses are marine flowering plants in class Monocotyledoneae that form vast meadows, and flower and seed underwater. As opposed to lower plants such as macroalgae (seaweeds), seagrasses are higher angiosperms that can live fully in seawater. They have terrestrial plant origins and reentered the sea millions of years ago (Orth et al., 2006). The globally recognized seagrasses have been divided into 6 families and 13 genera (~74 species), while China mainly contains 4 families and 10 genera (22 species), accounting for ~30% of the global seagrasses (Short et al., 2011; Huang et al., 2016). Seagrass beds are formed by large-scale seagrasses, which may consist of either a single or multiple species, and possess important ecological functions, such as stabilizing coastal sediments and thereby maintaining water quality, providing diverse habitats for several unique organisms, serving as nurseries for marine organisms (fish, crustaceans, and mollusks), and providing nutrients (N and P) and organic carbon to other parts of the oceans, including the deep sea, as well as contribute significantly to carbon sequestration (Suchanek et al., 1985; Hemminga and Duarte, 2000; Duarte et al., 2005; Orth et al., 2006).

The general trend of biological evolution is from low to high (from aquatic to terrestrial), but some unique evolutionary events occur during the evolutionary process, such as “second entry into water.” Some species have gradually changed from terrestrial to aquatic after their ancestors separated from their terrestrial relatives (sea otters, cetaceans, pinnipeds, and seagrasses). The ancestors of seagrasses separated from terrestrial monocotyledons 70–100 million years ago and have since adapted to submerged marine life (Waycott et al., 2007). During this process, a series of adaptive changes took place in their morphological structures and physiological ecology (Ye, 2002). As members of a polyphyletic group of plants, seagrasses live in the tidal and subtidal regions of marine habitats (Kato et al., 2003). Seagrasses are mainly found in the transition zone between the sea and terrestrial environments, which experiences some of the greatest environmental pressure on Earth. Owing to the long-term dual and interactive effects of the sea and terrestrial environments, there is ample opportunity to explore and uncover complex and diverse adaptive mechanisms and evolutionary information. However, to date, seagrass studies have mainly focused on species investigations, distribution monitoring, and ecological restoration (Short and Coles, 2001; Orth et al., 2006; Short et al., 2011). With the development of sequencing technology, hotspots in the field of evolutionary biology have aimed to reveal the evolutionary adaptive mechanisms of seagrasses at the molecular or genomic level. Olsen et al., conducted the first genomic study in 2016 on *Zostera marina* and found that *Z. marina* has lost genes for regulating stomatal opening, terpenoid and ethylene synthesis, ultraviolet and infrared sensing, and plant pigment that are common in terrestrial angiosperms for adapting to the marine environment. Additionally, *Z. marina* has acquired unique genes that regulate ion metabolism, nutrient absorption, gas exchange, and other life activities, which have allowed for the adaptation to the complex marine environment (high salt, high osmotic pressure, and low

light intensity) (Olsen et al., 2016; Williams, 2016). Moreover, in 2016, the complete genome of *Z. muelleri* was sequenced and it was found that, to adapt to the marine environment, genes related to hormone biosynthesis, signal transduction, and cell wall catabolism were either lost or modified (Lee et al., 2016).

Chloroplasts are multifunctional organelles in plant cells that play critical roles in photosynthesis and carbon fixation (Wicke et al., 2011; Daniell et al., 2016). The cpGenomes have a typical quadripartite structure with a pair of inverted repeats (IRs) separated by a small single-copy (SSC) and a large single-copy (LSC) region (Sugiura, 1992). Most angiosperm cpGenomes are remarkably conserved in terms of structure, gene content, and order (Wicke et al., 2011). Generally, plant cpGenomes are recombination-free, maternally inherited, and have low rates of nucleotide substitutions, which make them valuable sources of genetic markers for phylogenetic and population genetic analyses (Korpelainen, 2004; Ravi et al., 2008). Adaptive evolution is defined as the improved adaptability of a species to changing environmental conditions during the evolutionary process. Given the conservation of the cpGenome, studies on these genomes will enhance our understanding of plant adaptive evolution. To our knowledge, a comparative analysis on the cpGenomes of seagrasses that have adapted to the sea environment has not yet been performed.

In this study, we aimed to provide comprehensive insights into the evolution of the cpGenomes of several seagrass species. First, we sequenced the cpGenomes of three *Zosteraceae* species and conducted comparative cpGenome analyses using these three genomes and previously annotated cpGenomes of *Z. marina*. Then, we constructed a phylogenetic tree using the cpGenomes of all published seagrass genomes obtained from the NCBI database. Finally, we estimated positive selection for seagrass species and detected the type of selection pressure (positive, neutral, and purifying) to address whether they experienced different evolutionary forces compared to terrestrial monocotyledons.

MATERIALS AND METHODS

Sample Collection and Sequencing Assembly

Samples of three seagrass species (*Z. nigricaulis*, *Phyllospadix iwatensis*, and *Z. japonica*) were collected from their natural habitat with *Z. japonica* (120.6801 E, 37.93384 N) and *Phyllospadix iwatensis* (120.6476 E, 37.97221 N) collected from Yantai, China, and *Z. nigricaulis* (144.415392 E, 38.142656 S) collected from Geelong, Australia, respectively. All samples are currently stored at the Marine Ecology Laboratory of Ocean University of China. Total DNA was extracted from fresh leaves using TRIzol® Reagent (Invitrogen). Library construction for Illumina sequencing was performed using an Illumina TruSeq™ Nano DNA Sample Prep Kit. Total DNA was sequenced using an Illumina HiSeq 4000 platform (150 bp*2) (Shanghai BIOZERON Co., Ltd., Shanghai, China). Then, we removed potential low-quality reads from the raw reads using Trimmomatic 0.39 software (Bolger et al., 2014). NOVOPlasty v2.7.2 software (Dierckxsens et al., 2017) was used to assemble the cpGenomes

using *Z. marina* as the reference sequence. GapCloser software (Luo et al., 2012) was used to fill the remaining local inner gaps and correct the single base polymorphisms for the final assembly results. Finally, four junctions between the IRs and SSC/LSC regions were confirmed using the reference sequence.

Chloroplast Genome Annotation

The cpGenomes were annotated using GeSeq (Tillich et al., 2017) and annotation results were checked with BLAST and DOGMA (Wyman et al., 2004). Then, all protein-coding sequences were BlastP against the non-redundant (Nr) NCBI, SwissProt, KEGG, and COG databases for their functional annotations. Then, a circular map of the three seagrass species was obtained using Organellar Genome DRAW (Greiner et al., 2019). Finally, the newly obtained chloroplast genomes of three *Zosteraceae* species were submitted into the GenBank under accession numbers were MZ576842, MZ573775, MZ571509, respectively.

Comparative Chloroplast Genome Analysis

The cpGenome structures of the four *Zosteraceae* species (including *Z. marina*) were compared using mVISTA software (Frazer et al., 2004). Subsequently, we extracted all coding regions and intergenic spacers (IGSs) to examine the regions of divergence within the four *Zosteraceae* species for further phylogenetic analysis. The percentage of variable sites was calculated within each homologous region. Then, the expansion and contraction of the IR regions were analyzed by comparing the positions of the LSC/IR and SSC/IR junctions, and their adjacent genes using IRscope (Amiryousefi et al., 2018).

Phylogenetic Relationships

Phylogenetic analysis of eight seagrass cpGenomes from order Alismatales was performed using *Oryza minuta* as the outgroup. All shared plastid coding genes (PCGs) were concatenated into a super matrix and aligned using MUSCLE (Edgar, 2004). The evolutionary history was estimated using the maximum likelihood (ML) method based on the JTT + G + I + F model with 1,000 bootstrap replicates using MEGA X software (Kumar et al., 2018).

Adaptive Evolution Analysis

We compared the rate of non-synonymous (d_N) and synonymous (d_S) substitutions using the CODEML program, implemented by PAML v4.7 (Yang, 2007), to quantify the selective pressure. Ratios of d_N/d_S were calculated, where $\omega = 1$, $\omega > 1$, and $\omega < 1$ indicated neutral, positive, and purifying selection, respectively. We used the species phylogenetic tree obtained by previous studies (Davis et al., 2013; Petersen et al., 2016; Ross et al., 2016; Givnish et al., 2018) as the input tree file (Supplementary Figure 1).

To detect positively selected sites (PSSs) in eight seagrass species, we used two site-specific models (M7 and M8) in the CODEML program. A likelihood ratio test (LRT) with χ^2 distribution was performed to determine and compare nested models with a $p < 0.05$ significance threshold. Bayes

empirical Bayes (BEB) analysis was used to identify sites under positive selection with posterior probabilities > 0.80 (Ziheng et al., 2005). PSSs were identified using the Datamonkey web server (Pond and Frost, 2005) based on three methods, the FEL, SLAC, and MEME models, with a $p < 0.1$ significance threshold. In this study, positive selection sites detected by ≥ 3 methods were considered positive selection sites.

To determine whether each shared PCG faced a different evolutionary force in different habitats, Branch and branch-site models were conducted to detect positive selection in the eight seagrasses species, which were collectively set as the foreground branch. The LRT with χ^2 distribution was performed to determine and compare the nested models with a $p < 0.05$ significance threshold. BEB analysis was used to identify sites under positive selection with posterior probabilities > 0.80 .

RESULTS

Genomic Features of *Zosteraceae* Species

The cpGenome sizes of the three species and *Z. marina* of family *Zosteraceae* ranged from 143,877 bp in *Z. marina* to 152,726 bp in *P. iwatensis*, which were composed of four regions, including an LSC and SSC region separated by two IRs (Figure 1 and Table 1). The length differences of the genomes were mainly determined by the SSC length. The GC content of the cpGenome sequences ranged from 35.46 to 36.18% and the highest GC content was 36.18% in *P. iwatensis*, followed by *Z. japonica* (35.89%), *Z. nigricaulis* (35.88%), and *Z. marina* (35.46%). The number of genes present in the 4 plastomes of the *Zosteraceae* species was similar, except for *Z. marina*, which lost some genes and thus had a reduced gene number. There were 132 genes in *P. iwatensis*, 131 in *Z. nigricaulis*, 127 in *Z. japonica*, and 116 in *Z. marina*, which had the lowest number of genes. *Z. nigricaulis* and *P. iwatensis* had 38 tRNA genes, while *Z. japonica* and *Z. marina* had 34 and 30, respectively. All species had eight rRNA genes. The number of PCGs ranged from 78 to 86. Among all genes, there were 7–19 single intron genes, and *ycf3*, *clpP* and two *rps12* (except for *Z. marina*) had two introns each.

Boundary Regions and Comparative Analysis

There were 4 boundaries between the two IRs, and LSC and SSC in the cpGenome, namely, IRa-SSC (JSA), IRa-LSC (JLA), IRb-LSC (JLB), and IRb-SSC (JSB). When comparing the cpGenomes of the *Zosteraceae* species (Figure 2), we found that the four boundaries of the cpGenomes were relatively conserved. The JLB border was located between *rpl22* and *trnH*, except in *P. iwatensis*, which cut through *rps19*, resulting in a pseudogene. The JSB border was between *ycf1* and *trnL* in the 3 *Zostera* species, where *ycf1* was within the IRb. *P. iwatensis* had a 1,312-bp pseudogene. The JSA border was within *ndhF*, resulting in a 13–565-bp

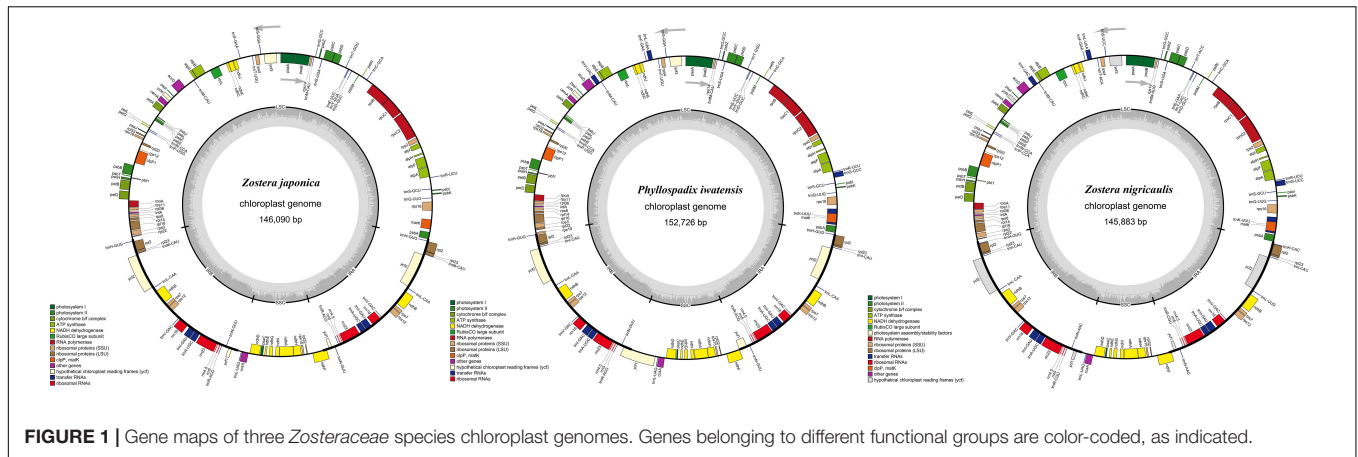


FIGURE 1 | Gene maps of three *Zosteraceae* species chloroplast genomes. Genes belonging to different functional groups are color-coded, as indicated.

TABLE 1 | Chloroplast genome features of four *Zosteraceae* species.

Species name	Genome size (bp)	LSC size (bp)	IR size (bp)	SSC size (bp)	GC (%)	Total no. of genes	PCGs No	tRNA No	rRNA No	Single intron	Double intron
<i>Zostera nigricaulis</i>	145,883	83,541	24,430	13,482	35.88	131	85	38	8	19	4
<i>Phyllospadix iwatensis</i>	152,726	84,753	25,167	17,639	36.18	132	86	38	8	19	4
<i>Zostera japonica</i>	146,090	83,664	24,628	13,169	35.89	127	85	34	8	13	4
<i>Zostera marina</i>	143,877	83,224	25,915	8,823	35.46	116	78	30	8	7	2

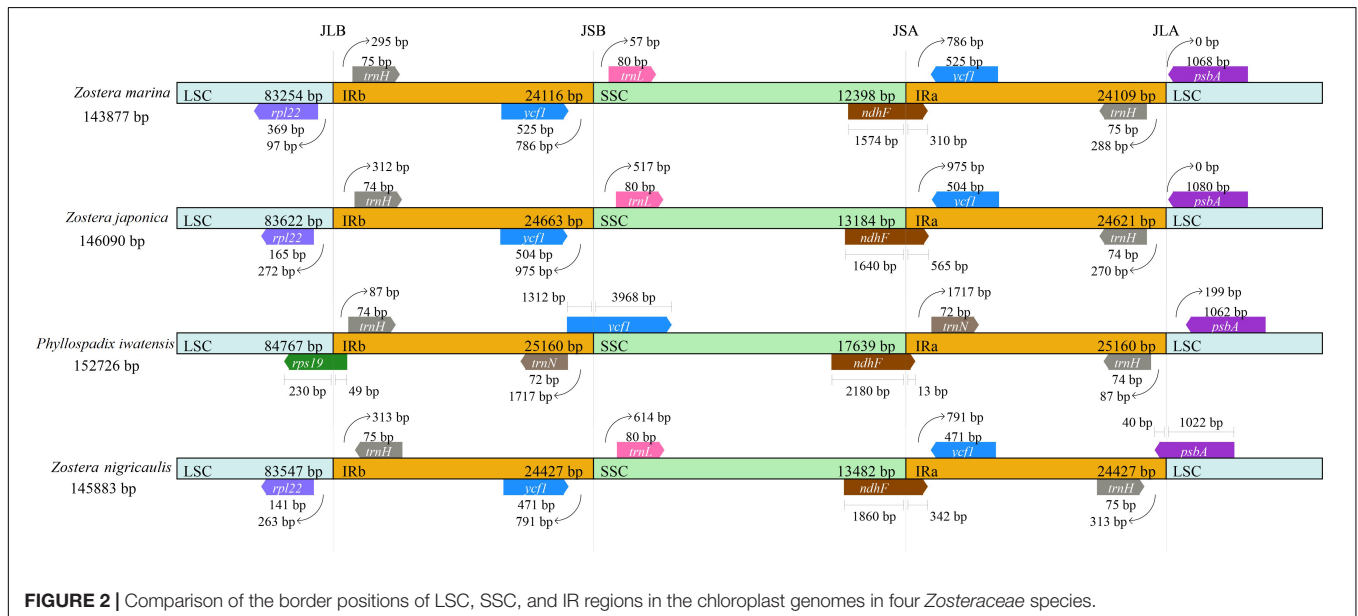


FIGURE 2 | Comparison of the border positions of LSC, SSC, and IR regions in the chloroplast genomes in four *Zosteraceae* species.

pseudogene in four species. Moreover, *psbA* was located entirely in the LSC and across the JLA border in *Z. japonica*, resulting in a 40-bp pseudogene in the IRa.

The mVISTA program was used to analyze the overall sequence identity of the cpGenome of the *Zosteraceae* species, using the *Z. japonica* annotation as a reference (Figure 3). The cpGenomes were conserved and displayed similar structures and gene orders. The divergence level of the non-coding regions was

higher than the coding regions. The IGS regions had the highest levels of divergence. Additionally, the LSC and SSC regions had a larger divergence than the IR regions (Figure 4).

Phylogenetic Relationships

To investigate the phylogenetic relationship of the seagrasses, we constructed an ML tree using MEGA X. As can be seen from the tree (Figure 5), the shared PCGs divided into three major clades:

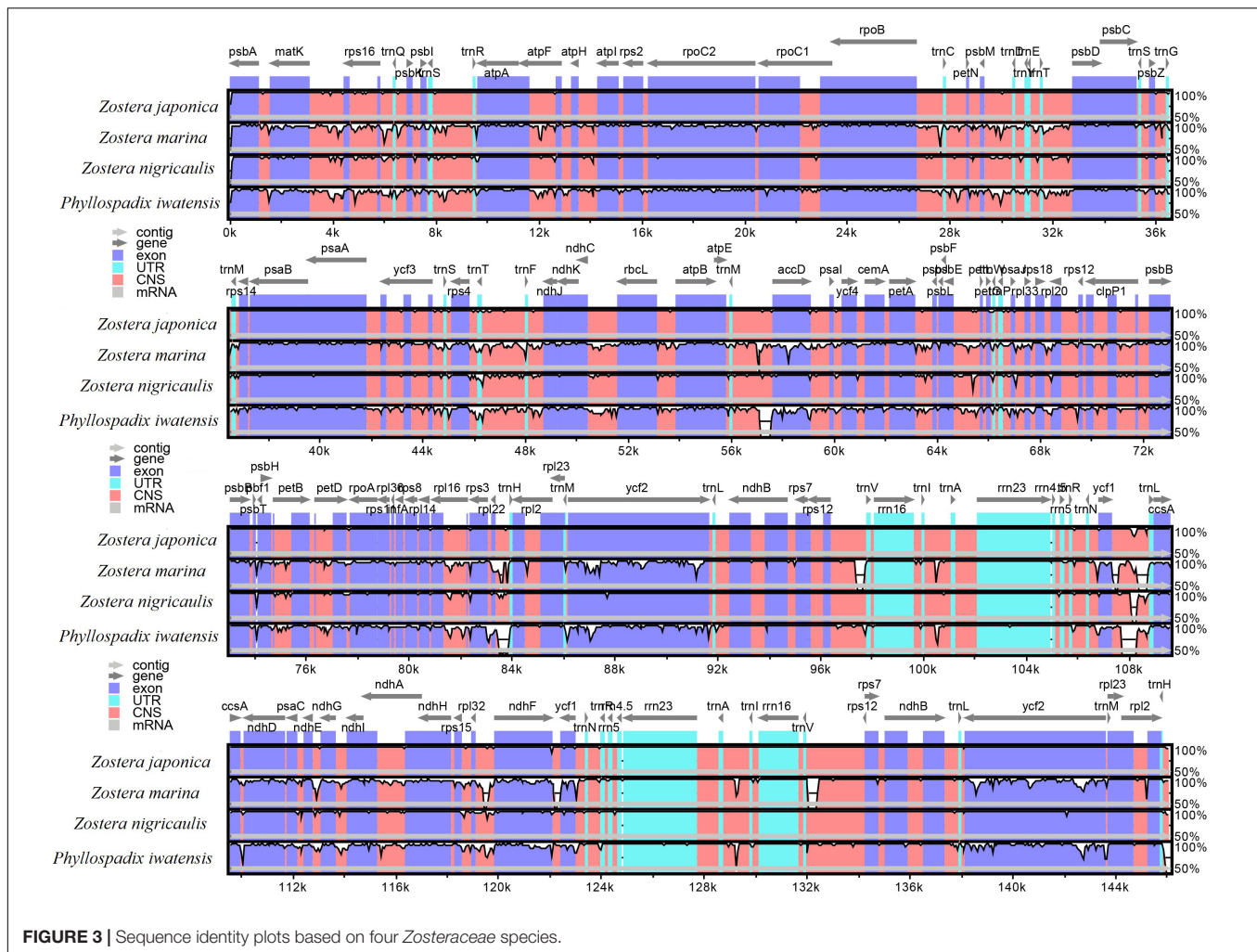


FIGURE 3 | Sequence identity plots based on four *Zosteraceae* species.

Hydrocharitaceae, *Zosteraceae*, and *Ruppiales*, among which, *Zosteraceae* and *Ruppiales* had a relatively close relationship. Moreover, *Z. marina* was a sister species to *Z. nigricaulis* and *Z. japonica* within *Zosteraceae*. These results showed that the phylogenetic proximity of all shared PCGs was closely related to the traditional taxonomic group (Ross et al., 2016).

Adaptive Evolution Analysis

Model M8 and the SLAC, FEL, and MEME methods were used to detect specific codons under selection in 59 single-copy shared genes of 8 seagrass species (Supplementary Table 1). Sites identified by ≥ 3 methods were regarded as robust candidates for positive selection sites. A total of nine genes with at least one common positively selected codon were identified, including 2, 2, and 13 common positively selected codons in *accD*, *clpP*, and *ycf2*, respectively.

The results of the Branch model analysis of all shared PCGs showed that the ω values ($\omega = 0.16933$) were significant ($p < 0.01$), indicating that at the whole-chloroplast protein level, monocotyledonous plants were strongly purified and selected for during their evolution (Table 2). Moreover, the

ω values were significantly higher in seagrass groups than in terrestrial monocotyledonous groups (Table 2), indicating that the chloroplast protein-encoding genes in seagrasses evolved more rapidly than in terrestrial monocotyledons. Furthermore, the analysis of 54 single-copy shared genes showed that 23 genes had significantly different evolutionary rates between the two groups and all evolved rapidly in seagrass groups, except *cemA* and *psbT* (Supplementary Table 2).

A total of 54 single-copy genes were also used for selective pressure estimation in the branch-site model analysis (Supplementary Table 3). Only Model A and Model A Null of *ccsA* were significant ($p = 0.022$). The BEB posterior probability of an amino acid site (213K) was > 0.80 .

DISCUSSION

Chloroplast Genome Features

In this study, the newly sequenced three *Zosteraceae* cpGenomes and previously published cpGenomes of *Z. marina* were roughly similar in size, gene order, and composition, and ranged in size from 143,877 to 152,726 Kb. The cpGenomes of the

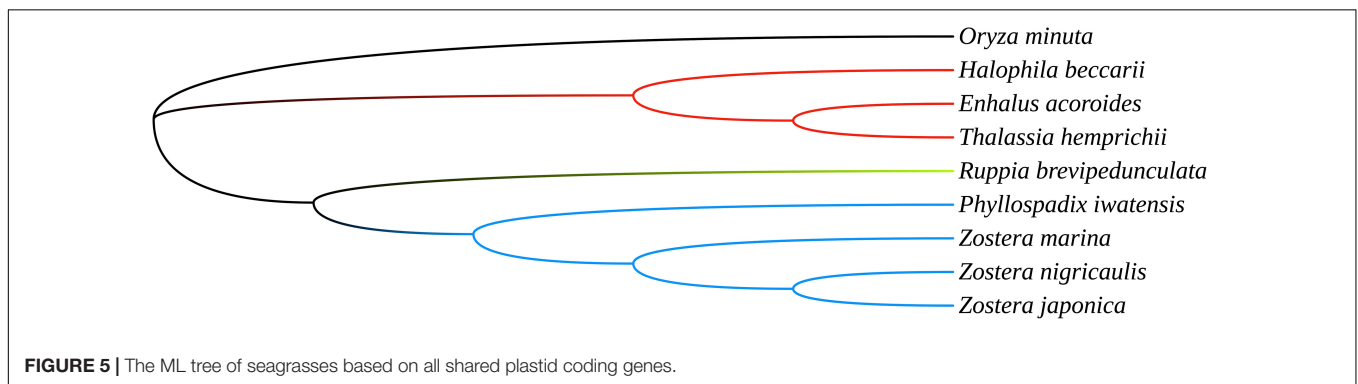
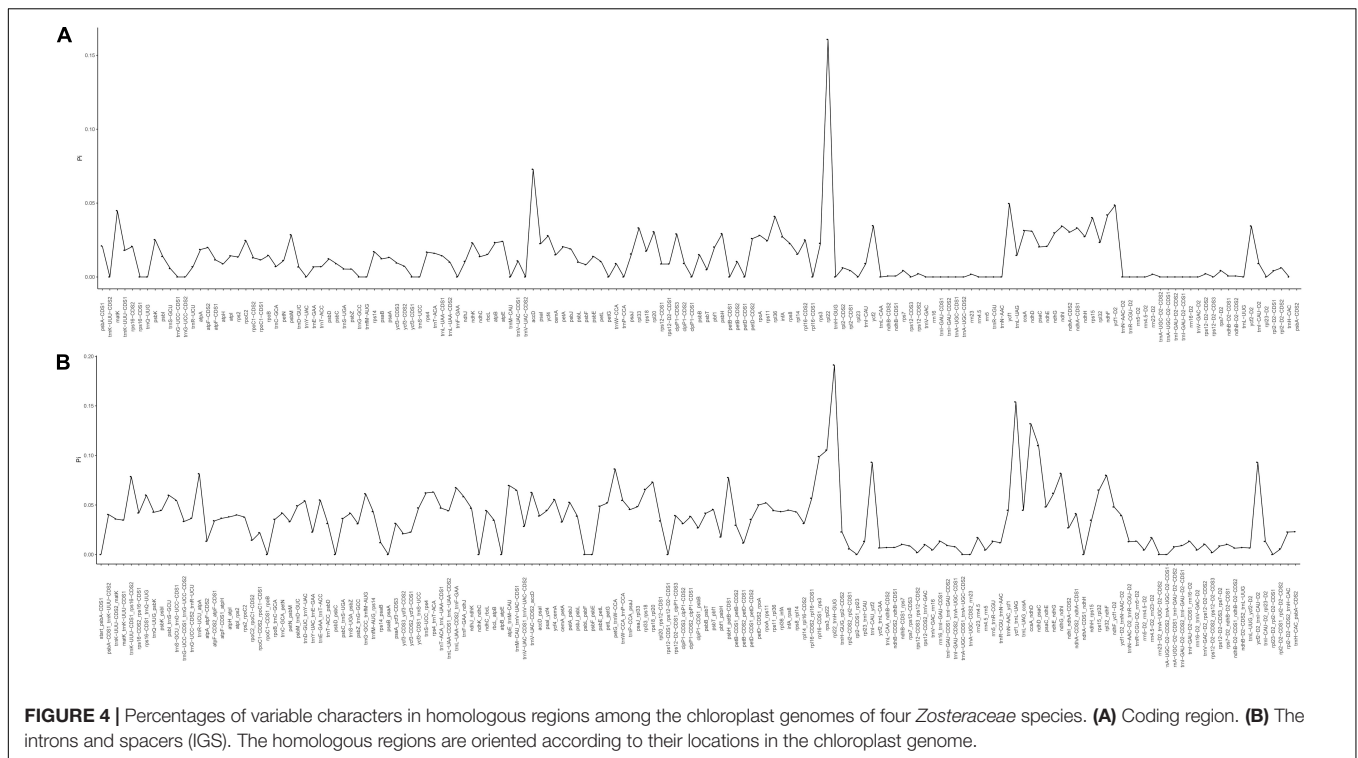


TABLE 2 | Selection pressure test results of all shared protein-coding genes in monocotyledons through a Branch model analysis.

Model	np	LnL	ω for branch	Model compared	P-values
A: All the branch has same ω	34	-176110.3	$\omega = 0.16933$		
B: All the branch has same $\omega = 1$	33	-184904.03	$\omega = 1$	B vs. A	$p < 0.01$
C: Clade of Seagrass has a ω_1 ; other clade has a ω_2	35	-175986.49	$\omega_1 = 0.21950$ $\omega_2 = 0.14204$	C vs. A	$p < 0.01$

4 species were comparable in size to other monocotyledons (Huotari and Korpelainen, 2012) and their sequences were highly conserved with no gene rearrangement events detected. Previous studies showed that gene loss was a common phenomenon among parasitic, submerged, and carnivorous plant plastomes (Wakasugi et al., 1994; Braukmann et al., 2009;

Wicke et al., 2011), and was usually associated with functional transfers to the nucleus (Fleischmann et al., 2011). Similar events were also detected in our study, where *rps19* was completely lost in *Zostera*, and Hydrocharitaceae lost several NDH-related genes (**Supplementary Table 4**). Plastid NDH-related genes code for most of the subunits of the plastid NADH dehydrogenase enzyme

complex that play a role in photooxidative stress responses (Martín and Sabater, 2010). Interestingly, previous studies have also shown a significant loss of NDH-related genes in the family of Hydrocharitaceae (Ross et al., 2016). Therefore, we speculate that this gene loss may be due to the adaptation to specific seagrass habitats and some genes may have been transferred to the nucleus. Previous studies showed that GC content is an important indicator of species affinity and GC skewness is an indicator of DNA lagging chains, leading chains, replication origin, and replication termini (Tillier and Collins, 2000; Necşulea and Lobry, 2007). Generally, in this study, plastomes were characterized by a low GC content in species from order Alismatales and the lowest GC content was detected in *Zosteraceae*. Moreover, the overall GC content of the IR region in *Zosteraceae* was higher than the LSC and SSC regions, a phenomenon that is common in other plants (Fan et al., 2018; Guo et al., 2018). Thus, we hypothesized that this may be caused by the presence of four RNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) in the IR region.

Sequence Variations

During cpGenome evolution, the expansion and contraction of IR boundaries allow for some genes to enter IR regions, while other genes can enter single-copy sequences, but the degree of sequence replication at the boundaries of each species is different (Yan et al., 2017; Fan et al., 2018; Wu et al., 2020). Our study showed that, although the cpGenomes of family *Zosteraceae* had a conserved structure, there were significant size differences caused mainly by IR boundary shifts (Figures 2, 3). For example, *ycf1* in *P. iwatensis* spanned the SSC/IRB region and spanned 3,968 bp in the SSC region, but was greatly reduced and located in the IR region of the other three species. Moreover, gene loss in *Zosteraceae* also caused IR boundary shifts. A previous study showed that the LSC/IRB border lies within the *rps19* and *trnH-GUG* gene cluster in many monocot plastomes (Wang et al., 2008). However, in our study, due to the loss of *rps19*, the boundary of three *Zosteraceae* species, except *P. iwatensis*, was located between *rpl22* and *trnH-GUG*. To determine the divergence hotspots, we compared the whole cpGenome sequences of the *Zosteraceae* species to compute the percentages of variable characters in coding and non-coding regions (Figure 4). Our results indicated that the proportion of variable sites was higher in the non-coding regions than the coding regions, which is in accordance with the results found for other taxa (Fan et al., 2018; Sun et al., 2020). Among the four regions of the cpGenome, the sequence variation in the IR region was the lowest, while the variation in the LSC region was the highest. Thus, we speculated that the highly conserved IR region may be related to the higher GC content. Considering the proportion and number of variable sites, we propose 17 (*rpl22*, *accD*, *rpl22-trnH-GUG*, *ycf1-trnL-UAG*, *ccsA-ndhD*, *ndhD-psaC*, *rps3-rpl22*, *rpl16-rps3*, *trnI-CAU-ycf2*, *ycf2-trnI-CAU*, *petG-trnW-CCA*, *ndhG-ndhI*, *trnR-UCU-atpA*, *rpl32-ndhF*, *trnK-UUU_rps16*, *psbH-petB*, and *rps18-rpl20*) of the most variable hotspot regions as candidate DNA barcodes for *Zosteraceae*, which will aid future studies on

the phylogenetic relationships and interspecies differences of seagrass species.

Adaptive Selection

In our study, we found clear signatures of positive selection sites in nine genes shared by the eight seagrass species, which were identified as robust candidate sites using three methods. These genes included two ATP subunit genes (*atpA* and *atpF*), two ribosomal subunit genes (*rps4* and *rpl20*), two DNA-dependent RNA polymerase genes (*rpoC1* and *rpoC2*), as well as *accD*, *clpP*, and *ycf2* (Supplementary Table 1). ATP synthase is essential for plant photosynthesis and is usually a product of two genetic systems in plants (Westhoff et al., 1985). Six ATP subunit genes (*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, and *atpI*) are encoded and synthesized in chloroplasts, and two genes exhibited site-specific selection in this study. Additionally, 16 genes were identified that encode ribosomal subunits, of which, two were under positive selection. Moreover, *rpoC* mainly encodes the β subunit of RNA polymerase. In this study, *rpoC1* and *rpoC2* were under positive selection, which may lead to alterations in cell wall metabolism, possibly as a result of altered transcription (Bisson et al., 2012; Cai et al., 2021). ACCase catalyzes malonyl-CoA formation from acetyl-CoA and bicarbonate in the first committed step of *de novo* fatty acid synthesis (Rawsthorne, 2002). The plastid, ACCase, is composed of four gene products, which were thought to be coordinately expressed, and *accD* is only encoded in the plastid genome (β -CT), while others are nuclear (Ke et al., 2000; Lee et al., 2004). In this study, we identified PSSs in *accD*, which may have played key roles in seagrass fatty acid biosynthesis. Additionally, *clpP* plays an important role in plant cells. The main function of its product is polypeptide degradation and it is a member of a gene family within the cpGenome that encodes clpP proteases (Clarke, 1999; Kuroda and Maliga, 2003). A previous study showed that *clpP* was under positive selection in Dipsacales species (Fan et al., 2018). We also found that *ycf2* had 13 sites under positive selection. This gene is the largest chloroplast gene reported in angiosperms and has become a useful gene for assessing sequence variations and evolutionary processes in plants (Drescher et al., 2000; Huang et al., 2010). Positive selection of *ycf2* was also found to be involved in the adaptation of other species (Fan et al., 2018; Zhong et al., 2019; Wu et al., 2020). However, owing to its unknown function, *ycf2* is an excellent candidate gene for future studies on the adaptive evolution of seagrass species. Moreover, these positively selected genes may have played key roles in seagrass adaptation to various environments.

Previous studies have shown that the cpGenome of seed plants has a lower evolutionary rate than nuclear genes (Drouin et al., 2008). In this study, the Branch model results showed that monocotyledonous plants experienced strong purifying selection, indicating that the cpGenomes of monocotyledonous plants were relatively conserved during evolution (Table 2). Furthermore, the Branch model analysis of the shared PCGs were significantly different ($p < 0.01$) between Model A and Model C (Table 2). These results suggest that the foreground and background branches have different evolutionary rates, indicating that seagrasses had a higher rate of evolution than

terrestrial monocotyledons. Seagrasses mainly grow in the transition zone between the marine and terrestrial environments. Due to long-term dual and mutual effects of the marine and terrestrial environments, their habitats undergo great energy fluctuations. With the change of the tides, they are periodically directly exposed to the air and susceptible to enhanced UV-B radiation. Previous studies have shown that environmental energy stimulates metabolism on many levels and it is known that energy-rich habitats are often characterized by higher evolutionary rates (Davies et al., 2004; Clarke and Gaston, 2006). Solar radiation, especially UV radiation, plays a direct mutagenic role and may accelerate molecular evolution (Rothschild, 1999; Willis et al., 2009). Thus, we speculated that the evolutionary differences between seagrasses and terrestrial monocotyledons may be because seagrasses faced greater selective pressure in their habitats. The analysis of single-copy shared genes also showed that most significantly different (21/23) genes evolved rapidly in seagrass groups, except *cemA* and *psbT* (Supplementary Table 2). Adaptive evolution is likely to occur in specific lineages. We conducted a branch-site model analysis to further detect whether selection was limited in seagrasses. We identified 1 positively selected site (213K) in *ccsA* (Supplementary Table 3), which mainly encodes a protein required for heme attachment to c-type cytochromes (Xie and Merchant, 1996), which may have allowed seagrasses to adapt to sea environments.

CONCLUSION

In this study, we sequenced the cpGenomes of 3 seagrass species, which revealed the cpGenome features between *Zosteraceae* (*Z. nigricaulis*, *P. iwatensis*, *Z. japonica*, and *Z. marina*). We also uncovered the phylogenetic relationship and evolution of single-copy shared genes alongside published seagrass sequences from the NCBI database. Our analyses revealed that all *Zosteraceae* species shared similar genome structures, gene orders, and compositions, and significant size differences were caused mainly by IR boundary shifts and gene loss. Site-specific selection analysis showed that some of the coding sites of nine chloroplast genes (*atpA*, *atpF*, *rps4*, *rpl20*, *rpoC1*, *rpoC2*, *accD*, *clpP*, and *ycf2*) underwent protein sequence evolution. Additionally, different levels of selection detected between seagrass and terrestrial monocotyledons lineages suggest their involvement in adaptation to sea environments. Phylogenetic results showed that *Z. marina* is a sister species to *Z. nigricaulis* and *Z. japonica* within

REFERENCES

- Amiryousefi, A., Hyvönen, J., and Poczar, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- Bisson, G. P., Mehaffy, C., Broeckling, C., Prenni, J., Rifat, D., Lun, D. S., et al. (2012). Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, *rpoB* mutant *Mycobacterium tuberculosis*. *J. Bacteriol.* 194, 6441–6452. doi: 10.1128/jb.01013-12
- Bolger, A. M., Marc, L., and Bjoern, U. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Zosteraceae, which is supported by the classification of this genus. These findings are valuable for further investigations on *Zosteraceae* cpGenomes and will serve as an excellent resource for future studies on seagrass adaptation to sea environments.

DATA AVAILABILITY STATEMENT

The chloroplast genomes of three *Zosteraceae* species data for this study were submitted into the GenBank under accession numbers were MZ576842, MZ573775, and MZ571509, respectively.

AUTHOR CONTRIBUTIONS

JC conceived, designed the study, and wrote the manuscript. JC and YZ performed the experiments. SS, SL, and MZ contributed materials and analysis tools. XT and YW revised the manuscript. All authors approved the final manuscript.

FUNDING

This research was funded by the National Key R&D Program of China (No. 2019YFD0901204), the NSFC-Shandong Joint Fund (No. U1806213), the National Natural Science Foundation of China (No. 42006144), and the NSFC-Shandong Joint Fund for Marine Ecology and Environmental Sciences (No. U1606404).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.741152/full#supplementary-material>

Supplementary Figure 1 | Phylogenetic trees of 17 monocotyledonous species.

Supplementary Table 1 | Positively selected sites of 59 single-copy genes shared by eight seagrass species.

Supplementary Table 2 | Selection pressure test results of shared protein-coding genes in seagrasses through a Branch model analysis.

Supplementary Table 3 | Selection pressure test results of shared protein-coding genes in seagrasses through a branch-site model analysis.

Supplementary Table 4 | Protein-coding gene losses in eight seagrass species.

- Braukmann, T. W. A., Kuzmina, M., and Stefanović, S. (2009). Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr. Genet.* 55, 323–337. doi: 10.1007/s00294-009-0249-7
- Cai, X.-L., Landis, J. B., Wang, H.-X., Wang, J.-H., Zhu, Z.-X., and Wang, H.-F. (2021). Plastome structure and phylogenetic relationships of *Styracaceae* (Ericales). *BMC Ecol. Evol.* 21:103. doi: 10.1186/s12862-021-01827-4
- Clarke, A., and Gaston, K. J. (2006). Climate, energy and diversity. *Proc. R. Soc. B Biol. Sci.* 273, 2257–2266.
- Clarke, A. K. (1999). ATP-dependent Clp proteases in photosynthetic organisms—a cut above the rest! *Ann. Bot.* 83, 593–599. doi: 10.1006/anbo.1999.0878

- Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 1–29.
- Davies, T. J., Savolainen, V., Chase, M. W., Moat, J., and Barraclough, T. G. (2004). Environmental energy and evolutionary rates in flowering plants. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 271, 2195–2200. doi: 10.1098/rspb.2004.2849
- Davis, J. L., Mcneal, J. R., Barrett, C. F., Chase, M. W., Cohen, J. I., Duvall, M. R., et al. (2013). Contrasting patterns of support among plastid genes and genomes for major clades of the monocotyledons. *Early Events Monocot Evol.* 83, 315–349. doi: 10.1017/cbo9781139002950.015
- Dierckxens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18.
- Drescher, A., Ruf, S., Calsa, T. Jr., Carrer, H., and Bock, R. (2000). The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22, 97–104. doi: 10.1046/j.1365-313x.2000.00722.x
- Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49, 827–831. doi: 10.1016/j.ympev.2008.09.009
- Duarte, C. M., Middelburg, J. J., and Caraco, N. (2005). Major role of marine vegetation on the oceanic carbon cycle. *Biogeosciences* 2, 1–8. doi: 10.5194/bg-2-1-2005
- Edgar, R. C. (2004). "MUSCLE: Multiple Sequence Alignment with Improved Accuracy and Speed", in: *Computational Systems Bioinformatics Conference*. New York: IEEE.
- Fan, W.-B., Wu, Y., Yang, J., Shahzad, K., and Li, Z.-H. (2018). Comparative chloroplast genomics of dipsacales species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front. Plant Sci.* 9:689. doi: 10.3389/fpls.2018.00689
- Fleischmann, T. T., Scharff, L. B., Alkatib, S., Hasdorf, S., Schöttler, M. A., and Bock, R. (2011). Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for redutive genome evolution. *Plant Cell* 23, 3137–3155. doi: 10.1105/tpc.111.088906
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279.
- Givnish, T. J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V. K., Saarela, J. M., et al. (2018). Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am. J. Bot.* 105, 1888–1910. doi: 10.1002/ajb.2.1178
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64.
- Guo, S., Guo, L., Zhao, W., Xu, J., Li, Y., Zhang, X., et al. (2018). Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii*. *Molecules* 23:246. doi: 10.3390/molecules23020246
- Hemminga, M. A., and Duarte, C. M. (2000). *Seagrass ecology*. Cambridge: Cambridge University Press.
- Huang, J. L., Sun, G. L., and Zhang, D. M. (2010). Molecular evolution and phylogeny of the angiosperm *ycf2* gene. *J. Syst. Evol.* 48, 240–248. doi: 10.1111/j.1759-6831.2010.00080.x
- Huang, X.-P., Jiang, Z.-J., Fan, H.-Q., Chen, X.-Y., Zhou, Y., Zhang, J.-P., et al. (2016). The nomenclature of the "algae" name of seagrasses in China. *Oceanol. Limnol. Sinica* 47, 290–294.
- Huotari, T., and Korpelainen, H. (2012). Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene* 508, 96–105. doi: 10.1016/j.gene.2012.07.020
- Kato, Y., Aioi, K., Omori, Y., Takahata, N., and Satta, Y. (2003). Phylogenetic analyses of *Zostera* species based on *rbcL* and *matK* nucleotide sequences: implications for the origin and diversification of seagrasses in Japanese waters. *Genes Genet. Syst.* 78, 329–342. doi: 10.1266/ggs.78.329
- Ke, J., Wen, T.-N., Nikolau, B. J., and Wurtele, E. S. (2000). Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme A carboxylase. *Plant Physiol.* 122, 1057–1072. doi: 10.1104/pp.122.4.1057
- Korpelainen, H. (2004). The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften* 91, 505–518. doi: 10.1007/s00114-004-0571-3
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kuroda, H., and Maliga, P. (2003). The plastid *clpP1* protease gene is essential for plant development. *Nature* 425, 86–89. doi: 10.1038/nature01909
- Lee, H., Golitz, A. A., Bayer, P. E., Jiao, Y., Tang, H., Paterson, A. H., et al. (2016). The genome of a southern hemisphere seagrass species (*Zostera muelleri*). *Plant Physiol.* 172, 272–283. doi: 10.1104/pp.16.00868
- Lee, S. S., Jeong, W. J., Bae, J. M., Bang, J. W., Liu, J. R., and Harn, C. H. (2004). Characterization of the plastid-encoded carboxyltransferase subunit (*accD*) gene of potato. *Mol. Cells* 17, 422–429.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.
- Martin, M., and Sabater, B. (2010). Plastid *ndh* genes in plant evolution. *Plant Physiol. Biochem.* 48, 636–645. doi: 10.1016/j.plaphy.2010.04.009
- Necşulea, A., and Lobry, J. R. (2007). A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.* 24, 2169–2179. doi: 10.1093/molbev/msm148
- Olsen, J. L., Rouzép, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., et al. (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* 530, 331–335.
- Orth, R. J., Carruthers, T. J. B., Dennison, W. C., Duarte, C. M., Fourqurean, J. W., Heck, K. L., et al. (2006). A global crisis for seagrass ecosystems. *BioScience* 56, 987–996.
- Petersen, G., Seberg, O., Cuenca, A., Stevenson, D. W., Thadeo, M., Davis, J. I., et al. (2016). Phylogeny of the Alismatales (Monocotyledons) and the relationship of *A. corus* (*A. corales?*). *Cladistics* 32, 141–159. doi: 10.1111/cla.12120
- Pond, S. L. K., and Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bti320
- Ravi, V., Khurana, J., Tyagi, A., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Syst. Evol.* 271, 101–122.
- Rawsthorne, S. (2002). Carbon flux and fatty acid synthesis in plants. *Progress Lipid Res.* 41, 182–196. doi: 10.1016/s0163-7827(01)00023-6
- Ross, T. G., Barrett, C. F., Soto Gomez, M., Lam, V. K., Henriquez, C. L., Les, D. H., et al. (2016). Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* 32, 160–178. doi: 10.1111/cla.12133
- Rothschild, L. J. (1999). The influence of UV radiation on protistan evolution. *J. Eukaryotic Microbiol.* 46, 548–555. doi: 10.1111/j.1550-7408.1999.tb06074.x
- Short, F. T., and Coles, R. G. (2001). *Global seagrass research methods*. Amsterdam: Elsevier.
- Short, F. T., Polidoro, B., Livingstone, S. R., Carpenter, K. E., Bandeira, S., Bujang, J. S., et al. (2011). Extinction risk assessment of the world's seagrass species. *Biol. Conserv.* 144, 1961–1971.
- Suchanek, T. H., Williams, S. L., Ogden, J. C., Hubbard, D. K., and Gill, I. P. (1985). Utilization of shallow-water seagrass detritus by Caribbean deep-sea macrofauna: $\delta^{13}C$ evidence. *Deep Sea Res. Part A Oceanogr. Res. Papers* 32, 201–214. doi: 10.1016/0198-0149(85)90028-7
- Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168.
- Sun, J., Wang, Y., Liu, Y., Xu, C., Yuan, Q., Guo, L., et al. (2020). Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. *Sci. Rep.* 10, 1–10.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11.
- Tillier, E. R., and Collins, R. A. (2000). The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *J. Mol. Evol.* 50, 249–257. doi: 10.1007/s002399910029
- Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci.* 91, 9794–9798. doi: 10.1073/pnas.91.21.9794
- Wang, R.-J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M., and Chaw, S.-M. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* 8:36. doi: 10.1186/1471-2148-8-36

- Waycott, M., Procaccini, G., Les, D. H., and Reusch, T. (2007). *Seagrass Evolution, Ecology and Conservation: A Genetic Perspective*. Netherlands: Springer.
- Westhoff, P., Alt, J., Nelson, N., and Herrmann, R. G. (1985). Genes and transcripts for the ATP synthase CF₀ subunits I and II from spinach thylakoid membranes. *Mol. General Genet.* 199, 290–299. doi: 10.1007/bf00330271
- Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Williams, S. L. (2016). Genomics: from sea to sea. *Nature* 530, 290–291.
- Willis, K., Bennett, K., and Birks, H. (2009). Variability in thermal and UV-B energy fluxes through time and their influence on plant diversity and speciation. *J. Biogeogr.* 36, 1630–1644. doi: 10.1111/j.1365-2699.2009.02102.x
- Wu, Z., Liao, R., Yang, T., Dong, X., Lan, D., Qin, R., et al. (2020). Analysis of six chloroplast genomes provides insight into the evolution of Chrysosplenium (Saxifragaceae). *BMC Genom.* 21:621. doi: 10.1186/s12864-020-07045-4
- Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352
- Xie, Z., and Merchant, S. (1996). The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes. *J. Biol. Chem.* 271, 4632–4639. doi: 10.1074/jbc.271.9.4632
- Yan, M., Moore, M. J., Meng, A., Yao, X., and Wang, H. (2017). The first complete plastome sequence of the basal asterid family Styracaceae (Ericales) reveals a large inversion. *Plant Syst. Evol.* 303, 61–70. doi: 10.1007/s00606-016-1352-0
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Ye, C.-J. (2002). Advances in the study on the marine higher plant eelgrass (*Zostera marina* L.) and its adaptation to submerged life in seawater. *Chinese Bull. Bot.* 19:184.
- Zhong, Q., Yang, S., Sun, X., Wang, L., and Li, Y. (2019). The complete chloroplast genome of the Jerusalem artichoke (*Helianthus tuberosus* L.) and an adaptive evolutionary analysis of the *ycf2* gene. *PeerJ* 7:e7596. doi: 10.7717/peerj.7596
- Ziheng, Y., Wong, W. S. W., and Rasmus, N. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen, Zang, Shang, Liang, Zhu, Wang and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.