

RESEARCH

Open Access



Chloroplast genomic comparison provides insights into the evolution of seagrasses

Jun Chen¹, Yu Zang², Shuai Shang¹, Zhibo Yang¹, Shuo Liang¹, Song Xue¹, Ying Wang^{1,3*} and Xuexi Tang^{1,3*}

Abstract

Background Seagrasses are a polyphyletic group of monocotyledonous angiosperms that have evolved to live entirely submerged in marine waters. Thus, these species are ideal for studying plant adaptation to marine environments. Herein, we sequenced the chloroplast (cp) genomes of two seagrass species (*Zostera muelleri* and *Halophila ovalis*) and performed a comparative analysis of them with 10 previously published seagrasses, resulting in various novel findings.

Results The cp genomes of the seagrasses ranged in size from 143,877 bp (*Zostera marina*) to 178,261 bp (*Thalassia hemprichii*), and also varied in size among different families in the following order: Hydrocharitaceae > Cymodoceaceae > Ruppiaceae > Zosteraceae. The length differences between families were mainly related to the expansion and contraction of the IR region. In addition, we screened out 2,751 simple sequence repeats and 1,757 long repeat sequence types in the cp genome sequences of the 12 seagrass species, ultimately finding seven hot spots in coding regions. Interestingly, we found nine genes with positive selection sites, including two ATP subunit genes (*atpA* and *atpF*), three ribosome subunit genes (*rps4*, *rps7*, and *rpl20*), one photosystem subunit gene (*psbH*), and the *ycf2*, *accD*, and *rbcl* genes. These gene regions may have played critical roles in the adaptation of seagrasses to diverse environments. In addition, phylogenetic analysis strongly supported the division of the 12 seagrass species into four previously recognized major clades. Finally, the divergence time of the seagrasses inferred from the cp genome sequences was generally consistent with previous studies.

Conclusions In this study, we compared chloroplast genomes from 12 seagrass species, covering the main phylogenetic clades. Our findings will provide valuable genetic data for research into the taxonomy, phylogeny, and species evolution of seagrasses.

Keywords Seagrass, Chloroplast genome, Repeats, Phylogenetic analysis, Adaptive evolution

*Correspondence:

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

¹ College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong, China

² Ministry of Natural Resources, Key Laboratory of Marine Eco-Environmental Science and Technology, First Institute of Oceanography, Qingdao, Shandong, China

³ Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao, Shandong, China

Background

Along with coral reefs and mangroves, seagrass beds are one of the three main marine ecosystems, providing important habitats for marine life. They have significant ecological functions, such as stabilizing coastal sediments, providing nursery grounds for juvenile fish, and sequestering carbon [1–4]. The currently recognized seagrasses consist of approximately 74 species of six families and 13 genera in the order Alismatales, accounting for less than 1% of all flowering plant species [5–7]. Les et al. (1997) detected three separate origins of seagrasses using the *rbcl* gene, which was confirmed in subsequent



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

studies [7–9]. Through independent evolutionary routes, seagrasses growing in intertidal and subtidal zones are characterized by similar environments, such as high salinity, low light, anaerobic soils, and extreme tides. Seagrasses have evolved shared traits such as salt tolerance, slender and soft leaves, carbon-concentrating mechanisms, and aerenchyma in the roots and rhizomes [10–14].

The unique evolutionary characteristics of seagrasses are similar to those of whales, which evolved from the land to aquatic environments [15]. The emergence of seagrasses is one of the most notable evolutionary transformations in the history of angiosperms. Recently, the broad application of technologies for genome sequencing has been demonstrated to be a valuable phylogenetic and evolutionary tool for revealing the genetic development and adaptive evolutionary mechanism of seagrasses to the marine environment [14, 16, 17]. For example, the first complete genome sequence and detailed genomic analysis of *Zostera marina* found that it had discarded several key innovative features of angiosperms, such as genes involved in stomatal and ethylene pathways, terpenoid synthesis, ultraviolet protection and far-infrared sensing, during its secondary entry into the sea, and also that it had recoded its cell wall components and expanded its sucrose synthesis and transport, ion transport and light-harvesting chlorophyll b-protein genes to adapt to the complex marine environment [17]. Meanwhile, Lee et al. (2016) found that genes related to hormone biosynthesis signal transduction and cell wall catabolism were also lost or modified in *Zostera muelleri* in the process of adapting to the marine environment [16]. Moreover, Lee et al. (2018) also sequenced the genomes of *Halophila ovalis* and explored the convergent evolutionary features with the above two seagrasses, finding that all three seagrasses lost genes related to ethylene and terpenoid biosynthesis while retaining those related to salt tolerance [14]. However, compared with higher terrestrial plants, genomics analysis in seagrasses is still in its infancy, and further research is urgently needed.

Plant chloroplasts are the main functional organs of plants, having been formed through endosymbiosis between early plants and cyanobacteria [18], and have multiple functions in plant cells, including playing a crucial role in photosynthesis and carbon fixation [19, 20]. Since the process of photosynthesis is influenced by three main factors (light irradiance and wavelength, carbon dioxide concentration, and temperature), it is reasonable to assume that the genetic basis of the cp genome changes in response to the conversion of living habitats. Chloroplasts are essential organelles in plant evolution due to their features. The cp genome is small and contains well-characterized features, including a

relatively stable gene content and conserved structural features such as a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeats (IRs) regions, together with the feature of a slow rate of nucleotide substitution, providing essential information to support comparative evolutionary research [19, 21, 22]. Adaptive evolution is defined as the improved adaptability of a species to changing environmental conditions. Given the conservation of the cp genome, observable alterations should be anticipated if adaptive evolution occurs in this molecule. Thus, a comparative analysis of the cp genomes of seagrasses should enhance our understanding of plant adaptation to the sea environment.

In the present study, we aimed to provide comprehensive insights into the evolution of cp genomes across prominent seagrass families. Here, we assembled and characterized the cp genome sequences of two seagrass species and compared them with 10 published seagrass cp genomes from four families, including Cymodoceaceae, Ruppiaceae, Zosteraceae, and Hydrocharitaceae. We also identified repeat sequences and positive selection, and reconstructed the phylogenetic relationships and molecular divergence dates for the major lineages within the seagrasses. These results will help develop genetic markers for seagrasses and provide an important theoretical basis for subsequent population genetic analysis and phylogenetic relationship analysis.

Results

Genomic features of seagrasses

This study analyzed 12 cp genomes covering the four families represented by seagrasses: Zosteraceae, Hydrocharitaceae, Ruppiaceae, and Cymodoceaceae. Among them, *Z. muelleri* and *H. ovalis* were assembled and annotated for the first time in this study (Fig. 1). All cp genomes had a typical four-part structure: a LSC, an SSC, and two IRs (Table 1). The length of the 12 cp genomes ranged from 143,877 bp to 178,261 bp and varied in size among the different families as follows: Hydrocharitaceae > Cymodoceaceae > Ruppiaceae > Zosteraceae. The IR was 24,399–44,815 bp in length, the LSC was 78,949–89,851 bp in length, and the SSC was 2,150–19,160 bp in length. The length differences between families were mainly related to the expansion and contraction of the IR region. The cp genomes of the seagrass species had 116–158 genes, including 78–122 protein-coding genes, eight ribosomal RNA genes (rRNA), and 30–42 transfer RNA genes (tRNA). The total GC content of the cp genomes was 35.5%–39.2%, and the GC content was more similar for species of the same family.

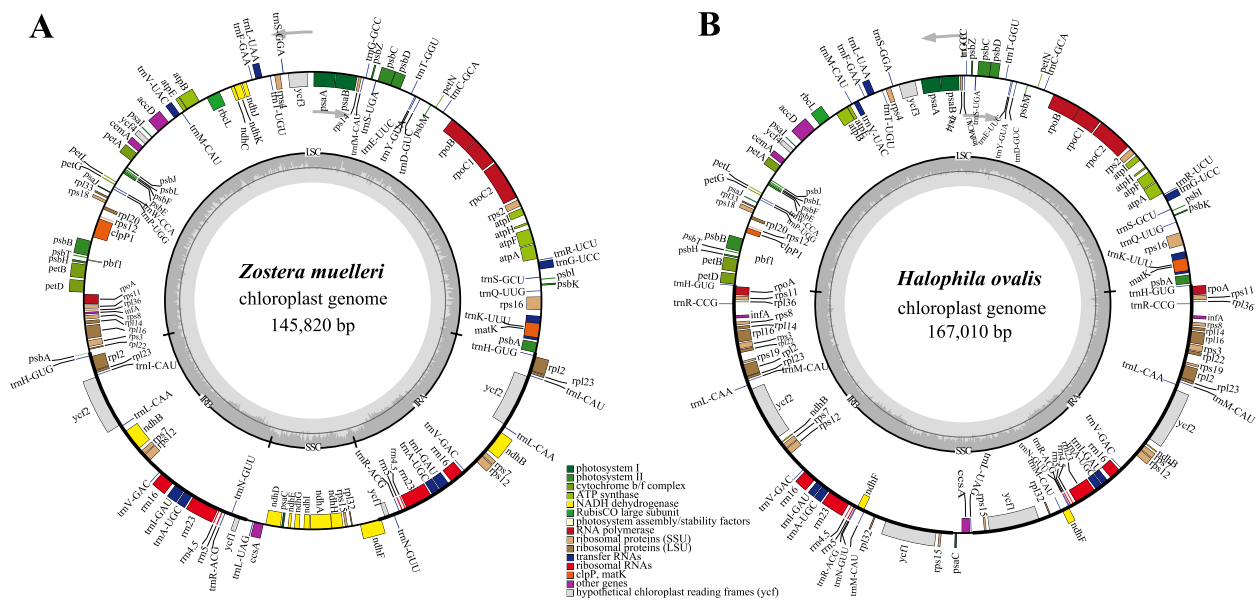


Fig. 1 Chloroplast genome gene map of *Halophila ovalis* and *Zostera muelleri*. Genes on the inside of the outer circle are transcribed clockwise, while those outside are transcribed counterclockwise

Repeat sequence analysis

In this study, a total of 2,751 simple sequence repeats (SSRs) were identified in the cp genome sequences of the 12 seagrass species (Fig. 2). Among them, the highest percentages of SSRs were mononucleotides (77.65%), followed by dinucleotides (12.40%), tetranucleotides (4.64%), trinucleotides (3.33%), pentanucleotides (1.22%), and hexanucleotides (0.76%) (Fig. 2E). Furthermore, we found that the Zosteraceae species had more SSRs, and A/T was the main mononucleotide SSR type in all species (Fig. 2B). In addition, we identified 1,757 long repeat sequence types in the cp genome sequences of the 12 seagrass species, including 1,182 tandem repeats, 279 palindromic repeats, and 296 dispersed repeats, respectively (Fig. 2C). The lengths of these long repeats were highly variable, with repeats of 1–20, 21–30, and 31–40 bp being more abundant among the three repeat types (Fig. 2D). Among them, Hydrocharitaceae species had more long repeat sequences (especially tandem repeats) and > 81 bp repeat types.

Comparative analysis

The overall sequence identity of the cp genomes of 12 seagrass species was visualized using the mVISTA program, using the *E. acoroides* annotated sequences as a reference. Figure 3 shows a genome-wide alignment with high sequence similarity (> 90% identity). The cp genomes of the same family showed higher similarity. The divergence level of the non-coding regions was higher than the

coding regions. Additionally, the LSC and SSC regions showed a higher level of sequence divergence than the IR regions. In addition, 59 coding regions were extracted to calculate the nucleotide variability (Supplementary Fig. 1), and the loci with the largest variation were *accD*, *clpP*, *infA*, *rpl22*, *rps15*, *rps18*, and *yef1*.

Phylogenetic relationships

To investigate the phylogenetic relationships of the seagrasses, we constructed a maximum likelihood (ML) tree using RAxML (Fig. 4). As indicated in the tree, the shared nucleotide-coding genes were divided into four major clades: Zosteraceae, Cymodoceaceae, Hydrocharitaceae, and Ruppiaceae, among which Cymodoceaceae and Ruppiaceae had a relatively close relationship. Moreover, *Z. muelleri* was more closely related to *Z. japonica*, followed by *Z. nigricaulis* and *Z. marina* within Zosteraceae. These results showed that phylogenetic proximity was associated with the traditional taxonomic group. The results of the MCMCtree analysis of species divergence times in seagrasses are also shown in Fig. 4. The results showed that the divergence time between *S. polyrhiza* and seagrasses was about 125.6 Mya (95% PHD = 117.7–133.9 Mya, calibration point = 128.0 Mya). The diversification of Hydrocharitaceae, Zosteraceae, Ruppiaceae, and Cymodoceaceae in Alismatales was about 65.54–101.62 Mya. Within Zosteraceae, the divergence between *Phyllospadix* and *Zostera* was approximately 38.29 Mya. The divergence time of *Z. marina* (mean age of 20.27 Mya) was much earlier than that of *Z. nigricaulis* (mean

Table 1 Chloroplast genome features of 12 seagrass species

Family	Species Name	Genome size (bp)	Reference	LSC size (bp)	IR size (bp)	SSC size (bp)	GC (%)	Total N.o of genes	PCGs N.o	tRNA N.o	rRNA N.o
Zosteraceae	<i>Zostera nigricaulis</i>	145,883	NC_058621.1	83,541	24,430	13,482	35.9	131	85	38	8
	<i>Phyllospadix iwataensis</i>	152,726	NC_058622.1	84,753	25,167	17,639	36.2	132	86	38	8
	<i>Zostera japonica</i>	146,090	NC_058623.1	83,664	24,628	13,169	35.9	127	85	34	8
	<i>Zostera marina</i>	143,877	NC_036014.1	83,224	25,915	8,823	35.5	116	78	30	8
	<i>Zostera muelleri</i>	145,820	In this study	83,524	24,399	13,498	35.9	131	85	38	8
Hydrocharitaceae	<i>Enhalus acoroides</i>	176,211	MN736636.1	89,851	42,105	2,150	38.4	142	95	39	8
	<i>Thalassia hemprichii</i>	178,261	NC_043774.1	83,691	44,815	4,940	39.2	158	122	28	8
	<i>Halophila beccarii</i>	168,585	NC_051970.1	80,881	41,487	4,730	38.5	132	88	36	8
	<i>Halophila ovalis</i>	167,010	In this study	78,949	42,898	2,265	38.8	140	90	42	8
Ruppiales	<i>Ruppia brevipedunculata</i>	158,943	MN736637.1	88,857	25,478	19,130	35.8	132	87	37	8
	<i>Ruppia sinensis</i>	158,897	MN233650.1	88,952	25,449	19,047	35.9	136	88	40	8
Cymodoceaceae	<i>Syringodium isoetifolium</i>	159,333	MZ325253.1	89,055	25,559	19,160	35.9	131	86	37	8

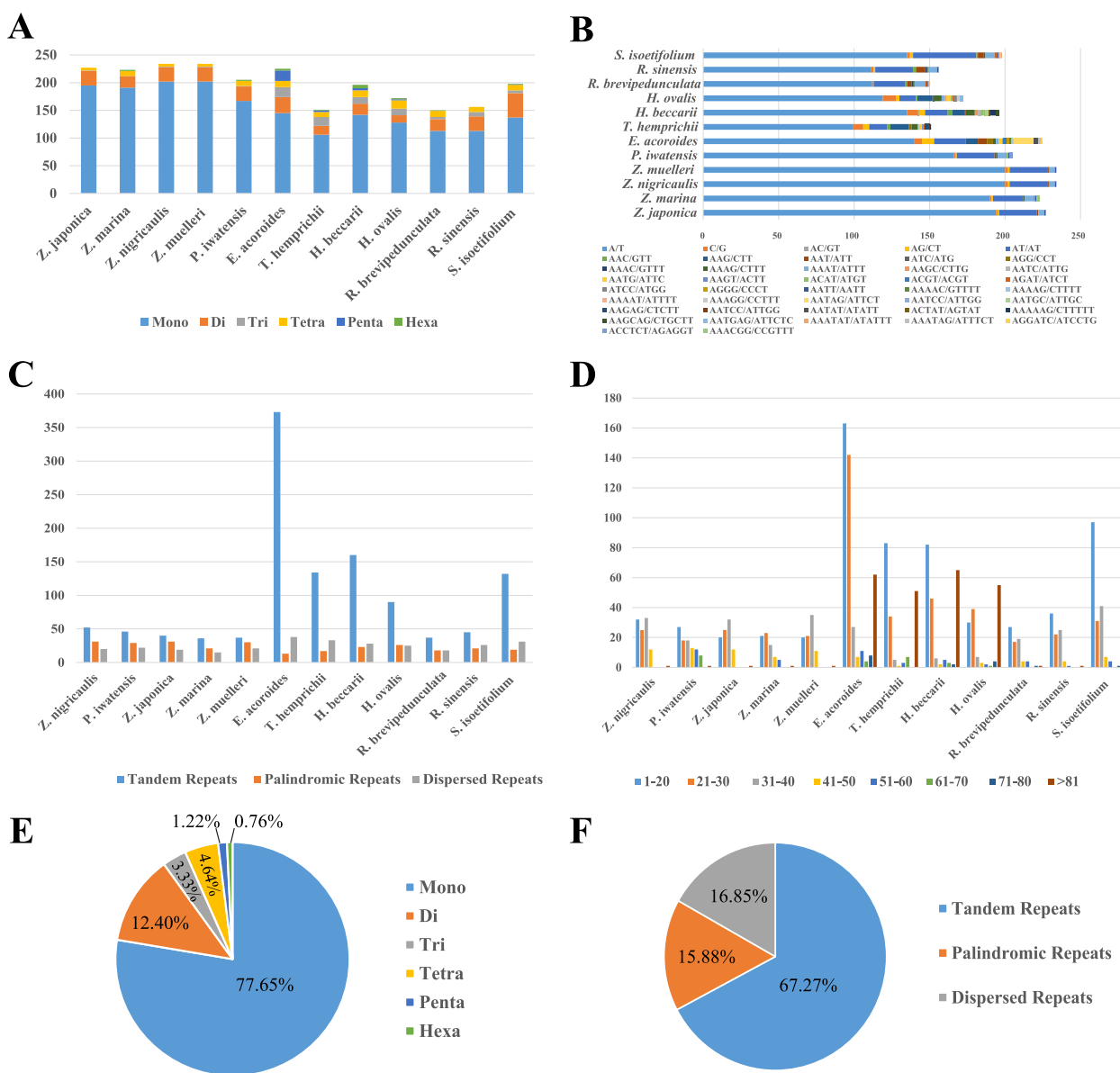


Fig. 2 Analysis of the repeated sequences of the 12 chloroplast genomes of seagrasses compared in this study. **A:** The number of SSRs in six SSR types; **B:** summary of different SSR repeat unit types; **C:** the number of repeat elements in three long repeat types; **D:** summary of long repeat types by length; **E:** distribution of SSR types; and **F:** distribution of long repeat types

age=7.97 Mya), and that of *Z. muelleri* and *Z. japonica* (mean age=3.54 Mya).

Adaptive evolution analysis

Using the M8 model and the Datamonkey web server SLAC, FEL, and MEME methods, a total of nine genes were detected with positive selection sites (Supplementary Table S1), including two ATP subunit genes (*atpA*

and *atpF*), two ribosome small subunit genes (*rps4* and *rps7*), one photosystem subunit gene (*psbH*), one ribosome large subunit gene (*rpl20*), and the *ycf2*, *accD*, and *rbcL* genes. Among them, the *ycf2* gene harbored eight sites under positive selection, with two in *accD*, while the other seven genes each had only one positively selected site.

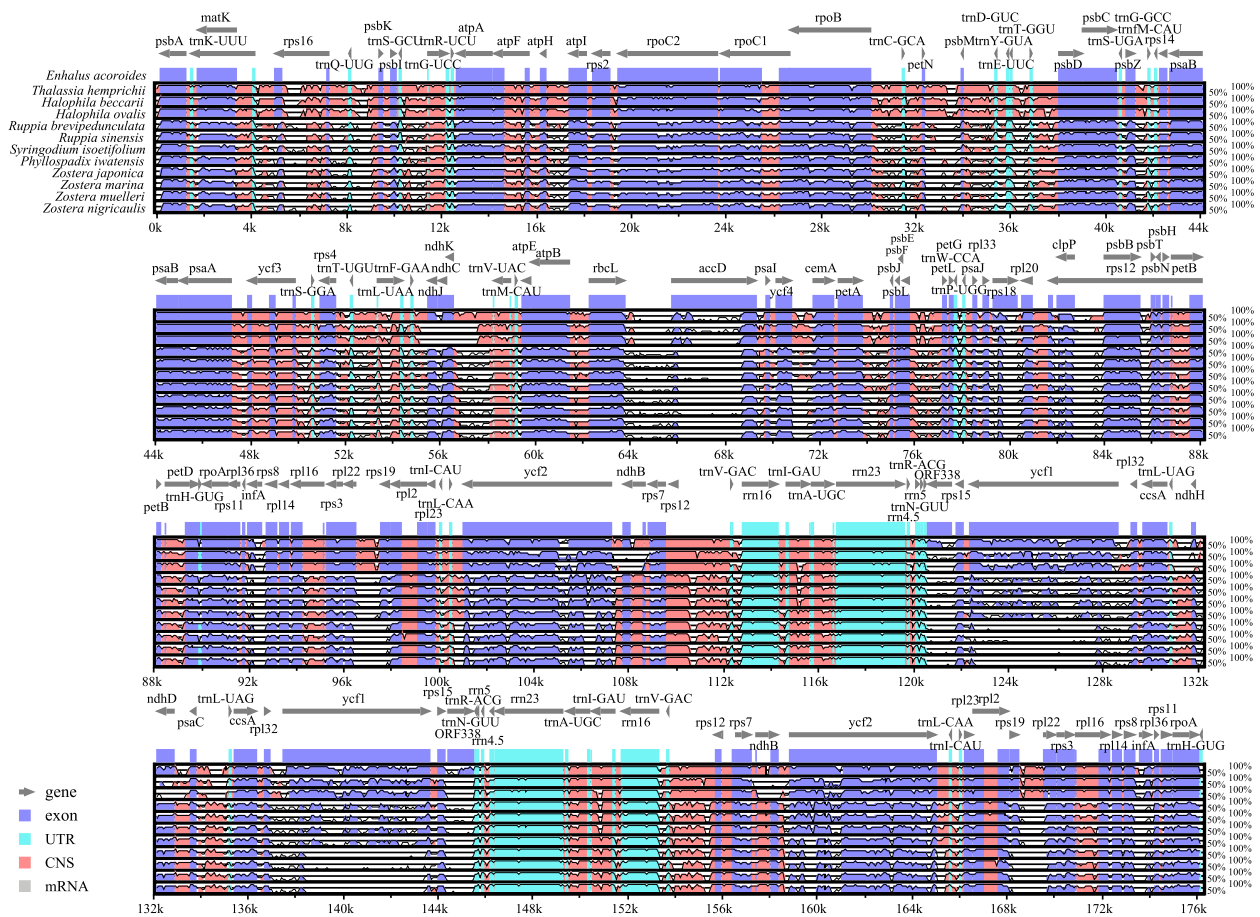


Fig. 3 Visualization of the alignment of 12 chloroplast genome sequences of seagrasses. The chloroplast genome of *Enhalus acoroides* was used as the reference. The Y-axis depicts percent identity to the reference genome (50–100%) and the X-axis depicts sequence coordinates within the chloroplast genome

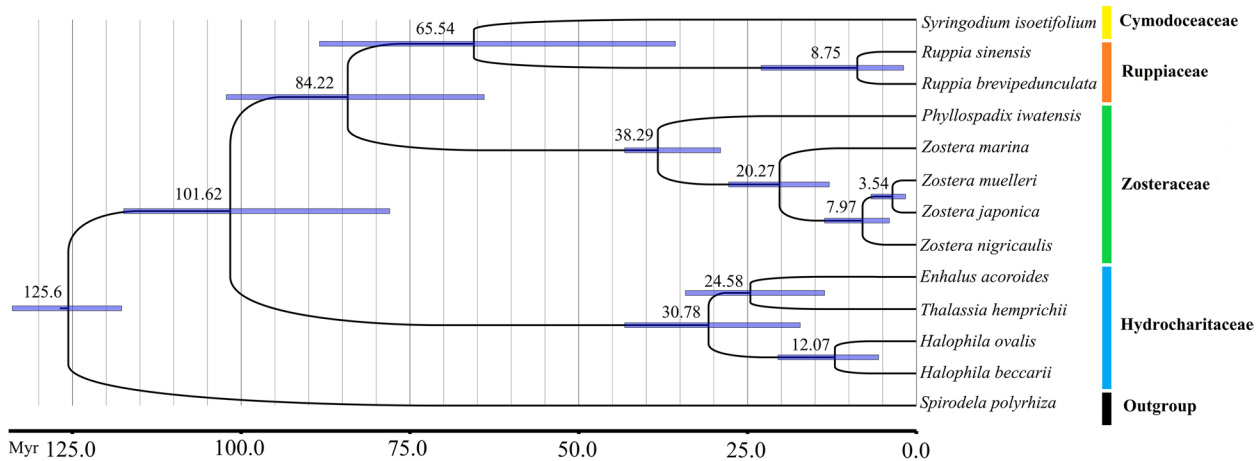


Fig. 4 Phylogenetic ML tree and time-calibrated phylogenetic tree of 59 single copy genes of 12 seagrasses

Discussion

Chloroplast sequence variation

The cp genomes of the 12 seagrasses in this study ranged in size from 143,877 to 178,261 bp, which is similar to most angiosperms [23]. However, the size of the cp genome within the seagrasses varied widely, with those of the species of Hydrocharitaceae being much larger than those of the other three families (Table 1). These differences in cp genome size may have been due to the expansion and contraction of the border positions between the IR and SSC regions [24, 25]. The GC content acts as a landmark for the physical location of functional elements in the genome [26] and is positively correlated with the rate of recombination and exon density [27]. Generally, cp genomes are characterized by a low GC content; however, in this study, Hydrocharitaceae had a higher GC content (38.4%–39.2%) than other seagrasses. Moreover, we used mVISTA to compare the whole cp of 12 species of seagrasses and used DnaSP to analyze the percentage of variable loci in 59 single-copy shared genes. Similar to previous results, the variation in noncoding regions was more significant than that of the coding regions [24, 28]. Genes with a relatively high mutation rate can be used as DNA barcodes to help distinguish between accessions within a given taxon and varieties in germplasm resources [29, 30]. We detected seven hot spots in coding regions, which can be used as candidate DNA barcodes for assessing the phylogenetic relationships and interspecific differences among seagrasses.

In addition, the 12 cp genomes of the seagrasses contained a high frequency of large repeats. Previous studies have suggested that larger and complex repeat sequences have played key roles in sequence rearrangements and cp genome evolution [31, 32]. In this study, we found that the Hydrocharitaceae genomes had the highest numbers of repeats and a significant correlation ($R=0.73$, $P=0.0065$) between the size of the cp genomes and the number of repeats (Supplementary Fig. 2). Furthermore, tandem repeats were the most abundant type of repeat in the Hydrocharitaceae cp genomes, which differs from the high content of dispersed repeats reported in other angiosperm lineages [33, 34]. From an evolutionary perspective, variations in repeated sequences among species are due to natural selection and adaptation by organisms to diverse environments [35]. Previous studies have shown that cp SSRs are dominated by A/T repeats, which contribute significantly to the AT richness of cp genomes [36]. In this study, SSRs in the seagrass cp genome also contained many AT units, and mononucleotide repeats accounted for 77.65% of the total SSRs. In addition, we found that dinucleotide repeats (AT/AT) were the most common of the different types in all the seagrasses, while pentanucleotide and hexanucleotide SSRs were rarely

found. The copy number variation of SSRs in the cp genome was highly polymorphic, and these seagrass variants can be used as molecular genetic markers for future studies in population genetics, phylogeography, phylogenetics, and species identification [37–39].

Adaptive selection

In our study, we identified nine genes with positive selection sites. Among them, ATP synthase is essential for plant photosynthesis and is usually a product of two genetic systems in plants [40]. In this study, we found that two ATP subunit genes (*atpA* and *atpF*) exhibited site-specific selection. Additionally, three genes (*rps4*, *rps7*, and *rpl20*) encoding ribosomal subunits were identified under positive selection. One photosystem II subunit gene (*psbH*) was also under positive selection. ACCase is a key enzyme in fatty acid biosynthesis [41]. The β -CT subunit of ACCase is encoded by *accD*, which is the only component of plant lipid metabolism known to be encoded by the plastid genome [42, 43]. We identified positively selected sites (PSSs) in *accD*, which may have played key roles in seagrass fatty acid biosynthesis. Additionally, *rbcL* provides all the catalytically essential residues of RuBisCO, a critical enzyme for both the reductive and oxidative photosynthetic carbon cycles. A previous study showed that many aquatic plants had acquired carbon-concentrating mechanisms to overcome the potentially low, fluctuating supply of CO₂ for underwater photosynthesis [44]. In particular, the *rbcL* gene evolved under positive selection in *Potamogeton* [45]. Although seagrasses can also use both CO₂ and HCO₃⁻ (bicarbonate) for photosynthetic carbon reduction [46], they live mainly in seawater with low CO₂ concentration, and so positive selection of this gene may be involved in increasing their CO₂ utilization efficiency. We also found that *ycf2* had eight sites under positive selection. This gene is the largest cp gene reported in angiosperms and is valuable for assessing sequence variations and evolutionary processes in plants [24, 47, 48]. Positive selection on the *ycf2* gene has also been found in the adaptive evolution of other species [28, 49, 50]. In short, these positively selected genes may play a key role in the adaptation of seagrasses to the marine environment.

Phylogenetic relationships

Seagrasses are a paraphyletic group of marine angiosperms that evolved in parallel three to four times from land plants back to the sea [7, 17]. Therefore, seagrass is a generic term for a variety of aquatic angiosperms and therefore represents an ecological group rather than a taxonomic group. In this study, the phylogenetic trees of the seagrasses based on the whole-genome nucleotide coding sequences of the cp genomes were clearly divided

into four major clades, representing Zosteraceae, Cymodoceaceae, Hydrocharitaceae, and Ruppiaceae, which is similar to previous analyses based on various molecular datasets [7, 24, 51–53]. Divergence time estimation dated the divergence of seagrasses and *S. polyrhiza* at around 125.6 Mya. This age is consistent with previous genome-based studies [17, 54]. The stem node ages of the seagrass lineages Zosteraceae and Cymodoceaceae/Ruppiaceae were at 84.22 Mya, which is slightly older than that in recent studies [52, 55]. The divergence time of *E. acoroides*/*T. hemprichii* and *Halophila* within Hydrocharitaceae was around 30.78 Mya. This age is slightly older than that reported by Li et al. (2012) at around 19.41 Mya [56], but is younger than that reported by Li et al. (2022) at around 34.64 Ma [52]. Among seagrasses, all species of Zosteraceae are seagrasses, whereas other families contain more than just seagrass species. Among Zosteraceae, the genus *Phyllospadix* has a clear taxonomic status owing to its morphological features and chromosome number that differs significantly from other species [57–59]. In this study, we also found that *P. iwatensis* was the first to diverge from the species of the Zosteraceae family in 38.29 Ma. This age is consistent with previous phylogenetic analyses and molecular clock estimates based on the cp *rbcL* and *matK* loci, which suggested that the family Zosteraceae emerged about 100 Ma, and the divergence of *Zostera* and *Phyllospadix* began around 36 Ma [60]. Fossil evidence indicates that seagrasses originated in the Late Cretaceous [61]. As higher angiosperms, seagrasses have an existing root system and can occupy a previously empty niche in shallow sedimentary shoreline marine systems. This may have been facilitated by the Cretaceous–Paleogene extinction event, which took place roughly 70–65 Mya and coincided with a lineage-specific whole-genome duplication in *Zostera* [17].

Conclusions

In this study, we sequenced the cp genomes of two seagrass species (*Z. muelleri* and *H. ovalis*) and revealed the cp genomic features with the available seagrass genomes obtained from the NCBI database. We screened out 2,751 SSRs and 1,757 long repeat sequence types in the cp genome sequences of the 12 seagrass species. We also identified nine positive selection genes and seven variable regions, which provide a reference for developing DNA markers and evaluating adaptive evolution in further studies of seagrass species. Phylogenetic and divergence time analysis based on the current data was generally consistent with previous studies. These findings will be valuable for further study of the cp genomes of seagrass species, and will provide valuable resources for studies of plant adaptation to marine environments.

Methods

Sampling, DNA extraction, and sequencing

In this study, two seagrasses, *Z. muelleri* and *H. ovalis*, were collected from their natural habitats, and the collection processes of both conformed to local and national regulations. The voucher specimens of *Z. muelleri* (voucher number: OUC-S120) and *H. ovalis* (voucher number: OUC-S121) were deposited in the Herbarium of Marine Ecology Laboratory, College of Marine Life Sciences, Ocean University of China (OUC). The samples were identified by Tang Xuexi, a Professor at OUC. Fresh leaves (100 mg) were preserved in silica gel immediately after sampling, after which TRIzol[®] reagent (Invitrogen, USA) was used in the laboratory to extract their total DNA following the manufacturer's protocol. The integrity, quality, and concentration of the DNA were determined by 1% agarose gel electrophoresis and a NanoDrop Spectrophotometer 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Illumina TruSeq[™] Nano DNA Sample Prep Kits were used for Illumina sequencing library construction, and the DNA was sequenced using an Illumina HiSeq 4000 platform (150 bp*2). After Illumina sequencing, approximately 63,351.8 Mb and 97,155.8 Mb of raw data for *Z. muelleri* and *H. ovalis* were generated, and these raw reads were QC-filtered and trimmed using the Trimmomatic 0.39 software [62]. In total, 61,803.7 Mb and 91,116.5 Mb clean data were obtained. Afterwards, *Z. marina* was used as a reference sequence to assemble its genome using the NOVOPlasty v2.7.2 software [63], and the GapCloser software [64] was used to fill in the remaining local internal gaps and correct for single-base polymorphisms. Finally, the starting position and orientation of the chloroplast assembly sequence were determined using the reference genome, and the possible partition structure of the chloroplast (LSC/IR/SSC) was determined to obtain the final chloroplast genome sequence.

Genome annotation and comparative analysis

The assembled cp genomes were annotated using GeSeq software [65], with the following parameters: (1) protein search identity, 60; (2) rRNA, tRNA, DNA search identity, 35; and (3) third-party tRNA annotators, the tRNAscan-SE software [66]. Then, the annotation results were checked with BLAST and DOGMA [67]. The circular maps of the cp genomes of the two seagrass samples were presented using Organellar Genome DRAW software [68]. Subsequently, two complete cp genomes were deposited in GenBank with the following accession numbers: *Z. muelleri* (OP611572) and *H. ovalis* (OP611573). Alignments of the 12 complete cp seagrass genome sequences (Table 1) were visualized using mVISTA [69].

Finally, all shared coding regions of the extracted 12 seagrass cp genomes were aligned separately using MUSCLE [70], and the nucleotide variability of each selected region was evaluated separately using DNASP v5.10 [71].

Repeat sequence analysis

Repeat sequences within the seagrass cp genome, including forward (F), reverse (R), palindrome (P), and complement (C), were searched using the online software REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [72], and identified with the following conditions: minimal size of 30 bp; 90% or greater sequence identity; and Hamming distance equal to 3. Tandem Repeats Finder v4.09 software (<https://tandem.bu.edu/trf/trf.html>) [73] was used to detect tandem repeats with > 6 bp repeat units. SSR markers were identified in the 12 seagrasses sequences using MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) [74] with a motif size of 1–6 nucleotides and thresholds of eight, five, four, three, three, and three repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively.

Phylogenetic relationships and divergence time estimation

Phylogenetic analysis of the 12 cp genomes of the seagrass species was performed with *Spirodela polyrhiza* as the outgroup. All shared nucleotide-coding sequences were concatenated into a super matrix and aligned with MUSCLE [70]. A phylogenetic tree was produced using the ML method based on the GTRGAMMA model with 1,000 bootstrap replicates using RAxML [75]. MCMC-tree implemented in the PAML 4.7 package was used to estimate the speciation time [76]. The obtained ML tree was used as the input tree file for the analysis process. Three calibration points (*E. acoroides* vs. *T. hemprichii*: 8.59–34.64 MYA, *Zostera* vs. *P. iwatensis*: 15.29–42.00 MYA, and Zosteraceae vs. P. Hydrocharitaceae: 72.70–117.00 MYA) derived from the TimeTree database (<http://www.timetree.org/>) were applied to constrain the divergence times of the nodes.

Adaptive evolution analysis

The CODEML program in the PAML 4.7 package [76] was used to calculate the rate of nonsynonymous substitutions (dN) and synonymous substitutions (dS) of single-copy protein coding genes (PCGs). The dN/dS ratio, also known as the ω value, was used to measure the rate of gene evolution, where ω values greater than 1, equal to 1, and less than 1 represent positive selection, neutral selection and purifying selection, respectively. The site model allows for different ω values for different sites in the same sequence. The M7 (null hypothesis: $0 < \omega < 1$) and M8 (alternative hypothesis: $\omega > 1$) models were used to detect PSSs on 12 seagrass species. In the likelihood

ratio test (LRT) results of the two models, the alternative hypothesis M8 model was accepted if $p < 0.05$, and the null hypothesis M7 model was accepted if this was not true. With $p < 0.05$, the PSSs detected using the M8 model were considered as potential PSSs when the Bayes empirical Bayes (BEB) posterior probability was > 0.90 [77]. Meanwhile, PSSs were identified using the Data Monkey Web Server (<http://www.datamonkey.org/>) [78] based on three methods, namely the FEL, SLAC, and MEME models, with $P < 0.1$. In this study, positive selection sites detected using at least three or more of the methods above were considered to be positive selection sites.

Abbreviations

cp	Chloroplast
LSC	Large single copy
SSC	Small single copy
IRs	Inverted repeats
SSR	Simple sequence repeat
ML	Maximum likelihood
dN	Nonsynonymous substitutions
dS	Synonymous substitutions
LRT	Likelihood ratio test
BEB	Bayes empirical Bayes
rRNA	Ribosomal RNA genes
tRNA	Transfer RNA genes
PCGs	Protein coding genes
PSSs	Positively selected sites

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04119-9>.

Additional file 1: Supplementary Figure 1. Percentages of variable characters in protein-coding regions among the 12 chloroplast genomes of the seagrasses.

Additional file 2: Supplementary Figure 2. Relationships between seagrass chloroplast genome sizes and the number of repeats.

Additional file 3: Table S1. Positively selected sites of 59 single-copy genes shared by twelve seagrass species.

Acknowledgements

We are grateful to the editors and reviewers for their valuable suggestions, which helped us improve the manuscript.

Authors' contributions

JC, YZ and XXT collected and identified plant material and conducted the project. JC conducted the experiments, analyzed the data and drafted the manuscript. ZBY, LS and SX supervised the study. SS, YW and XXT revised the manuscript. All authors have read and approved the final draft of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (42206117, 42176154, 42006144), the NSFC-Shandong Joint Fund (U1806213), and the National Key R&D Program of China (2019YFD0901204).

Availability of data and materials

All newly sequenced chloroplast genomic sequences of the two seagrasses in this study can be downloaded from the National Center for Biotechnology Information (NCBI) under accession numbers OP611572 (<https://www.ncbi>.

nml.nih.gov/nuccore/OP611572) and OP611573 (<https://www.ncbi.nlm.nih.gov/nuccore/OP611573>). Information on the chloroplast genomic sequences of the other 10 seagrasses used for the combined analysis can also be downloaded from GenBank and found in Table 1.

Declarations

Ethics approval and consent to participate

We complied with all relevant institutional, national and international guidelines for the collection of *Z. muelleri* and *H. ovalis* with permissions from Ocean University of China.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interest.

Received: 6 November 2022 Accepted: 14 February 2023

Published online: 22 February 2023

References

- Suchanek TH, Williams SL, Ogden JC, Hubbard DK, Gill IP. Utilization of shallow-water seagrass detritus by Caribbean deep-sea macrofauna: $\delta^{13}\text{C}$ evidence. *Deep Sea Res Part A Oceanographic Res Papers*. 1985;32(2):201–14.
- Hemminga MA, Duarte CM. *Seagrass ecology*. Cambridge: Cambridge University Press; 2000.
- Duarte CM, Middelburg JJ, Caraco N. Major role of marine vegetation on the oceanic carbon cycle. *Biogeosciences*. 2005;2(1):1–8.
- Orth RJ, Carruthers TJ, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, Hughes AR, Kendrick GA, Kenworthy WJ, Olyarnik S. A global crisis for seagrass ecosystems. *Bioscience*. 2006;56(12):987–96.
- Short FT, Polidoro B, Livingstone SR, Carpenter KE, Bandeira S, Bujang JS, Calumpong HP, Carruthers TJ, Coles RG, Dennison WC. Extinction risk assessment of the world's seagrass species. *Biol Cons*. 2011;144(7):1961–71.
- Huang X-P, Jiang Z-J, Fan H-Q, Chen X-Y, Zhou Y, Zhang J-P, Li W-T, Zhang P-D, Zheng F-Y, Han Q-Y. The nomenclature of the "algae" name of seagrasses in China. *Oceanologia et Limnologia Sinica*. 2016;47(1):290–4.
- Les DH, Cleland MA, Waycott M. Phylogenetic studies in Alismatidae. II: evolution of marine angiosperms (seagrasses) and hydrophyly. *Systematic Botany*. 1997;22:443–463.
- Janssen T, Bremer K. The age of major monocot groups inferred from 800+ rbcL sequences. *Bot J Linn Soc*. 2004;146(4):385–98.
- Larkum AW, Orth RJ, Duarte CM. Seagrasses: biology, ecology and conservation. *Phycologia*. 2006;45(5):5.
- Beer S, Bjork M, Hellblom F, Axelsson L. Inorganic carbon utilization in marine angiosperms (seagrasses). *Funct Plant Biol*. 2002;29(3):349–54.
- Borum J, Pedersen O, Kotula L, Fraser MW, Statton J, Colmer TD, Kendrick GA. Photosynthetic response to globally increasing CO₂ of co-occurring temperate seagrass species. *Plant, Cell Environ*. 2016;39(6):1240–50.
- Ye C, Zhao K. Osmotically active compounds and their localization in the marine halophyte eelgrass. *Biol Plant*. 2003;46(1):137–40.
- Waycott M, Procaccini G, Les DH, Reusch T. *Seagrass Evolution, Ecology and Conservation: A Genetic Perspective*. Netherlands: Springer; 2007.
- Lee H, Golicz AA, Bayer PE, Severn-Ellis AA, Chan CK-K, Batley J, Kendrick GA, Edwards D. Genomic comparison of two independent seagrass lineages reveals habitat-driven convergent evolution. *J Exp Bot*. 2018;69(15):3689–702.
- Lambers H, Chapin FS, Pons TL. *Plant physiological ecology*. New York: Springer; 2008.
- Lee H, Golicz AA, Bayer PE, Jiao Y, Tang H, Paterson AH, Sablok G, Krishnaraj RR, Chan CK-K, Batley J. The genome of a southern hemisphere seagrass species (*Zostera muelleri*). *Plant physiology*. 2016;172(1):272–83.
- Olsen JL, Rouzé P, Verhelst B, Lin Y-C, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*. 2016;530(7590):331–5.
- Howe CJ, Barbrook AC, Koumandou VL, Nisbet RER, Symington HA, Wightman TF. Evolution of the chloroplast genome. *Philos Trans R Soc Lond B Biol Sci*. 2003;358(1429):99–107.
- Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011;76(3):273–97.
- Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016;17(1):1–29.
- Sugiura M. The chloroplast genome. *Plant Mol Biol*. 1992;19(1):149–68.
- Korpelainen H. The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften*. 2004;91(11):505–18.
- Zhang T, Fang Y, Wang X, Deng X, Zhang X, Hu S, Yu J. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. *PLoS ONE*. 2012;7(1):e30531.
- Chen J, Zang Y, Shang S, Liang S, Zhu M, Wang Y, Tang X. Comparative chloroplast genomes of *Zosteraceae* species provide adaptive evolution insights into seagrass. *Front Plant Sci*. 1992;2021:12.
- Wang W, Messing J. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS ONE*. 2011;6(9):e24670.
- Zhang L, Kasif S, Cantor CR, Broude NE. GC/AT-content spikes as genomic punctuation marks. *Proc Natl Acad Sci*. 2004;101(48):16855–60.
- Freudenberg J, Wang M, Yang Y, Li W. Partial correlation analysis indicates causal relationships between GC-content, exon density and recombination rate in the human genome. *BMC Bioinformatics*. 2009;10(1):1–10.
- Fan W-B, Wu Y, Yang J, Shahzad K, Li Z-H. Comparative chloroplast genomics of dipsacales species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front Plant Sci*. 2018;9:689.
- Mehmood F, Shahzadi I, Waseem S, Mirza B, Ahmed I, Waheed MT. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics*. 2020;112(1):581–91.
- Jiang L, Li M, Zhao F, Chu S, Zha L, Xu T, Peng H, Zhang W. Molecular identification and taxonomic implication of herbal species in genus *Corydalis* (Papaveraceae). *Molecules*. 2018;23(6):1393.
- Milligan BG, Hampton JN, Palmer JD. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol Biol Evol*. 1989;6(4):355–68.
- Cavaliere-Smith T. Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr Biol*. 2002;12(2):R62–4.
- Wu Y, Liu F, Yang D-G, Li W, Zhou X-J, Pei X-Y, Liu Y-G, He K-L, Zhang W-S, Ren Z-Y. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Front Plant Sci*. 2018;9:376.
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*. 2008;66(4):350–61.
- Britten RJ, Kohne DE. Repeated Sequences in DNA: Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science (New York, NY)*. 1968;161(3841):529–40.
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE*. 2012;7(5):e36869.
- Aradhya MK, Potter D, Gao F, Simon CJ. Molecular phylogeny of Juglans (*Juglandaceae*): a biogeographic perspective. *Tree Genet Genomes*. 2007;3(4):363–78.
- Wang S, Shi C, Gao L-Z. Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. *PLoS ONE*. 2013;8(9):e73946.
- Xue J, Wang S, Zhou SL. Polymorphic chloroplast microsatellite loci in *Nelumbo* (*Nelumbonaceae*). *Am J Bot*. 2012;99(6):e240–4.

40. Westhoff P, Alt J, Nelson N, Herrmann RG. Genes and transcripts for the ATP synthase CF₀ subunits I and II from spinach thylakoid membranes. *Mol Gen Genet MGG*. 1985;199(2):290–9.
41. Rawsthorne S. Carbon flux and fatty acid synthesis in plants. *Prog Lipid Res*. 2002;41(2):182–96.
42. Ke J, Wen T-N, Nikolau BJ, Wurtele ES. Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme A carboxylase. *Plant Physiol*. 2000;122(4):1057–72.
43. Lee SS, Jeong WJ, Bae JM, Bang JW, Liu JR, Harn CH. Characterization of the plastid-encoded carboxyltransferase subunit (accD) gene of potato. *Mol Cells*. 2004;17(3):422–9.
44. Maberly SC, Madsen TV. Freshwater angiosperm carbon concentrating mechanisms: processes and patterns. *Funct Plant Biol*. 2002;29(3):393–405.
45. Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K. Molecular adaptation of rbcL in the heterophyllous aquatic plant *Potamogeton*. *PLoS ONE*. 2009;4(2):e4633.
46. Beer S. Photosynthesis and photorespiration of marine angiosperms. *Aquat Bot*. 1989;34(1–3):153–66.
47. Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J*. 2000;22(2):97–104.
48. Huang JL, Sun GL, Zhang DM. Molecular evolution and phylogeny of the angiosperm *ycf2* gene. *J Syst Evol*. 2010;48(4):240–8.
49. Zhong Q, Yang S, Sun X, Wang L, Li Y. The complete chloroplast genome of the Jerusalem artichoke (*Helianthus tuberosus* L.) and an adaptive evolutionary analysis of the *ycf2* gene. *PeerJ*. 2019;7:e7596.
50. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, Liu H. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics*. 2020;21(1):1–14.
51. Ross TG, Barrett CF, Soto Gomez M, Lam VK, Henriquez CL, Les DH, Davis JI, Cuenca A, Petersen G, Seberg O. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics*. 2016;32(2):160–78.
52. Chen L-Y, Lu B, Morales-Briones DF, Moody ML, Liu F, Hu G-W, Huang C-H, Chen J-M, Wang Q-F. Phylogenomic analyses of Alismatales shed light into adaptations to aquatic environments. *Mol Biol Evol*. 2022;39(5):msac079.
53. Cuenca A, Petersen G, Seberg O, Davis JI, Stevenson DW. Are substitution rates and RNA editing correlated? *BMC Evol Biol*. 2010;10(1):1–15.
54. An D, Zhou Y, Li C, Xiao Q, Wang T, Zhang Y, Wu Y, Li Y, Chao D-Y, Messing J. Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of *Spirodela*. *Proc Natl Acad Sci*. 2019;116(38):18893–9.
55. Coyer J, Hoarau G, Kuo J, Tronholm A, Veldsink J, Olsen J. Phylogeny and temporal divergence of the seagrass family Zosteraceae using one nuclear and three chloroplast loci. *Syst Biodivers*. 2013;11(3):271–84.
56. Chen L-Y, Chen J-M, Gituru RW, Wang Q-F. Generic phylogeny, historical biogeography and character evolution of the cosmopolitan aquatic plant family Hydrocharitaceae. *BMC Evol Biol*. 2012;12(1):1–12.
57. Kuo J, Den Hartog C. Seagrass taxonomy and identification key. *Glob Seagrass Res Methods*. 2001;33:31–58.
58. Kuo J. Chromosome numbers of the Australian Zosteraceae. *Plant Syst Evol*. 2001;226(3):155–63.
59. Tanaka N, Kuo J, Omori Y, Nakaoka M, Aioi K. Phylogenetic relationships in the genera *Zostera* and *Heterozostera* (Zosteraceae) based on *matK* sequence data. *J Plant Res*. 2003;116(4):273–9.
60. Kato Y, Aioi K, Omori Y, Takahata N, Satta Y. 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*. 2003;78(5):329–42.
61. Larkum AW, Waycott M, Conran JG. Evolution and biogeography of seagrasses. In: *Seagrasses of Australia*. Cham: Springer; 2018. pp. 3–29.
62. Bolger AM, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. 2014;15:2114–20.
63. Dierckxens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45(4):e18–e18.
64. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012;1(1):2047–2217X–2041–2018.
65. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11.
66. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955–64.
67. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organelle genomes with DOGMA. *Bioinformatics (Oxford, England)*. 2004;20(17):3252–5.
68. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. *Nucleic Acids Res*. 2019;47(W1):W59–64.
69. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(suppl_2):W273–9.
70. Edgar RC. MUSCLE: Multiple Sequence Alignment with Improved Accuracy and Speed. In: *Computational Systems Bioinformatics Conference*. 2004.
71. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*. 2009;25(11):1451–2.
72. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42.
73. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
74. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics (Oxford, England)*. 2017;33(16):2583–5.
75. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*. 2014;30(9):1312–3.
76. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
77. Ziheng Y, Wong WSW, Rasmus N. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 2005;22(4):1107–18.
78. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*. 2005;21(10):2531–3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

