

Chloroplast genome assembly of *Thalassia testudinum* from the Caribbean Sea, and its evolutionary analysis with related genera

Abner J. Colón Colón

Inter American University of Puerto Rico Barranquitas Campus

Gerardo Laureano Rosario

Inter American University of Puerto Rico Barranquitas Campus

Dariel A. Vázquez Fuentes

Inter American University of Puerto Rico Barranquitas Campus

Rosalinda Aybar Batista

Inter American University of Puerto Rico Barranquitas Campus

Ramón E. Rivera Vicéns

Inter American University of Puerto Rico Barranquitas Campus

Ángel R. Núñez Marrero

anunez@br.inter.edu

Inter American University of Puerto Rico Barranquitas Campus

Article

Keywords: Seagrass, Chloroplast genome, *Thalassia testudinum*, Evolutionary changes, Phylogenetic analysis

Posted Date: August 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9784196/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 4th, 2026.
See the published version at <https://doi.org/10.1038/s41598-026-73952-3>.

Abstract

Seagrass meadows, composed of submerged marine angiosperms, provide a critical framework for studying biodiversity and ecological adaptations. The widely distributed genus *Thalassia* holds significant promise for sustainable environmental development. Moreover, a notable gap remains in the analysis of the *Thalassia testudinum* from the Caribbean, which is essential for accurately evaluating evolutionary distances between *Thalassia* species. To address this, a recent study utilized Minion long-read sequencing to analyze the chloroplast (cp) genome of *T. testudinum* collected from Guánica, Puerto Rico. In addition, a comprehensive genomic cp comparison to uncover evolutionary relationships with other *T. testudinum* and *T. hemprichii* populations. The sequenced Guánica cp genome was 179,257 base pairs in size, containing 141 protein-coding genes, and comparisons revealed a high degree of genetic conservation across *Thalassia* species. Furthermore, Simple Sequence Repeats (SSRs) analysis highlighted strong similarities among geographically related populations, with genetic variability primarily isolated to mononucleotide compositions and specific dinucleotide and tetranucleotide repeat motifs. The phylogenetic placement within the Hydrocharitaceae family confirmed the relation of *T. testudinum* of Guánica to Florida. Ultimately, these findings establish a foundational understanding of the evolutionary relationships of *T. testudinum* cp from Puerto Rico to other *Thalassia* species.

Introduction

Seagrass meadow ecosystems are marine environments dominated by angiosperms (flowering plants) that form dense communities in shallow marine waters worldwide^{1,2}. They contribute to the retention and stabilization of sediments by dissipating wave energy, provide food and habitat for various marine species, and improve water quality and carbon sequestration^{3,4}. Far from being a monophyletic group, seagrasses constitute a paraphyletic assemblage of angiosperms that, through certain evolutionary events, colonized the ocean from terrestrial ancestors⁵⁻⁷. Common seagrass species found in the Caribbean include *Thalassia testudinum*, *Syringodium filiforme*, *Halodule wrightii*, *Halophila decipiens*, and *Halophila stipulacea*^{8,9}. However, *Thalassia testudinum* is the most abundant seagrass in the region^{3,4,10}. This species has been classified mainly using its morphological features, such as long, linear leaves with round tips (up to 2cm), which emerge from a strong, articulated rhizome buried in the soft substrate from approximately 25 meters deep onwards in calm waters⁹.

Since seagrasses form the basis for diverging time estimates and detailed biogeographic analyses, the Hydrocharitaceae family, composed of 14 genera and approximately 136 species, has been the subject of recent studies employing modern approaches to analyze the diversity of its species¹¹. Additionally, studies have investigated the phylogenetic relationship and genetic diversity of *Thalassia* species (i.e., *T. hemprichii* and *T. testudinum*) in various regions of the world^{4,5}. In one study, the analysis of the chloroplast genome of *T. hemprichii* allowed a comparison with 12 other species of seagrasses in China to understand their evolutionary adaptations⁵. One of the findings indicated that the complete chloroplast genome of *T. hemprichii* is 178,261 base pairs (bp) in size. Furthermore, the study revealed

nine chloroplast genes with positive selection sites, including two ATP subunit genes (*atpA* and *atpF*), three ribosome subunit genes (*rps4*, *rps7*, and *rpl20*), and four photosystem genes *ycf2*, *accD*, *psbH*, and *rbcL*.

In the case of *T. testudinum*, the study⁴ focused on its chloroplast genome from a sample collected in Florida, United States (USA). The assembled chloroplast genome of 179,858 bp in size was used to analyze evolutionary adaptations to marine environments. ATP synthesis genes, such as *AtpA* and *AtpF*, were identified, as well as ribosome subunit genes (*rps4*, *rps7*, and *rpl20*), which are essential for protein synthesis and seagrass development. In addition, a subunit gene of the *psBH* photosystem was identified, along with other key genes (*ycf2*, *accD*, and *rbcL*) that are involved in metabolic pathways and adaptations that reinforce the resilience of seagrasses and their habitat⁴.

All the aforementioned studies were done with *Thalassia* spp. from places far from the Caribbean. Furthermore, the identification of *Thalassia* spp. by morphological and single gene barcoding can result in wrong species assignment due to its potential phenotypic plasticity, introgression, incomplete lineage sorting, or horizontal gene transfer^{4,5,8,12}. Thus, this study evaluates the chloroplast genome of *Thalassia* sp. from the Caribbean (south coast of Puerto Rico) and compares genetic divergence with that of *T. testudinum* from Florida, USA. In addition, genomic profiles and evolutionary patterns between *T. testudinum* and *T. hemprichii* are compared, as are the associations between genetic variability. This study helps to elucidate whether the populations present in the Caribbean correspond to the *T. testudinum* species and whether they exhibit genetic differences from other *Thalassia* spp.

Results

The *Thalassia* sp. sample was collected in southwestern Puerto Rico, and based on evaluation of morphological features, such as leaf shape and cell arrangement, the collected marine plant was presumptively identified as *Thalassia testudinum* (Fig. 1A). In the context of non-surface characteristics, confocal microscopy enabled an in-depth appreciation of the inner cellular structures, showing good tissue preservation with distinguishable chloroplasts (Fig. 1B&C). Therefore, surface characterization using the SEM provided a robust morphological analysis of the collected *Thalassia* sp., highlighting detailed surface arrangement and details, including diatoms (Fig. 1D&E).

Sequencing of the Guánica chloroplast yielded 980,725 reads, a mean length of 3,911.1 base pairs (bp), and an N50 read length of 6,904 bp, for a total output of 3.86 Gb. The assembled chloroplast genome has the typical four-part structure, comprising one large single copy (LSC), one small single copy (SSC), and two inverted repeats (IRs) (Fig. 2). The cp genome length was 179,257 bp, larger than that of the other two *Thalassia* genomes (Table 1). The IR length ranged from 128,422 to 179,257 bp, and the SSC length ranged from 128,423 to 133,322 bp. In addition, the total GC content of the cp genome was 39%, and MinION sequencing provided an average coverage depth of approximately 200x (Fig. S1).

This study analyzed the structural characteristics of the chloroplast genome from four *Thalassia* sp. from different geographical regions. Therefore, Table 1 presents a correlation between samples of *T. testudinum* from Guánica (Puerto Rico) and Florida (USA) and two *T. hemprichii* from China, then *T. hemprichii* Ch1 (NC_043774) and *T. hemprichii* Ch2 (MN478365). The total genome size ranged from 178,261 bp in *T. hemprichii* Ch1 and 178,438 bp in *T. testudinum* Ch2 to 179,858 bp in *T. testudinum*, with *T. testudinum* from Guánica being slightly smaller (179,257 bp). The number of annotated protein coding genes (PCGs) varied substantially: *T. testudinum* from Puerto Rico showed a similar gene annotation (141), in contrast to *T. testudinum* of Florida (142 genes) and *T. hemprichii* Ch1 (140 genes), and *T. hemprichii* Ch2 (144 genes), respectively. Discrepancies were observed in the size of the inverted repeat (IR) regions, whose lengths ranged from 24,882 bp in the Florida sample to 50,835 bp in the sample of Guánica. The small single-copy regions (SSC) showed considerable variation, which was markedly reduced in *T. hemprichii* Ch1 (4,940 bp) compared to the *T. hemprichii* Ch2 (4,939 bp) and *T. testudinum* from Guánica (4,899 bp).

Table 1

Structural diversity in the chloroplast genomes of *Thalassia* spp. across biogeographic regions.

Feature	<i>T. testudinum</i> , Guánica, PR	<i>T. testudinum</i> , Florida, USA	<i>T. hemprichii</i> , China (Ch1)	<i>T. hemprichii</i> China (Ch2)
bp	179,257	179,858	178,261	178,438
PCGs	141	142	140	144
IR	45,934	24,882	44,815	46,672
LSC	82,487	87,349	83,691	83,634
SSC	4,899	15,096	4,940	4,939

PCGs – Protein Coding Genes, not including tRNA and rRNA, based on the CPStool gene summary.

The chloroplast gene content of *Thalassia testudinum* from Guánica, Puerto Rico, closely resembles that of Florida *T. testudinum* and both *T. hemprichii* chloroplast genomes, especially in key photosynthetic genes. All genomes include the complete Photosystem I set (*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*) and most genes related to Photosystem II, ATP synthase, cytochrome *b/f*, *ccsA*, *rbcL*, ribosomal proteins, RNA polymerase, and other conserved functions. Some differences are evident between *T. testudinum* and *T. hemprichii*. Both *T. testudinum* genomes have *psbE*, but *T. hemprichii* Ch1 lacks it and instead has *psbN*. In the NADH dehydrogenase group, *T. testudinum* from Guánica and Florida shares a similar gene profile, including *ndhD* and *ndhG*, while *T. hemprichii* Ch1 lacks these genes. The conserved open reading frame (ORF) regions also differ: *T. hemprichii* Ch1 contains *ycf3* and *ycf4*, whereas *T. testudinum* genomes mainly have *ycf1* and *ycf2*. Notably, the Guánica chloroplast genome differs specifically in ORF content, missing *orf106*, which is present in Florida and both *T. hemprichii* genomes (Table S1).

The principal coordinate analysis (PCoA) of the genomic data demonstrates a clear differentiation among *Thalassia* species. The first axis (PCoA1) accounts for 99% of the variation and distinctly separates *T. hemprichii* China Ch1 from all other samples (Fig. 3A). Conversely, *T. hemprichii* China Ch2, *T. testudinum* Florida, and *T. testudinum* Guánica cluster more closely along this axis. Furthermore, the shared chloroplast genes grouped by functional category clearly distinguish between the Large Single Copy (LSC) and Inverted Repeat (IR) regions. Genes associated with photosynthesis and energy production (Photosystem I, Photosystem II, Cytochrome b/f, and ATP synthase) are exclusively located in the LSC, comprising 5, 13, 6, and 6 genes, respectively. Similarly, Rubisco (1) and tRNA (1) are confined to the LSC. NADH dehydrogenase genes are evenly distributed, with 4 in the LSC and 4 in the IRs. Ribosomal subunit genes are more prevalent in the IRs; small ribosomal proteins are represented by 5 in the LSC and 7 in the IRs, whereas large ribosomal proteins are present in 2 in the LSC and 7 in the IRs. The RNA polymerase genes predominantly reside in the LSC, with 3 shared genes compared to 1 in the IRs. The rRNA genes are exclusively located in the IRs, with 3 shared genes. Other shared genes are distributed between the two regions, with 3 in the LSC and 3 in the IRs for other coding genes, and 7 in the LSC and 10 in the IRs for other or unclassified genes (Fig. 3B).

A total of 8 chloroplast gene groups (*atp*, *ndh*, *pet*, *psa*, *psb*, *rpl*, *rpo*, *rps*) were analyzed by estimating non-synonymous (K_a), synonymous (K_s) substitution rates, and their K_a/K_s ratio (ω) from *T. testudinum* Guánica to contrast with the other *Thalassia* species. In Fig. 4, panels A and B show that most gene groups have low K_a and K_s values. However, the PET gene group (subunits of the cytochrome b/f complex), which is found in contrast between *T. testudinum* Guánica and *T. testudinum* Florida, has an average value of K_a higher than 0.6. The same gene group shows an average K_a of 0.3 against species from China. The *psb* (photosystem II) and *rpl* (large ribosomal subunits) genes exhibit more uniform K_a and K_s values across species. Therefore, in panel C, the most complex patterns are observed. The PET group, in contrast to *T. testudinum* Guánica and *Thalassia testudinum* Florida, remains close to the threshold of $\omega = 1$. On the other hand, at threshold C, contrasting *T. testudinum* Guánica with *Thalassia hemprichii* ch2 and *Thalassia testudinum* Florida, the group of *rps* (Small ribosomal units) genes contrasting *T. testudinum* Guánica with the two species from China exceeds the average value of 1.0. However, the groups of genes that were not mentioned, in contrast to all *T. testudinum* Guánica, remained below the average value of 1.

The analysis of the SSRs present in the chloroplast genomes of the different species of the genus *Thalassia* reveals patterns of variability between *T. testudinum* Florida and *T. hemprichii*, as well as between graphically isolated populations (Fig. 5). However, differences were identified in the frequency and type of simple repeats, particularly in monomers, where variations in the composition of nitrogenous bases are observed. The samples of *T. hemprichii* present a highly similar profile. In the same way, the variants of *T. testudinum* from Florida and the species sequenced in Guánica, Puerto Rico, show a high concordance. In general, pentanucleotide (penta) and hexanucleotide (Hexa) SSRs are present uniformly in all the species analyzed, showing no differences in number or sequence. However, dinucleotide (di) and tetranucleotide (tetra) motifs exhibit differences between species.

The results of the relative use of synonymous codons (RSCU; Fig. 6) indicate that 32 codons were used more frequently with RSCU > 1. It is observed that 27 of these 32 codons possess A/U at the third nucleotide position, except for UAG, AGG, GGG, UCC, and UUG. In contrast, most codons ending in G/C showed RSCU values less than 1. Of the 32 codons that possess A/U, AGG, GGG, AGA, ACA, GCA, and a different codon, UUG possesses an RSCU intensity > 1 in the *T. testudinum* Guánica. Only the AGA codon shows an RSCU intensity > 1 in the four *T. testudinum*. The A/U bias (RSCU > 1) in the third position of the codons was further evidenced by the A/U contents of the codons, with a mean value of 50% in the third position of the codon. 32 codons showed codon values that there is no bias of RSCU = 1.00 and RSCU < 1.00.

Phylogenetic analysis based on complete chloroplast genome sequences reveals evolutionary patterns that are highly congruent with the current taxonomic classification of the cp genome family Hydrocharitaceae. (Fig. 7) The close position of *Thalassia testudinum* Florida and *T. testudinum* Guánica formed a well-supported clade, with a bootstrap value of 100%. *A. thaliana* was enabled as an outgroup to root the phylogenetic tree. Bootstrap support values show 88% in *Thalassia hemprichii*.

Discussion

The complete chloroplast genome of *Thalassia testudinum* from Guánica, Puerto Rico (Fig. 2) was sequenced and yielded an average depth of approximately 200X, which was sufficient for high-quality assembly (Fig. S1). The Genome length was 179,257bp in contrast to the other chloroplasts of *T. testudinum* Florida and *T. hemprichii* from China, Ch1 and Ch2 showed that the Guánica *T. testudinum* had a remarkable similarity in genome length with the *T. testudinum* Florida. A total of protein-coding genes (PCG) 141 genes and 39% GC (Table 1) from Guánica and 142 genes and 39% GC from Florida (Table 1)⁴. Conversely, the *T. hemprichii* chloroplasts showed similarities in genome size and in the IR, LSC, and SSC regions. The difference in chloroplasts between species could be given by geographical location of the *Thalassia* spp^{4,5,13–15}. On the other hand, a difference was observed in the SSC region, which in the Guánica genome reaches 4,899bp, while the Florida genome extends to 15,096bp. This, in addition, is observed as a notable difference in the IR, which in Guánica constitutes about 45,934bp, almost twice as much as the Florida species. Also, IR shifts are associated with functional and evolutionary implications due to changes in gene relocation, shifts in IR boundaries, substitution rates, and selective patterns; what *T. testudinum* from Guánica could be representing an adaptive genomic reorganization that is linked to specific environmental pressures of the insular Caribbean¹⁶. This structural diversity raises the relevance of *T. testudinum*⁴, which is attributed to a signal of population differentiation and the correlation between gene reorganization due to environmental changes.

Comparison between the sequenced *T. testudinum* Guánica and *T. testudinum* Florida, and *T. hemprichii* China (Table 1S, Fig. 3A & B) revealed patterns of functional divergence that could be linked to regional environmental pressures. Several key genes involved in photosynthesis were conserved in the chloroplast genome of Puerto Rican *T. testudinum*. Specifically, the Photosystem II subunits *psbB*, *psbC*, and *psbF*, the ATP synthase subunits *atpA*, *atpB*, and *atpF*, and the NADH dehydrogenase subunit *ndhA*

were found in Guánica, as well as in Florida *T. testudinum* and the *T. hemprichii* species (Fig. 6). This conservation is significant given the geographic and ecological differences, indicating that the Puerto Rican chloroplast genome retains the core photosynthetic machinery with only minor variations in certain genes and ORF composition. In Guánica *T. testudinum*, the only absence in the ORF *orf106* shows a specific difference from other *Thalassia* spp. Any genomic divergence is notable given the geographic and ecological context: while *T. hemprichii* in China inhabits tropical environments with lower seasonal thermal variability⁹, *T. testudinum* populations in Florida¹⁷ Puerto Rico faces congruent and more intense fluctuations in temperature, pH, and salinity. According to previous studies, these abiotic factors can gradually change the biochemical characteristics of *Thalassia* grasslands, affecting gene expression and the stability of key metabolic pathways, and resulting in a change in the conservation response in living habitats¹⁸. Studies have shown that genomic plasticity in seagrass can be mediated by purifying and relaxed selection mechanisms, depending on the environmental context^{9,19}. Consequently, the structural changes in the plastidial genome of the Puerto Rican *T. testudinum* might be a potential adaptation to maintain photosynthetic efficiency under stress. These changes can be visualized in the presence-absence clustering (Fig. 6A), which shows that *T. testudinum* Guánica is close to *T. testudinum* Florida but still distinct. From an evolutionary perspective, gene loss and structural changes are not necessarily adaptive disadvantages. Rather, in extreme or marginal environments like Puerto Rico, this reduction in the gene repertoire can favor ecological specialization, allowing the seagrass to survive in suboptimal conditions^{4,20}.

A comparison with mutations in other species (Fig. 4) reveals that *T. testudinum* species exhibits contrasting diversity in non-synonymous (K_a) and synonymous (K_s) substitutions, and in the K_a/K_s (ω) ratio, relative to other *Thalassia* spp. The observed variability resulted in evolutionary disparity across the *atp*, *ndh*, *pet*, *psa*, *psb*, *rpl*, *rpo*, and *rps* gene groups. Most genes exhibited purifying selection ($K_a/K_s < 1$), ensuring the elimination of harmful genetic variations. This process preserves genetic integrity while allowing for adaptation to the local environment^{21,22}. In contrast, the *pet* gene group (Fig. 4A) showed signs of positive selection. Specifically, derived traits in this group appear more frequent in the Guánica population compared to Florida, differing from the patterns observed in the Chinese species. This suggests that natural selection is favoring specific adaptations in the Guánica population to enhance survival. These genes had mutations that improve their function, which shows a higher frequency of these genes in their population environment^{21,23}. In addition, the genetic diversity observed in these gene groups was primarily driven by silent mutations, changes in the DNA sequence that do not alter the amino acid sequence (Fig. 4B). Unlike the patterns seen in other *Thalassia* lineages, the Guánica species exhibited a distinct and stable profile of these synonymous changes. This suggests that while the protein identity remains conserved, the preference for optimal synonymous codons may be modulating protein folding rates and tertiary structure. Consequently, this mechanism likely facilitates adaptation to the local marine environment by optimizing gene function without compromising genomic integrity^{6,24,25}. However, the K_a/K_s ratio of genes such as *rpl* and *rps* has changed as evolutionary pressure occurred. The K_a/K_s ratio change for *rpl* is viewed in *T. testudinum* Guánica with *T. testudinum* Florida, and *T. hemprichii* Ch1 and Ch2. In addition, the *pet* and *rps* genes maintained a constancy of $K_a/K_s \sim 1.0$,

which demonstrates a relative relaxation of the purifying pressure, that is, they have remained conserved in the face of the evolutionary pressures that have occurred over the years²⁶.

The number of SSRs was identified from the complete chloroplast genome of *T. testudinum* sequenced in Guánica, P.R., while most SSRs were found in the LSC region and rarely in the IR regions. The SSRs showed a slight variation in *T. testudinum* Guánica (Fig. 5). The diversity of SSRs in the *Thalassia* species sequenced, in contrast to the other *Thalassia* species, reflects coherence in the composition of the nitrogenous bases, conservation of repetitive elements, and relentlessness of evolutionary pressure present^{16,27,28}. Although the variations are subtle, they are sufficient to distinguish species and reinforce SSRs as markers in chloroplast genomes related to functional conservation in the environment, which frames the *T. testudinum* Guánica as highly conserved¹².

RSCU values (RSCU) were calculated using CodonW in MEGA and showed that the patterns of use in all *Thalassia* species reflected a notable A/U trend in the third position (Fig. 6). In particular, the *T. testudinum* Guánica, Puerto Rico, exhibits intensification of codon use compared to the other *Thalassia* species. The bias of A/U codons in the third position, with an average of 50% in the genomes analyzed, has crucial functional implications, since they tend to form less rigid secondary structures in the mRNA, which facilitates translation and reduces energy expenditure. In addition, it is highlighted that, in marine environments, where metabolic efficiency is critical, optimization can represent an important adaptive advantage²⁹. Also, we analyzed that codon pairing influenced mRNA stability, tRNA availability, and protein folding efficiency³⁰. When comparing our results, there is evidence of functional convergence in the preferential use of A/U-terminated codons³¹. However, specific modulations were also observed according to the genus and the environment, which we can mention, reinforcing the idea that codon bias in chloroplasts is evolutionarily conserved, but can adapt to local pressures. A clear example is in the order of preference in the encoding of leucine (UUA > CUU > UUG > CUA > CUG > CUC) in these genera, which suggests a functional hierarchy influenced by molecular and evolutionary factors. Our results revealed a higher intensity in specific A/U codons, which suggests the local adaptation that is associated with variations in gene expression, photosynthetic efficiency, and tolerance to stress-causing factors, which reinforces the use of codon bias as an evolutionary marker.

The phylogenetic tree was constructed based on the nucleotide-coding sequences of the cp genomes in the main clade of the family Hydrocharitaceae (Fig. 7). The phylogenetic tree showed robust resolution in almost all nodes with a bootstrap greater than 90%, suggesting a high confidence in the inferred evolutionary relationships. The clades that group *T. testudinum* Florida and *T. testudinum* Guánica presented values of 100% of Bootstrap, and the same with *T. hemprichii* from China, so the monophyly of the genus *Thalassia* was supported within the plastidial context. The position of *A. thaliana* as an outgroup helped to have the key reference point that contextualized the evolutionary transition to the aquatic environment. This phylogenetic distance between the outgroup and marine clades underscored the magnitude of adaptive changes that have occurred in plastidial genomes, including modifications in the use of codons, Ka/Ks, and SSRs². Specifically, *T. testudinum* Guánica contrasts with *T. testudinum*

Florida, which suggests a close evolutionary relationship but also suggests that the species present in Puerto Rico is *Thalassia testudinum*. This relationship can be correlated with the unique patterns, reinforcing the hypothesis of local adaptation and functional divergence within the conserved lineage^{4,32}. This phenomenon has been previously demonstrated in other seagrass species^{5,7,12,15,17}.

A comparative analysis of the chloroplast genome of *Thalassia testudinum* Guánica with three other *Thalassia* spp. chloroplast genomes revealed a highly conserved gene set, with some minor yet informative differences. The Guánica genome clustered closely with the Florida *T. testudinum* genome based on gene presence and absence, confirming their close relationship within the same species. However, Guánica lacks *orf106*, which is present in the Florida *T. testudinum* and both *T. hemprichii* chloroplast genomes. The presence of positively selected genes and SSRs variation further suggests that while the Guánica chloroplast genome maintains the core functional structure of *T. testudinum*, it also exhibits localized genomic differences that may indicate population-level divergence. Overall, these results suggest that geographic separation can lead to subtle variation in the chloroplast genome of *T. testudinum*, while the essential photosynthetic and housekeeping genes remain highly conserved.

Methods

Sample collection. *Thalassia* sp. was collected from Tamarindo Beach, located in Guánica, Puerto Rico (17.95391°N, 66.85055°W), in about 1–3 feet of shallow waters. The collection of the *Thalassia* specie sample was conducted with the authorization of the Department of Natural and Environmental Resources (Departamento de Recursos Naturales y Ambientales; DRNA), permit No. O-VS-PVS15-SJ-01672-01062026. The plant material (*Thalassia testudinum*) was formally identified by Abner J. Colón Colón based on established morphological characteristics and taxonomic descriptions reported for Caribbean seagrasses. A voucher specimen was deposited in the MAPR Herbarium of the University of Puerto Rico, Mayagüez Campus, under accession number 403445 (barcode MAPR43893). Samples were transported on ice at 4°C and stored at -20°C until processing. Before isolation and extraction, samples were washed several times to remove contaminants and epiphytes.

Scanning electron microscopy (SEM) and confocal microscopy (CM). Fresh collected *Thalassia* sp. sample was cropped in size of approximately 2 cm. After the first part, one part of the tissue was added to 1 mL of PFA 1% solution and waited for 1 hour. Imaging parameters consisted of 405 nm excitation lasers to detect intrinsic autofluorescence signals from the biological tissue. Images were acquired with an exposure time of approximately 2000 ms and frame averaging set to enhance signal quality. For scanning electron microscopy (SEM), protocols^{33–35} were carried out with modifications. A sample of 2 cm tissue from *Thalassia* sp. was placed in a fixing solution (2% formaldehyde in PBS buffer) for 1 hour at refrigeration temperature (approximately 4°C). Afterwards, the sample was washed three times with PBS for 15–30 minutes each and washed three additional times with distilled water (dH₂O) under the same time conditions. Following the washing steps, the *Thalassia* sample was dehydrated through a low-high ethanol gradient of 25%, 50%, 75%, 85%, 95%, and finally 100%, 15 minutes each solution. The sample was then left to dry at room temperature for 15–30 minutes. After dehydration and room

temperature drying, the sample was mounted on a metal stub over a carbon sticker. Subsequently, the sample was vacuum dried for 30 minutes until the pressure reached 8 Pa. Then, the sample was coated with a thin layer of gold ranging between 7 and 12 nm using the Luxur Gold coater. Sample imaging started at 5 kV beam intensity to visualize detailed surface structures, followed by 10 kV and 15 kV to explore deeper surface details. Magnification ranges between 2,950 – 10,500 $\times$, while the resolution was set to 1920 x 1200, to ensure high-quality outputs.

DNA extraction and sequencing. For the chloroplast isolation, the central vein of the *Thalassia* spp. leaves were removed. Chloroplast extraction was performed following the protocols^{36,37}, with modifications to the preparation of grinding (0.5mL), cell suspension solutions (0.5mL), and centrifugation (1,000 x g for 7 minutes at 4°C). Briefly, the suspension solution was composed of (0.33 M sorbitol, 2 mM EDTA, 1 mM magnesium chloride, and 50 mM HEPES; pH = 7.6), and the grinding solution (0.33 M sorbitol, 10 mM sodium pyrophosphate, 4 mM magnesium chloride, 2 mM ascorbic acid; pH = 6.5). Chloroplast DNA extraction was performed using the DNeasy® Plant Pro-Kit (QIAGEN, Hilden, DE). DNA sequencing was performed using the MinION platform from Oxford Nanopore Technologies with the SQK-NBD114.24 barcoding kit (Native Barcoding Kit 24 V14) and R10.4.1 flow cell (FLO-MIN114). Base calling was performed using Dorado v1.0 with the high-quality option.

Genome assembly and comparative analysis. Nanoplot v1.45.0 was used to evaluate the quality of nanopore reads⁴⁰. Removal of low-quality reads was performed by Chopper⁴¹. To remove any residual contamination from the chloroplast isolation, a meta-assembly was performed using Flye v2.9.6 with the meta option⁴². To classify and separate contigs into categories (e.g., bacterial, organelles, etc.) Tiara v1.0.3 was used⁴². Using the publicly available *T. hemprichii* chloroplast genome (NC_043774.1³⁴) and the Tiara plasmid output as a reference for mapping to the nanopore raw reads using minimap2 v2.26⁴². The final chloroplast assembly was performed using the ptGAUL v1.0.5 program with *T. hemprichii* as a reference⁴³. For annotation of the chloroplast genome, the GeSeq web server from Chlorobox was used with standard parameters⁴⁴. To perform robust chloroplast comparison and determine key similarities and differences between *Thalassia* spp., three additional *Thalassia* chloroplast genomes, one of *T. testudinum* and two from *T. hemprichii*, were used⁶. Each *Thalassia* chloroplast underwent the same annotation process. The software package CPStools v3.0 was used for chloroplast and gene comparisons⁴³, and for Ka/Ks calculations between the *Thalassia* sp. chloroplasts. SSRs identification in the *Thalassia* chloroplast genomes was performed by the web server MISA-web version number⁴³. Differences in relative synonymous codon usage were determined using the RSCUcaller R package v1.0⁴⁴. A total of 26 chloroplast genomes from the Hydrocharitaceae family, and *Arabidopsis thaliana* as an outgroup, were used for a maximum likelihood phylogenetic analysis. Mafft v7.525 was used to produce genome alignments⁴⁵, and IQTree3 was used with a bootstrap of 1000 with the MFP option^{17,46,47}. The resulting trees were visualized and annotated in iTOL v6, where *Thalassia* lineages were highlighted, and bootstrap values were displayed at the corresponding nodes^{4,47}.

Declarations

Acknowledgements

This research was conducted at the Institute of Sustainable Biotechnology at the Inter- American University of Puerto Rico, Barranquitas (IAUPR-BR). We gratefully acknowledge Dr. Christopher A. Sambolín Pérez for his collaboration in reviewing the SEM methodology. A special thanks to the curator Dr. Jeanine Vélez Gavilán for her valuable assistance in the deposition a voucher specimen of *Thalassia testudinum* in the MAPR Herbarium.

Author contributions

The plant material (*Thalassia testudinum*) collection and formal identification were conducted by A.J.C.C. *Thalassia* chloroplast extraction and DNA extraction were carried out by A.J.C.C. & D.A.V.F. R.A.B. conducted the long-read Minion sequencing and provided experimental support. G.L.R. and A.J.C.C. carried out chloroplast assembly and annotation, and performed the analysis of the results, writing, revision, proofreading, and editing of the manuscript. General advice regarding the study was provided by R.E.R.V and A.R.N.M. Funding acquisition was by A.R.N.M. It is noted that all authors are aware of this study, and the respective authors reviewed the manuscript comprehensively.

Funding

The research and open access were supported by the Department of Science and Technology at the IAUPR-BR, in collaboration with the U.S. Department of Education, DHHS TITLE V, Award No. P031S220125.

Competing Interest

The authors have no conflicts of interest or competing interests to disclose.

Additional information

To ensure transparency and reproducibility, all sequencing data generated during this work have been made publicly available. The sequenced MinION long reads were deposited in the National Center for Biotechnology Information (NCBI) under the Sequence Read Archive (SRA) with the accession number SRX33179284 and Bio-Project accession number PRJNA1460811. The cp genome, annotation, and phylogenetic data produced in this study were deposited under the figshare repository with the following DOI: 10.6084/m9.figshare.32140951.

References

1. Van Dijk, K., Bricker, E., Van Tussenbroek, B. I. & Waycott, M. Range-wide population genetic structure of the Caribbean marine angiosperm *Thalassia testudinum*. *Ecol. Evol.* **8**, 9478–9490

- (2018).
2. Gutiérrez, R. A., Green, P. J., Keegstra, K. & Ohlrogge, J. B. Phylogenetic profiling of the Arabidopsis thalianaproteome: what proteins distinguish plants from other organisms? *Genome Biol.* **5**, R53 (2004).
 3. Medina-Gómez, I., Madden, C. J., Herrera-Silveira, J. & Kjerfve, B. Response of *Thalassia testudinum* Morphometry and Distribution to Environmental Drivers in a Pristine Tropical Lagoon. *PLoS ONE.* **11**, e0164014 (2016).
 4. Ma, X. et al. Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nat. Plants.* **10**, 240–255 (2024).
 5. Chen, J. et al. Chloroplast genomic comparison provides insights into the evolution of seagrasses. *BMC Plant. Biol.* **23**, 104 (2023).
 6. Kelchner, S. A. The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics. *Ann. Mo. Bot. Gard.* **87**, 482 (2000).
 7. Olsen, J. L. et al. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**, 331–335 (2016).
 8. Nguyen, X. V. et al. New insights into DNA barcoding of seagrasses. *Syst. Biodivers.* **13**, 496–508 (2015).
 9. Rojas, J. S. V. Crecimiento y asignación de biomasa radicular de *Thalassia testudinum* y *Syringodium filiforme*, en praderas marinas monoespecíficas y mixtas de la región de Barú, Caribe suroccidental (2020).
 10. Ugarelli, K. et al. Microbiomes of *Thalassia testudinum* throughout the Atlantic Ocean, Caribbean Sea, and Gulf of Mexico are influenced by site and region while maintaining a core microbiome. *Front. Microbiol.* **15**, 1357797 (2024).
 11. Ulrich, S. et al. Origin and Early Evolution of Hydrocharitaceae and the Ancestral Role of Stratiotes. *Plants* **13**, 1008 (2024).
 12. Lucas, C., Thangaradjou, T. & Papenbrock, J. Development of a DNA Barcoding System for Seagrasses: Successful but Not Simple. *PLoS ONE.* **7**, e29987 (2012).
 13. Zhang, W. et al. Comparative analysis of 17 complete chloroplast genomes reveals intraspecific variation and relationships among *Pseudostellaria heterophylla* (Miq.) Pax populations. *Front. Plant. Sci.* **14**, 1163325 (2023).
 14. Lian, C. et al. Comparative analysis of chloroplast genomes reveals phylogenetic relationships and intraspecific variation in the medicinal plant *Isodon rubescens*. *PLoS ONE.* **17**, e0266546 (2022).
 15. Yong, Y., Wang, Y., Wang, D., Yuan, X. & Zhang, Q. The organelle genomes of the endangered seagrass *Zostera caespitosa* reveal sequence divergences, massive gene transfer, and uncommon RNA editing types. *Front. Plant. Sci.* **16**, 1550467 (2025).
 16. Zhu, B. et al. Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PLoS ONE.* **16**, e0248556 (2021).

17. Ma, X. et al. Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nat. Plants*. **10**, 240–255 (2024).
18. Viana, I. G., Moreira-Saporiti, A. & Teichberg, M. Species-Specific Trait Responses of Three Tropical Seagrasses to Multiple Stressors: The Case of Increasing Temperature and Nutrient Enrichment. *Front. Plant. Sci.* **11**, 571363 (2020).
19. Mohamed, S. I. A., Elsayed, G. H., Shaffai, E., Yahya, A., Mettwally, W. S. & S. M. M. & A. In-vitro study of cytotoxic and apoptotic potential of *Thalassia hemprichii* (Ehren.) Asch. And *Enhalus acoroides* (L.f.) Royle against human breast cancer cell line (MCF-7) with correlation to their chemical profile. *BMC Complement. Med. Ther.* **24**, 244 (2024).
20. Lee, H. et al. Genomic comparison of two independent seagrass lineages reveals habitat-driven convergent evolution. *J. Exp. Bot.* **69**, 3689–3702 (2018).
21. Cvijovic, I., Good, B. H. & Desai, M. M. The Effect of Strong Purifying Selection on Genetic Diversity.
22. Erixon, P. & Oxelman, B. Whole-Gene Positive Selection, Elevated Synonymous Substitution Rates, Duplication, and Indel Evolution of the Chloroplast *clpP1* Gene. *PLoS ONE*. **3**, e1386 (2008).
23. Shi, L. et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **47**, W65–W73 (2019).
24. Qin, Q. et al. Comparative analysis of chloroplast genomes reveals molecular evolution and phylogenetic relationships within the Papilionoideae of Fabaceae. *BMC Plant. Biol.* **25**, 157 (2025).
25. Smith, D. R. Mutation Rates in Plastid Genomes: They Are Lower than You Might Think. *Genome Biol. Evol.* **7**, 1227–1234 (2015).
26. Yang, Q. et al. Chloroplast Genome Evolution in Four Montane Zingiberaceae Taxa in China. *Front. Plant. Sci.* **12**, 774482 (2022).
27. Sundaray, J. K. et al. Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. *Aquacult. Int.* **24**, 1089–1102 (2016).
28. Yu, Y. et al. Extraction and analysis of high-quality chloroplast DNA with reduced nuclear DNA for medicinal plants. *BMC Biotechnol.* **24**, 20 (2024).
29. Rubalcaba, J. G. Metabolic responses to cold and warm extremes in the ocean. *PLoS Biol.* **22**, e3002479 (2024).
30. Berg, M. D. & Brandl, C. J. Transfer RNAs: diversity in form and function. *RNA Biol.* **18**, 316–339 (2021).
31. Yu, Y. et al. Extraction and analysis of high-quality chloroplast DNA with reduced nuclear DNA for medicinal plants. *BMC Biotechnol.* **24**, 20 (2024).
32. Li, Z. Z. et al. Plastome phylogenomics and historical biogeography of aquatic plant genus *Hydrocharis* (Hydrocharitaceae). *BMC Plant. Biol.* **22**, 106 (2022).
33. Sambolín Pérez, C. A. et al. Biochemical and Molecular Characterization of *Musa* sp. Cultured in Temporary Immersion Bioreactor. *Plants* **12**, 3770 (2023).

34. Goldstein, J. I. et al. Variable Pressure Scanning Electron Microscopy (VPSEM). in *Scanning Electron Microscopy and X-Ray Microanalysis* 173–185 (Springer New York, New York, NY, doi:10.1007/978-1-4939-6676-9_12. (2018).
35. Pathan, A. K., Bond, J. & Gaskin, R. E. Sample preparation for SEM of plant surfaces. *Mater. Today*. **12**, 32–43 (2010).
36. Bhattacharya, O., Ortiz, I. & Walling, L. L. Methodology: an optimized, high-yield tomato leaf chloroplast isolation and stroma extraction protocol for proteomics analyses and identification of chloroplast co-localizing proteins. *Plant. Methods*. **16**, 131 (2020).
37. Leiva-Mora, M., Alvarado-Capó, Y., Acosta-Suárez, M., Cruz-Martín, M. & Roque, B. Protocolo para el aislamiento de cloroplastos intactos de hojas de plantas de *Musa* spp. obtenidas por cultivo in vitro y evaluación de la actividad fotosintética.
38. De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
39. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
40. Karlicki, M., Antonowicz, S. & Karnkowska, A. Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350 (2022).
41. Zhou, W. et al. Plastid Genome Assembly Using Long-read Data (ptGAUL).
42. Tillich, M. et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
43. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
44. Maździarz, M., Zając, S., Pauksztó, Ł. & Sawicki, J. RSCUcaller: an R package for analyzing differences in relative synonymous codon usage (RSCU). *BMC Bioinform.* **26**, 141 (2025).
45. Katoh, K. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
46. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
47. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).

Figures

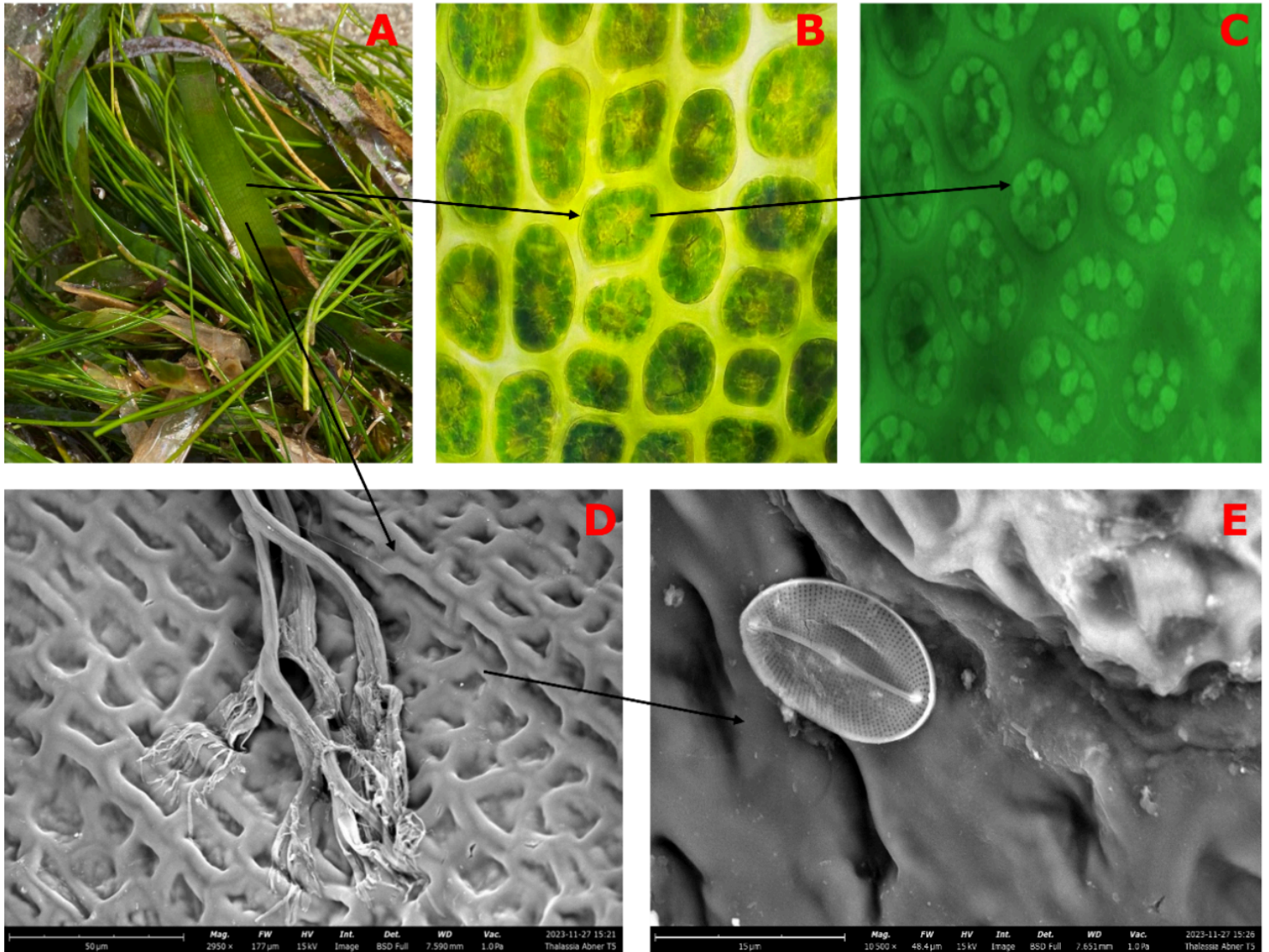


Figure 1

(A) *Thalassia* sp. from the ecosystem recollection zone; (B) Morphological characteristics at optical light microscopy 100X; (C) confocal fluorescence microscopy with 20X; (D and E) Scanning Electron Microscopy (SEM) images of anatomical structures of *Thalassia* sp.

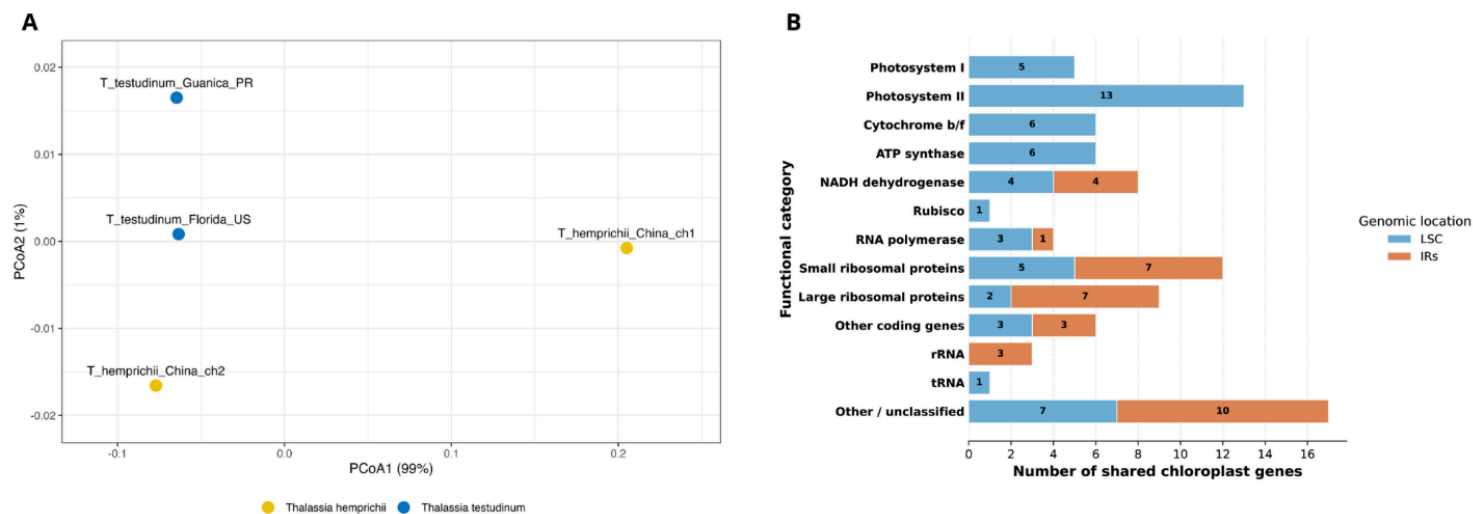


Figure 3

Presence-absence of shared genes arranged by functional category in the chloroplast genomes of *Thalassia* spp. (A) Principal component analysis in the presence and absence of shared genes. (B) Genomic location of shared in chloroplast genome architecture, colored code: blue (LSC) and brown (IRs).

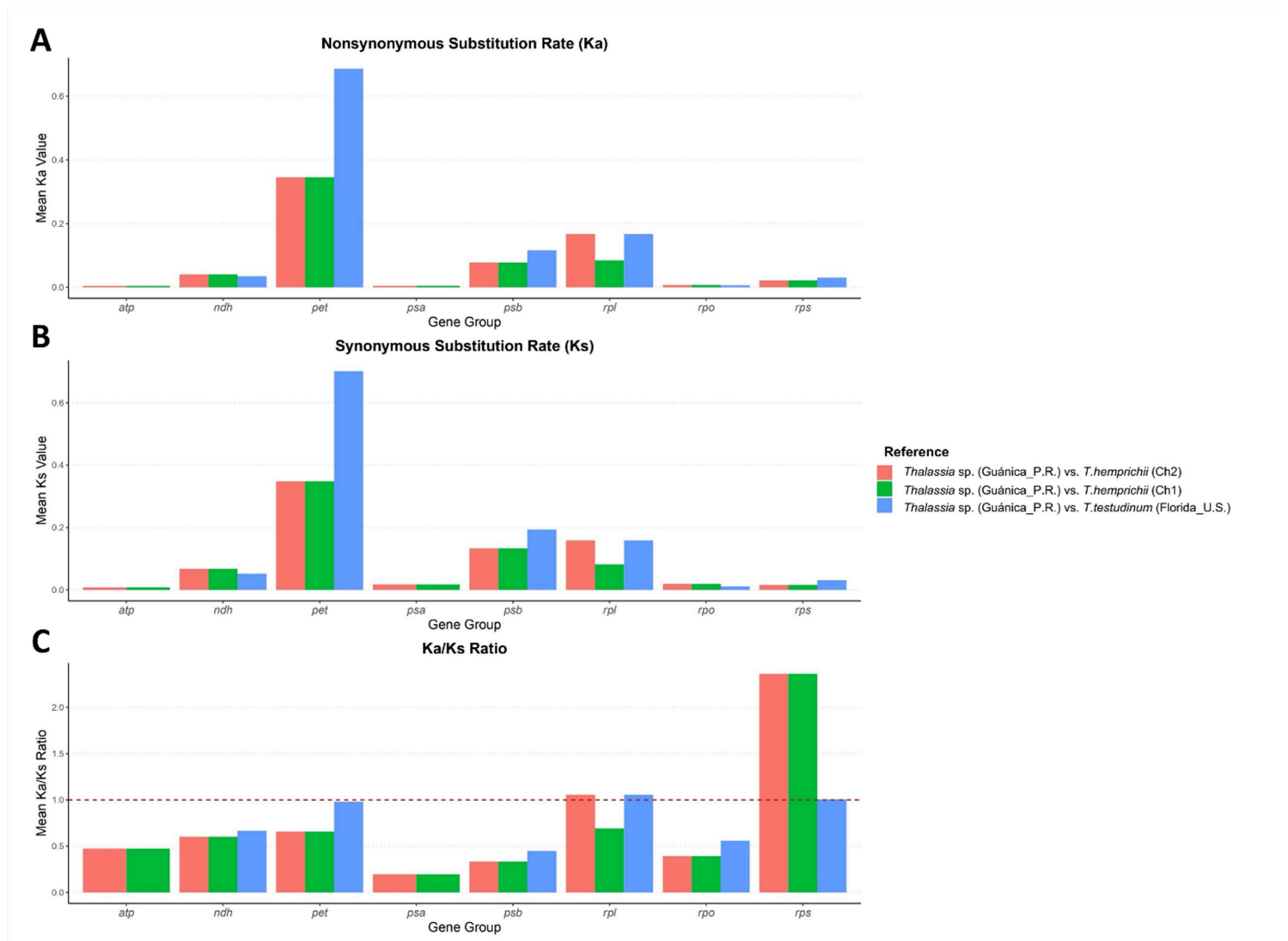


Figure 4

Analysis of substitution and evolutionary precision in the chloroplast genes of *Thalassia* spp. The comparison of rates of non-synonymous (Ka) (A) and synonymous (Ks) substitutions (B), as well as the Ka/Ks ratio (C), in chloroplast genes grouped by gene (*atp*, *ndh*, *pet*, *psa*, *psb*, *rpl*, *rpo*, and *rps*) is observed. The colors show the contrast differences of the *Thalassia* spp. The orange color represents *Thalassiasp.* (Guánica, P.R.) in contrast to *T. hemprichii* Ch2, green represents *Thalassiasp.* in comparison to *T. hemprichii* Ch1; and blue represents *T. testudinum* (Guánica, PR) in comparison to *T. testudinum* (Florida, U.S.).

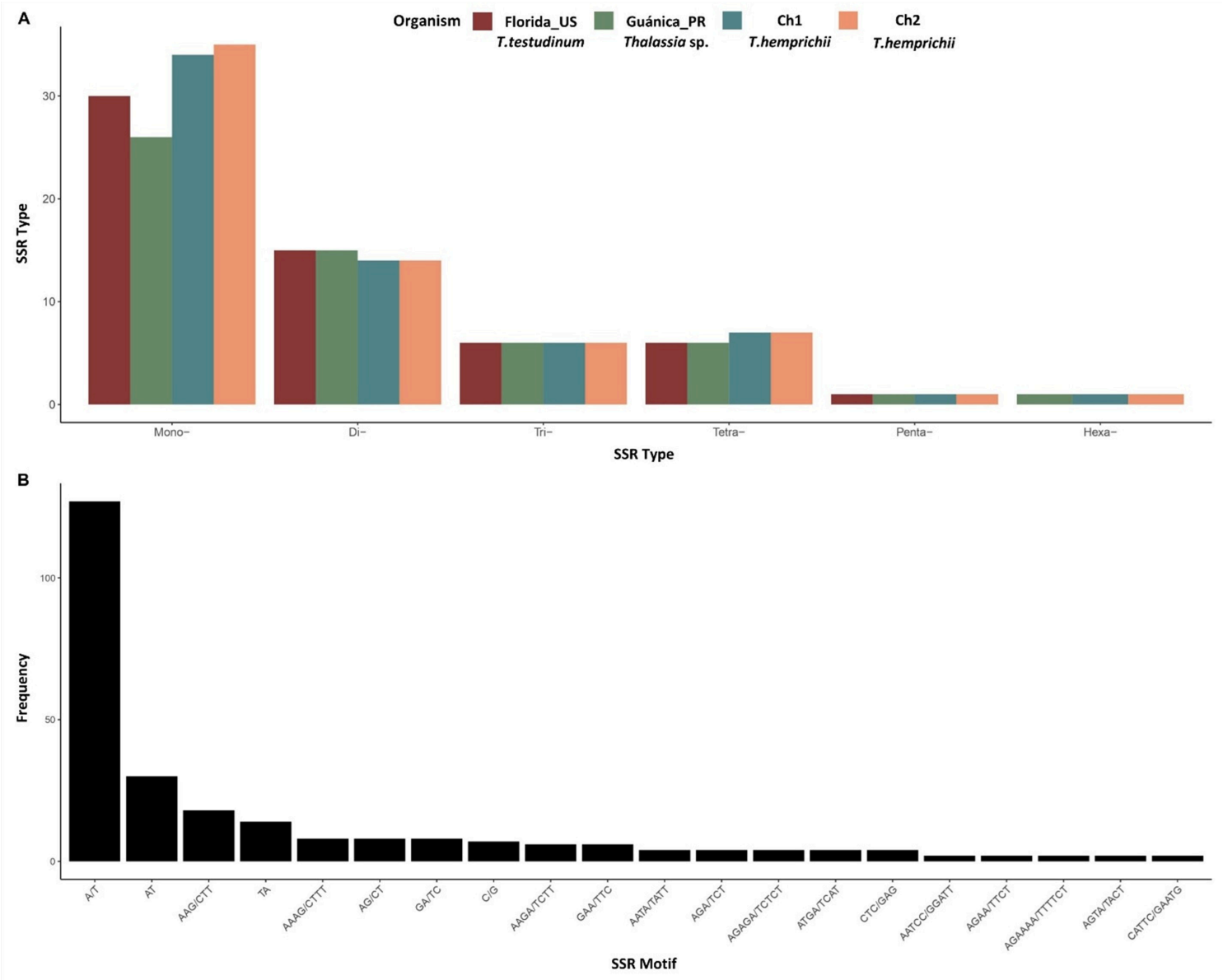


Figure 5

Comparative analysis of Simple Sequence Repeats (SSRs) in *Thalassia* spp. computed using PAML: Florida, U.S., Guánica, P.R., and China.

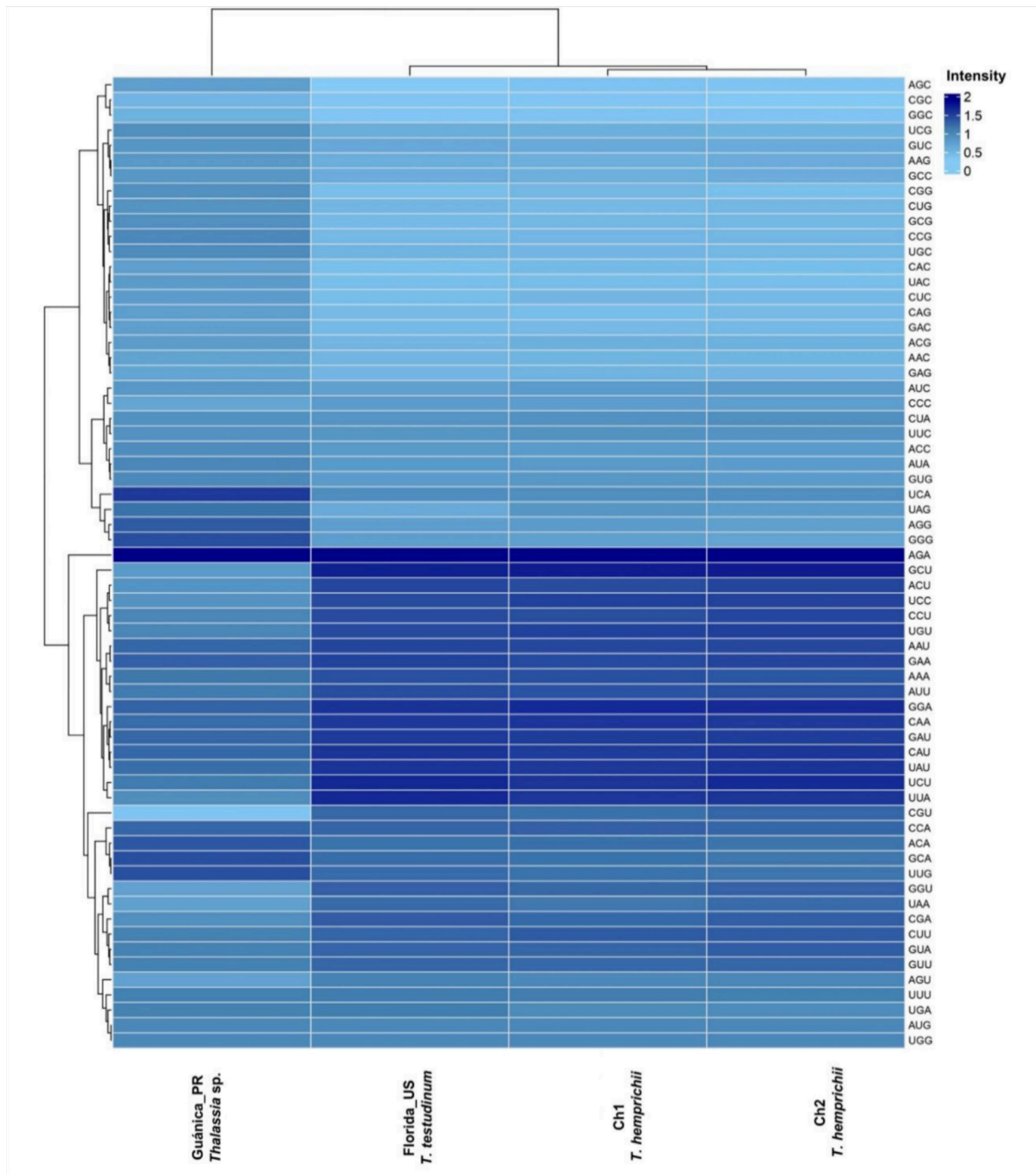


Figure 6

Intensity of codons in different *Thalassia* spp. The colors range from pale cerulean (0) to deep navy blue (2), indicating high similarity levels among codons. The samples are categorized into *Thalassia testudinum* (Guánica), *Thalassia testudinum* (Florida), *Thalassia hemprichii* (China: Ch1), and *Thalassia hemprichii* (China: Ch2).

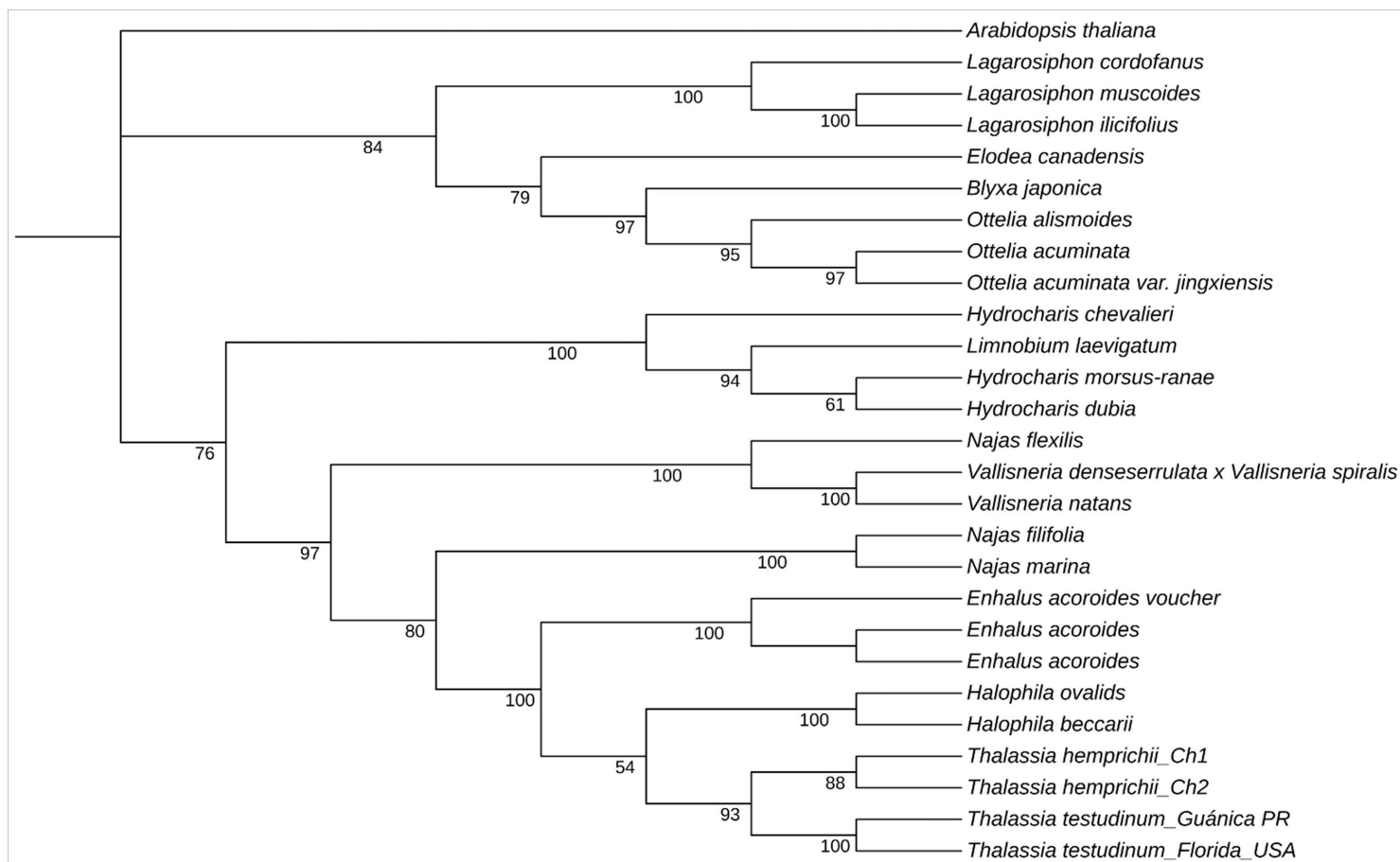


Figure 7

Maximum likelihood phylogenetic tree inferred with IQ-TREE, encompassing *Thalassia* spp. (Florida: US, Puerto Rico: PR, and China: Ch1 and Ch2), outgroup species, and chloroplast within the Hydrocharitaceae family

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ColonAbnerSupplementaryInformation.docx](#)