Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How do environments and reproductive systems affect the genetic diversity of seagrasses ?

Jun Chen

c.jevol@zju.edu.cn

Zhejiang University https://orcid.org/0000-0003-3187-7668

Xinyue Teng

Zhejiang University https://orcid.org/0009-0008-6531-7080

Pan Li

Zhejiang University https://orcid.org/0000-0002-9407-7740

Ruirui Fu

Xinjie Jin

Martin Lascoux

Uppsala University

Article

Keywords: Seagrasses, environment, sexual system, genetic diversity, nearly neutral theory

Posted Date: September 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10611218/v1

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

Additional Declarations: There is **NO** Competing Interest.

How do environments and reproductive systems affect the genetic diversity of seagrasses ?

Xinyue Teng[1], Pan Li[1,3], Ruirui Fu[1,5], Xinjie Jin[2], Martin Lascoux[4], Jun Chen[1,*]

[1] Systematic & Evolutionary Botany and Biodiversity Group, the State Key Laboratory for Vegetation Structure, Function and Construction, College of Life Sciences, Zhejiang University, Hangzhou, 310058 China

[2] College of Life and Environmental Science, Wenzhou University, Wenzhou, 325035, China

[3] Key Laboratory of Biodiversity and Environment on the Qinghai-Tibetan Plateau, Ministry of Education, School of Ecology and Environment, Xizang University, Lhasa 850000, China

[4] Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, SE-75236 Sweden

[5] Department of Forestry, College of Forestry, Northwest A&F University, Yangling, 712100 China

[*] Corresponding author: cjevol@zju.edu.cn

Abstract

Seagrasses occupy a vast geographical area and provide important ecosystem services. They are characterized by a large diversity of reproductive systems and environments. Some species are dioecious while others are hermaphrodites or monoecious. Some are fully marine and others live in freshwater environments. Hence, seagrasses are a perfect group of species to assess the relative importance of reproductive system and environment in shaping genetic diversity and test how these factors affect predictions of the Nearly Neutral Theory of molecular evolution (NNT). In the present study we sequenced the transcriptome of 127 individuals from 27 species of the order Alismatales. A well-supported phylogeny confirmed that evolutionary transition from freshwater to marine life occurred multiple times. Species from freshwater environments were, on average, more genetically diverse than those from marine environments. This difference was, in large part, due to the low diversity of dioecious species in marine environments. The latter could be due to pollen limitation and ensuing Allee effect and/or biased sex ratio. Selection was more efficient in freshwater environments than in marine ones, as would be expected under the nearly neutral theory (NNT) and the quantitative fit between data and the NNT prediction was surprisingly good.

Keywords: Seagrasses, environment, sexual system, genetic diversity, nearly neutral theory

Introduction

In spite of the availability of massive amount of genomic data, understanding what shape genetic diversity remains challenging. This may not come as a surprise considering the large number of factors influencing genetic diversity (Ellegren & Galtier, 2016). Some of these factors are related to the properties of the organism: for instance, genetic factors, e.g. recombination and mutation, or life history traits, e.g. reproductive system. The reproductive system can itself be further subdivided into sexual system (e.g. dioecy vs. monoecy) and mating system (e.g. selfing vs. outcrossing) both of which are expected to have an important influence on genetic diversity. Others factors are environmental and ecological, for instance, geological or climatic events, and interaction with other species. To this day, the nearly neutral theory of molecular evolution (Ohta, 1973), because of its elegant simplicity and its clear and testable predictions, remains the main theoretical framework under which we may attempt to identify the factors influencing genetic diversity. The nearly neutral theory posits that there is a fraction of new mutations of such weak effect that their fixation probabilities are determined by both selection and random genetic drift (Ohta, 1973). The rate of fixation of such mutations is expected to depend on the effective population size of species (Ohta, 1992; Ohta & Gillespie, 1996), which in turn depends on most of the factors listed above, such as selection, reproductive system, generation time or population size (Caballero, 2020). For example, Chen *et al.* (2017) demonstrated the strong influence of mating system and generation time on the efficacy of selection: selfing species and long-term perennials both accumulated more slightly deleterious mutations than outcrossing annuals. Similarly, sexual system also influences genetic diversity and, for example, dioecy is associated with high diversity and higher selection efficacy in the *Silene* genus (Muyle *et al.*, 2021). Previous tests of the nearly neutral theory showed a rather good fit to the data when species from different genus were considered (Chen *et al.*, 2017;

Bastian *et al.*, 2025). However, the fit was weaker when species were more closely related and biologically similar (Mackintosh *et al.*, 2019) and species with small effective population sizes fitted better than those with large ones (Chen *et al.*, 2020), possibly because of the larger part played by favourable mutations in the latter.

In the present study, we focus on the factors shaping genetic diversity in seagrasses. Seagrasses are angiosperms living in estuarine or marine environments. They evolved from terrestrial plants some 70-100 mya (Gomez *et al.*, 2015). Today, seagrasses occupy a vast area (estimated to be from 600,000 km^2 to 1,6 million km^2) and provide important ecosystem services (https://ocean.si.edu/ocean-life/plants-algae/seagrass-and-seagrass-beds). This alone would make them worth studying. But what makes them particularly interesting from an evolutionary point of view is the diversity of their reproductive systems and of the contrasted environments in which they live. Some species are fully marine while others live in freshwater environments. Many seagrass species live in depths of 1 to 3 meters in the intertidal zone, but some seagrasses (*Halophila decipiens*) have been found at depths of up to 85 meters (Den Hartog, 1970). Marine seagrasses have transitioned from a freshwater environment to a submerged saline environment on at least three independent occasions (Ma *et al.*, 2024). Based on the comparison of a few selected gene families, Ma *et al.* (2024) argued that the transition from the submerged-freshwater to submerged-marine environment did not require major changes, even if seagrass species did, of course, experience some significant changes such as gain of salt tolerance and loss of stomata.

However, Ma *et al.* (2024) only considered a limited number of species and therefore did not consider the diversity of reproductive systems that characterize seagrasses. Indeed, in both fully marine and freshwater environments the sexual system varies and seagrasses can be

monoecious, hermaphrodites or dioecious. Pollination can be hydrophilous or aerial (in *Enhalus* species, for instance), asexual growth is common and some species may form seed banks. In general, the reproductive system is known to have a profound effect on effective population size and consequently on the efficacy of selection, recombination and mutation. We would therefore expect to find a large variation in genetic diversity among seagrass species, as well as major differences in the distribution of fitness effects of new mutations (DFE). The impact of reproductive system on genetic diversity is also likely to be strongly affected by the environment with intertidal and subtidal species being *a priori* exposed to a much more variable environment than fully marine species. Therefore, if the variability in the environment is associated to large fluctuations in population size, one would expect species from the intertidal and subtidal zones to have lower genetic diversity than those from the full marine environment. On the other hand, pollination may be less successful in marine than in intertidal environments, especially for dioecious species due to rarity of either sex or pollen dispersal problems (Larkum *et al.*, 2006).

In the present study we first generated Iso-seq full length transcriptome references for 21 species of seagrasses representing the core Alismatids clade of Alismatales and estimated nucleotide diversity from RNA-Seq data of 27 species of seagrasses for both synonymous and non-synonymous sites. We then used these data to search for the main determinants of genetic diversity and test simple predictions from the Nearly Neutral Theory of molecular evolution (Ohta, 1973), with special emphasis on the effect of sexual system and environment.

Results

Phylogeny and life history traits of Alismatales

In total, 27 species were collected from 10 families of the Alismatales order and a ML phylogenetic tree was inferred. Seventeen species lived in a freshwater environment and ten were marine species. The phylogenetic tree was based on 2,079 orthologous genes and had strong bootstrap support (**Fig. 1A**). Alismatales diverged into Araceae and the core Alismatids around 96.4 Mya. The latter further split into two major clades about 86.4 Mya, one including Aponogetonaceae, Posidoniaceae, Cymodoceaceae, Ruppiaceae, Potamogetonaceae, and Zosteraceae, and the other including Alismataceae, Butomaceae, and Hydrocharitaceae. In general, the topology is consistent with that obtained with chloroplast genes by Les *et al.* (1997) (Note that Najadaceae is now merged into Hydrocharitaceae) and that based on combined transcriptomic and genomic data (Chen *et al.*, 2022). The phylogeny confirmed that evolutionary transition from freshwater to marine life occurred multiple times. The topology of the phylogenetic tree inferred from 127 individuals is consistent with the topology of species tree (**Fig. S1**).

All marine seagrasses of this study are submersed and water-pollinated except *Enhalus acoroides*, which is the only known marine species so far that does aerial surface pollination (Döring, 2022). Eight species are monoecious, 10 are dioecious, and 9 are hermaphroditic (**Fig. 1A**). *Najas minor* and *N. marina* are annual (Haynes, 1979) and the remaining 25 species are reported to be perennial. See **Table S4** for all details.

The effects of reproductive system and habitat on genetic diversity and DFE

In this analysis, we excluded *Spirodela polyrhiza* as it could be a lab strain due to its extremely low genetic diversity and high GC3 content (**Table S4**). To avoid bias from paralogs, for each gene family one of the longest genes, with length over 500bp, was chosen in each species for all population genetic analyses.

Reproductive system traits had a strong impact on nucleotide diversity (**Table 1**). We first compared species with different pollination modes. All estimates of nucleotide diversity are reported per base pair. The median synonymous nucleotide diversity, π_4, was almost twice as large for species with aerial pollination than for species with hydrophile pollination (3.83×10^{-3} and 1.85×10^{-3}, respectively). Sexual system too influenced nucleotide diversity, with π_4 equal to 3.39×10^{-3}, 1.85×10^{-3}, and 4.25×10^{-3} for dioecious, monoecious, and hermaphroditic species, respectively. However, the largest and most significant genetic diversity difference laid between freshwater and marine species with a median π_4 of 3.97×10^{-3} and 6.67×10^{-4}, respectively (p-value = 5.55×10^{-3}, corrected for the phylogeny). This difference can be attributed to the dioecious species since the marine dioecious species had much lower genetic diversity than the freshwater dioecious ones (median π_4= 2.65×10^{-4} versus 4.23×10^{-3}, p-value = 0.0102. **Fig. 1B**). A lower diversity was also observed in both marine monoecious and hermaphrodite species than their freshwater counterparts although the difference was not significant (**Table 1**). In general, the phylogenetic signal on genetic diversity was negligible (Pagel's $\lambda = 7.33\times10^{-5}$, LR test p-value = 1 when comparing to $\lambda = 0$) while habitat difference explained 26.9% of the total variance of genetic diversity and the sexual system explained 9.44% of the variance.

A low value of the ratio of nonsynonymous to synonymous nucleotide diversity, π_0/π_4, indicates that selection against deleterious mutations is efficient. The ratio π_0/π_4, was significantly lower in freshwater species than in marine ones (0.248 versus 0.347, p-value = 6.08×10^{-4}). The largest difference between freshwater and marine environments was again observed for dioecious species (0.201 versus 0.431, respectively. p-value = 1.62×10^{-7}, **Fig. 1C**). In contrast, similar π_0/π_4 ratio values were observed in marine and freshwater monoecious

160 species (0.307 and 0.283, respectively) while marine hermaphrodites had higher π_0/π_4 than
161 freshwater ones (0.407 versus 0.274), though the difference was non-significant after
162 correction for phylogeny. In conclusion, selection appears to be more efficient at purging
163 slightly deleterious mutations in freshwater than in marine environment. To ensure that our
164 results were not affected by species-specific expressed genes, we also estimated both π_4 and
165 π_0/π_4 with the 2,079 low copy number orthologous genes that were used to construct the
166 Alismatales phylogenetic tree. Estimates of π_4 and π_0/π_4 obtained with both datasets were
167 highly correlated (Pearson's correlation test, r = 0.993, p-value < 2.2×10^{-16} for π_4, and r =
168 0.981, p-value < 2.2×10^{-16}), and species group comparisons led to the same conclusions.
169
170 For five marine and five freshwater species with a sample size of 10 to 26 alleles, we also
171 estimated the distribution of fitness effects of new mutations (DFE) using a gamma distribution
172 for deleterious and neutral mutations and an exponential distribution for beneficial mutations
173 (**Table 2**). A low value of the shape parameter of the gamma distribution, β, indicates a large
174 proportion of slightly deleterious mutations and that purging does not occur when the effective
175 population size increases. With, on average, a slightly smaller β (0.11 versus 0.153, p-value =
176 0.998), the marine species had a significantly larger proportion of neutral and slightly
177 deleterious ($N_e s \in [-1, 0]$) than the freshwater species (0.346 versus 0.227, p-value =
178 0.00112). In contrast, the proportion of strongly deleterious ($N_e s \in [-\infty, -10]$) and strongly
179 beneficial mutations ($N_e s \in [1, \infty]$) was larger in freshwater species than in marine ones (0.661
180 versus 0.534, p-value = 0.0172; 8.42×10^{-3} versus 7.18×10^{-5}, p-value = 0.00185, respectively.
181 **Fig. S2**). The selective strength of beneficial mutations (S_b) was much stronger in freshwater
182 species than in marine species (16.34 versus 0.174, p-value = 0.471) though not significant.
183 **Table S4** gives all DFE parameter estimates.
184

185 ***The effects of generation time, ploidy, inbreeding, and GC3 content***

186 In this study, two species were annual (*Najas minor* and *Najas marina*) and three were
187 tetraploid (*Najas minor*, *Potamogeton crispus*, and *Potamogeton distinctus*). There was no
188 significant difference between annual and perennial species, nor between diploid and tetraploid
189 species for the above statistics.

190

191 The inbreeding coefficient F_{IS} varied from -0.55 in *Ruppia maritima* to 0.54 in *Najas minor*
192 with a median of -0.034. On the average, freshwater species had a very similar inbreeding
193 coefficient to marine species (median = -0.034 versus -0.043). While monoecious species had
194 positive F_{IS} = 0.126, dioecious and hermaphrodite ones, on the other hand, had negative F_{IS} =
195 -0.0101 and -0.159, respectively (**Table 3**). But none of these differences were significant.
196 Selfing rate was also calculated for hermaphrodite and monoecious species. The monoecious
197 species had higher selfing rates than hermaphroditic species (0.244 versus 0.0677) and the
198 freshwater species also had higher selfing rates than marine species (0.174 versus 0.0531).
199 Neither F_{IS} nor selfing rate had any significant effect on π_4, π_0/π_4 or DFE parameters.

200

201 The distribution of Guanine and Cytosine (G/C) content in the third codon position, GC3,
202 varied from 0.38 in *Zostera marina* and 0.39 in *Z. muelleri* to 0.48 in *Colocasia esculenta*
203 (*Spirodela polyrhiza* was excluded from this analysis because of its extremely high GC3
204 content, 0.68). The median value of GC3 was 0.455 in dioecious species, and was slightly
205 higher than the median value in both monoecious (0.448) and hermaphroditic species (0.417).
206 There was no significant difference in GC3 between marine and freshwater species (0.44 versus
207 0.434). Neither did we find a correlation between GC3 and F_{IS} or selfing rate.

208

As GC3 content is generally positively associated to the population recombination rate in plants (Marais, 2003), we expected a positive correlation between π_4 and GC3 and a negative correlation between π_0/π_4 and GC3. However, the correlation varied with sexual system, i.e., a significant interaction between GC3 and sexual system was found for both π_4 and π_0/π_4 (**Fig S3**). In dioecious species, GC3 had a positive effect on π_4 (linear regression coefficient = 31.03, *p*-value = 2.157×10^{-5}). On the contrary, GC3 had a negative effect on π_4 in hermaphroditic species (coef. = -38.47, *p*-value = 1.011×10^{-9}). No significant effect of GC3 on π_4 was found in monoecious species. π_0/π_4 was negatively correlated with GC3 in dioecious species (coef. = -4.65, *p*-value=0.0498) while a positive correlation was observed in hermaphroditic species (coef. = 6.087, *p*-value = $1{,}01\times10^{-3}$). No significant correlation was identified between π_0/π_4 and GC3 in monoecious species.

The effects of census size and historical N_e

The species distribution range was estimated as the number of observation sites for each species in 1×1 grid of latitude and longitude. A significantly positive log-log linear correlation was observed between species distribution range and genetic diversity after correction for phylogeny: the wider the species range, the higher the genetic diversity (coef. = 0.48, R^2 = 0.207, *p*-value = 0.0143). We further used the inverse of leaf size as an estimate of plant density and multiplied that with the distribution range to obtain a proxy for species census size (N_c), which in turn was significantly correlated to π_4 (coef. = 0.44, R^2=0.433, *p*-value = 2.75×10^{-5}, **Fig. 2A**).

To evaluate the effect of historical N_e on genetic diversity, trajectories of historical effective population size changes were first estimated using Stairway Plot 2 based on the SFS of 4-folded sites. For marine dioecious species, historical population sizes were rather stable but on average

much smaller (1,000 ~ 10,000) than those of freshwater species except *Halophila ovalis*, which had a large and stable population size around 300,000. In contrast, the freshwater dioecious species had effective population sizes around 5×10,000~2×100,000 about 20,000 years ago (**Fig. 3A**). Thereafter the population sizes of *Najas marina* and *Vallisneria natans* remained stable while that of *Hydrilla verticillata* decreased to 6000 and *Ottelia cordata* and *Egeria densa* continued decreasing to 200 about 100 years ago. The effective population sizes of both monoecious and hermaphrodite species decreased to the same extent (**Fig. 3B, C**). We then calculated the harmonic mean of historical effective population sizes as an estimate of N_e for each species (**Table 3**). The median historical N_e of marine dioecious species (N_e= 3,141) was much smaller than that of the freshwater dioecious species (N_e= 18,627). For both monoecious and hermaphroditic species, estimates of historical N_e were at a very similar level for the marine and freshwater species, varying from 10,700 to 17,800 (**Table 3**). Finally, we observed a good fit of the regression of genetic diversity on historical N_e (coef. = 1.26, *p*-value < 2.2×10^{-16}, R^2 = 0.901, **Fig. 2B**).

The π_0/π_4 ratio was negatively correlated to the distribution range (coef. = -0.116, *p*-value = 0.00571, R^2 = 0.241), the proxy of N_c (coef. = -0.107, *p*-value = 1.72×10^{-6}, R^2 = 0.499), and historical N_e (coef. = -0.21, *p*-value = 1.4e-9, R^2 = 0.634. **Fig. 2C, D**).

Differences in evolutionary rate between freshwater and marine species

K_a/K_s was calculated based on the M_{free} model, which allows K_a/K_s ratio to vary freely for each phylogenetic branch, and averages across all 2,079 ortholog genes after filtering out genes with $K_a/K_s > 5$ or $K_s = 0$. Overall, K_a/K_s values were significantly lower in freshwater species than in marine species, regardless of the sexual system or pollination mode (K_a/K_s = 0.0998 and 0.145 for dioecious, 0.08 and 0.134 for monoecious, and 0.0882 and 0.123 for

hermaphroditic, *p*-values < 5e-8, **Fig. 4A).** A significantly negative log-log linear correlation was observed between mean K_a/K_s and the proxy of N_c (coef. = -0.08327, R^2 = 0.492, *p*-value = 1.66e-4; **Fig. 4C**). Similarly, $K_a/K_s$was also significantly and negatively correlated with historical N_e (coef. = -0.133, R^2 = 0.281, *p*-value = 0.0045; **Fig. 4D**).

By comparing an evolutionary model allowing two different K_a/K_s ratios for marine and freshwater species (M_3) to a model with a single K_a/K_s across all species (M_0), we identified 298 OGs evolving at different rates between marine and freshwater species (i.e., the M_3 model had a significantly better performance than the M_0 model). Of these OGs, 240 had smaller K_a/K_s ratios in freshwater species (median = 0.077 versus 0.184, **Table S5**). They were functionally enriched in genes related to metal ion homeostasis, photomorphogenesis, stomatal movement, drought tolerance, and so on (FDR < 0.05, **Fig. 4B**). On the other hand, the freshwater species had larger K_a/K_s values in 58 OGs (**Table S6**), which were functionally enriched in GO terms such as response to high light intensity, negative regulation of ethylene-activated signaling pathway, and response to low light intensity stimulus (p-value < 0.05). However, no significant enriched GO terms were retained after corrected for FDR. Notably, the ABI five binding protein 3 (AFP3) has a K_a/K_s ratio significantly larger in the freshwater species (1.053 versus 0, FDR = 0.024). AFPs are negative regulators interacting with key components of ABA signaling pathway, including MAP kinases (Lynch *et al.*, 2026).

Difference in gene copy number and expression between freshwater and marine species

Eight OGs were present in freshwater species but were either absent or not expressed in all marine species (**Table S7**). In particular, gene PDV2 is essential for chloroplast division, CLC-E encodes a chloride channel protein in thylakoid membrane and regulates the light reaction of photosynthesis, and HSA32 encodes a heat-stress-associated protein in pollen tube and seed.

In total, 39 OGs had larger copy numbers in freshwater species than in marine species (**Table S8**, **Fig. 5A**). Seven of them (e.g., gene AGT, PSBR) have functions related to photosynthesis and photorespiration and eight (e.g., GOX1, CRPK1, TCF1) are functionally related to plant response to various external stimuli, such as acidic PH, aluminum / cadmium ion, freezing, salt stress, insect, fungus, herbivore, oxygen level, and in ammonia and sulfate assimilation, and in vitamin E biosynthesis as well. Vitamin E is essential for plant response to stress, e.g., to protect photosystem and maintain the stability of chloroplast under thermal, light or drought stress (Collin *et al.*, 2008; Hasanuzzaman *et al.*, 2014; Niu *et al.*, 2022). On the other hand, in 25 OGs, marine species had larger copy numbers than freshwater species (**Table S9, Fig. 5A**). Six OGs (e.g., PAE8, HBT) are functionally enriched in biological processes related to in root meristem specification, root cap development, cell wall organization, and pectin biosynthesis while two (P5CS2 and SOS1) are related to hyperosmotic salinity response, potassium / sodium ion import, regulation of intracellular pH. Adaptation to the marine environment does not only require salt tolerance, but also elasticity and structural strength of the cell wall due to strong water currents (Ma *et al.*, 2024). No genes were found that are specifically expressed in marine species in this study.

Additionally, we identified 74 and 51 OGs with gene expression level significantly higher in marine and freshwater species, respectively ($|\log_2 FC| > 1$ and FDR < 0.05; **Fig. 5B, Table S10**). 46 of them were functionally related to plant tolerance to drought, salt, heat, clod, UV, and oxidative stress while nine were photosynthesis related, including chloroplast development and positioning. Genes PUB21 and GME had the highest expression level in marine species (FC = 126, FDR = 4.87e-4; FC = 2948, FDR = 3.63e-4, respectively). PUB21 negatively regulates drought, salinity stress, and immune responses in plant. While the stress signals can up-regulate the expression of PUB21, the silencing of PUB21 significantly enhances resistance to those

stresses (Yang *et al.*, 2022; Yi *et al.*, 2024). Gene GME is involved in ascorbic acid (vitamin C) biosynthesis, which has a vital protective role in plant tolerance to various abiotic stresses. The photosystem of marine seagrass is more vulnerable to the damage of high-energy wavelength of light due to the loss of stomata and photoreceptor genes during its complex evolution (Olsen *et al.*, 2016; Zhao *et al.*, 2021). An elevated ascorbic acid content is widely observed in seagrass and can protect its photosystem from reactive oxygen species (ROS) (Zhao *et al.*, 2021; Thiraviyam *et al.*, 2025). Gene CESA1 had the highest expression level in freshwater species but was barely expressed in marine species (mean TPM = 117.9 versus 0.0083, FDR = 0.0047). The gene plays a crucial role in the synthesis of cellulose and hemicellulose within plant cell wall that can prevent breakage due to physical forces and is important for the evolution of land plants (Graham *et al.*, 2000). Interestingly, for all but seven differentially expressed genes, regardless of their expression levels, K_a/K_s values in marine species were significantly higher than in freshwater species (median = 0.155 versus 0.0899, respectively).

Discussion

Seagrasses live in contrasted environments and have diverse reproductive systems. The main aim of the present study was to investigate to what extent these two factors affected seagrasses genetic diversity. The short answer is that environment most certainly matters and, furthermore, interacts with the reproductive system. More specifically, diversity is higher in freshwater than in marine environments and this is principally due to the large difference in nucleotide diversity of dioecious species between the two environments. Below we shall discuss these results and their implications for the nearly neutral theory of molecular evolution. But before that, we shall discuss the relationship between standing genetic diversity and other factors.

Genetic diversity is associated to historical effective population size and current census number

Many factors have been invoked to explain the variation in nucleotide diversity and selection efficacy, including life history traits (e.g. propagule size, mating system), ecological characteristics (e.g. distribution range, environment factors) and/or genomic architecture (e.g. genome size, recombination rate) (Romiguier *et al.*, 2014; Corbett-Detig *et al.*, 2015; Buffalo, 2021; Clark *et al.*, 2023). All these factors can eventually lead to differences in either demography (census and historical effective population size) or selective strength (e.g. DFE, hitch-hiking effect) (reviewed in Charlesworth & Jensen, 2022).

In this study, the largest difference in nucleotide diversity was observed between dioecious species from marine and freshwater environments, respectively. However, in contrast to previous studies (Chen, et al. 2017; Burgarella, et al. 2024) effects of mating system (selfing versus outcrossing), pollination agent, or generation time were negligible. This could simply be due to the fact that our study lacks species with high selfing rate, even for hermaphroditic species. Our sampling strategy also limited the effect of clonal reproduction. Similarly, most species in the present study are perennial and diploid; only two species *Najas minor* and *N. marina* are annual (Haynes, 1979) and three species, *Potamogeton crispus*, *P. distinctus* and *Najas minor* are possibly tetraploid (Kaplan *et al.*, 2013; Ito *et al.*, 2017).

Somewhat surprisingly, a crude estimate of census size based on distribution range size and leaf size is a rather good proxy of genetic diversity in our study. Similarly, the harmonic mean of historical effective population size change shows a significant association to both synonymous nucleotide diversity, π_4, and selection efficacy, π_0/π_4. In plants, (Milesi *et al.*,

2024) studied the Ne-trajectories of seven tree species with large distribution ranges. They did not estimate the correlation between past N_e and the current diversity but showed that Ne-trajectories could be clustered into three groups reflecting the biology and geographical distribution of the species. For example, the three boreal species (*Picea abies*, *Pinus sylvestris* and *Betula pendula*) belonged to the same group of species with very large distribution ranges across Eurasia while *Quercus petraea* and *Fagus sylvatica* formed another group with relatively smaller distributions and similar ice ages refugial locations. In animals, Wilder *et al.* (2023) showed that species with smaller historic effective population size had a higher probability of extinction than larger ones, even though it should be noted that ecological factors were better predictors of extinction risks. The dependency of current levels of nucleotide diversity on ancient demographic events captured by effective population size is not surprising as nucleotide diversity takes time to accumulate and, under a pure drift model, and assuming an effective population size of 10,000, it will take 4,000 generations for a new allele to reach a frequency of 0.1, given that it does (Kimura & Ohta, 1973; Slatkin & Rannala, 2000). The association with the current census number is, in this respect, more surprising, and may suggest that the census population size did not change drastically recently. In any case, our study has explicitly pointed out the importance of historical demography for the maintenance of genetic diversity.

Fluctuating, yet higher, Ne in estuarine than in marine environments

Freshwater dioecious species experienced stronger natural selection than marine dioecious ones. One possible cause could be that freshwater environments are more variable than marine ones. Although intertidal marine species have to deal with stronger water currents, higher salinity and temporary drought, freshwater species basically live in a more or less terrestrial environment and have to cope with frost, drought, UV damage, heavy metal contamination,

384 various pathogens and herbivores and a host of other stresses (Wassens *et al.*, 2017; Singh *et*
385 *al.*, 2023; Daumal *et al.*, 2024). To start to assess whether a more variable environment was a
386 plausible cause of more intense selection, we retrieved annual temperature experienced by each
387 species across its distribution range during the 1960-2021 period (**Fig. 6**). Though the
388 maximum temperature varied where tropical species live under a higher temperature than
389 temperate ones, the maximum temperature fluctuation was much higher in freshwater than in
390 marine environment. When calculating temperature fluctuation, we limited the lowest
391 temperature to zero for all species, which is rather reasonable for marine species as they are
392 submerged. However, for freshwater species living in temperate regions, frost could be a major
393 challenge in the winter while long-term drought and extremely high temperatures are a major
394 threat in the summer. Previous studies have shown that temporal fluctuating selection on
395 multiple genes can have a pervasive effect on the maintenance of genetic variation across
396 genomes of many species (Rees & Ellner, 2019; Kelly, 2022; Johnson *et al.*, 2023; Strickland
397 *et al.*, 2024).
398
399 On the other hand, we also found lower K_a/K_s ratio in in the freshwater species for all three
400 sexual systems after correction for phylogeny. This could also be explained by the difference
401 of environmental stability between marine and freshwater habitats. Many of the selected genes
402 were functionally related to resistance to drought, heat, or disease and with rather drastic gene
403 expression differentiation between the freshwater and marine species. However, bias towards
404 high expression genes may exist by using transcriptomic data although the distributions of
405 K_a/K_s did not show any strong correlation with the expression level.
406
407 ***The possible effects of sex ratio and sex biased fitness***

408 The main factor contributing to differences in genetic diversity between species in the present
409 study is the habitat. But once this is accounted for, the second most important factor is the
410 reproductive system, with dioecy standing apart from monoecious and hermaphrodite species.
411 More specifically, dioecious species have a very low diversity in marine environment, but a
412 similar level to hermaphrodite species in freshwater environment. Studying 15 Silene species,
413 Muyle *et al.* (2021) observed that dioecious species were more genetically diverse than Silene
414 species with other sexual systems. They attributed this difference to an Allee effect in dioecious
415 species. Dioecious species cannot decline below a given threshold without risking declining to
416 extinction. Hence, conditioning on survival, this Allee effect contributes to the maintenance of
417 genetic diversity in dioecious species. It is unclear why such a mechanism would operate in
418 freshwater and not in marine environment. Instead, the observed collapse of genetic diversity
419 for dioecious species living in marine environment could be due to factors associated to
420 pollination or to an increase in sex ratio bias. Biased sex ratios have indeed been observed in
421 seagrasses from intertidal and shallow subtidal environments (Williams, 1995; Shelton, 2008;
422 Shelton, 2010; Buckel *et al.*, 2012; Nguyen *et al.*, 2018). For example, Shelton (2008, 2010)
423 investigated the marine seagrasses *Phyllospadix scouleri* and *P. serrulatus* (surf grasses),
424 which have some of the most extreme female-biased sex ratios documented (>90% female).
425 In these two species, sex-ratios were driven by male-biased mortality during the life cycle. It
426 was speculated that the higher mortality in males than in females could possibly be related to
427 (i) a weaker attachment of male plants to their substrate or other cause of differential mortality
428 or (ii) the presence of young sex chromosomes in these species and an exposure of deleterious
429 mutations in the heterochromatic sex. Without further data it is hard to tell apart these different
430 hypotheses but we note that, in our study, marine dioecious species did have a much higher
431 π_0/π_4 ratio than freshwater ones (0.40 vs 0.20). In most seagrass species, the presence of sexual
432 chromosomes has not been ascertained and the primary sex ratio is not known. It is therefore

433 hard to know whether one would expect environment to contribute to the bias and the observed
434 difference between freshwater and marine environment. Pollen limitation could also contribute
435 to a smaller effective population size and to low genetic diversity. While sex ratio was also
436 strongly female-biased in *Phyllospadix torreyi* (Williams, 1995), this did not seem to have led
437 to pollen limitation. However, a recent literature survey suggests that pollen limitation may be
438 a common Allee effect in marine hydrophylous seagrasses, though not in freshwater ones (Van
439 Tussenbroek *et al.*, 2016). Seed- or fruit-set of marine hydrophilous plants was observed to be
440 generally lower than in hydrophilous freshwater species. Whether this is also the case in the
441 species studied here remains to be seen, but even though it should be pointed out that in some
442 seagrass species both sex-ratio and pollen limitation have been shown to vary across
443 populations (Nguyen *et al.*, 2018), this certainly is an interesting aspect to pursue.
444
445 ***The Nearly Neutral Theory holds in seagrasses***
446 One of the predictions of the Nearly Neutral Theory is that the log values of the efficacy of
447 selection (π_0/π_4) and effective population size should be correlated and the slope of this
448 regression should be equal to the shape parameter of the DFE, β (Welch *et al.*, 2008). In the
449 present study, the absolute value of the slope r (0.21) was indeed very close to the mean
450 estimate of β (0.206). As noted above, the Nearly Neutral Theory held quite well in previous
451 tests based on a diverse array of species, whether those were based on population data (Chen
452 *et al.*, 2017; Chen *et al.*, 2020) or on phylogenetic data (Bastian *et al.*, 2025) but less well for
453 within genus comparisons when species were biologically quite similar (Mackintosh *et al.*
454 (2019) compared 38 species of European butterflies (Papilionoidea) with identical mating
455 systems and environments, at least relative to the diversity observed among seagrasses). In the
456 present case, differences in sexual system and environment are large enough to generate
457 significant differences in the DFE across species and allow one to estimate the slope of the

458 relationship between π_0/π_4 and N_e. Chen et al. (2017) used both plants and animals and
459 Bastian et al. (2025) focused on 150 placental mammals. The present study provides a further
460 test of the Nearly Neutral Theory, in seagrasses this time, and therefore extend the generality
461 of the mechanism proposed by Ohta to explain levels of genetic diversity across species. As
462 pointed by Bastian et al. (2025), the good agreement of the data with the nearly neutral theory
463 suggests that the effective population size does capture the effect of the numerous variables
464 that impact genetic variation and the efficacy of selection.
465

Material & Methods

Plant material sampling

In total, 127 individuals from 27 species of the order Alismatales were included in this study. All samples were collected in China and the number of individuals per species varied from three to 13 (**Table S1**). In order to avoid collecting clones, we kept a distance over 500 meters between samples for marine species. For freshwater species we collected one individual sample per location if possible or otherwise collected multiple samples separated by a distance over 100 meters. Fresh leaves were sampled and immediately frozen in liquid nitrogen and subsequently stored at −80 °C before sequencing.

PacBio full-length transcriptomic sequencing and assembly

For 21 species, full-length isoform sequencing using PacBio SMRT technology was applied to generate reference transcriptomes (see **Table S2**). Frozen leaves were ground and total RNA was extracted in TRIzol-CTAB-LiCl reagent on dry ice following the protocol provided by the manufacturer. The integrity of the RNA was determined with the Agilent 2100 Bioanalyzer (Agilent Technologies) and agarose gel electrophoresis. The purity and concentration of the RNA were determined with the Nanodrop (Thermo Fisher Scientific) and Qubit (Thermo Fisher Scientific). Only high-quality RNA sample (OD260/280=1.8~2.2, OD260/230≥2.0, RIN≥8, >1 μg) was used to construct sequencing library. About 300ng total RNA was reverse transcribed into cDNA and amplified by using NEBNext® Single Cell/Low Input cDNA Synthesis& Amplification Module and Iso-Seq Express Oligo Kit. cDNAs were purified by ProNex Beads and were used to construct the library by SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences) with damage repaired, end repaired, A-tailling and ligated to sequencing adapters. The SMRTbell template was annealed to sequencing primer and bound

to polymerase and sequenced on the PacBio Sequel II platform using Sequel II Binding Kit 2.0 (Pacific Biosciences).

Raw reads were processed through the isoseq3 pipeline (https://github.com/pacificbiosciences/isoseq/, accessed 2025/04/09). Circular consensus sequences were generated therefrom. Subsequently, these sequences were classified into full-length and non-full-length reads by inspecting the poly(A) signal, as well as 5' and 3' adaptors. Consensus isoforms were identified from full length non-chimera sequences (FLNCs) and refined using non-full-length reads to yield high quality isoforms (HQ, with predicted accuracy >0.99). The Illumina RNAseq data from the same individual were employed to correct the PacBio sequences, with the correction carried out in lordec (http://atgc.lirmm.fr/lordec/, accessed 2025/04/09) (Salmela & Rivals, 2014). The HQ full-length polished consensus transcripts were then processed using cogent v.8.0.0 (https://github.com/Magdoll/Cogent, accessed 2025/04/09) to identify gene families and reconstruct the coding genome in the absence of a reference genome. Finally, TransDecoder v.5.5.0 (Haas *et al.*, 2013) was employed to generate a final dataset of full-length transcriptomes for 21 species.

For four species, *Colocasia esculenta*, *Posidonia australis*, *Spirodela polyrhiza*, and *Zostera muelleri*, published reference genomes were downloaded from NCBI (**Table S1**).

RNAseq transcriptomic sequencing and SNP calling

Of the 127 samples, 39 were RNAseq reads accessed from NCBI SRA, and 88 were Illumina RNAseq reads (2×150 bp) generated in this study (**Table S1**). RNA libraries were constructed

using the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) in accordance with the manufacturer's guidelines. Sequencing was performed on an Illumina Novaseq 6000 platform, generating 150bp paired end reads (Novogene, Tianjin, China).

Quality control of raw RNAseq reads was carried out using Fastp v.0.23.1 (Chen *et al.*, 2018) with parameters '-g -q 5 -u 50 -n 15 -l 150 --min_trim_length 10 --overlap_diff_limit 1 --overlap_diff_percent_limit 10'. Then an in-house Perl script was used to filter out the reads with extremely high GC-content that could be contaminated by algeas. We employed hisat2 v.2.2.1 (Kim *et al.*, 2019) with default settings to align the clean RNAseq reads of each individual to the reference genome or the full-length transcriptomes of their own respective species. It is noteworthy that the short reads of *Aponogeton lakhonensis* were aligned to the full-length transcriptome of *Aponogeton crispus*, and the short reads of *Potamogeton crispus* were aligned to the full-length transcriptome of *Potamogeton distinctus*. Subsequently, Samtools v.2.2.1 (Danecek *et al.*, 2021) was utilized to convert Sequence Alignment/Map (SAM) files into Binary Alignment/Map (BAM) files, sort the BAM files, and eliminate duplicate reads. Only the properly paired reads aligned to a unique position were retained.

Individual genotypes were generated using HaplotypeCaller and merged with GenomicsDBImport by means of GATK v.4.5.0.0 (McKenna *et al.*, 2010). Finally, genotypes were determined using GenotypeGVCFs for both variant and non-variant sites. Hard-filtering was applied to variant sites based on the genotype summary statistics distributions of 'QUAL < 30', 'QD <10', 'ExcessHet > 3' of each species. Variant and non-variant sites containing any missing genotypes were also filtered out.

Orthologue identification

Orthofinder v.2.5.5 (Emms & Kelly, 2019) was applied to search for orthologous groups (OGs) using the longest isoforms in 25 species (4 genomes and 21 full-length transcriptomes) and in total 21,078 OGs were obtained. Pairwise similarity search using BLASTP (Altschul *et al.*, 1990) was performed between all species pairs within each OG and only the best transcripts with e-value < 1e-10 were retained. In total, 2,079 orthologous transcripts were retrieved. For *Aponogeton lakhonensis*, *Potamogeton crispus* and all individual samples of RNAseq dataset, orthologous transcripts were generated by replacing the corresponding reference sites with species specific genotypes, respectively. Protein sequences were then translated from orthologous transcripts for each individual and used for the phylogenetic inference.

Inference of phylogenetic tree of Alismatales

The phylogenetic tree of 27 species in Alismatales was inferred using both concatenation and coalescent approaches. For the concatenation method, MAFFT v.7 (Katoh *et al.*, 2002) and trimal v.1.4 (Capella-Gutierrez *et al.*, 2009) were used to align the 2,079 orthologous protein coding sequences and remove gaps with the parameter setting '-gt 0.2 -st 0.001 -cons 85'. Subsequently, a supermatrix and a partition file were generated using the R package 'phylotools' (Revell, 2024). The phylogenetic tree was inferred with the Maximum-likelihood (ML) method implemented in IQTREE v.2.1.4 (Minh *et al.*, 2020) with 1000 bootstrap replicates to assess the reliability of the tree topology. The best nucleotide substitution model was chosen based on BIC value of each model and was carried out automatically by ModelFinder with the parameter '-m MFP'. The coalescent species tree was inferred with the following two steps: 1) 1,390 orthologous gene that were present in at least three species were used to generate individual gene trees by IQ-TREE; 2) a species tree was then reconstructed

based on all gene trees using the coalescent method implemented in ASTRAL v.5.7.8 (Zhang *et al.*, 2018). Similarly, the phylogenetic trees based on 127 individuals were also generated based on both methods using the ML and coalescent methods. Two outgroup species (*Colocasia esculenta* and *Spirodela polyrhiza*) were used to root the phylogenetic trees.

Divergence time estimation

The divergence time of each node in the concatenated ML phylogenetic tree was estimated using the MCMCTree program implemented in the software PAML v.4.10.7 (Yang, 2007), which uses an approximate likelihood calculation to improve the computation speed (dos Reis & Yang, 2011). Four fossil calibration points (**Table S3**) were used to recalibrated the molecular clock: (1) 33.9-56 Mya for the Crown for Potamogetonaceae and Zosteraceae; (2) 81.0-82.0 Mya for the crown of Potamogetonaceae, Zosteraceae, Posidoniaceae, Ruppiaceae, Cymodoceaceae and Aponogetonaceae; (3) 33.9-56 Mya for the crown for Hydrilla and Vallisneria; (4) 66.0-100.5 Mya for the crown for Araceae.

The ML species tree was set as the topological constraint and the parameter for the molecular clock was set to 2 (this allows for independent rates in each branch). The MCMC simulations were run twice, the first 200,000 iterations were discarded as burn-in and every 10 iterations were sampled until a total of 200,000 samples were collected. The recalibrated species tree was visualized in TVBOT v.2.6.1 (Xie *et al.*, 2023) with the divergence time in 95% highest posterior density (HPD) intervals for all nodes.

Population genetic analyses

586 To avoid the biases caused by possible paralogs, we performed all population genetic analyses
587 on the longest transcript for each orthologous group of each species and transcripts with a
588 length short than 500bp were excluded. Pairwise genetic difference (π) was calculated at 0-
589 fold and 4-fold sites, a ratio of π_0/π_4 and GC content at the third position of codons were also
590 calculated for each transcript using in-house perl and R scripts. Each 0-fold and 4-fold sites
591 were randomly divided into two groups, one for calculation of π and the other for calculation
592 of π_0/π_4 ratio. We repeated this step for 1,000 iterations and took the average values of π_4, π_0
593 and π_0/π_4 for each species. We also estimated π and π_0/π_4 values using the above protocols
594 with 2,079 orthologous genes that were used to infer phylogenetic tree.
595
596 To polarize the site frequency spectrum, the two closest species were chosen as outgroups.
597 Orthologous sequences were first aligned as protein sequences and then amino acids were
598 translated back into nucleotide codons. The inference of ancestral state was performed using
599 'est-sfs' (Keightley & Jackson, 2018). The distribution of fitness effects of new mutations
600 (DFE) was estimated for each species based on unfolded SFS using polyDFE v.2.0. (Tataru &
601 Bataillon, 2019). The BFGS method was chosen for parameter optimization and parameters of
602 basin-hopping algorithm was adjusted to maximize the log likelihood and minimize the
603 gradient for each run. DFE parameters were estimated and include ancestral state error (eps_an),
604 selective coefficients for deleterious and beneficial mutations (S_d, S_b), shape parameter (β),
605 and proportions of beneficial mutations (p_b). Numbers of fixed sites with outgroups at 4-fold
606 and 0-fold sites were provided to estimate the proportion of fixed adaptive mutations (α) with
607 $S_b > 1$.
608
609 Stairway plot 2 was applied to estimate historical N_e changes for each species based on
610 unfolded SFS. For perennial species, we set the mutation rate (μ) to be 2e-8 per site per

generation and generate time (g) to be 10 years. *Najas marina* and *N. minor* were recognized as annual (https://plantatlas2020.org/, accessed 2025/04/09) (Les *et al.*, 2015) and both annual and perennial populations were reported in *Ottelia cordata* (Huang *et al.*, 2020; Shen *et al.*, 2021). For these three species, Stairway plot 2 was also run with μ = 2e-9 per site per generation and g = 1 year. To estimate harmonic mean of historical N_e, we filtered out Stairway plot 2 result with time < 100 years or > 1e6 years. Time (t) and $N_{(t)}$ estimates were taken for the harmonic mean calculation following the equation:

$$\frac{1}{N_{mean}} = \frac{1}{T_{total}} \sum_{i=1}^{n} \frac{t_i}{N_i}$$

Where t_i is the length of ith time tile; N_i is the effective population size over that time tile; T_{total} is the total length of time tiles. To capture the major changes of N_e, time tiles were merged if the fold change of N_e was less than two and an arithmetic mean N_e was taken over those tiles before the harmonic mean N_e was calculated across all time tiles.

The inbreeding coefficient F_{IS} was calculated for each individual and averaged for each species using Plink v.1.9 (Chang *et al.*, 2015) based on observed and expected homozygous genotypes. VCFtools v.4.0 (Danecek *et al.*, 2011) was used to calculate Tajima's *D* and Bcftools v.1.16 (Danecek *et al.*, 2011) was used to calculate runs of homozygosity (ROH) for each transcript. Finally, a random forest sequential model implemented in ROH-selfing was run to estimate selfing rate for monoecious and hermaphroditic species based on ROH, Tajima's *D* and F_{IS} coefficient following the authors' recommendation (Zeitler & Gilbert, 2024).

Life history traits, distribution ranges, census sizes and temperature data collection

For each species, we gathered data on three continuous and four discrete life history traits. Specifically, information on plant height, maximum leaf width, maximum leaf length and life cycles was sourced from databases FRPS (http://www.iplant.cn/frps, accessed 2025/04/09),

China Wetland Plant Database (https://zgsdzw.com/, accessed 2025/04/09) and Tropicos (https://tropicos.org/home, accessed 2025/04/09), and plant habitat, sex and pollination were collected from the research of Les *et al.* (1997).

The coordinates of observed occurrence of each species were retrieved from the database of Global Biodiversity Information Facility (GBIF, https://www.gbif.org/, accessed 2025/04/09). Only the occurrence of GBIF data based on human observation was kept. The coordinates of each record were projected to a grid of 1° × 1° in latitude and longitude and the distribution range of each species was calculated by counting the total number of grids covered. The proxy of census size was calculated by dividing the distribution range by the maximum leaf length following the method of Roberts and Josephs (2025).

The average maximum and minimum temperatures (°C) for each month from 1960 to 2021 were downloaded from WorldClim (Fick & Hijmans, 2017). The minimum water temperature was set to zero even if the recorded air temperature was far below zero. The mean and standard error of maximum and minimum temperatures of one species was taken by averaging across all 1° × 1° grids of its distribution range. The annual maximum temperature fluctuation of one species was obtained by subtracting the annual minimum temperature from the annual maximum temperature and averaging across its distribution range.

Statistical analyses

All statistical analyses were carried out in R. For the regression tests corrected for phylogenetic sampling bias, a generalized linear mixed model with phylogeny as the random effect was applied using the r function 'pglmm_compare' implemented in the package 'phyr' (Ives *et al.*, 2023). Model comparisons and R2 calculation were carried out using r package 'rr2' (Ives,

2018). We excluded *Aponogeton crispus*, *A. lakhonensis*, and *Ottelia cordata* for the estimation of distribution ranges and census sizes due to too few occurrences. *Spirodela polyrhiza* was only used for phylogenetic inference but excluded for all other analyses due to its extremely high homozygosity and low genetic diversity (π_4 = 6.35e-5), which suggests that its RNAseq data from public database may not be sampled from the natural population.

Estimates of evolutionary rate K_a/K_s

We used the maximum likelihood method to estimate the evolutionary rate for each species ($\omega = K_a/K_s$) as implemented in CodeML program included PAML v.4.10.7 (Yang, 2007). 2,079 OGs were translated into protein sequences before alignment and then reverse translated back to CDS sequences using ParaAT (Zhang *et al.*, 2012). To compare the evolutionary rate of each group of species, we first established a null model with one ω value for all species (M_0). Three hierarchical models were also established: 1) a constant ratio model (M_0) with one ω_0 value for all branches. 2) a three-ratio model (M_3) with one ω_{in} value for all internal branches, and one ω_{ma} for all external branches leading to marine species, one ω_{fw} for all external branches to freshwater species; and 3) a free-ratio model that allows ω to vary freely for every branch (M_{free}). To access ω values between groups of species for each gene, we compared M_3, and M_{free} models to M_0 and calculated the significance of rejecting M_0 using the log-likelihood ratio test. False discovery rate (FDR) was performed to correct for multiple testing.

Gene copy number and expression differentiation

For each OG, we counted the number of gene models and compared the presence or absence and the copy number difference between species groups using PhyloGLM test (Lai *et al.*, 2023) followed the method described in Levy et al (Levy *et al.*, 2017).

RNAseq short reads aligned to the longest gene model of each OG were counted and normalized as TPM for all 127 individuals. R function 'pglmm_compare' was used to compare gene expression level and calculate p-value between pairs of species groups with the individual phylogenetic tree taken as the random effect. And an FDR-corrected *p*-value threshold of 0.05 and fold change larger than two was applied to select differentiation expressed genes between species groups of different habitats, sexual systems, and pollination modes, respectively.

Data availability

All transcriptomic sequencing data used in this study have been deposited in NCBI SRA database and the Genome Warehouse in National Genomic Data Center with the project ID PRJCA050827. All accession numbers can be found in **Table S1**.

Code availability

All codes used for the main analyses in this study are accessible from https://github.com/cjevol/seagrass.

Author contributions: J.C. designed the research; P.L., X.J., J., R.R., F. and X.Y., T. collected samples; X.Y, T. and J.C. performed data analyses. X.Y, T., M.L. and J.C wrote the paper.

Acknowledgements

This project was supported by National Natural Science Foundation of China (32371689), the National Natural Science Foundation of China (32570239), the Research Funds for the State Key Laboratory for Vegetation Structure, Function and Construction (VegLab2025003), and the Key R&D Program of Zhejiang (2024C03244) granted to J.C. We thank Mingyue Ye, Ying Liu, Yuxiang, Zhu from Zhejiang University, Hong Zhao from Shandong University, Langxing Yuan from Chinese Academy of Tropical Agricultural Sciences, Xiaowei Xin from Shandong Drug And Food Vocational College, for helping with sample collection.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.

Bastian M, Enard D, Lartillot N. 2025. Empirical validation of the nearly neutral theory at divergence and population genomic scale using 150 mammals genomes. *bioRxiv*: 2025.2004.2017.649326.

Buckel CA, Blanchette CA, Warner RR, Gaines SD. 2012. Where a male is hard to find consequences of male rarity in the surfgrass Phyllospadix torreyi. *Marine Ecology Progress Series* **449**: 121-132.

Buffalo V. 2021. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. *eLife* **10**: e67509.

Caballero A. 2020. *Quantitative Genetics*. Cambridge: Cambridge University Press.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15): 1972-1973.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**(1): s13742-13015-10047-13748.

Charlesworth B, Jensen JD. 2022. How Can We Resolve Lewontin's Paradox? *Genome Biology and Evolution* **14**(7).

Chen J, Glemin S, Lascoux M. 2017. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution* **34**(6): 1417-1428.

Chen J, Glemin S, Lascoux M. 2020. From Drift to Draft: How Much Do Beneficial Mutations Actually Contribute to Predictions of Ohta's Slightly Deleterious Model of Molecular Evolution? *Genetics* **214**(4): 1005-1018.

Chen L-Y, Lu B, Morales-Briones DF, Moody ML, Liu F, Hu G-W, Huang C-H, Chen J-M, Wang Q-F. 2022. Phylogenomic Analyses of Alismatales Shed Light into Adaptations to Aquatic Environments. *Molecular Biology and Evolution* **39**(5).

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17): i884-i890.

Clark A, Koc G, Eyre-Walker Y, Eyre-Walker A. 2023. What Determines Levels of Mitochondrial Genetic Diversity in Birds? *Genome Biology and Evolution* **15**(5).

Collin VC, Eymery F, Genty B, Rey P, Havaux M. 2008. Vitamin E is essential for the tolerance of Arabidopsis thaliana to metal-induced oxidative stress. *Plant, Cell & Environment* **31**(2): 244-257.

Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *Plos Biology* **13**(4): e1002112.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156-2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2).

Daumal MM, Oosterhuis D, Verhofstad MJJM, Erkens RHJ, Peeters ETHM. 2024. The impact of drought duration on two Potamogeton species with different growth forms. *Aquatic Sciences* **86**(3): 73.

Den Hartog C. 1970. *The seagrasses of the world*. Amsterdams: North-Holland Publishing Company.

Döring M. 2022. *English Wikipedia - Species Pages*: Wikimedia Foundation.

dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**(7): 2161-2172.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nature Reviews Genetics* **17**(7): 422-433.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**(1): 238.
Fick S, Hijmans R. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**.
Gomez B, Daviero-Gomez V, Coiffard C, Martín-Closas C, Dilcher DL. 2015. Montsechia, an ancient aquatic angiosperm. *Proceedings of the National Academy of Sciences* **112**(35): 10985-10988.
Graham LE, Cook ME, Busse JS. 2000. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci U S A* **97**(9): 4535-4540.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**(8): 1494-1512.
Hasanuzzaman M, Nahar K, Fujita M 2014. Chapter 12 - Role of Tocopherol (Vitamin E) in Plants: Abiotic Stress Tolerance and Beyond. In: Ahmad P, Rasool S eds. *Emerging Technologies and Management of Crop Stress Tolerance*. San Diego: Academic Press, 267-289.
Haynes RR. 1979. Revision of North and Central American Najas (Najadaceae). *SIDA, Contributions to Botany* **8**(1): 34-56.
Huang W, Han S, Xing Z, Li W. 2020. Responses of leaf anatomy and CO2 concentrating mechanisms of the aquatic plant Ottelia cordata to variable CO2. *Frontiers in Plant Science* **11**.
Ito Y, Tanaka N, Gale SW, Yano O, Li J. 2017. Phylogeny of Najas (Hydrocharitaceae) revisited: Implications for systematics and evolution. *TAXON* **66**(2): 309-323.
Ives A, Dinnage R, Nell LA, Matthew H, Li D. 2023. phyr: Model Based Phylogenetic Analysis. *https://daijiang.github.io/phyr/*.
Ives AR. 2018. R^{2}s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs. *Systematic Biology* **68**(2): 234-251.
Johnson OL, Tobler R, Schmidt JM, Huber CD. 2023. Fluctuating selection and the determinants of genetic variation. *Trends in Genetics* **39**(6): 491-504.
Kaplan Z, Jarolímová V, Fehrer J. 2013. Revision of chromosome numbers of Potamogetonaceae: a new basis for taxonomic and evolutionary implications. *Preslia* **85**(4): 421–482.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**(14): 3059-3066.
Keightley PD, Jackson BC. 2018. Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. *Genetics* **209**(3): 897-906.
Kelly JK. 2022. The genomic scale of fluctuating selection in a natural plant population. *Evolution Letters* **6**(6): 506-521.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**(8): 907-915.
Kimura M, Ohta T. 1973. The age of a neutral mutant persisting in a finite population. *Genetics* **75**(1): 199-212.
Lai J, Zhu W, Cui D, Mao L. 2023. Extension of the glmm.hp package to zero-inflated generalized linear mixed models and multiple regression. *Journal of Plant Ecology* **16**(6): rtad038.

Larkum A, Orth R, Duarte C. 2006. *Seagrasses: Biology, Ecology and Conservation*. The Netherlands: Springer.

Les DH, Cleland MA, Waycott M. 1997. Phylogenetic Studies in Alismatidae, II: Evolution of Marine Angiosperms (Seagrasses) and Hydrophily. *Systematic Botany* **22**(3): 443-463.

Les DH, Peredo EL, Tippery NP, Benoit LK, Razifard H, King UM, Na HR, Choi H-K, Chen L, Shannon RK, et al. 2015. Najas minor (Hydrocharitaceae) in North America: A reappraisal. *Aquatic Botany* **126**: 60-72.

Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al. 2017. Genomic features of bacterial adaptation to plants. *Nat Genet* **50**(1): 138-150.

Lynch TJ, Erickson McNally BJ, Losic T, Lindquist J, Finkelstein R. 2026. ABI5-binding proteins are substrates of key components in the ABA core signaling pathway affecting seeds. *Plant Physiology* **200**(1): kiaf674.

Ma X, Vanneste S, Chang J, Ambrosino L, Barry K, Bayer T, Bobrov AA, Boston L, Campbell JE, Chen H, et al. 2024. Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nature Plants* **10**(2): 240-255.

Mackintosh A, Laetsch DR, Hayward A, Charlesworth B, Waterfall M, Vila R, Lohse K. 2019. The determinants of genetic diversity in butterflies. *Nature Communications* **10**(1): 3466.

Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics* **19**(6): 330-338.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**(9): 1297-1303.

Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, Cavers S, Fady B, Faivre-Rampant P, Gonzalez-Martinez SC, et al. 2024. Resilience of genetic diversity in forest trees over the Quaternary. *Nature Communications* **15**(1): 8538.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**(5): 1530-1534.

Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, Silva A, Bataillon T, Widmer A, Glémin S, et al. 2021. Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus Silene. *Molecular Biology and Evolution* **38**(3): 805-818.

Nguyen HM, Kleitou P, Kletou D, Sapir Y, Winters G. 2018. Differences in flowering sex ratios between native and invasive populations of the seagrass Halophila stipulacea. *Botanica Marina* **61**: 337-342.

Niu Y, Zhang Q, Wang J, Li Y, Wang X, Bao Y. 2022. Vitamin E synthesis and response in plants. *Front Plant Sci* **13**: 994058.

Ohta T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. *Nature* **246**(5428): 96-98.

Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology, Evolution, and Systematics* **23**(Volume 23): 263-286.

Ohta T, Gillespie JH. 1996. Development of Neutral and Nearly Neutral Theories. *Theoretical Population Biology* **49**(2): 128-142.

Olsen JL, Rouze P, Verhelst B, Lin YC, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F, et al. 2016. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**(7590): 331-335.

Rees M, Ellner SP. 2019. Why So Variable: Can Genetic Variance in Flowering Thresholds Be Maintained by Fluctuating Selection? *The American Naturalist* **194**(1): E13-E29.
Revell LJ. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**: e16505.
Roberts MD, Josephs EB. 2025. k-mer-based diversity scales with population size proxies more than nucleotide diversity in a meta-analysis of 98 plant species. *Evolution Letters* **9**(4): 434-445.
Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**(7526): 261-263.
Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**(24): 3506-3514.
Shelton AO. 2008. SKEWED SEX RATIOS, POLLEN LIMITATION, AND REPRODUCTIVE FAILURE IN THE DIOECIOUS SEAGRASS PHYLLOSPADIX. *Ecology* **89**(11): 3020-3029.
Shelton AO. 2010. The origin of female-biased sex ratios in intertidal seagrasses (Phyllospadix spp.). *Ecology* **91**(5): 1380-1390.
Shen Y, Lei J, Song X, Ren M. 2021. Annual Population Dynamics and Their Influencing Factors for an Endangered Submerged Macrophyte (Ottelia cordata). *Frontiers in Ecology and Evolution* **9**.
Singh V, Singh N, Rai SN, Kumar A, Singh AK, Singh MP, Sahoo A, Shekhar S, Vamanu E, Mishra V. 2023. Heavy Metal Contamination in the Aquatic Ecosystem: Toxicity and Its Remediation Using Eco-Friendly Approaches. *Toxics* **11**(2).
Slatkin M, Rannala B. 2000. Estimating allele age. *Annu Rev Genomics Hum Genet* **1**: 225-249.
Strickland K, Matthews B, Jónsson ZO, Kristjánsson BK, Phillips JS, Einarsson Á, Räsänen K. 2024. Microevolutionary change in wild stickleback: Using integrative time-series data to infer responses to selection. *Proceedings of the National Academy of Sciences* **121**(37): e2410324121.
Tataru P, Bataillon T. 2019. polyDFEv2.0: testing for invariance of the distribution of fitness effects within and across species. *Bioinformatics* **35**(16): 2868-2869.
Thiraviyam P, Aafreen, Kannan K, Ganapathy D, Sivaperumal P. 2025. Evaluation of antioxidant and antibacterial activities of seagrass Cymodocea serrulata. *Discover Plants* **2**(1): 68.
Van Tussenbroek BI, Soissons LM, Bouma TJ, Asmus R, Auby I, Brun FG, Cardoso PG, Desroy N, Fournier J, Ganthy F, et al. 2016. Pollen limitation may be a common Allee effect in marine hydrophilous plants: implications for decline and recovery in seagrasses. *Oecologia* **182**(2): 595-609.
Wassens S, Ning N, Hardwick L, Bino G, Maguire J. 2017. Long-term changes in freshwater aquatic plant communities following extreme drought. *Hydrobiologia* **799**(1): 233-247.
Welch JJ, Eyre-Walker A, Waxman D. 2008. Divergence and Polymorphism Under the Nearly Neutral Theory of Molecular Evolution. *Journal of Molecular Evolution* **67**(4): 418-426.
Wilder AP, Supple MA, Subramanian A, Mudide A, Swofford R, Serres-Armero A, Steiner C, Koepfli KP, Genereux DP, Karlsson EK, et al. 2023. The contribution of historical processes to contemporary extinction risk in placental mammals. *Science* **380**(6643): eabn5856.
Williams SL. 1995. Surfgrass (Phyllospadix Torreyi) Reproduction: Reproductive Phenology, Resource Allocation, and Male Rarity. *Ecology* **76**(6): 1953-1970.

Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. 2023. Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res* **51**(W1): W587-W592.

Yang Y, Karthikeyan A, Yin J, Jin T, Ren R, Fang F, Cai H, Liu M, Wang D, Li K, et al. 2022. The E3 Ligase GmPUB21 Negatively Regulates Drought and Salinity Stress Response in Soybean. *Int J Mol Sci* **23**(13).

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**(8): 1586-1591.

Yi SY, Nekrasov V, Ichimura K, Kang SY, Shirasu K. 2024. Plant U-box E3 ligases PUB20 and PUB21 negatively regulate pattern-triggered immunity in Arabidopsis. *Plant Mol Biol* **114**(1): 7.

Zeitler L, Gilbert KJ. 2024. Using Runs of Homozygosity and Machine Learning to Disentangle Sources of Inbreeding and Infer Self-Fertilization Rates. *Genome Biology and Evolution* **16**(7): evae139.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**(Suppl 6): 153.

Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun* **419**(4): 779-781.

Zhao W, Zhang QS, Tan Y, Liu Z, Ma MY, Wang MX, Luo CY. 2021. Photoprotective roles of ascorbate and PSII cyclic electron flow in the response of the seagrass Zostera marina to oxygen-evolving complex photoinactivation. *Photosynthetica* **59**(4): 600-605.

Tables

Table 1 The genetic diversity (π_4) and selection efficacy (π_0/π_4) for sexual system depending on the habitat and pollination.

$\pi_4(\times 10^{-3})$ / $\frac{\pi_0}{\pi_4}$	Habitat		Pollination	
sex	Freshwater	Marine	Hydrophilous	Aerial
Dioecious	4.23 (3.99, 4.48) / 0.201 (0.194, 0.208)	0.26 (0.22, 0.31) / 0.431 (0.339, 0.538)	2.67 (2.55, 2.8) / 0.237 (0.230, 0.248)	4.23 (3.99, 4.48) / 0.194 (0.182, 0.207)
Monoecious	2.28 (2.18, 2.38) / 0.283 (0.262, 0.399)	1.81 (1.7, 1.92) / 0.307 (0.290, 0.344)	1.83 (1.72, 1.94) / 0.307 (0.276, 0.335)	2.71 (2.61, 2.81) / 0.310 (0.282, 0.327)

Hermaphrodite	5.64 (5.42, 5.86) / 0.274 (0.244, 0.285)	2.18 (2.07, 2.3) / 0.407 (0.326, 0.493)	2.18 (2.07, 2.3) / 0.407 (0.326, 0.493)	5.64 (5.42, 5.86) / 0.274 (0.244, 0.285)

948 Note: 95 % confident intervals were shown in parentheses. Estimates of π_4 are per bp.

949

Table 2 The estimates of DFE parameters for marine and freshwater species.

habitat	β	S_b	$N_e s$ $[-\infty, -10]$	$N_e s$ $[-10, -1]$	$N_e s$ $[-1, 0]$	$N_e s$ $[0, 1]$	$N_e s$ $[1, \infty]$
Marine	0.11	0.174	0.534	0.0981	0.346	0.0226	7.18e-5
freshwater	0.153	16.34	0.661	0.0969	0.227	9.06e-4	8.42e-3
p-value	0.988	0.471	0.0172	0.373	0.00112	0.355	0.00185

Table 3 The median values of inbreeding coefficient, GC3 content, census size proxy, and historical N_e for marine and freshwater species of three sexual types

Marine / freshwater	F_{IS}	Selfing ratio	GC3	N_c	N_e (10^3)
Dioecious	0.0287 / -0.217	n.a. / n.a	0.45 / 0.458	0.273 / 13.8	3.14 / 18.6
Monoecious	-0.396 / 0.185	0.183 / 0.355	0.386 / 0.452	0.818 / 1.4	11.3 / 16.2
Hermaphrodite	-0.159 / -0.331	0.0478 / 0.0697	0.429 / 0.414	2.65 / 1.04	10.8 / 17.8

n.a.: not applicable

Figures

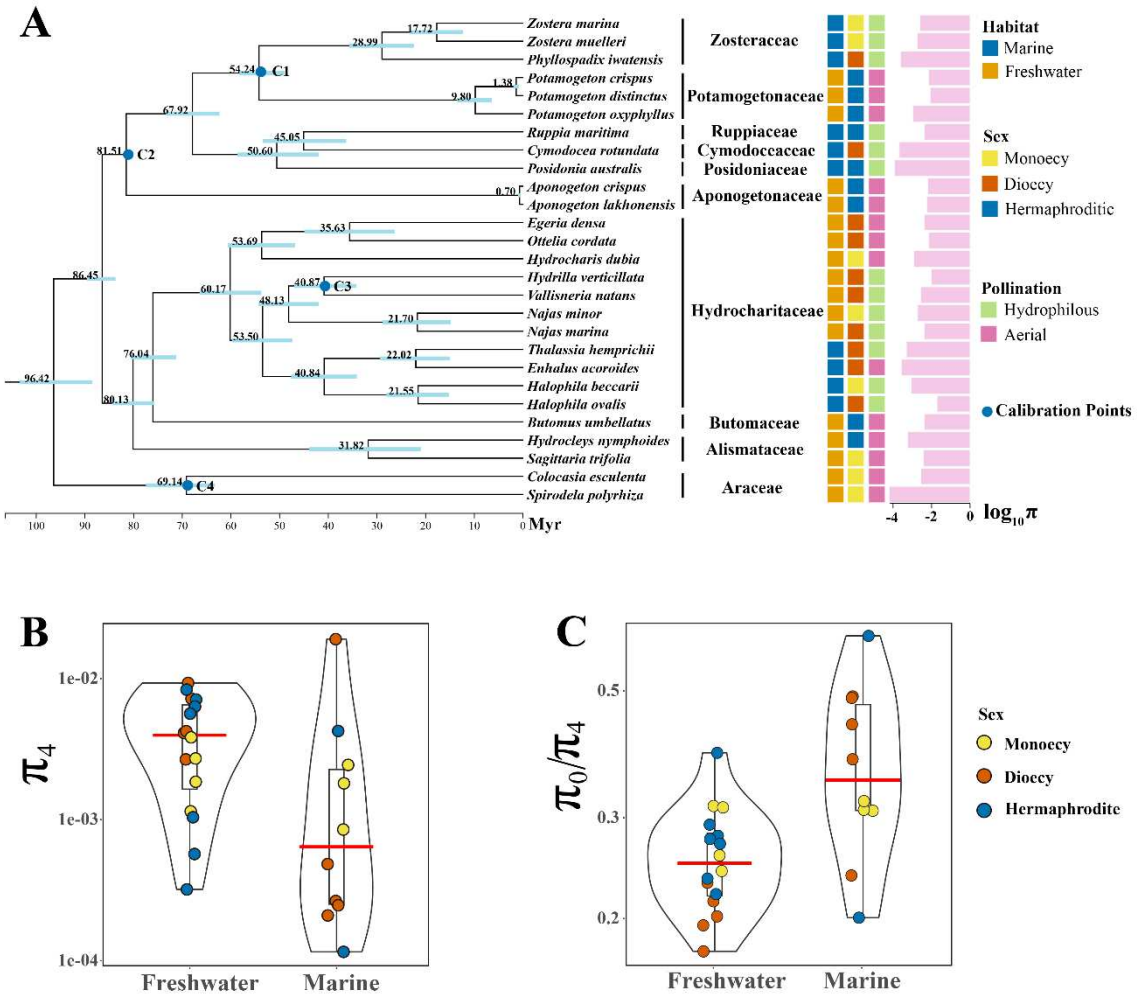


Fig. 1 The phylogenetic tree of 27 Alismatales species and difference between marine and freshwater species in genetic diversity. (**A**) The phylogenetic tree of 27 Alismatales species inferred based on 2,079 orthologous gene families. Blue bars give the 95% HD of divergence time and the figures above the bars are mean estimates of divergence time. The axis below the tree is the time in millions of year BP. (**B**) The distributions of genetic diversity π_4 and (**C**) efficacy of selection, π_0/π_4 ratio between freshwater and marine species. Each dot represents the mean value for a species and the red line is the median and the box is the interquartile range.

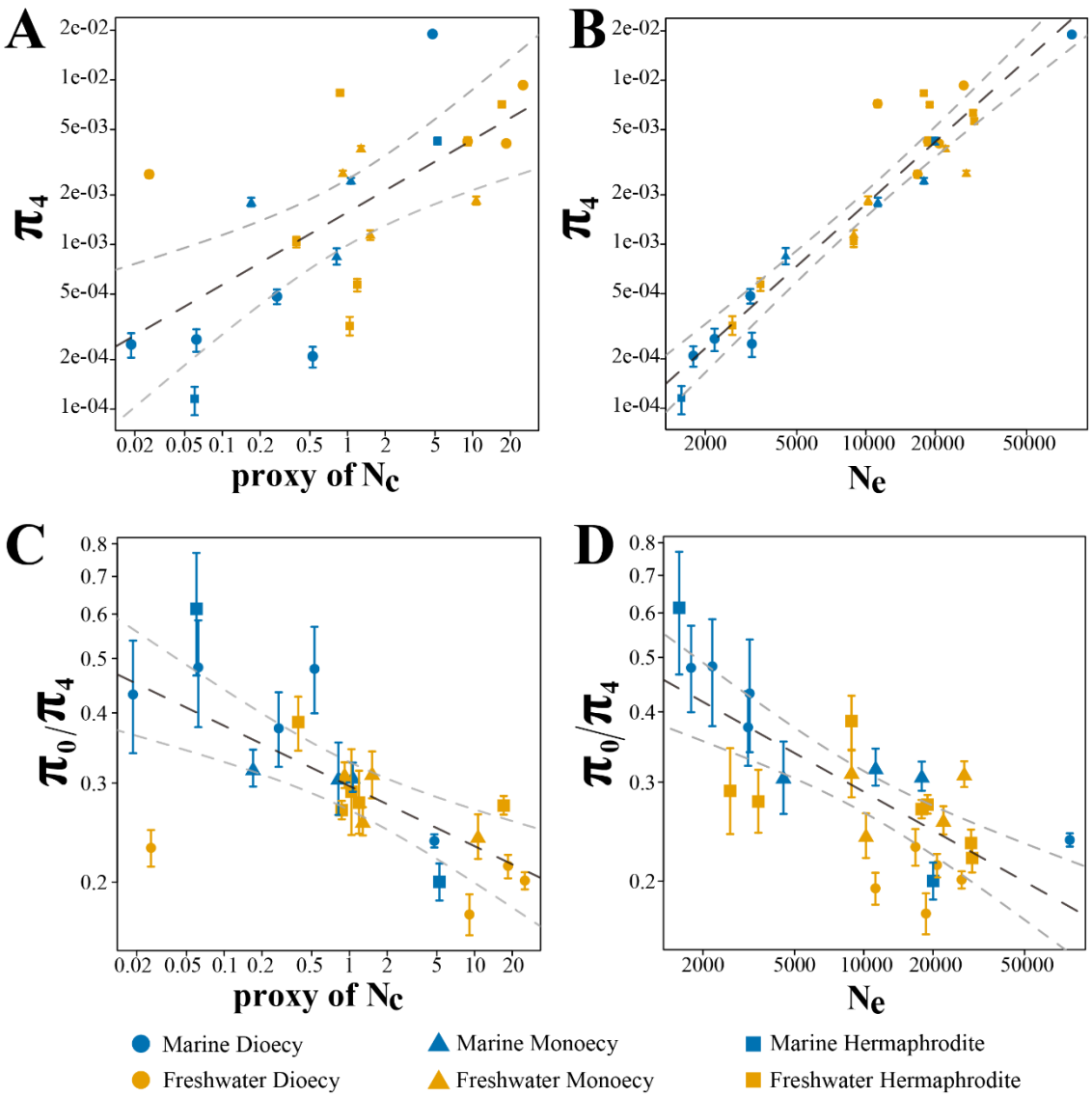


Fig. 2 Correlation between genetic diversity, the selection efficacy and census size or historical population size. (**A**) The log-log linear regression of genetic diversity at 4-fold site (π_4) on the proxy of census size (N_c); (**B**) The regression of π_4 on historical population size (N_e); (**C**) The log-log linear regression of selection efficacy (π_0/π_4) on the proxy of N_c; (**D**) the regression of π_0/π_4 on N_e. The bars represented 95% confidence intervals and the dots represented the medians.

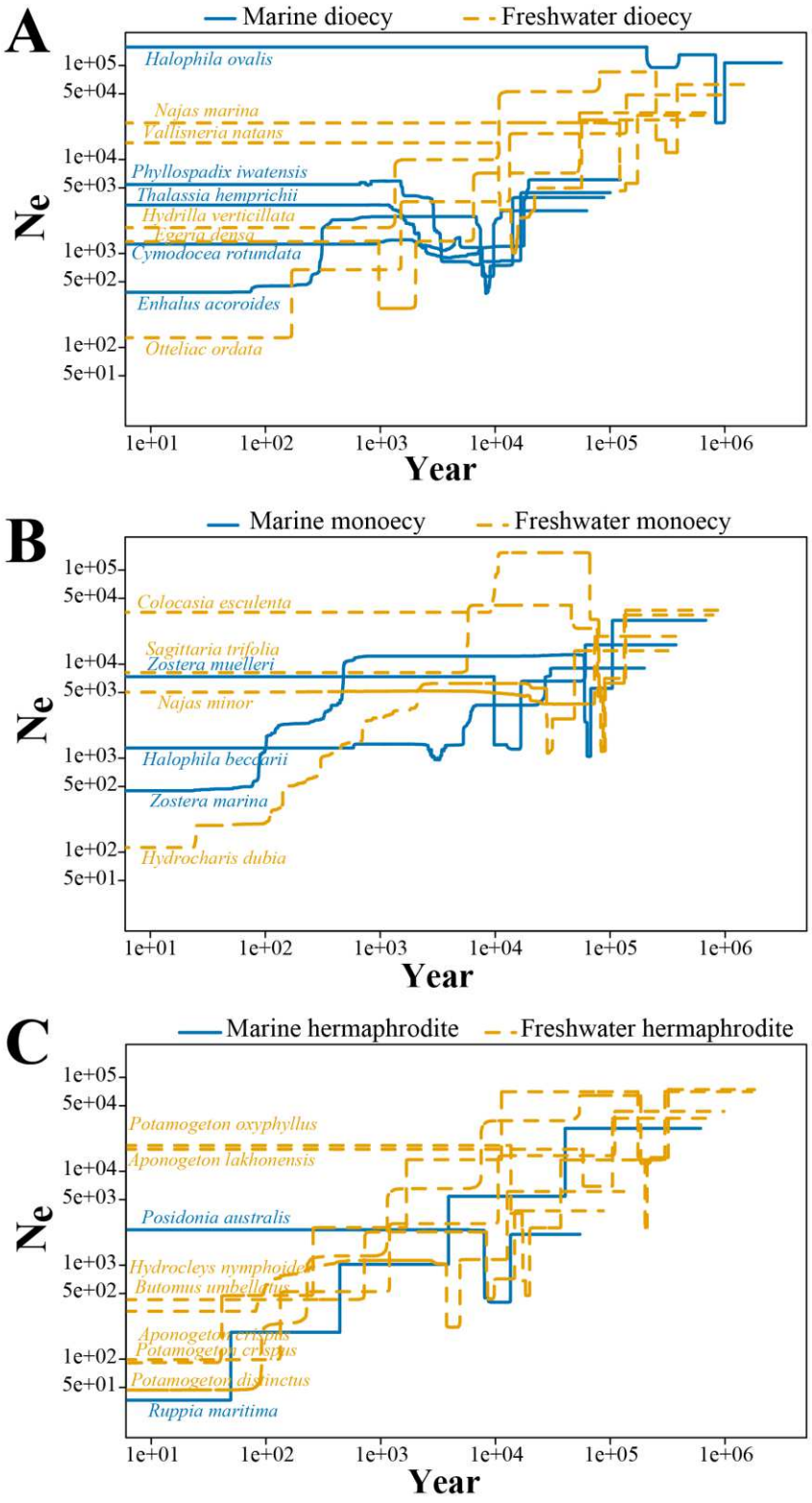


Fig. 3 Estimates of historical population size change curves using Stairway plot 2. (**A**) dioecious, (**B**) monoecious, and (**C**) hermaphroditic species for marine (blue line) and freshwater (yellow dotted line) species.

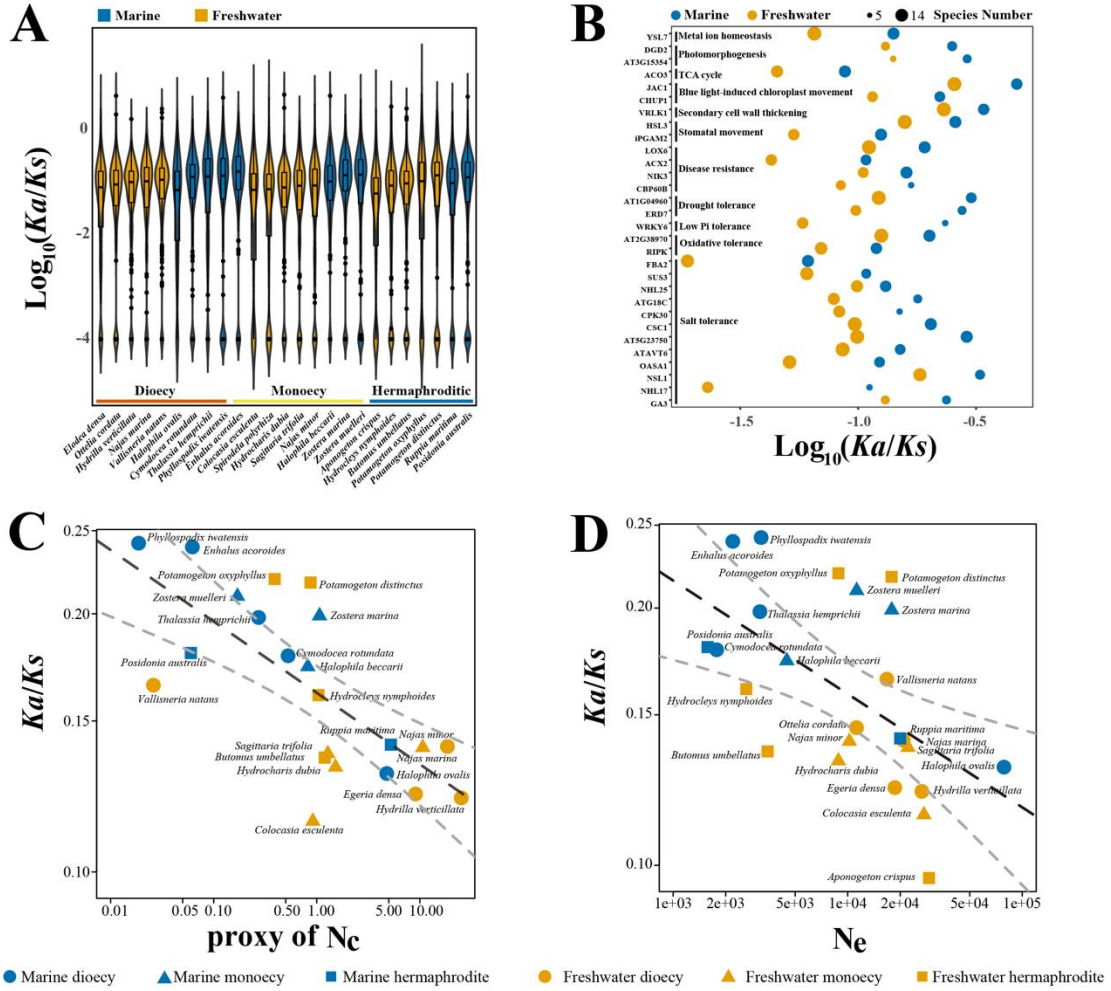
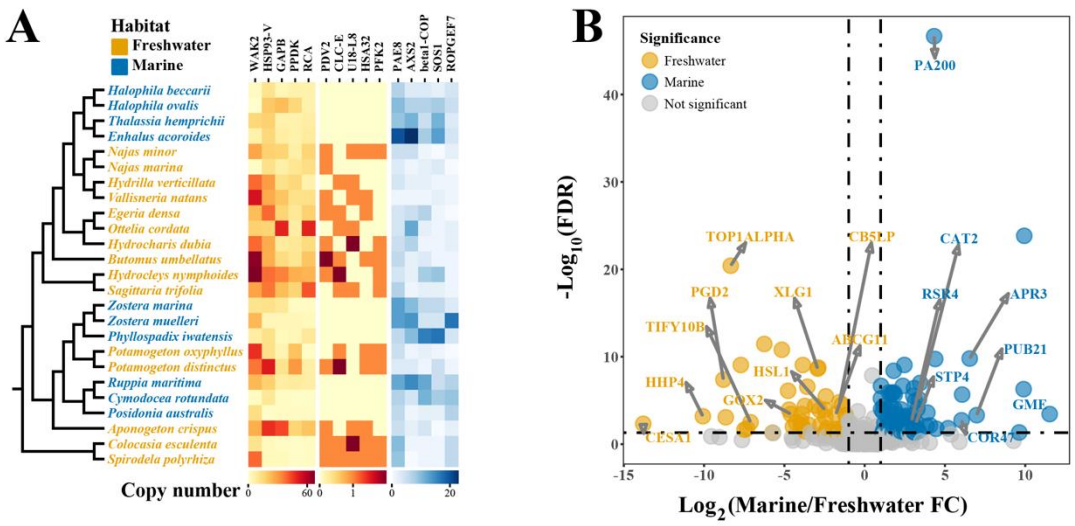
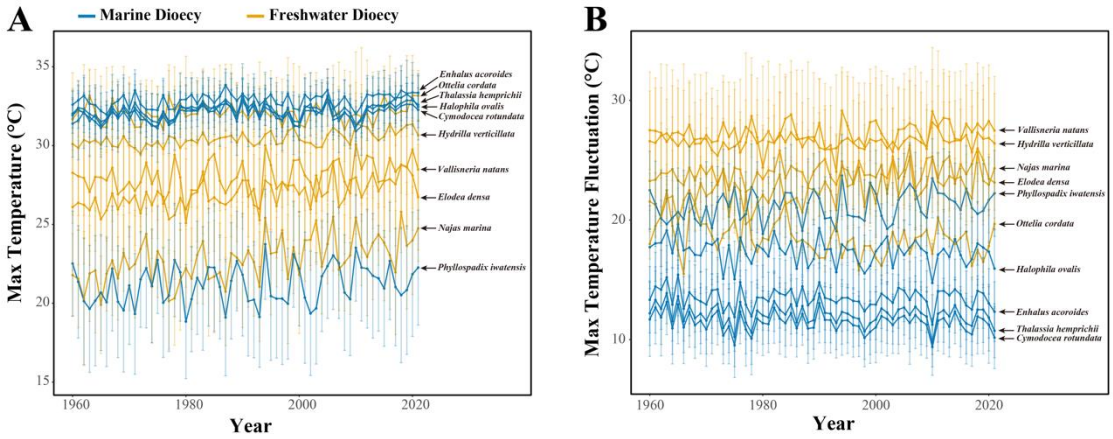


Fig. 4 Difference in evolutionary rates between marine and freshwater seagrasses. (**A**) The distribution of K_a/K_s values in log scale for 2079 orthologous groups (OGs) in 25 seagrass species. The average K_a/K_s was lower in freshwater than in marine species. (**B**) Examples of functionally enriched genes with K_a/K_s significantly lower in freshwater species than in marine species. (**C, D**) The log-log linear regression of K_a/K_s on the proxy of census size (N_c) and historical effective population size (N_e).



997
998 **Fig. 5 Difference in gene copy and expression between freshwater and marine species.** (**A**)
999 An example of genes with larger copy numbers in either freshwater species (red-yellow
1000 heatmaps) or marine species (blue-white heatmaps); (**B**) The volcano plot showing
1001 differentially expressed genes between freshwater and marine seagrasses. Genes have higher
1002 expression level in freshwater species (yellow dots) or in marine species (blue dots) with the
1003 fold change larger than two and FDR smaller than 0.05 (the dash line).
1004
1005



1006
1007 **Fig. 6 Temperature difference between marine and freshwater dioecious species**. (**A**) The
1008 mean and standard error bars of annual maximum temperature and (**B**) maximum temperature
1009 fluctuation experienced by dioecious species in freshwater (yellow) and marine (blue) across
1010 their distribution ranges from 1960 to 2021.
1011
1012

Table legends

Table 1 The genetic diversity (π_4) and selection efficacy (π_0/π_4) for sexual system depending on the habitat and pollination.

Table 2 The estimates of DFE parameters for marine and freshwater species of three sexual types.

Table 3 The median values of inbreeding coefficient, GC3 content, census size proxy, and historical N_e for marine and freshwater species of three sexual types

Figure legends

Fig. 1 The phylogenetic tree of 27 Alismatales species and difference between marine and freshwater species in genetic diversity. (**A**) The phylogenetic tree of 27 Alismatales species inferred based on 2,079 orthologous gene families. Blue bars give the 95% HD of divergence time and the figures above the bars are mean estimates of divergence time. The axis below the tree is the time in millions of year BP. (**B**) The distribution of genetic diversity π_4 and (**C**) efficacy of selection, π_0/π_4 ratio between freshwater and marine species. Each dot represents the mean value for a species and the red line is the median and the box is the interquartile range.

Fig. 2 Correlation between genetic diversity, the selection efficacy and census size or historical population size. (**A**) The log-log linear regression of genetic diversity at 4-fold site (π_4) on the proxy of census size (N_c); (**B**) The regression of π_4 on historical population size (N_e); (**C**) The log-log linear regression of selection efficacy (π_0/π_4) on the proxy of N_c; (**D**) the regression of π_0/π_4 on N_e. The bars represented 95% confidence intervals and the dots represented the medians.

Fig. 3 Estimates of historical population size change curves using Stairway plot 2. (**A**) dioecious, (**B**) monoecious, and (**C**) hermaphroditic species for marine (blue line) and freshwater (yellow dotted line) species.

Fig. 4 Difference in evolutionary rates between marine and freshwater seagrasses. (**A**) The distribution of K_a/K_s values in log scale for 2079 orthologous groups (OGs) in 25 seagrass

species. The average K_a/K_s was lower in freshwater than in marine species. (**B**) Examples of functionally enriched genes with K_a/K_s significantly lower in freshwater species than in marine species. (**C, D**) The log-log linear regression of K_a/K_s on the proxy of census size (N_c) and historical effective population size (N_e).

Fig. 5 Difference in gene copy and expression between freshwater and marine species. (**A**) An example of genes with larger copy numbers in either freshwater species (red-yellow heatmaps) or marine species (blue-white heatmaps); (**B**) The volcano plot showing differentially expressed genes between freshwater and marine seagrasses. Genes have higher expression level in freshwater species (yellow dots) or in marine species (blue dots) with the fold change larger than two and FDR smaller than 0.05 (the dash line).

Fig. 6 Temperature difference between marine and freshwater dioecious species. (**A**) The mean and standard error bars of annual maximum temperature and (**B**) maximum temperature fluctuation experienced by dioecious species in freshwater (yellow) and marine (blue) across their distribution ranges from 1960 to 2021.

Supplementary Figures

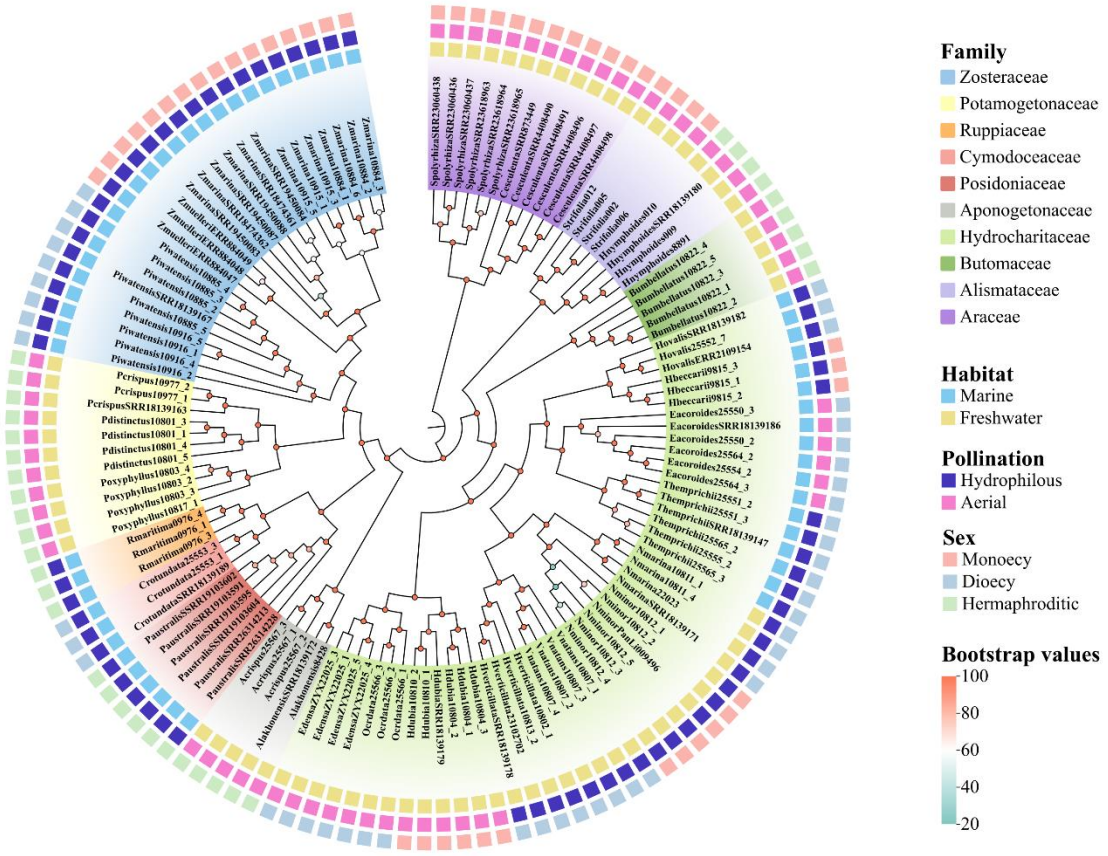


Fig. S1 The ML tree based on 127 individuals. Dots in the nodes of phylogenetic tree represent the bootstrap supporting scores and the colored squares from the inner to the outer layers represent the family, habitat, pollination, and sexual system for each species, respectively.

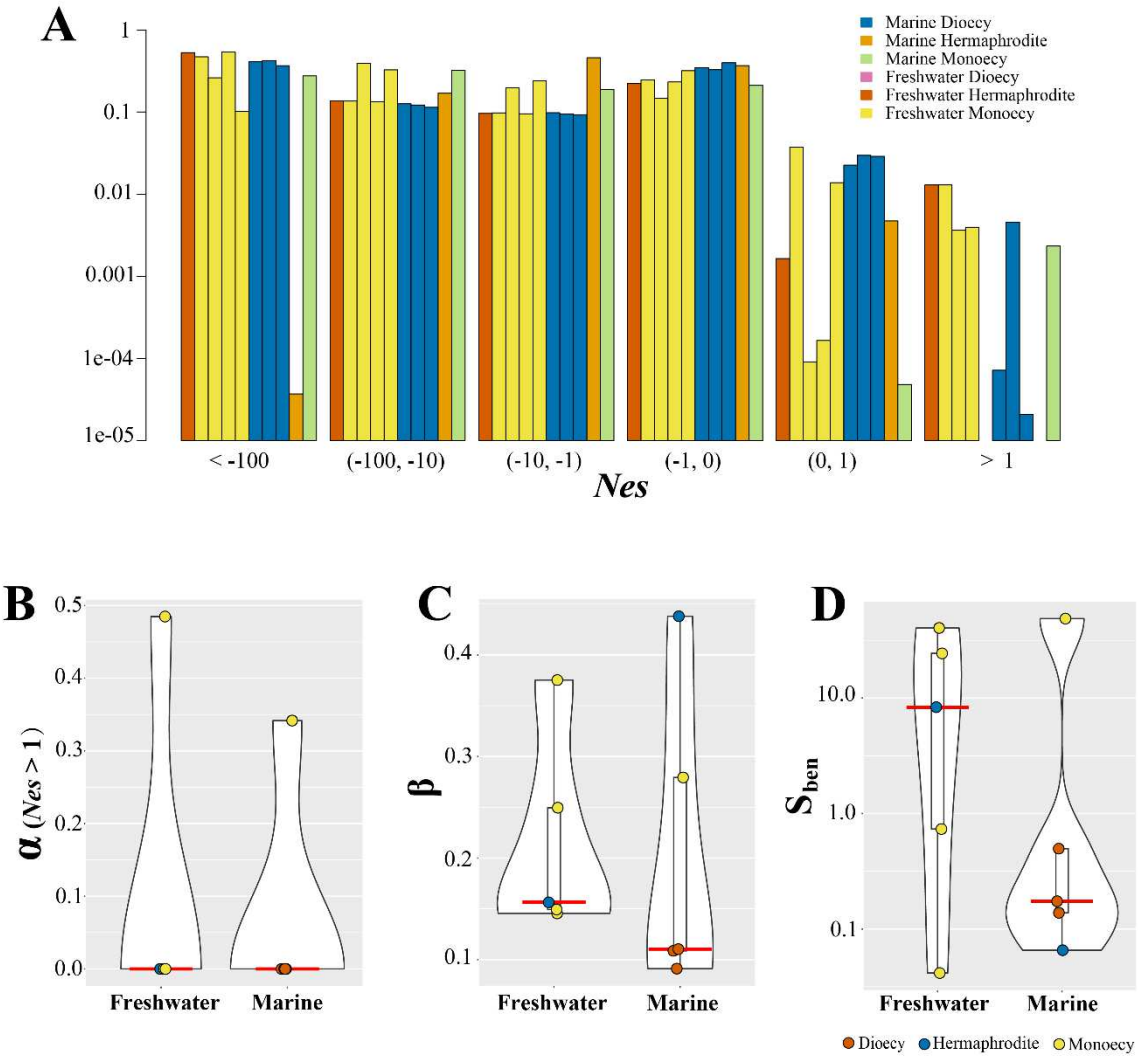


Fig. S2 The distributions of fitness effects of new mutations (DFE) for 10 selected seagrass species. (**A**) Histograms for the proportions of mutations under various selective strength; (**B**) The violin plots for proportions of fixed adaptive mutations (α), (**C**) the shape parameter (β), and (**D**) the selective strength for beneficial mutations.

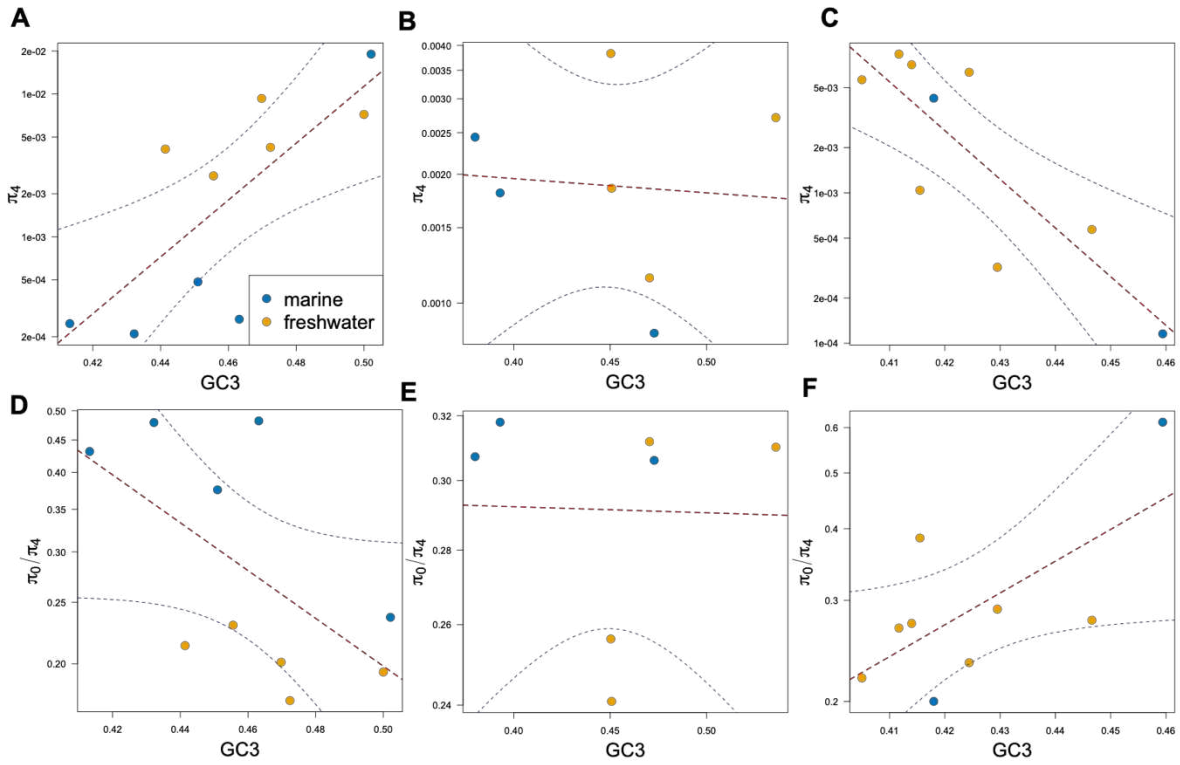


Fig. S3 The correlation between π_4 or π_0/π_4 and GC3 content. Dots in (**A**) and (**D**) represent dioecious species, (**B**) and (**E**) for monoecious species and (**C**) and (**F**) for hermaphroditic species. Blue dots show marine species and orange dots show freshwater species. Red lines represent the linear predicted curve and the blue lines are 95% confident intervals.

Supplementary table legends

Table S1 Sample collection information for each species

Table S2 Genomic or transcriptomic statistics for each species

Table S3 Details of the four fossils used as calibration points in divergence time estimation of Alismatales

Table S4 Traits, distribution ranges, and population summary statistics for all species

Table S5 The list of genes with K_a/K_svalues significantly smaller in freshwater species

Table S6 The list of genes with K_a/K_svalues significantly higher in freshwater species

Table S7 The list of genes present in freshwater species but absent in marine species

Table S8 The list of genes with larger number of copy number in freshwater species

Table S9 The list of genes with larger number of copy number in marine species

Table S10 The list of genes with expression levels significantly different between freshwater and marine species

Supplementary figure legends

Fig. S1 The ML tree based on 127 individuals. Dots in the nodes of phylogenetic tree represent the bootstrap supporting scores and the colored squares from the inner to the outer layers represent the family, habitat, pollination, and sexual system for each species, respectively.

Fig. S2 The distributions of fitness effects of new mutations (DFE) for seagrasses based on sexual system and living environment. (**A**) Histograms for the proportions of mutations under various selective strength from $N_e s < -100$ to $N_e s > 1$; (**B**) The violin plots for proportions of fixed adaptive mutations (α), **(C)** the shape parameter (β)**,** and **(D)** the selective strength for beneficial mutations.

Fig. S3 The correlation between π_4 or π_0/π_4 and GC3 content. Dots in (**A**) and (**D**) represent dioecious species, (**B**) and (**E**) for monoecious species and (**C**) and (**F**) for hermaphroditic species. Blue dots show marine species and orange dots show freshwater species. Red lines represent the linear predicted curve and the blue lines are 95% confident intervals.

1154

1155

Supplementary Files

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

- SeagrassSupplementaryFigures.pdf
- SeagrassSupplementaryTables.xlsx