



# Phylogenomic Analyses of Alismatales Shed Light into Adaptations to Aquatic Environments

Ling-Yun Chen,<sup>†,1,2</sup> Bei Lu,<sup>†,1,3</sup> Diego F. Morales-Briones <sup>†,4,5</sup> Michael L. Moody,<sup>6</sup> Fan Liu,<sup>1</sup> Guang-Wan Hu,<sup>1</sup> Chien-Hsun Huang,<sup>\*,7</sup> Jin-Ming Chen <sup>\*,1</sup> and Qing-Feng Wang<sup>\*,1</sup>

<sup>1</sup>Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden/Core Botanical Garden, Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China

<sup>2</sup>Department of Resources Science of Traditional Chinese Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China

<sup>3</sup>College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup>Department of Plant and Microbial Biology, University of Minnesota, 140 Gortner Laboratory, 1479 Gortner Avenue, Saint Paul, MN 55108, USA

<sup>5</sup>Systematics, Biodiversity and Evolution of Plants, Ludwig-Maximilians-Universität München, Menzinger Str. 67, 80638 Munich, Germany

<sup>6</sup>Department of Biological Sciences, University of Texas at El Paso, 500 West University Ave, El Paso, TX 79968, USA

<sup>7</sup>State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Biodiversity Sciences and Ecological Engineering, Institute of Plant Biology, Institute of Biodiversity Sciences, School of Life Sciences, Fudan University, Shanghai 200433, China

\*Corresponding authors: E-mails: huang\_ch@fudan.edu.cn; jmchen@wbcas.cn; qfwang@wbcas.cn.

†These authors contributed equally to this work.

Associate editor: Emma Teeling

## Abstract

Land plants first evolved from freshwater algae, and flowering plants returned to water as early as the Cretaceous and multiple times subsequently. Alismatales is the largest clade of aquatic angiosperms including all marine angiosperms, as well as terrestrial plants. We used Alismatales to explore plant adaptations to aquatic environments by analyzing a data set that included 95 samples (89 Alismatales species) covering four genomes and 91 transcriptomes (59 generated in this study). To provide a basis for investigating adaptations, we assessed phylogenetic conflict and whole-genome duplication (WGD) events in Alismatales. We recovered a relationship for the three main clades in Alismatales as (Tofieldiaceae, Araceae) + core Alismatids. We also found phylogenetic conflict among the three main clades that was best explained by incomplete lineage sorting and introgression. Overall, we identified 18 putative WGD events across Alismatales. One of them occurred at the most recent common ancestor of core Alismatids, and three occurred at seagrass lineages. We also found that lineage and life-form were both important for different evolutionary patterns for the genes related to freshwater and marine adaptation. For example, several light- or ethylene-related genes were lost in the seagrass Zosteraceae, but are present in other seagrasses and freshwater species. Stomata-related genes were lost in both submersed freshwater species and seagrasses. Nicotianamine synthase genes, which are important in iron intake, expanded in both submersed freshwater species and seagrasses. Our results advance the understanding of the adaptation to aquatic environments and WGDs using phylogenomics.

**Key words:** Alismatales, phylogenetic conflict, adaptation, whole-genome duplication, marine angiosperms.

## Introduction

Land plants evolved from freshwater algae about half a billion years ago (Tena 2020). To adapt to terrestrial environments, land plants acquired genes dealing with intense ultraviolet light, red light, and water evaporation (Han et al. 2019; Kong et al. 2020). Some extant members of major vascular plant lineages have returned to the water as obligates: ferns and lycophytes (e.g., *Azolla*, Isoetaceae), gymnosperms (e.g., *Retrophyllum minus*), and angiosperms (e.g., *Utricularia*, *Ceratophyllum*, and water lily). The

earliest unequivocal fossils of angiosperms date to the Early Cretaceous (Friis et al. 2010) and some of the earliest angiosperms are aquatic. For example, Nymphaeales have been identified from 125 to 115 Ma (Friis et al. 2010) and Ceratophyllales from the late Albian (c. 113 Ma; Wang and Dilcher 2018). Aquatic angiosperms constitute only 1–2% of extant angiosperms, but are found in approximately 17% of families representing about 100 evolutionarily independent origins (Cook, 1990). Compared with terrestrial angiosperms, aquatic angiosperms occupy distinctive and stressful ecological habitats characterized by low light

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

levels, reduced carbon and oxygen availability, and mechanical damage through wave exposure. Aquatic angiosperms have adapted to the requirements of aquatic habitats with life-forms notably different from terrestrial plants, generally being divided into emergent, floating-leaved, or submersed forms, the latter most extreme (Les et al. 1997). Survival in aquatic habitats has resulted in adaptations specific to low oxygen concentrations. For example, the aerenchyma in roots, stems, and leaves improves the capture and transportation of oxygen (Jung et al. 2008; Nowak et al. 2010). At the metabolic level, aquatic plants might have enhanced glycolytic fluxes and ethanolic fermentation (Billet et al. 2018).

Aquatic angiosperms comprise freshwater and marine species (often called seagrasses), the latter evolved from the former (Les et al. 1997). A few studies have investigated plant adaptation to freshwater at the genomic level. For example, An et al. (2019) found that disease-resistant genes, which protect plants from pathogens, are not only amplified but also highly expressed in *Spirodela polyrhiza*. Compared with freshwater species, marine species live in environments that have a higher concentration of salt but a lower concentration of iron (Duarte et al. 1995; Anton et al. 2018). Several studies have investigated the genes related to the adaptation from freshwater to seawater. Using genomic data, Golicz et al. (2015), Olsen et al. (2016), and Lee et al. (2016) found that seagrasses lost genes associated with ethylene, terpenoid biosynthesis, ultraviolet damage, and stomatal differentiation. Seagrasses have a constitutive formation of aerenchyma, possibly obviating the need for ethylene production, which is a ubiquitous response to inundation in terrestrial plants to form aerenchyma (Golicz et al. 2015). On the other hand, seagrasses regained genes related to salinity adaptation (Olsen et al. 2016; Lee et al. 2018). However, these studies compared genomic data of seagrasses with those of terrestrial or floating aquatic plants, but not freshwater submersed plants. As the environments for seagrasses and submersed freshwater plants are somewhat similar, a direct comparison between these life-forms in the two environments may provide a clearer understanding of the loss or gain of genes in aquatic angiosperms and identify genes that are important for adaptation to saline environments. In addition, only a few seagrasses and freshwater species were used in previous studies and the common or unique strategies of aquatic plant adaptation have not been investigated in a more comprehensive and phylogenomic context.

Alismatales, with a worldwide distribution, is the second oldest clade of monocots (Eguchi and Tamura, 2016). It consists of core Alismatids, Araceae, and Tofieldiaceae. Among them, core Alismatids includes 12 families (Alismataceae, Butomaceae, Hydrocharitaceae, Scheuchzeriaceae, Aponogetonaceae, Juncaginaceae, Maundiaceae, Zosteraceae, Potamogetonaceae, Posidoniaceae, Ruppiaceae, and Cymodoceaceae), 58 genera, and c. 593 species (The Plant List: <http://www.theplantlist.org/>; The Angiosperm Phylogeny Group 2016). Araceae

includes 117 genera and c. 3,368 species and Tofieldiaceae includes four genera and 29 species (The Plant List). Alismatales comprises the largest clade of aquatic angiosperms representing all aquatic life-forms: emergent (e.g., *Alisma*), floating-leaved (e.g., *Spirodela*), and submersed (e.g., *Potamogeton*, *Najas*). It also encompasses all the marine angiosperms (Zosteraceae, Posidoniaceae, Hydrocharitaceae, and Cymodoceaceae), which comprise three lineages with independent origins from submersed freshwater life-forms (Les et al. 1997). Araceae and Tofieldiaceae also contain a combination of both terrestrial and aquatic taxa. All this makes Alismatales the ideal group for studying aquatic plant adaptations.

Although several studies have investigated the phylogeny of Alismatales, phylogenetic uncertainty remains throughout the clade. One of the recalcitrant relationships is among the three main clades within Alismatales. Analyses using chloroplast (cp) and mitochondrial markers supported the relationship (Tofieldiaceae [Araceae, core Alismatids]) (Iles et al. 2013; Luo et al. 2016; Petersen et al. 2016; Ross et al. 2016; Li et al. 2019). However, studies using cp and/or nuclear markers supported the relationship (Araceae [Tofieldiaceae, core Alismatids]) (Azuma and Tobe, 2011; Nauheimer et al. 2012; Hertweck et al. 2015; Li et al. 2021). Conflicting phylogenetic relationships also exist within Alismataceae, Potamogetonaceae, and Hydrocharitaceae (Tanaka et al. 1997; Les et al. 2006; Chen et al. 2012; Ross et al. 2016). Investigating the evolutionary history of Alismatales is essential to assess adaptations to aquatic environments.

Whole-genome duplication (WGD; also known as polyploidy) has had a profound influence on plant radiations and environmental adaptations (Van de Peer et al. 2017), such as plant function under salinity (Chao et al. 2013; Yang et al. 2014). WGD events have been identified in two Alismatales lineages: 1) *Zostera* (Zosteraceae), which underwent a WGD around 65 Ma (Olsen et al. 2016) and 2) in Araceae, which had WGDs at *Spirodela* and *Pinellia* separately or at their ancestor (Wang et al. 2014; Ren et al. 2018; Qiao et al. 2019). Only *Zostera*, *Alisma*, and Araceae have been used to evaluate WGDs in Alismatales and the ages of these events are estimated to be older than the stem nodes of the respective lineages where WGD was identified (e.g., the WGD of *Zostera* [65 Ma] is older than the stem node age of Zosteraceae [58 Ma; Hertweck et al. 2015]). The age inconsistency implies that positions of the WGDs need further assessment. Up to now, studies have not comprehensively explored WGDs within Alismatales and it is still unclear if WGD events are involved in the adaptation to aquatic environments.

Here, we used a data set that included 91 transcriptomes (59 generated in this study) and four genomes. The data set represents 13 of the 14 families within Alismatales (The Angiosperm Phylogeny Group 2016), 40 of the 57 genera within core Alismatids, 17 of the 117 genera within Araceae, and three of the four genera within Tofieldiaceae (The Plant List). The only unsampled family

is Ruppiaceae, which has a consistent phylogenetic position (e.g., Li et al. 2021; Ross et al. 2016) and has a submerged life-form, similar to members of the family Potamogetonaceae. The data set also represents terrestrial, freshwater, and marine species. By conducting orthology inference, phylogenetic analyses, and genome duplication inference, we aimed to 1) explore phylogenetic conflict in Alismatales, 2) examine WGD events in Alismatales, especially those potentially involved in adaptation to aquatic environments, and 3) investigate genes associated with adaptation to aquatic and high salinity environments.

## Results and Discussion

A workflow showing data sources, analytical steps, and results is provided in [supplementary fig. S1, Supplementary Material](#) online, for details.

### Data Processing and Phylogenetic Analyses

Reads of 59 newly generated transcriptomes were deposited at the NCBI Sequence Read Archive (BioProject: PRJNA809041; [supplementary table S1, Supplementary Material](#) online). The number of filtered read pairs ranged from 4.9 to 31.5 million. The final number of “monophyletic outgroup” (MO) orthologs with 95 taxa was 1,005. Fifty-three coding sequences (CDSs) for 92 samples were also assembled.

RAxML (concatenation) and ASTRAL (coalescent-based) phylogenetic analyses were applied to both the nuclear and cp data sets. The topologies inferred using RAxML ([fig. 1](#)) and ASTRAL ([fig. 2](#)) with nuclear orthologs are similar except for the position of a *Potamogeton* species. Both analyses resulted in strong support (Bootstrap [BS] = 100 or ASTRAL local posterior probability [LPP] = 1) for most clades. Both analyses recovered the relationship (Tofieldiaceae [Araceae, core Alismatids]), however, Araceae + core Alismatids lacks BS support (BS = 0 and LPP = 0.99; [fig. 1](#)). There is also low LPP support for Potamogetonaceae ([fig. 1](#)).

The overall support inferred from cp data is lower than that inferred from the nuclear data set, although most clades have BS > 90 or LPP > 0.90 ([supplementary fig. S2-1 and S2-2, Supplementary Material](#) online). The topologies inferred from nuclear orthologs and cp data are similar. All analyses supported (Tofieldiaceae [Araceae, core Alismatids]). Phylogenetic inconsistencies between the nuclear orthologs and cp data also exist for *Echinodorus* within Alismataceae and (*Hydrocharis*, *Limnobium*) within Hydrocharitaceae. Species-level phylogenetic inconsistency was also common within Potamogetonaceae. Almost all the nodes with inconsistencies have BS < 90 or LPP < 0.90 ([supplementary fig. S2-1 and S2-2, Supplementary Material](#) online).

We combined cp CDSs assembled from RNA sequencing (RNA-seq) reads and CDSs extracted from cp genomes for phylogenetic analyses. The topology of our cp trees ([supplementary fig. S2-1 and S2-2, Supplementary](#)

[Material](#) online) was similar to that inferred from cp genomes (Ross et al. 2016). For example, *Hydrocharis* + *Limnobium* was clustered with a clade formed by *Thalassia*, *Enhalus*, *Halophila*, *Vallisneria*, and *Najas*. Since many plant plastomes are heavily RNA-edited, combining DNA sequences with RNA-edited transcripts has been shown to cause problems in phylogenetic analyses (Ramaswami et al. 2013). Therefore, we used only the nuclear phylogeny for further analyses in our study.

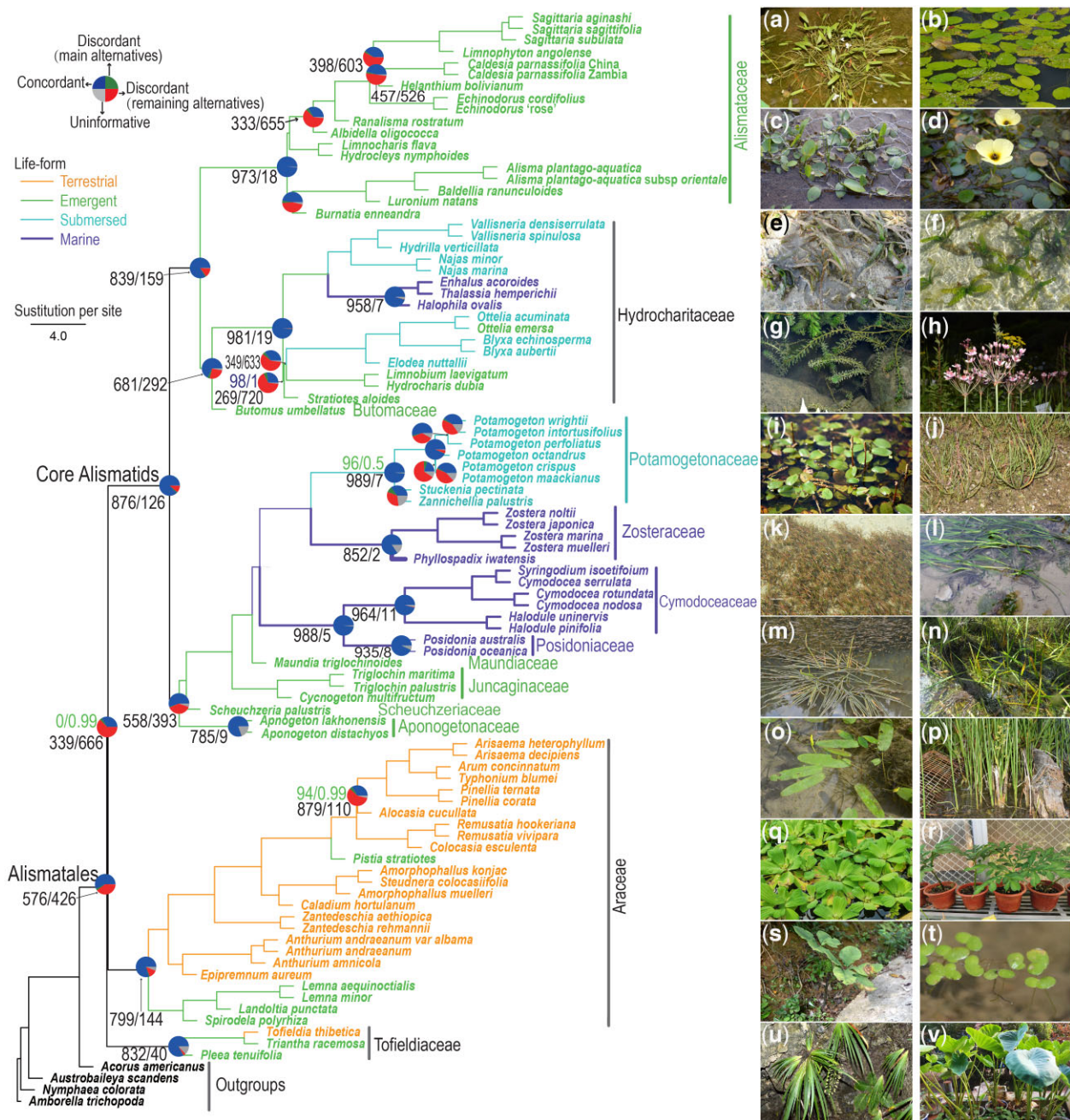
Overall, the phylogenetic analyses supported monophyly for all families of Alismatales for which multiple accessions were sampled (i.e., Alismataceae, Aponogetonaceae, and Araceae) and the Core Alismatids, consistent with previous studies (Les et al. 1997; Iles et al. 2013; Luo et al. 2016; Petersen et al. 2016; Ross et al. 2016).

### Phylogenetic Conflict

To investigate phylogenetic conflict, we calculated the concordant and conflicting bipartitions, and internode certainty all (ICA; Salichos et al. 2014) values using PhyParts. ICA value quantifies the degree of conflict in a given internode. An ICA value close to 1 indicates strong concordance in the bipartition of interest, while an ICA value closer to 0 indicates equal support for one or more conflicting bipartitions. A negative value indicates that the internode of interest conflicts with one or more bipartitions that have a higher frequency, and an ICA value close to -1 indicates the absence of concordance for the bipartition of interest (Salichos et al. 2014). We also calculated Quartet Concordance (QC), Quartet Differential (QD), and Quartet Informativeness (QI) scores using Quartet Sampling (QS; Pease et al. 2018). A QC value close to 1 indicates all concordant quartets, a QC close to 0 indicates equivocal concordant/discordant quartets, and a QC smaller than 0 indicates discordant quartets are more frequent than concordant quartets. No QD indicates no alternative topologies (i.e., QC = 1), a QD close to 1 indicates that the two alternative topologies appear in equally frequency, and a QD close to 0 indicates one of the two alternative topologies is favored. A QI value close to 1 indicates all replicates are informative, while a value close to 0 indicates all replicates were uncertain (Pease et al. 2018).

The monophyly for most families of Alismatales was highly supported with at least 80% of informative gene trees being concordant ([fig. 1](#)), ICA > 0.8, and full QS support (1/-/1) ([supplementary fig. S3, Supplementary Material](#) online). Conflict was detected for the clade formed by core Alismatids and Araceae (number of concordant/conflicting bipartitions is 339/666; ICA = 0.3; QS score = -0.24/0.14/0.99). Conflict was also detected for some other clades, for example, the two sub-clades within Alismataceae and the clades involving *Stratiotes* (ICA = 0.43, QS = 0.042/0/1) and *Limnobium* + *Hydrocharis* (ICA = 0.5, QS = -0.14/0.67/0.99) within Hydrocharitaceae ([supplementary fig. S3, Supplementary Material](#) online).

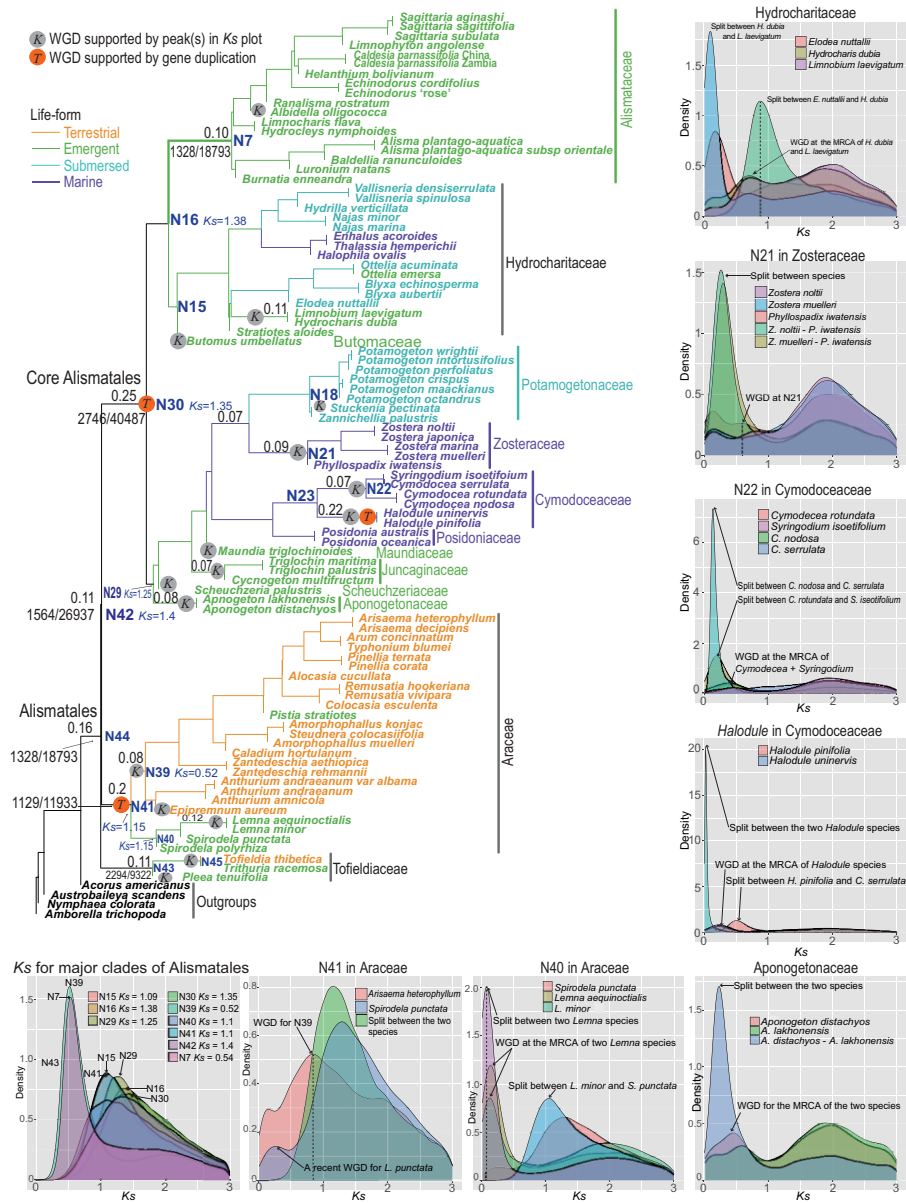
To detect possible reticulation on the nodes with large phylogenetic conflict, we performed species network



**FIG. 1.** Maximum likelihood phylogeny of Alismatales inferred with RAxML from the concatenated 1,005-nuclear gene matrix. The phylogeny is congruent with the topology inferred with ASTRAL (Fig. 2) except for *Potamogeton octandrus*. All nodes have maximum support (RAxML BS = 100/ASTRAL LPP = 1) unless noted near the branches (green). Pie charts represent gene trees for concordant bipartitions (blue), the most frequent alternative topology (green), remaining alternatives (red), and those uninformative for nodes (BS  $\leq$  50%). Numbers below branches represent the number of concordant/discordant genes trees for some nodes. Pie charts and numbers are shown only for nodes discussed in the main text. Plant photos were taken by Lingyun Chen, Jinming Chen, Kuo Liao, and Qingfeng Wang. (a) *Sagittaria subulata*, (b) *Albidella oligococca*, (c) *Luronium natans*, (d) *Hydrocleys nymphoides*, (e) *Enhalus acoroides*, (f) *Thalassia hemprichii*, (g) *Hydrilla verticillata*, (h) *Butomus umbellatus*, (i) *Potamogeton distinctus*, (j) *Triglochin maritima*, (k) *Cymodocea rotundata*, (l) *Zostera marina*, (m) *Cycnogeton multifructum*, (n) *Maundia triglochoides*, (o) *Aponogeton lakhonensis*, (p) *Scheuchzeria palustris*, (q) *Pistia stratiotes*, (r) *Amorphophallus konjac*, (s) *Arisaema lobatum*, (t) *Lemna minor*, (u) *Tofieldia thibetica*, (v) *Colocasia esculenta*. High-resolution plant photos are available at the figshare: <https://doi.org/10.6084/m9.figshare.16967767>.

searches using PhyloNet v3.8.2 (Than et al. 2008) with a Bayesian Markov chain Monte Carlo algorithm. We carried out phylogenetic network analyses for Alismatales, Alismataceae, Hydrocharitaceae, Potamogetonaceae, and Araceae. PhyloNet analyses resulted in five independent

and convergent runs for Alismatales, Araceae, Hydrocharitaceae, and Potamogetonaceae, and four runs for Alismataceae (supplementary fig. S4, Supplementary Material online). The maximum posterior probability (MPP) networks with inheritance probabilities ( $\gamma$ ) were

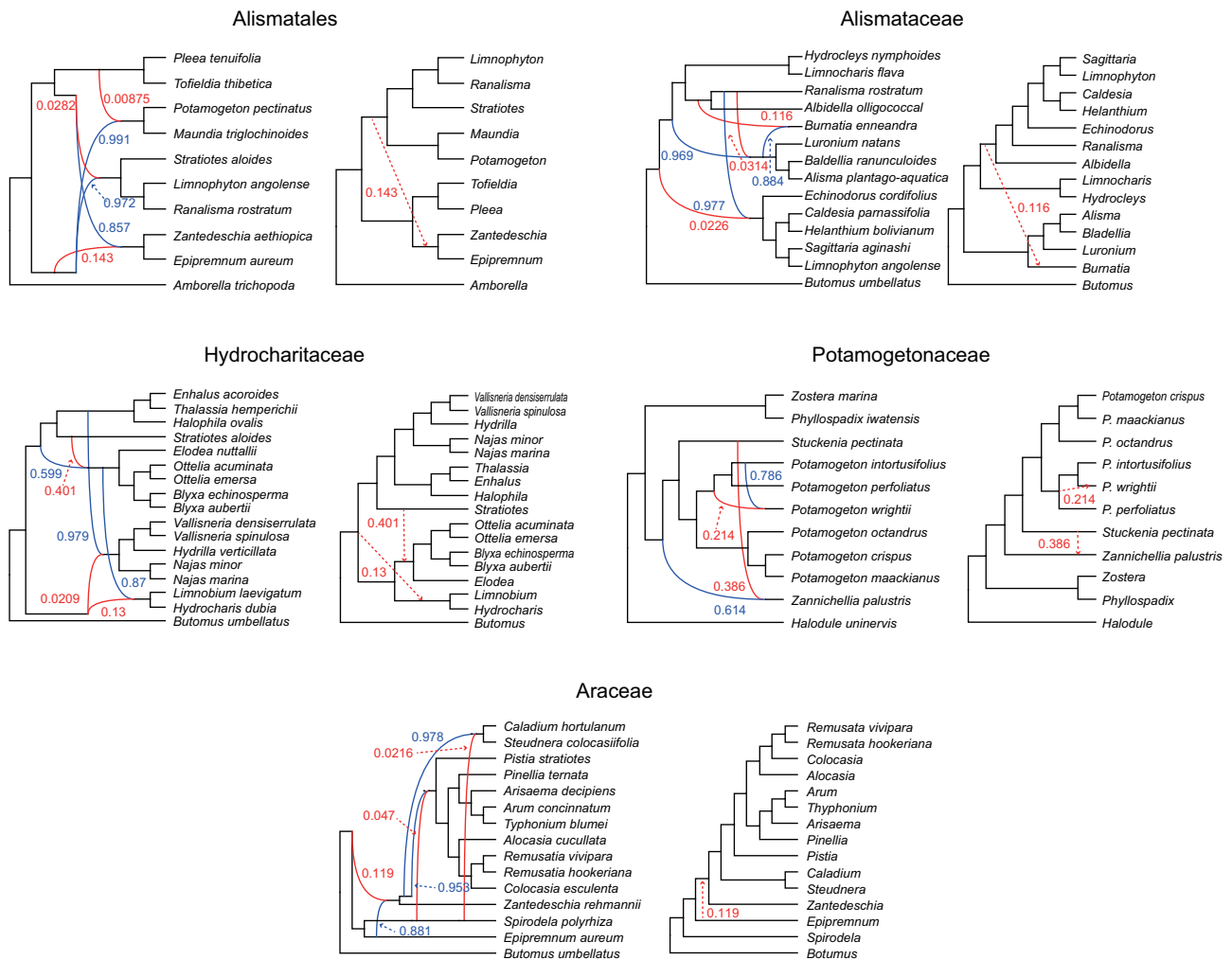


**Fig. 2.** ASTRAL species tree inferred from 1,005 nuclear genes trees. Percentages next to nodes denote the proportion of duplicated genes (shown when it is  $\geq 7\%$ ) when mapping gene clusters to the ASTRAL species tree. Numbers before and after the slash (/) represent the number of gene duplications shared by the taxa and the total ancestral gene number before a WGD event experienced along the branch as inferred from software Tree2GD. The number of gene duplications is shown only when  $\geq 1000$ . Whole-genome duplications inferred from Ks distribution are labeled on the tree. The density plots of within-species Ks and between-species Ks are shown for seven representative clades. Node numbers correspond to those in [supplementary fig. S6, Supplementary Material online](#), for details. Ks distribution for all species of Alismatales is present in [supplementary fig. S8, Supplementary Material online](#), for details. MRCA, most recent common ancestor.

obtained after summarization of converged runs ([fig. 3](#)). Inheritance probabilities  $\gamma \approx 0.5$  represent hybridization,  $\gamma \approx 0.1$  introgression, and a small  $\gamma (< 0.05)$  incomplete lineage sorting (ILS) ([Solís-Lemus et al. 2017](#)). We discuss PhyloNet results of each of the five lineages below.

**Alismatales:** The MPP topology from PhyloNet recovered the relationship (core Alismatids [Araceae, Tofieldiaceae]) (H1; [fig. 3](#)). The approximately unbiased (AU) tests do not show support for any of the three possible topologies for Alismatales as 626 of the 798 were equivocal ([fig. 4a](#)). However, H1 was more frequent when counting the number of trees supporting each

alternative topologies by likelihood support ([fig. 4b](#)). Moreover, H1 was also more frequent when counting the number of trees using BS values ([fig. 4c](#)). The other two topologies for Alismatales were not supported: 1) The topology of hypothesis H0 (Tofieldiaceae, [Araceae, core Alismatids]) was recovered in phylogenetic analyses using both nuclear and cp data ([fig. 1, supplementary fig. S2, Supplementary Material online](#)). However, the clade Araceae + core Alismatids lacked BS support ([fig. 1](#)) and had low ICA and QS scores ([supplementary fig. S3, Supplementary Material online](#)). PhyloNet found that Araceae inherited about 14.3% ([fig. 3](#)) of its genome

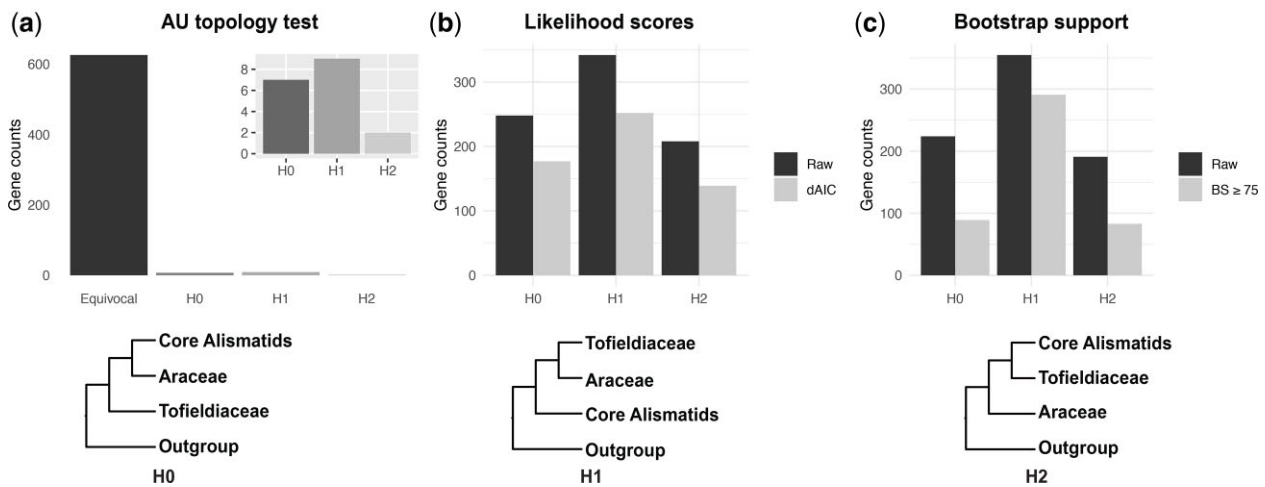


**FIG. 3.** MPP species networks of the five reduced-taxon data set inferred with PhyloNet. Numbers next to curved lines indicate inheritance probabilities ( $\gamma$ ) for each hybrid node. Trees next to each network represent the networks showing only reticulation events with a minor edge with  $\gamma \geq 0.1$ .

from the most recent common ancestor (MRCA) of core Alismatids. Given the  $\gamma$  value, introgression explains the close relationship between Araceae and core Alismatids (H0) in our results and previous studies (Iles et al. 2013; Ross et al. 2016; Luo et al. 2016; Petersen et al. 2016). 2) The hypothesis H2, (Araceae [Tofieldiaceae, core Alismatids]), was recovered but weakly supported in previous studies. Namely, the nodal support for Tofieldiaceae + core Alismatids was low, BS < 75 in Azuma and Tobe (2011), Nauheimer et al. (2012), and Liang et al. (2019). The PhyloNet analyses found that core Alismatids inherited 0.875% of its genome from Tofieldiaceae and 2.82% of its genome from the MRCA of Tofieldiaceae + core Alismatids. Given the low  $\gamma$  values, ILS best explains the close relationship between Tofieldiaceae and core Alismatids (H2). 3) A polytomy hypothesis for Alismatales was rejected using the ASTRAL (CITATION) as polytomy. Given the low nodal support and short branch length for the clade core Alismatids + Araceae and AU test results, the evolution of these three clades of Alismatales may be the result of a rapid radiation,

which makes resolution of the topology difficult. In conclusion, our AU analyses slightly favored topology H1 for Alismatales, but there is phylogenetic conflict regarding the three major clades of Alismatales that could be due to introgression and ILS.

**Alismataceae:** The major phylogenetic inconsistency within Alismataceae is the position of *Burnatia*. Using three cp markers, *Burnatia* was recovered as sister to *Hydrocleys* + *Limnocharis* (Chen et al. 2012). However, using additional cp markers, *Burnatia* was recovered as sister to a clade comprising *Alisma*, *Baldellia*, *Damasonium* and *Luronium* (Lehtonen 2017; Li et al. 2022), similar to our results (figs. 1 and 2). PhyloNet inferred three possible reticulations in Alismataceae (fig. 3). Two of them have minor edges with small  $\gamma$  (0.0226 and 0.0314, fig. 3), which supports ILS. The third reticulation revealed that *Burnatia* inherited about 11.6% of its genome from the MRCA of *Ranalisma* + *Albidella*. Moreover, *Ranalisma* and *Albidella* were phylogenetically close to *Hydrocleys* + *Limnocharis* (fig. 1). The close relationship between *Burnatia* and *Hydrocleys* + *Limnocharis* (Chen et al. 2012)



**FIG. 4.** Topology test results for the three main clades within Alismatales. Gene counts for topology tests of the three main clades using the AU test (a), RAxML likelihood scores (b), and BS support values (c).

could be due to introgression from the MRCA of *Ranalisma* + *Albidella* to *Burnatia*.

**Hydrocharitaceae:** Relationships among the major clades of Hydrocharitaceae have been inconsistent across multiple studies. The family comprised two major clades when *rbcl* and *matK* were used, one included *Halophila*, *Thalassia*, *Enhalus*, *Najas*, *Hydrilla*, and *Vallisneria*, another included *Egeria*, *Elodea*, *Blyxa*, *Ottelia*, *Lagarosiphon*, *Stratiotes*, and *Limnobium* + *Hydrocharis* (Tanaka et al. 1997). However, Chen et al. (2012) found that *Limnobium* + *Hydrocharis* clustered with the former clade, contradicting Tanaka et al. (1997). Les et al. (1997; 2006) recovered *Limnobium* + *Hydrocharis* to be sister to all other Hydrocharitaceae, but with poor support. In our ASTRAL analyses, *Limnobium* + *Hydrocharis* showed conflict (ICA = 0.5, QS = 0.14/0.67/0.99) (supplementary fig. S3, Supplementary Material online). The PhyloNet analysis recovered one reticulation with  $\gamma = 0.13$ , suggesting introgression from the MRCA of Hydrocharitaceae to the *Limnobium* + *Hydrocharis* ancestor. Thus, introgression possibly led to the phylogenetic inconsistency in the placement of *Limnobium* + *Hydrocharis*. *Stratiotes* was recovered as sister to all other Hydrocharitaceae taxa (Chen et al. 2012; Ross et al. 2016; Li et al. 2020), contradicting Tanaka et al. (1997), which clustered it with the clade formed by *Egeria*, *Elodea*, *Blyxa*, *Ottelia*, and *Lagarosiphon*. In our analyses, the clade including *Stratiotes* also showed conflict (ICA = 0.43, QS = 0.042/0/1). PhyloNet recovered a hybridization event ( $\gamma = 0.599$  and 0.401) between *Stratiotes* and the MRCA of the clade comprising *Elodea*, *Ottelia*, and *Blyxa*, which can explain the phylogenetic inconsistency of *Stratiotes*. Overall, phylogenetic conflict regarding *Limnobium* + *Hydrocharis* and *Stratiotes* can be explained by introgression and hybridization, respectively.

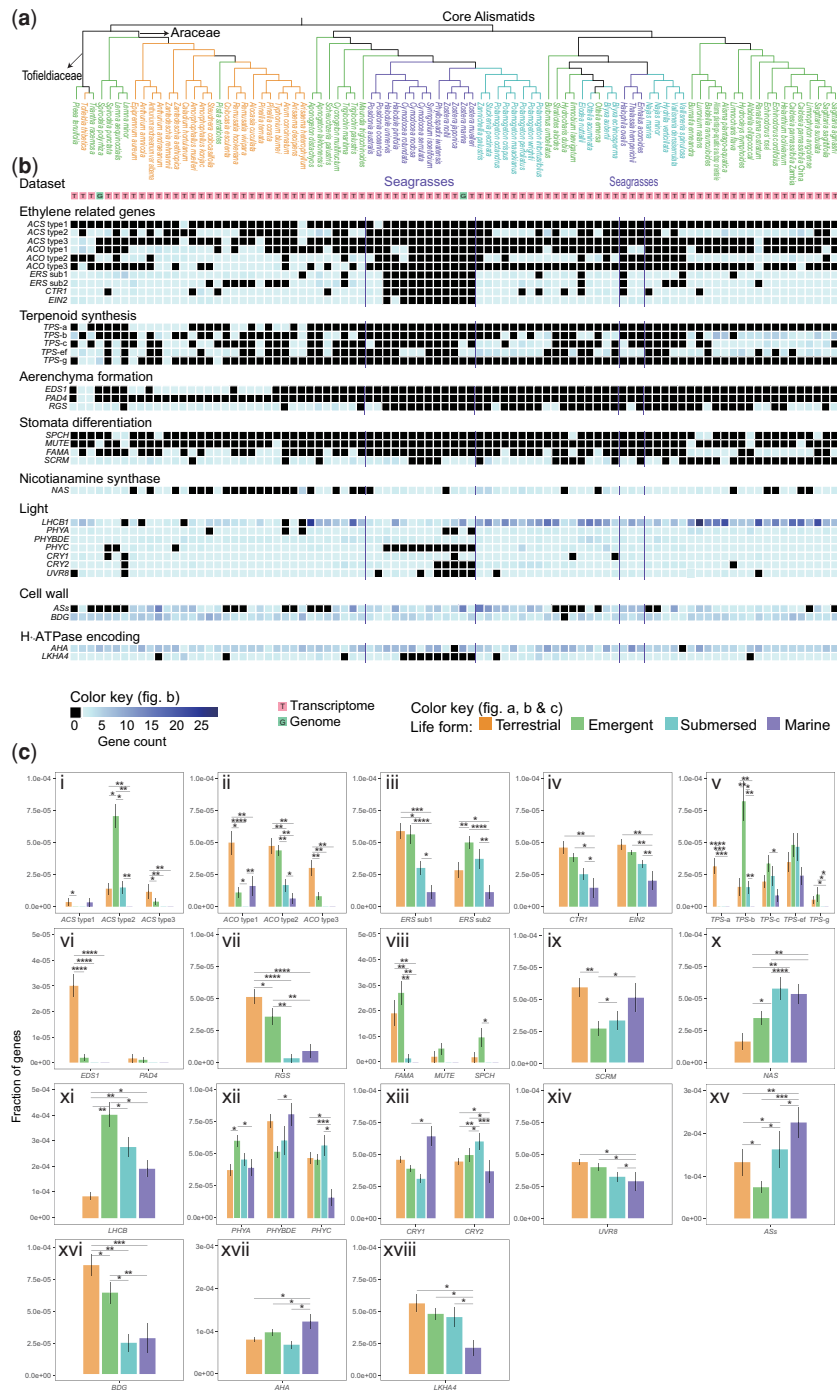
**Potamogetonaceae:** This is a taxonomically complex group with several hybrid taxa reported. Based on internal transcribed spacer and cp data, *Potamogeton wrightii* was reported as a parent lineage of multiple species, including

*Potamogeton biwaensis* (Iida et al. 2018) and *Potamogeton anguillanus* (Du et al. 2010). Our PhyParts and QS analyses suggested strong phylogenetic conflict within Potamogetonaceae (high proportion of discordant bipartitions, fig. 1; low ICA and QS scores, supplementary fig. S3, Supplementary Material online). The PhyloNet analyses found that *P. wrightii* is a hybrid between *Potamogeton in tortusifolius* and an extinct or unsampled taxon ( $\gamma = 0.214$ ; fig. 3). Moreover, *Zannichellia palustris* was recovered as a potential hybrid between *Stuckenia pectinata* and an extinct or unsampled taxon ( $\gamma = 0.386$ ). The phylogenetic conflict could be attributed to hybridization, but determining the extent of hybridization will require more comprehensive sampling.

**Araceae:** Introgression was detected (fig. 3) from an extinct/unsampled Araceae ancestor to the clade that included all of the sampled taxa except *Epipremnum* and *Spirodela* ( $\gamma = 0.119$ ). Based on cp markers and a few nuclear markers, introgression has been detected for Araceae, such as among *Cryptocoryne* species (Jacobsen and Ørgaard 2019). These results suggest that introgression affected relationships at the genus-level, which has not previously been reported in Araceae.

### Divergence Time Estimation

To get a timescale for Alismatales diversification, we performed a BEAST analysis using eight fossils and one secondary calibration (supplementary table S2, Supplementary Material online). Divergence time estimation dated the crown age of Alismatales around 135.7 Ma (95% highest posterior density: 119.4–153.9; supplementary fig. S5, Supplementary Material online). This age is slightly older than that in recent studies, such as 124 Ma in Givnish et al. (2018) and 122 Ma in Hertweck et al. (2015), but is younger than that of Li et al. (2019) at around 148 Ma. The stem node age of the seagrass clade within Hydrocharitaceae was around 62.8 Ma. The stem node ages of the seagrass lineages



**Fig. 5.** Phylogeny, gene copy counts and gene fraction of genes involved in adaptation to marine environments across Alismatales. (a) Species tree of Alismatales with the relationship for the three main clades manually fixed as (core Alismatids [Araceae, Tofieldiaceae]). (b) Gene copy counts for each species. (c) Gene fraction for each life-form. Gene fraction for a species was calculated by the number of corresponding genes divided by the number of total genes in a species. The gene fraction in each life-form is the average gene fraction of species in each life-form. Stars above columns represent the number of the four statistical tests (HyperG and PhyloGLM tests based on gene presence/absence data and gene copy-number data separately) with significant differences ( $P$ -value < 0.05).

Zosteraceae and Cymodoceaceae + Posidoniaceae were at 51.7 Ma (43.0–60.7) and 67.3 Ma (56.8–78.4), respectively. These ages are consistent with previous studies. For example, divergence time estimation with fossil calibrations found the mean stem node age of Zosteraceae at 53.5 Ma (Hertweck et al. 2015). However, these estimated ages are younger than the oldest seagrass fossils reported from the Late Cretaceous (Larkum et al. 2018), which could be

attributed to the view that some seagrasses have gone extinct over evolutionary time (Larkum et al. 2018).

### WGDs are Common Across Alismatales

To infer WGDs, we first calculated the proportion and number of duplicated genes for each node in the species tree using the tree-based methods of Yang et al. (2018);



*map\_dups\_mrca.py*) and Tree2GD. A high proportion of duplicated genes at a node could be due to a WGD at that node (type I duplication; Huang et al. 2016). However, a WGD that happened at one subclade within a node, but was lacking from another, could also lead to the detection of a high proportion of gene duplication at that node (type II or type III duplication; supplementary table S3, Supplementary Material online; Huang et al. 2016). Compared with type II and type III duplication, type I is more likely a real WGD (Huang et al. 2016). Tree2GD calculated the proportion of genes supporting each of the three types. For tree-based analyses, we used  $\geq 20\%$  duplicated genes (Yang et al. 2018) as a cut-off to identify WGD. Type of duplication (Huang et al. 2016) and evidence from the literature were also considered.

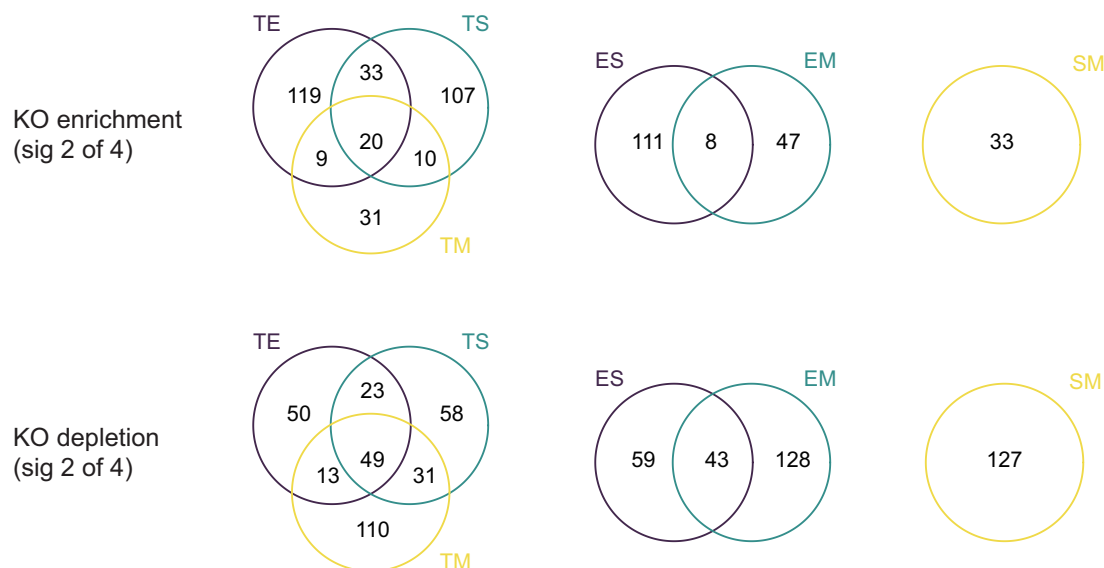
Results showed that three clades have  $\geq 20\%$  duplicated genes: core Alismatids (N30, 25%/2746), Araceae (N41, 20%/1129), and *Halodule* (Cymodoceaceae, 22%/NA; fig. 2 and supplementary fig. S6, Supplementary Material online). Core Alismatids had a higher proportion of type I duplications than the other two types (supplementary table S3, Supplementary Material online), which implies a WGD could have occurred (supplementary table S3, Supplementary Material online) at that node. Araceae has a higher proportion of type III duplication than type I. To verify the candidate WGD event at the MRCA of core Alismatids, Zosteraceae, and Araceae respectively, we reconstructed a species tree using 99 samples, then mapped 3,493 gene clusters to the tree and counted the number of duplicated gene clusters at each node of Alismatales using PhyPart (Smith et al. 2015). The results showed that 663 of 3,493 gene clusters were duplicated at the MRCA of the core Alismatids (supplementary fig. S7, Supplementary Material online). Among the 663 gene clusters, 148 (22.3%) gene pairs have syntenic evidence in *Zostera marina*. For the WGD at Zosteraceae (N21), 205 duplicated gene clusters were found, and 128 (62.4%) gene pairs have syntenic evidence in *Z. marina*. For the WGD at Araceae (N41), 552 duplicated gene clusters were found, and 203 (36.8%) gene pairs have syntenic evidence in *Sorbus intermedia* (supplementary fig. S7, Supplementary Material online). The WGD at the MRCA of Araceae was also reported by Wang et al. (2014) and An et al. (2019), which suggested that two rounds of WGDs in *Spirodela* occurred at c. 95 Ma, consistent with the age of Araceae (mean stem node age: 128.0 Ma; mean crown node age: 82.3 Ma). A WGD could have occurred at the stem node of *Halodule*, which had 22% duplicated genes, although the type of WGD could not be calculated given only two taxa in this clade. Overall, our results suggested there are three candidate WGD events, 1) the MRCA of core Alismatids, 2) the MRCA of Araceae, and 3) the MRCA of *Halodule*.

As an alternative method to infer WGDs, we also plotted the distribution of synonymous substitution rate ( $K_s$ ). We considered a  $K_s$  peak as a candidate WGD if it was clear in within-species  $K_s$  plot and between-species  $K_s$  plot. For

an obscure peak, evidence from the literature was required. The analyses suggested 16 candidate WGD events (fig. 2, supplementary fig. S8, Supplementary Material online). No  $K_s$  peak was identified for the candidate WGD at core Alismatids (N30) and Araceae (N41). The failure to detect the WGD at these nodes could be due to saturation effects in  $K_s$  estimation (Rabier et al. 2014). We discuss the major candidate WGDs identified based on  $K_s$  distribution below.

In total, 16 WGD events were identified using the  $K_s$  method. A  $K_s$  peak around 0.7 (supplementary table S4, Supplementary Material online) was recovered for 20 of 21 species at node N39 within Araceae (supplementary fig. S8-3, Supplementary Material online). The peak at 0.7 is younger than the diversification peak of Araceae ( $K_s = 1.15$ ), but older than the diversification of N39 ( $K_s = 0.52$ ; supplementary table S5, Supplementary Material online). N39 has 8% duplicated genes (fig. 2) and type I accounts for the highest proportion of duplication. Therefore, there should be a WGD for N39. This WGD was supported by a WGD in *Pinellia* (within the N39 clade) identified by Ren et al. (2018). The  $K_s$  plot recovered a clear peak of 0.15 for *Lemna minor* and *Lemna aequinoctialis* (Araceae), older than the split between the two species ( $K_s = 0.1$ ), but younger than the split between *Lemna* and *Spirodela* ( $K_s = 1.1$ ). *Lemna* has 12% duplicated genes (fig. 2). Therefore, a WGD event may have occurred for *Lemna*, consistent with the finding that at least one WGD occurred for the ancestor of *Lemna* after the split between *Lemna* and *Spirodela* (Van Hoeck et al. 2015). We also detected candidate WGD events for *Albidella olligococcal* (Alismataceae; supplementary fig. S8-1; supplementary table S4, Supplementary Material online), *Limnobium laevigatum* + *Hydrocharis dubia*, *Butomus umbellatus* (Butomaceae), *Stuckenia pectinata* (Potamogetonaceae), *Maundia triglochinosoides* (Maundiaceae), *Cycnogeton multifructum* (Juncaginaceae; supplementary fig. S8-2, Supplementary Material online), *Scheuchzeria palustris* (Scheuchzeriaceae), Aponogetonaceae, and Tofieldiaceae (supplementary fig. S8-3, Supplementary Material online). These WGD events were not detected when using tree-based methods, which may in part be because these methods need more comprehensive sampling.

Among the 16 WGD events identified using  $K_s$ , three occurred for seagrass lineages, namely Zosteraceae and Cymodoceaceae. Olsen et al. (2016) suggested that a WGD for *Zostera* (Zosteraceae) occurred around 65 Ma. Our analyses recovered a peak ( $K_s \approx 0.6$ ) for three of the six species in Zosteraceae. Between-species  $K_s$  plots showed that the peak is at the node of Zosteraceae (N21, 9% duplicated genes), which has a stem node age of 48–65 Ma. Therefore, the WGD at Zosteraceae identified here could correspond to the WGD at *Zostera* reported by Olsen et al. (2016). Cymodoceaceae has two candidate WGD events. One is at the MRCA of *Syringodium* + *Cymodocea* (N22), supported by a peak ( $K_s \approx 0.45$ ) for all four species in the clade and 7% duplicated genes at the node (supplementary fig. S8-2,



**Fig. 6.** Numbers of KOs shared among life-forms. For example, the terrestrial (T) life-form was compared with the three aquatic life-forms (emergent [E], submersed [S], and marine [M]) to identify the KOs enriched in each of the three life-forms separately. Next, the number of enriched KOs shared among the three aquatic life-forms was shown in the middle of the venn. “Sig 2 of 4” means two of the four statistical tests are significantly different ( $P$ -value  $< 0.05$ ).

Supplementary Material online). Another is at  $K_s \approx 0.25$ , positioned at the MRCA of the *Halodule* clade and supported by 22% duplicated genes. Reconstruction of the evolution of chromosome numbers for marine angiosperms suggested that WGD events had occurred in *Zosteraceae* and *Cymodoceae* (Silva et al. 2021).

In summary, we identified 18 candidate WGD events (supplementary table S4, Supplementary Material online). Among them, two (N30 and N41) were identified using tree-based methods, 15 were identified using  $K_s$ , and one (*Halodule*) was identified using both methods. WGD is believed to have evolutionary consequences for adaptation and is well documented for plants (Hollister 2015). It has been thought to have particular importance for adaptation to novel environments (Van de Peer et al. 2017). Recent genomic studies have provided empirical evidence for WGDs contributing to adaptation and diversification of mangroves that grow in aquatic high saline environments, such as tropical coastal swamps (Xu et al. 2017; Feng et al. 2021) as well as other novel or harsh environments for plants (Feng et al. 2020, Wang et al. 2019). The frequency of WGDs found among the Alismatales in this study provides further evidence that they could have contributed to aquatic habitat adaptations and the multiple inferred events among the seagrasses supports the role of WGD in the adaptation to high salinity extreme environments (Feng et al. 2021). However, WGD events do not appear to be associated with specific life-forms among the aquatic lineages, as all four life-forms in Alismatales have WGD events. More broad analyses using a phylogenomic context across hydrophyte lineages will provide more insight into the role of WGD in these adaptations.

### Genes Involved in Aquatic Adaptation

Previous studies have reported genes possibly related to the adaptation to aquatic environments (Olsen et al. 2016; Han et al. 2019). The evolutionary patterns of gene loss or gain within different aquatic plants remain poorly known. To investigate the evolutionary patterns of these genes (supplementary table S6, Supplementary Material online), we compared the numbers of gene copies for eight gene groups in Alismatales species under a phylogenetic context in relation to the four life-forms found in Alismatales. These results largely corroborate the results of Olsen et al. (2016) regarding the seagrass *Z. marina*. However, our in-depth phylogenetic sampling of Alismatales, incorporating multiple submersed and marine life-form plants, provides evidence that there is a great deal of variation within and among life-forms regarding gene gain and loss and some genes previously considered to be lost or directly related to adaptation to the aquatic environment should be reconsidered.

- 1) Ethylene is a critical plant hormone that controls various physiological responses (Broekgaarden et al. 2015). The biosynthesis and reception of ethylene mainly involve the genes *ACS*, *ACO*, *ETR1*, *ERS1*, *ETR2*, *ERS2*, and *EIN4*, *CTR1*, and *EIN2*. Olsen et al. (2016) attributed the loss of ethylene-related genes in *Z. marina* to the loss of stomata and the absence of insect herbivores in the sea. Our analyses found that all of the targeted ethylene-related genes were lost in *Z. marina* and *Zostera muelleri*, consistent with Olsen et al. (2016), and loss of these genes was generally more common among seagrasses

than freshwater taxa with notable exceptions in *Posidonia* and two Hydrocharitaceae genera (fig. 5b). The only freshwater taxon, *Elodea nuttallii*, with comparable losses is known for tolerance to brackish water (Thouvenot and Thiébaud 2018) (fig. 5b and c-i–iv). The loss of these genes is not notable in freshwater taxa in comparison with terrestrial Alismatales, suggesting it is not directly related to stomatal loss as previously suggested (Olsen et al. 2016).

There are seagrass pathogens in the sea (Sullivan et al. 2018) and animals feeding on seagrasses, such as snails (Rotini et al. 2018). The ethylene in seagrasses could respond to pathogens and grazers. It is also possible that ethylene plays a role in seagrass growth and development. Overall, complete loss of ethylene-related genes is not found in all seagrasses, but the loss is common across most lineages and more are lost in seagrasses than freshwater taxa (supplementary table S7, Supplementary Material online).

- 2) Terpenoids play an important role in plant–insect and plant–pathogen interactions, as well as plant development (Keeling and Bohlmann 2012). The terpene synthase (*TPS*) gene family examined here, comprises five subfamilies, a, b, c, e/f, and g (Jiang et al. 2019). *TPS* subfamily a and g were lost in all core Alismatids representing an ancestral state for the clade. Compared with terrestrial plants, subfamilies b, c, and e/f are more common in the emergent life-form plants (fig. 5b and c-v). This follows a similar pattern to the ethylene-related ACS type 2 and ERS subclade 2 gene (fig. 5b). Both ethylene and terpenoids are essential for a response to stress. Emergent plants need to be able to adapt to extreme shifts in the environment during their lifetime and be able to regulate metabolic processes under both inundated and emergent conditions. The retention of subfamilies b, c, and e/f and more ethylene-related genes in emergent plants could be attributed to the conditions they live in. This result is corroborated by a study in *Ranunculus*, the emergent aquatic plant *Ranunculus sceleratus* has more genes under positive selection than its terrestrial and submersed relatives (Zhao et al. 2016).
- 3) Aerenchyma-related genes. Enhanced Disease Susceptibility (*EDS1*) and Phytoalexin Deficient 4 (*PAD4*) are regulators for lysigenous aerenchyma formation by programmed cell death in *Arabidopsis* (Mühlenbock et al. 2008). Regulator of G protein signaling (*RGS*) is an inducer for aerenchyma formation in the absence of ethylene in maize (Steffens and Sauter 2010). Aerenchyma in aquatic plants is present constitutively in leaves and roots as an adaptation, thus *EDS1*, *PAD4*, and *RGS* may be expected to have an important role in the submersed and marine life-form. Our results showed that *EDS1* was present in 15 of the 21 terrestrial species of Alismatales, but were lost in 68 of the 70 aquatic samples of Alismatales (fig. 5b). *PAD4* was only present in two species of Tofieldiaceae. *RGS* was present in 20 of the 21 terrestrial species of Alismatales, but were lost in 44 of the 70 aquatic samples of the same group (fig. 5b). Interestingly, the Kyoto encyclopedia of genes and genomes (KEGG) orthology (abbreviated as KO) terms involving aerenchyma were enriched in submerged life-forms in our KO analyses. These results imply that the formation of aerenchyma in aquatic plants of Alismatales may be independent of *EDS1*, *PAD4*, and *RGS*, and these genes may be more important for an induced response rather than the constitutive production of aerenchyma and therefore lost.
- 4) Stomata-related genes. Stomata control transpiration and regulate gas exchange necessary for photosynthesis. *SPCH*, *MUTE*, and *FAMA* genes (together referred to as *SMF*), belong to the bHLH super-family and control the stomatal differentiation from meristematic mother cells (MMCs) to guard cells (GCs) (Pires and Dolan 2010; Ran et al. 2013). *ICE/SCRM* genes, also belong to bHLH; they control stomatal development and are also involved in the cold acclimation response and freezing tolerance (Pires and Dolan 2010). While stomata are ubiquitous in terrestrial and aquatic plants with emergent or floating leaves, they are absent from submersed leaves. We found that gene fraction for *SMF* genes decreased (or the genes were lost) in the submersed and marine life-form (fig. 5c-viii), which can be attributed to the absence of stomata. *SMF* genes were also absent from some species that have stomata, such as *S. polyrhiza*, a species with whole genomic data accessed in this study. The gene loss can be attributed to the multifunctionality of *SMF* genes, as a single *SMF* gene could control the specification from MMC to GC in early land plants (Ran et al. 2013). Alternatively, the function of *SMF* genes could have been supplemented by the *ICE/SCRM* genes (Ran et al. 2013; Olsen et al. 2016), which have a higher gene fraction in the submersed and marine life-forms than in the emergent life-form (fig. 5c-ix).
- 5) Nicotianamine, a metabolite synthesized in plants, has a critical role under iron and zinc deficiency in plants (Klatte et al. 2009; Schuler et al. 2012). Our results showed that the gene fraction of nicotianamine synthase (*NAS*) genes increased in submersed plants and seagrasses (fig. 5b and c-x). For example, 14 of the 16 marine species sampled have 1–2 gene copies (supplementary table S7, Supplementary Material online), while no copies were found in most terrestrial or emergent life-forms. The iron concentration of seawater is only 1% of that in rivers and groundwater (Firdaus et al. 2014) and iron deficiency was found to be common in Caribbean seagrasses, Red Sea seagrasses, and macroalgae (Duarte et al. 1995; Anton et al. 2018), which restricted their

productivity. Although the reason(s) for the increase of NAS copies in submersed plants is unclear, the increase in seagrasses could be attributed to the iron deficiency in seawater.

- 6) Light-related genes: The light-harvesting chlorophyll a/b-binding (LHCB) protein 1 belongs to the LHCB super-family, which regulates photosynthetic light-harvesting and several nonphotochemical processes (Xu et al. 2012). Our analyses suggest that *LHCB1* genes expanded in *Z. marina* as reported by Olsen et al. (2016). We also found *LHCB1* expanded at the MRCA of core Alismatids, however, it contracted among most other marine lineages (fig. 5b and c-xi; supplementary table S7, Supplementary Material online). These results do not support the enhancement of these gene families due to low light alone, but provide insight into variation in adaptive strategies for marine lineages.

Land plants evolved phytochromes, cryptochromes, and UVR8 to sense and capture light and overcome intense ultraviolet light in the terrestrial condition (Han et al. 2019). Our analyses suggested that *phytochrome C* is absent in most seagrasses (except *Posidonia* and Hydrocharitaceae), whereas *cryptochrome 2* (*CRY-2*) and *UVR-8* are absent in *Zostera*, as reported by Olsen et al. (2016), but present in most other seagrasses (fig. 5b and c-xiii & xiv). Olsen et al. (2016) attributed the loss of these genes to the differences in light absorption by plants through water. However, our results suggested that these genes are widely retained in submersed plants and seagrasses, despite their underwater life-form. Further exploring the effects of plant depth and diffusion of light through high salinity water will be important to explore gene function regarding these environments.

- 7) Cell wall-related genes. Sulfated polysaccharides (SPs) are cell wall components. They are mainly found in algae (McCandless and Craigie, 1979), and lost in land plants (Olsen et al. 2016). The SPs are synthesized by aryl sulfotransferases (AS). Olsen et al. (2016) found that *Z. marina* regained the aryl sulfotransferase genes as an adaptation to the marine environment. We found that all three major clades within Alismatales have regained AS genes (fig. 5b and c-xv), suggesting these genes could be involved in the adaptation to aquatic environments in general. The AS genes were also found in all seagrass lineages sampled here, corroborating the chemical studies that SPs exist in the seagrass lineages Cymodoceaceae, Hydrocharitaceae, Zosteraceae, and Ruppiaceae (Maeda et al. 1966; Aquino et al. 2011). SPs have also been identified in freshwater plants such as *Eichhornia crassipes*, *Hydrocotyle bonariensis*, and *Nymphaea ampla* (Dantas-Santos et al. 2012). No SPs have been reported from the freshwater plants within the core Alismatids. More chemical studies are needed to detect if these plants have SPs.

The cuticle covers aerial surfaces as a plant adaptation to terrestrial habitats (Kong et al. 2020). The biosynthesis of cuticles involves more than 32 genes (Kong et al. 2020). We found there is no significant difference in copy-number for these genes across Alismatales except for the *BODYGUARD* (*BDG*) gene. The *BDG* encodes an epidermis-specific cell wall-located protein that is required for the formation of the cuticle in *Arabidopsis* (Kurdyukov et al. 2006). The *BDG* has contracted in the submersed and marine life-forms (fig. 5c-xvi), which can be attributed to the much reduced or absent cuticle found in these plants (Enríquez et al. 2014).

- 8) Osmoregulation.  $H^+$ -ATPases (AHA) are ion pumps that use the energy of ATP hydrolysis to reduce  $Na^+$  influx, which plays a major role in salt stress tolerance. Olsen et al. (2016) found the gene encoding AHA expanded in *Z. marina* and our results suggested this is a common strategy in seagrasses, as the gene copy-number increased in all the four seagrass clades (supplementary table S7, Supplementary Material online and fig. 5c-xvii).

Although a comprehensive phylogenetic approach was used to investigate evolutionary adaptation to the aquatic habit in this study, some genes that appear “lost” might be the product of under-expression or some other mechanisms of gene expression suppression. Further studies with more genomic data are needed to investigate these adaptations.

### KO Terms Related to Aquatic Environment Adaptations

Analyses were conducted to discover KO terms not previously reported to be involved in the adaptation to aquatic environments. We investigated KOs that significantly enriched or depleted in the emergent, submersed and marine life-form separately (fig. 6). Most of the results are concordant with the gene evolution analyses. For example, KO terms involved in nicotianamine synthesis were enriched in the submersed life-form and the marine life-form; KO terms involving ethylene reception, phytochrome, or photosystem were depleted in the submersed and marine life-form. Moreover, we identified some depleted or enriched KOs in aquatic life-forms (e.g., elongation factor 3; supplementary table S8, Supplementary Material online). These KOs have not been considered in aquatic adaptation in previous studies.

Compared to the terrestrial life-form, the three aquatic life-forms shared 20 enriched KO terms (fig. 6; supplementary table S9a, Supplementary Material online), which could be related to adaptation to the aquatic condition. For example, a KO term involving *CYP4X* (cytochrome P450 family 4) was enriched, although little is known about the function in plants of this KO or the others discovered. The submersed and marine life-form shared ten enriched KOs that are not shared by the emergent life-form. Among the ten KOs, one is involved in

nicotianamine synthesis that plays a critical role under conditions of iron and zinc deficiency in plants (Klatte et al. 2009; Schuler et al. 2012). Another encodes the elongation factor 3, which has multiple functions in plant development and cell wall biosynthesis. Compared to the emergent life-form, the submersed and marine life-form shared eight enriched KOs (supplementary table S9b, Supplementary Material online). One of the KOs is the CTF4 (chromosome transmission fidelity protein). The ortholog of CTF4 in rice sustains normal cell cycle progression (Zhang et al. 2021). The enrichment of this KO could be attributed to the well-developed aerenchyma in the submersed and marine life-forms. Compared to the freshwater submersed life-form, the marine life-form had 33 enriched KOs (supplementary table S9c, Supplementary Material online). For example, three of the KOs were related to anthocyanidin metabolism and CYP86A1 encodes the long-chain fatty acid omega-monooxygenase. The function of these KOs in plants needs further investigation to determine how they may relate to marine environment adaptation.

Compared to the terrestrial life-form, the three aquatic life-forms shared 49 depleted KOs (fig. 6; supplementary table S9d, Supplementary Material online). The submersed and marine life-form shared 31 KOs that are not shared by the emergent life-form. Compared to the emergent life-form, the submersed and marine life-form shared 43 depleted KOs (supplementary table S9e, Supplementary Material online). Compared to the submersed life-form, the marine life-form has 127 depleted KOs (supplementary table S9f, Supplementary Material online). Five of the depleted KOs in marine life-form plants involve ethylene reception (*ETR*, *ERS*, *EIN2*, and *EIN3*) or phytochrome, which is consistent with Olsen et al. (2016) and Lee et al. (2018). The depletion in the marine life-form also involves five photosystem KOs (*psaH*, *psbO*, *psbW*, *psbY*, and *psbR*), likely related to the light condition in the marine environment. The difference in light quality and intensity between submerged freshwater and marine habitats has not received much attention, in relation to gene function. The results of this research suggest differences in these habitats could have consequences for the evolution of photosystem gene function.

### Why Are Submersed Marine Angiosperms Only Found in Alismatales?

The angiosperms that grow entirely submersed in marine environments are all from Alismatales (Larkum et al. 2018). Here, we explore the possible reasons that contributed to this phenomenon: 1) The challenge of transitioning from a terrestrial to a marine environment should require a greater array of adaptations than transitioning from freshwater to seawater (Vermeij and Dudley 2000). Freshwater plants have advantages over terrestrial plants regarding the adaptation to the marine environment. Although the transition from terrestrial to sea occurred in mangroves (Takayama et al. 2021), they do not grow

entirely submersed, but live in coastal saline or brackish water with emergent leaves, thus reducing the need for specialized adaptations related to photosynthesis and nutrient transport. 2) Aquatic angiosperms could be divided into two categories: aquatic taxa in terrestrial orders and aquatic orders (Nymphaeales, Ceratophyllales, Acorales, and Alismatales; Du and Wang 2016). During the Late Cretaceous, the Tethys seas provided vast shallow water areas (Larkum et al. 2018), which are suitable for the origination of seagrasses. On the contrary, most of the aquatic taxa in terrestrial orders originated during the Late Cenozoic and later, thus are younger than the existence of the Tethys seas (e.g., aquatic *Ranunculus* [He et al. 2021]). Therefore, the aquatic taxa in terrestrial orders had fewer chances of adaptation to the sea. 3) Alismatales has the largest number of aquatic species and genera among all the aquatic groups, thus having more chances of adaptation to the marine environment. For example, although Ceratophyllales originated during the Cretaceous, it includes only four species and one family (The Plant List). Moreover, compared with Nymphaeales, Ceratophyllales, and Acorales, Alismatales is relatively diverse in form. For example, Alismatales have great diversity in leaf morphology (ribbon-like, broad and circular, and short linear) and reproductive system (dioecy, monoecy, hermaphroditism, and dioecy and monoecy; Flora of China: <http://www.efloras.org/>, accessed February 2022). All of the seagrasses except for *Halophila* have ribbon-like leaves, which are not common among other submersed aquatic lineages. 4) Alismatales, with its broad diversity, may have had a better chance of having the genomic predisposition for adaptation to the marine environment. For example, many of the genes hypothesized as important for marine environments can be found in submersed forms of freshwater Alismatids, such as the increase in Nicotianamine genes and *LHCB1*. The combination of functional gains and losses could lead to the specific genomic conditions for evolution to a marine environment that is then phylogenetically constrained to this lineage. Overall, a combination of these factors likely explains why all Angiosperms that grow in a submersed marine environment are all from Alismatales. Genomes of some early evolving aquatic plants such as *Ceratophyllum demersum* and *Euryale ferox* (Yang et al. 2020), and mangrove plants (Feng et al. 2021) have been recently sequenced and can be included to better elucidate the marine and freshwater adaptation in future studies.

## Materials and Methods

### Sampling and Sequencing

Ninety-one Alismatales samples from 89 species (*Alisma plantago-aquatica* and *Caldesia parnassifolia* were each represented by two samples) were included. Twenty-one of the Alismatales samples were terrestrial, 37 were emergent/floating aquatic, 17 were freshwater submersed, and 16 were seagrasses. Following the Angiosperm Phylogeny Group IV (The

Angiosperm Phylogeny Group 2016), *Amborella trichopoda* (Amborellaceae), *Nymphaea colorata* (Nymphaeaceae), *Austrobaileya scandens* (Austrobaileiaceae), and *Acorus americanus* (Acoraceae) were included as outgroups. Of the 95 samples, 4 were from published whole-genome sequencing data, 32 were RNA-seq reads accessed from NCBI SRA, and 59 were RNA-seq reads ( $2 \times 150$  bp) generated in this study (supplementary table S1, Supplementary Material online).

### Transcriptomic Data Processing and Assembly

Raw read processing and organelle and over-represented reads removal followed Morales-Briones et al (2021). After assembly with Trinity v2.8.5 (Haas et al. 2013), low-quality transcripts and chimeric transcripts were removed before clustering into putative genes. Finally, transcripts were translated with TransDecoder v5.5.0 (accessed from <https://github.com/TransDecoder>, September 2019) and transcript redundancy was assessed. See Supplementary Material online for detailed methods.

### Homology Inference, cp Genes Assembly, and Phylogenetic Analyses

To infer homologs, an all-by-all BLASTN search was performed for CDSs of all 95 taxa. Then putative homolog groups were clustered using MCL v1.37 (van Dongen 2000). Orthologs were inferred using the MO method in Yang and Smith (2014). A maximum likelihood (ML) tree using the concatenated orthologs was built using RAxML v8.2.12 (Stamatakis 2006; all ML trees in this study were inferred using RAxML with the GTRGAMMA model and 200 rapid BS replicates unless noted). A species tree was inferred from gene trees of orthologs with ASTRAL v5.7.3 (Zhang et al. 2018). Moreover, a species tree was built with 99 samples, which included genomic data of six Alismatales species, transcriptomes, and five outgroups. This species tree was used to verify possible WGD events, accompanied by syntenic analyses (Supplementary Material online).

We accessed cp genomes for 13 species of Alismatales from NCBI GENBANK (supplementary table S1, Supplementary Material online). We also assembled cp CDSs using organelle reads from RNA-seq. Phylogenetic trees were inferred using RAxML and ASTRAL separately. See Supplementary Material online for detailed methods.

### Assessing Phylogenetic Conflict and Phylogenetic Networks

To investigate phylogenetic conflict, each of the nuclear gene trees was rooted using *Amborella trichopoda*. Then, each of the rooted trees, skipping nodes with  $BS \leq 50\%$ , was mapped against the ASTRAL tree using PhyParts (Smith et al. 2015) to calculate the concordant/conflicting bipartitions and the ICA values. To distinguish the nodes lacking support from conflicting support, QS (Pease et al. 2018) with 100 replicates was carried out with the

ASTRAL tree and the concatenated alignment of all the nuclear orthologs.

To infer species networks that account for both reticulation (e.g., hybridization) and ILS, Bayesian inference was carried out with PhyloNet for five groups, namely Alismatales (includes 10 species), Alismataceae (14), Hydrocharitaceae (17), Potamogetonaceae (11), and Araceae (19). See Supplementary Material online for detailed methods.

### Phylogenetic Hypothesis Testing for Alismatales

Phylogenetic analyses detected strong conflict for the relationship between core Alismatids, Araceae, and Tofieldiaceae. To test which topology is the most favored, we carried out AU tests (Shimodaira 2002). The 798 gene trees from PhyloNet analyses of Alismatales were used. Site-wise log-likelihood scores for each of three alternative topologies of Alismatales were calculated using RAxML. Then, AU tests were carried out using CONSEL (Shimodaira and Hasegawa 2001). As complementary analyses to the AU tests, we counted the number of genes supporting each of the three topologies using likelihood scores from RAxML. We also counted the number of genes supporting each of the three topologies using raw BS values and the BS values  $\geq 75\%$  in gene trees separately.

To test if the discordance among the three main clades could be explained by polytomies instead of dichotomies, we carried out the quartet-frequency-based polytomy tests by Sayyari and Mirarab (2018) as implemented in ASTRAL. We filtered the 1,005 orthologs to include only orthologs that were present in all nine Alismatales species and two outgroups. In this way, 602 orthologs were extracted. ML gene trees were inferred using RAxML separately. Then, the polytomy test was performed using gene trees with raw BS values. Because this test is sensitive to gene tree error, we performed a second polytomy test using gene trees whose branches with BS values  $\leq 75\%$  were collapsed.

### Divergence Time Estimation

Divergence time was estimated using BEAST v2.5.2 (Bouckaert et al. 2014). The topology of Alismatales was fixed based on the ML tree inferred from the 1,005-nuclear ortholog matrix. Eight fossil calibration points and one secondary point were applied (supplementary table S2, Supplementary Material online). The top 40 clock-like orthologs were extracted from the nuclear orthologs and used for divergence time estimation. See Supplementary Material online for detailed methods.

### Inference of Genome Duplication

To investigate potential WGD events, we applied tree-based and  $K_s$  methods. 1) Putative WGDs were assessed by counting the proportion of duplicated genes at nodes. First, ML trees were constructed for the homologs inferred from MCL. SH-like support was assessed using the SH-like test (Anisimova et al. 2011) with RAxML. Then 3,492 gene

clusters including  $\geq 30$  taxa and SH-like support  $> 80$  were extracted. Finally, gene duplication events were recorded by mapping the rooted gene clusters to the ASTRAL tree of Alismatales. These analyses were done using scripts from Yang et al. (2018). Putative WGDs were also detected using Tree2GD v1.0.37 (accessed from <https://github.com/Dee-chen/Tree2gd> in December 2020). To reduce computing time, 48 samples were selected to keep the main clades from the tree of the full data set and used for the Tree2GD analysis. 2) Peptide (PEP) and CDS from paralogous pairs were used to plot within-species  $K_s$  distribution using the pipeline in Yang et al. (2015).  $K_s < 0.02$  were filtered. According to preliminary analyses, most Alismatales species have a peak at  $K_s \approx 0.05$  or 0.1. We ignored this peak for most species, as it is difficult to distinguish between a recent WGD and a recent retrotransposition event (Tiley et al. 2018). To determine the position of a  $K_s$  peak, between-species  $K_s$  was plotted using the pipeline in Wang et al. (2019) with minor changes. The  $K_s$  for nine backbone nodes of Alismatales were estimated using multiple species pairs in each node (supplementary table S5, Supplementary Material online). A  $K_s$  plot showed that the stem node of Alismatales corresponded to a peak of  $K_s \approx 1.45$  (supplementary fig. S8-1a, Supplementary Material online). Hence, we ignored the peaks earlier than 1.45 for all species, as they may correspond to WGDs earlier than Alismatales. Finally, three candidate WGD events were verified using syntenic genes in genome (see Supplementary Material online for detailed methods).

### Identification of KOs and Genes Related to Adaptation to Aquatic Environments

We prepared two data matrices for these analyses. 1) KO annotation for PEPs/genes of all 91 Alismatales samples was obtained using eggNOG-Mapper v2 (Cantalapiedra et al. 2021). Each query was searched against eggNOG sequences using diamond (*-m diamond*) to fix the taxonomic scope in plants (*-tax\_scope 33090*). 1,718,373 (70.12%) genes out of 2,450,555 predicted proteins were successfully annotated using the KO database (Kanehisa et al. 2016). Then, PEPs/genes were clustered according to the KO annotation. Last, a matrix including gene copy-number per KO term was generated. 2) According to Olsen et al. (2016) and Han et al. (2019), seven groups of genes involved in ethylene reception and reaction, terpenoid synthesis, aerenchyma formation, stomata differentiation, light, cell wall, and  $H^+$ -ATPase were selected to examine the evolution in Alismatales (supplementary table S6, Supplementary Material online). Leukotriene A4 hydrolase (LTA4H) and nicotianamine synthase (NS), selected from our KO analyses, were also tested. A database consisting of PEPs of the 95 samples and a few other species, such as *Arabidopsis*, was generated. For each of these genes, at least one sequence was used as a “bait” to search against the database with SWIPE v2.1.0 (Rognes 2011). A maximum of 12 hits were kept for each species. Candidate genes were filtered by using PFAM domains or InterProScan domains. After that,

the candidate genes were aligned using MAFFT v7.407 (Katoh and Standley 2013), and low occupancy columns were trimmed using Phyutility v.2.7.1 (*-clean 0.01*). An ML tree for each of these genes was built using RAXML. Subdivisions for each gene family were assigned according to positions of outgroups and previous studies. Finally, a Hypergeometric test (Hyperg) and PhyloGLM test (Levy et al. 2018) were applied to check if a KO or gene family is significantly associated with the life-form. Two versions of Hyperg and PhyloGLM tests were used, based on either gene presence/absence data (hypergbn and phyloglmbn) or gene copy-number data (hypergcn and phyloglmcn) (Levy et al. 2018). Hyperg ignores the phylogenetic structure of a data set, while PhyloGLM takes it into account (Levy et al. 2018).

### Conclusion

This study investigated the source of phylogenetic conflict within Alismatales, and adaptations to aquatic environments. The species network inference and hypothesis testing recovered a relationship: ([Tofieldiaceae, Araceae], core Alismatids). Phylogenetic conflict among the three main clades could be attributed to introgression and incomplete lineage sorting. Our analyses also suggested 18 candidate WGD events in Alismatales, including one at the MRCA of core Alismatids. We found that common strategies in freshwater submersed species and seagrasses included loss of genes involved in terpenoid synthesis and stomatal differentiation (SMF subfamily) and expansion of genes involved in nicotianamine and sulfated polysaccharides synthesis. Another adaptive strategy is the expansion of  $H^+$ -ATPases genes in seagrasses. Lineage-specific strategies included the loss of light-related genes (harvest, sense, and UV resistance) in Zosteraceae that were retained in other seagrasses. We also identified several KOs that might be related to adaptation to aquatic environments.

The molecular mechanisms for adaptation to novel environments, such as freshwater and marine aquatic habitats, are complex. This study sheds light on the potential genomic mechanism involved in the evolution of early monocots and adaptation to aquatic environments. While a wide sampling of aquatic lineages and habits here provides means to assess evolutionary patterns within Alismatales, a more complete sampling of Araceae, in particular, can better assess the phylogenetic relationship within Alismatales. Moreover, future studies investigating aquatic adaptation could benefit from sampling a wider array of plant groups with multiple independent origins of aquatic taxa.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (grant number

XDB31000000), the National Natural Science Foundation of China (grant numbers. 31670226 and 32070231), and the Sino-Africa Joint Research Center. We appreciate Duoyuan Chen, Ning Wang, and Tao Shi for suggestions on WGD analyses; Susanne S. Renner, Kuo Liao, Zhi-Zhong Li, and Yu Cao for plant sampling; Samuli Lehtonen for plant identification; the three anonymous reviewers for their constructive suggestions.

## Author Contributions

L.-Y.C., J.-M.C., and Q.-F.W. designed research; L.-Y.C., M.L.M., F.L., G.-W.H., J.-M.C., and Q.-F.W. contributed to the taxon sampling and sequencing; L.-Y.C. and D.F.M.-B. performed DNA/RNA data processing and phylogenetic analyses; L.-Y.C. and J.-M.C. performed analyses on WGD analyses and molecular dating; B.L. and L.-Y.C. performed analyses on gene evolution; L.-Y.C., D.F.M.-B., and B.L. prepared the figures and tables. L.-Y.C. drafted the manuscript. All authors revised the manuscript.

## Data Availability

The RNA-seq reads underly this article are available on the GenBank SRA databases under the accession numbers provided in [supplementary table S1](#). High-resolution plant photos, assembled transcriptomes, sequence alignments, phylogenetic trees, and KO terms are available at the figshare: <https://doi.org/10.6084/m9.figshare.16967767>.

## References

The Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* **181**:1–20.

An D, Zhou Y, Li C, Xiao Q, Wang T, Zhang Y, Wu Y, Li Y, Chao D-Y, Messing J, et al. 2019. Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of *Spirodela*. *Proc Natl Acad Sci U S A.* **116**:18893–18899.

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* **60**:685–699.

Anton A, Hendriks IE, Marba N, Krause-Jensen D, Garcias-Bonet N, Duarte CM. 2018. Iron deficiency in seagrasses and macroalgae in the Red Sea is unrelated to latitude and physiological performance. *Front Mar Sci.* **5**:74.

Aquino RS, Grativol C, Mourão PA. 2011. Rising from the sea: correlations between sulfated polysaccharides and salinity in plants. *PLoS One* **6**:e18862.

Azuma H, Tobe H. 2011. Molecular phylogenetic analyses of Tofieldiaceae (Alismatales): family circumscription and intergeneric relationships. *J Plant Res.* **124**:349–357.

Billet K, Genitoni J, Bozec M, Renault D, Barloy D. 2018. Aquatic and terrestrial morphotypes of the aquatic invasive plant, *Ludwigia grandiflora*, show distinct morphological and metabolomic responses. *Ecol Evol.* **8**:2568–2579.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp Biol.* **10**: e1003537.

Broekgaarden C, Caarls L, Vos IA, Pieterse CM, Van Wees SC. 2015. Ethylene: traffic controller on hormonal crossroads to defense. *Plant Physiol.* **169**:2371–2379.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* **38**:5825–5829.

Chao D-Y, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* **341**:658–659.

Chen L-Y, Chen J-M, Gituru RW, Wang Q-F. 2012. Generic phylogeny, historical biogeography and character evolution of the cosmopolitan aquatic plant family Hydrocharitaceae. *BMC Evol Biol.* **12**:30.

Cook CDK. 1990. *Aquatic plant book*. Hague: SPB Academic Publishing.

Dantas-Santos N, Gomes DL, Costa LS, Cordeiro SL, Costa MSSP, Trindade ES, Franco CRC, Scortecchi KC, Leite EL, Rocha HAO. 2012. Freshwater plants synthesize sulfated polysaccharides: heterogalactans from water hyacinth (*Eichornia crassipes*). *Int J Mol Sci.* **13**:961–976.

Du Z-Y, Wang Q-F. 2016. Phylogenetic tree of vascular plants reveals the origins of aquatic angiosperms. *J Syst Evol.* **54**: 342–348.

Du Z-Y, Yang C-F, Chen J-M, Guo Y-H. 2010. Identification of hybrids in broad-leaved *Potamogeton* species (Potamogetonaceae) in China using nuclear and chloroplast DNA sequence data. *Plant Syst Evol.* **287**:57–63.

Duarte CM, Martín M, Margarita G. 1995. Evidence of iron deficiency in seagrasses growing above carbonate sediments. *Limnol Oceanogr.* **40**:1153–1158.

Eguchi S, Tamura MN. 2016. Evolutionary timescale of monocots determined by the fossilized birth-death model using a large number of fossil records. *Evolution* **70**:1136–1144.

Enríquez S, Schubert N. 2014. Direct contribution of the seagrass *Thalassia testudinum* to lime mud production. *Nat Commun.* **5**:3835.

Feng X, Li G, Xu S, Wu W, Chen Q, Shao S, Liu M, Wang N, Zhong C, He Z, et al. 2021. Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*. *New Phytol.* **231**:2346–2358.

Feng C, Wang J, Wu L, Kong H, Yang L, Feng C, Wang K, Rausher M, Kang M. 2020. The genome of a cave plant, *Primulina huaijiensis*, provides insights into adaptation to limestone karst habitats. *New Phytol.* **227**:1249–1263.

Firdaus ML, Alwi W, Trinoveldi F, Rahayu I, Rahmidar L, Warsito K. 2014. Determination of chromium and iron using digital image-based colorimetry. *Procedia Environ Sci.* **20**:298–304.

Friis EM, Pedersen KR, Crane PR. 2010. Diversity in obscurity: fossil flowers and the early history of angiosperms. *Philos Trans R Soc Lond B Biol Sci.* **365**:369–382.

Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VK, Saarela JM, Sass C, Iles WJ, De Sousa DJL, Leebens-Mack J, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am J Bot.* **105**: 1888–1910.

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

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

Han X, Chang X, Zhang Z, Chen H, He H, Zhong B, Deng XW. 2019. Origin and evolution of core components responsible for monitoring light environment changes during plant terrestrialization. *Mol Plant.* **12**:847–862.



- He J, Lyu R, Luo Y, Xiao J, Xie L, Wen J, Li W, Pei L, Cheng J. 2021. A phylotranscriptome study using silica gel-dried leaf tissues produces an updated robust phylogeny of Ranunculaceae. *BioRxiv*. <https://doi.org/10.1101/2021.07.29.454256>
- Hertweck KL, Kinney MS, Stuart SA, Maurin O, Mathews S, Chase MW, Gandolfo MA, Pires JC. 2015. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Bot J Linn Soc*. **178**:375–393.
- Hollister JD. 2015. Polyploidy: adaptation to the genomic environment. *New Phytol*. **205**: 1034–1039.
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol*. **33**:394–412.
- Iida S, Ashiya M, Kadono Y. 2018. The hybrid origin of *Potamogeton biwaensis* Miki, an endemic submerged plant in Lake Biwa. *Japan Aquat Bot*. **150**:23–26.
- Iles WJ, Smith SY, Graham SW, Wilkin P, Mayo S. 2013. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. In: Wilkin P, Mayo SJ, editors. *Early events in monocot evolution*. Cambridge: Cambridge University Press. p. 1–28.
- Jacobsen N, Ørgaard M. 2019. Natural hybridization – recombination – an ever-ongoing process. *Thai For Bull (Botany)* **47**:19–28.
- Jiang S-Y, Jin J, Sarojam R, Ramachandran S. 2019. A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol Evol*. **11**: 2078–2098.
- Jung J, Lee SC, Choi H-K. 2008. Anatomical patterns of aerenchyma in aquatic and wetland plants. *J Plant Biol*. **51**:428–439.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. **44**:D457–462.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**:772–780.
- Keeling CI, Bohlmann J. 2012. Plant terpenoids. In: Natanya Civjan, editor. *Natural products in chemical biology*. Hoboken: John Wiley & Sons, Inc. p. 1–10.
- Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P. 2009. The analysis of *Arabidopsis* nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. *Plant Physiol*. **150**:257–271.
- Kong L, Liu Y, Zhi P, Wang X, Xu B, Gong Z, Chang C. 2020. Origins and evolution of cuticle biosynthetic machinery in land plants. *Plant Physiol*. **184**:1998–2010.
- Kurdyukov S, Faust A, Nawrath C, Bär S, Voisin D, Efremova N, Franke R, Schreiber L, Saedler H, Métraux J-P, et al. 2006. The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. *Plant Cell* **18**:321–339.
- Larkum AW, Waycott M, Conran JG. 2018. Evolution and biogeography of seagrasses. In: Larkum AWD, Kendrick GA, Ralph PJ, editors. *Seagrasses of Australia-structure, ecology and conservation*. Switzerland AG: Springer International Publishing. p. 3–29.
- Lee H, Golitz AA, Bayer PE, Jiao Y, Tang H, Paterson AH, Sablok G, Krishnaraj RR, Chan C-KK, Batley J, et al. 2016. The genome of a southern hemisphere seagrass species (*Zostera muelleri*). *Plant Physiol*. **172**:272–283.
- Lee H, Golitz AA, Bayer PE, Severn-Ellis AA, Chan C-KK, Batley J, Kendrick GA, Edwards D. 2018. Genomic comparison of two independent seagrass lineages reveals habitat-driven convergent evolution. *J Exp Bot*. **69**:3689–3702.
- Lehtonen S. 2017. Splitting *Caldesia* in favour of *Albidella* (Alismataceae). *Aust Syst Bot* **30**:64–69.
- Les DH, Cleland MA, Waycott M. 1997. Phylogenetic studies in Alismatales. II: evolution of marine angiosperms (seagrasses) and hydrophyly. *Syst Bot*. **22**:443–463.
- Les DH, Moody ML, Soros CL. 2006. A reappraisal of phylogenetic relationships in the monocotyledon family Hydrocharitaceae (Alismatales). *Aliso* **22**:211–230.
- Levy A, Gonzalez IS, Mittelviehhaus M, Clingenpeel S, Paredes SH, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al. 2018. Genomic features of bacterial adaptation to plants. *Nat Genet*. **50**:138–150.
- Li Z-Z, Lehtonen S, Martins K, Gichira AW, Wu S, Li W, Hu G-W, Liu Y, Zou C-Y, Wang Q-F, et al. 2020. Phylogenomics of the aquatic plant genus *Ottelia* (Hydrocharitaceae): implications for historical biogeography. *Mol Phylogenet Evol*. **152**:106939.
- Li Z-Z, Lehtonen S, Martins K, Wang Q-F, Chen J-M. 2022. Complete genus-level plastid phylogenomics of Alismataceae with revisited historical biogeography. *Mol Phylogenet Evol*. **166**:107334.
- Li H-T, Luo Y, Gan L, Ma P-F, Gao L-M, Yang J-B, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, et al. 2021. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol*. **19**:232.
- Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants* **5**:461–470.
- Liang J, Ma Q, Yang Z. 2019. The first complete chloroplast genomes of two Alismataceae species, and the phylogenetic relationship under order Alismatales. *Mitochondrial DNA Part B* **4**:122–123.
- Luo Y, Ma P-F, Li H-T, Yang J-B, Wang H, Li D-Z. 2016. Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. *Genome Biol Evol*. **8**: 932–945.
- Maeda M, Koshikawa M, Nisizawa K, Takano K. 1966. Cell wall constituents, especially pectic substance of a marine phanerogam *Zostera marina*. *Bot Mag*. **79**:422–26.
- McCandless E, Craigie J. 1979. Sulfated polysaccharides in red and brown algae. *Annu Rev Plant Physiol*. **30**:41–53.
- Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, Timoneda A, Yim WC, Cushman JC, Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae sl. *Syst Biol*. **70**:219–235.
- Mühlenbock P, Szechynska-Hebda M, Ptaszczyca M, Baudo M, Mateo A, Mullineaux PM, Parker JE, Karpińska B, Karpiński S. 2008. Chloroplast signaling and *LESION SIMULATING DISEASE1* regulate crosstalk between light acclimation and immunity in *Arabidopsis*. *Plant Cell* **20**:2339–2356.
- Nauheimer L, Metzler D, Renner SS. 2012. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytol*. **195**:938–950.
- Nowak JS, Ono J, Cronk QC. 2010. Anatomical study of an aquatic mustard: *Subularia aquatica* (Brassicaceae). *Aquat Bot*. **93**:55–58.
- Olsen JL, Rouzé P, Verhelst B, Lin Y-C, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F, et al. 2016. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**:331–335.
- Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am J Bot*. **105**:385–403.
- Petersen G, Seberg O, Cuenca A, Stevenson DW, Thadeo M, Davis JJ, Graham S, Ross TG. 2016. Phylogeny of the Alismatales (Monocotyledons) and the relationship of Acorus (Acorales?). *Cladistics* **32**:141–159.
- Pires N, Dolan L. 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol*. **27**:862–874.
- Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol*. **20**:1–23.
- Rabier C-E, Ta T, Ané C. 2014. Detecting and locating whole-genome duplications on a phylogeny: a probabilistic approach. *Mol Biol Evol*. **31**:750–762.

- Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. 2013. Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods* **10**:128–132.
- Ran J-H, Shen T-T, Liu W-J, Wang X-Q. 2013. Evolution of the bHLH genes involved in stomatal development: implications for the expansion of developmental complexity of stomata in land plants. *PLoS One* **8**:e78997.
- Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole-genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant*. **11**: 414–428.
- Rognes T. 2011. Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics* **12**:1–11.
- Ross TG, Barrett CF, Soto Gomez M, Lam VK, Henriquez CL, Les DH, Davis JJ, Cuenca A, Petersen G, Seberg O, et al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* **32**: 160–178.
- Rotini A, Tibbetts IR, Migliore L, Rossini RA. 2018. The trade-off between digestibility and phenol content influences the food choice of the obligate seagrass-feeding neritid snail *Smaragdia souverbiana*. *J Molluscan Stud.* **84**:12–18.
- Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol Biol Evol.* **31**:1261–1271.
- Sayyari E, Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* **9**:132.
- Schuler M, Rellán-Álvarez R, Fink-Straube C, Abadía J, Bauer P. 2012. Nicotianamine functions in the phloem-based transport of iron to sink organs, in pollen development and pollen tube growth in *Arabidopsis*. *Plant Cell* **24**:2380–2400.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* **51**:492–508.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246–1247.
- Silva SL, de Carvalho R, Magalhães KM. 2021. Chromosomal evolution in seagrasses: is the chromosome number decreasing? *Aquat Bot.* **173**:103410.
- Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* **15**:1–15.
- Solis-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Mol Biol Evol.* **34**:3292–3298.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–2690.
- Steffens B, Sauter M. 2010. G proteins as regulators in ethylene-mediated hypoxia signaling. *Plant Signal Behav.* **5**:375–378.
- Sullivan BK, Trevathan-Tackett SM, Neuhauser S, Govers LL. 2018. Review: Host-pathogen dynamics of seagrass diseases under future global change. *Mar Pollut Bull.* **134**:75–88.
- Takayama K, Tateishi Y, Kajita T. 2021. Global phylogeography of a pantropical mangrove genus *Rhizophora*. *Sci Rep.* **11**:7228.
- Tanaka N, Setoguchi H, Murata J. 1997. Phylogeny of the family Hydrocharitaceae inferred from *rbcl* and *matK* gene sequence data. *J Plant Res.* **110**:329–337.
- Tena G. 2020. From algae to land plants. *Nat Plants* **6**:594–594.
- Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**:322–16.
- Thouvenot L, Thiébaud G. 2018. Regeneration and colonization abilities of the invasive species *Elodea canadensis* and *Elodea nuttallii* under a salt gradient: implications for freshwater invasibility. *Hydrobiologia* **817**:193–203.
- Tiley GP, Barker MS, Burleigh JG. 2018. Assessing the performance of Ks plots for detecting ancient whole-genome duplications. *Genome Biol Evol.* **10**:2882–2898.
- Van de Peer Y, Mizrahi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nat Rev Genet.* **18**:411–424.
- Van Dongen SM. 2000. *Graph clustering by flow simulation*. Netherlands: University of Utrecht.
- Van Hoesck A, Horemans N, Monsieurs P, Cao HX, Vandenhove H, Blust R. 2015. The first draft genome of the aquatic model plant *Lemna minor* opens the route for future stress physiology research and biotechnological applications. *Biotechnol Biofuels* **8**:1–13.
- Vermeij GJ, Dudley R. 2000. Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biol J Linn Soc.* **70**:541–554.
- Wang H, Dilcher DL. 2018. A new species of *Donlesia* (Ceratophyllaceae) from the Early Cretaceous of Kansas, USA. *Rev Palaeobot Palynol.* **252**:20–28.
- Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo M, Lomsadze A, Borodovsky M, Kerstetter R, Shanklin J, et al. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenuous reduction fast growth and aquatic lifestyle. *Nat Commun.* **5**:1–13.
- Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C, Mikenas J, et al. 2019. Evolution of Portulacaceae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Mol Biol Evol.* **36**:112–126.
- Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y, et al. 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoraceae) revealed by whole-genome sequencing. *Natl Sci Rev.* **4**:721–734.
- Xu Y-H, Liu R, Yan L, Liu Z-Q, Jiang S-C, Shen Y-Y, Wang X-F, Zhang D-P. 2012. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*. *J Exp Bot.* **63**:1095–1106.
- Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, Smith SA. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* **217**:855–870.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK-S, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol.* **32**:2001–2014.
- Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol Biol Evol.* **31**:3081–3092.
- Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T, Zhang L, Shen X, Meng F, et al. 2020. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat Plants* **6**:215–222.
- Yang C, Zhao L, Zhang H, Yang Z, Wang H, Wen S, Zhang C, Rustgi S, von Wettstein D, Liu B. 2014. Evolution of physiological responses to salt stress in hexaploid wheat. *Proc Natl Acad Sci U S A.* **111**:11882–11887.
- Zhang C, Ogilvie HA, Drummond AJ, Stadler T. 2018. Bayesian inference of species networks from multilocus sequence data. *Mol Biol Evol.* **35**:504–517.
- Zhang P, Zhu C, Geng Y, Wang Y, Yang Y, Liu Q, Guo W, Chachar S, Riaz A, Yan S, et al. 2021. Rice and *Arabidopsis* homologs of yeast CHROMOSOME TRANSMISSION FIDELITY PROTEIN 4 commonly interact with Polycomb complexes but exert divergent regulatory functions. *Plant Cell* **33**:1417–1429.
- Zhao S-Y, Chen L-Y, Wei Y-L, Wang Q-F, Moody ML. 2016. RNA-seq of *Ranunculus sceleratus* and identification of orthologous genes among four *Ranunculus* species. *Front Plant Sci.* **7**:732.