

Macroevolutionary dynamics in the transition of angiosperms to aquatic environments

Andrea S. Meseguer¹ , Rubén Carrillo¹, Sean W. Graham²  and Isabel Sanmartín¹ 

¹Real Jardín Botánico de Madrid (RJB), CSIC, 28014 Madrid, Spain; ²Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Summary

Author for correspondence:
Andrea S. Meseguer
Email: asanchezmeseguer@gmail.com

Received: 11 June 2021
Accepted: 27 February 2022

New Phytologist (2022) **235**: 344–355
doi: 10.1111/nph.18100

Key words: angiosperms, aquatic plants, extinction, macroevolution, phylogeny, speciation, terrestrial plants.

- Angiosperm lineages in aquatic environments are characterized by high structural and functional diversity, and wide distributions. A long-standing evolutionary riddle is what processes have caused the relatively low diversity of aquatic angiosperms compared to their terrestrial relatives.
- We use diversification and ancestral reconstruction models with a comprehensive > 10 000 genus angiosperm phylogeny to elucidate the macroevolutionary dynamics associated with transitions of terrestrial plants to water.
- Our study reveals that net diversification rates are significantly lower in aquatic than in terrestrial angiosperms due to lower speciation and higher extinction. Shifts from land to water started early in angiosperm evolution, but most events were concentrated during the last *c.* 25 million years. Reversals to a terrestrial habitat started only 40 million years ago, but occurred at much higher rates. Within aquatic angiosperms, the estimated pattern is one of gradual accumulation of lineages, and relatively low and constant diversification rates throughout the Cenozoic.
- Low diversification rates, together with infrequent water transitions, account for the low diversity of aquatic angiosperms today. The stressful conditions and small global surface of the aquatic habitat available for angiosperms are hypothesized to explain this pattern.

Introduction

‘Land’ plants (Embryophyta) – including angiosperms, gymnosperms, bryophytes, lycophytes and ferns – evolved from a single lineage that adapted to life on terrestrial habitats *c.* 450 million years ago (Ma) (Graham, 1993). This event required major changes in anatomy, physiology and reproductive biology, and led to the most successful radiation of all photosynthetic organisms recorded: *c.* 329 000 species of land plants vs *c.* 44 000 algae (Guiry, 2012; Christenhusz & Byng, 2016). Conversely, the reverse shift – the adaptation to aquatic lifestyles by terrestrial plants – is a frequent evolutionary transition in plants, having occurred in numerous lineages across the embryophyte tree of life (TOL), and especially among flowering plants (angiosperms) (Sculthorpe, 1967). The number of transitions of angiosperm lineages from land to water has not been quantified based on a robust phylogeny, although *c.* 89 clades were estimated to have aquatic species (including 33 aquatic families, 407 aquatic genera and a larger number of aquatic species within otherwise predominantly terrestrial lineages) (Cook, 1990, 1999). Within angiosperms, the aquatic flora has a greater richness within monocots than in other angiosperm clades (Sculthorpe, 1967; Grace, 1993), suggesting the existence of a phylogenetic constraint (i.e. a phylogenetic signal) in the appearance of the aquatic lifestyle.

Aquatic angiosperm lineages display extraordinary levels of morphological, reproductive, physiological, biochemical and ecological disparity, having adapted to a variety of aquatic lifestyles, including marine, freshwater and brackish water, though the majority of aquatic angiosperms occur in freshwater habitats (Sculthorpe, 1967; Philbrick & Les, 1996; Eckert *et al.*, 2016). Most aquatic angiosperms also exhibit wide geographic distributions (*c.* 60% occur on more than one continent), low indices of endemism and high levels of geographic overlap in ecologically divergent species (Sculthorpe, 1967; Cook, 1983, 1985). Aquatic angiosperms display elevated frequencies of clonal fragmentation and other types of asexual reproduction, favoring rapid multiplication and dispersal of propagules, which are associated with a remarkable ability to colonize distant lands and environmentally different settings (Sculthorpe, 1967; Hutchinson, 1975; Barrett *et al.*, 1993; Van Groenendael *et al.*, 1996; Santamaría, 2002; Eckert *et al.*, 2016). These characteristics, their morphological and ecological disparity, and the geographic overlap among species that are partitioned ecologically, have led some scientists to associate the transition to an aquatic lifestyle in angiosperms with adaptive radiation. For example, the remarkable variation in morphology, reproductive strategies and life history traits in the aquatic family Pontederiaceae has been linked to ecological specialization and adaptation to different aquatic ecosystems (Barrett & Graham, 1997).

In light of this, a long-standing evolutionary enigma in the transition of angiosperms to aquatic environments is the observed low levels of diversity of aquatic lineages compared to their terrestrial relatives (Sculthorpe, 1967). Today, aquatic species represent a small percentage of all angiosperms (*c.* 1–2%) (Cook, 1999) – although the number of cryptic species has steadily increased with the use of molecular tools (Sokoloff *et al.*, 2019), especially among marine pelagic forms (Knowlton, 2000); diversity levels also decline from amphibious and emergent species to fully aquatic, submerged hydrophilous taxa (Barrett *et al.*, 1993). Moreover, aquatic angiosperm species frequently show moderate levels of overall genetic diversity and low levels of within-population variation (Zeidler *et al.*, 1994; Jordan *et al.*, 1996), which might be linked to the predominance of clonal multiplication, limited sexual reproduction and frequent founder effects (Eckert *et al.*, 2016). The low species diversity in aquatic angiosperm clades runs against the classic notion of an adaptive radiation (Schluter, 2000). In the latter, ecological opportunity – the colonization of a new environmental niche – leads to accelerated rates of diversification, followed by a diversification slowdown as species saturate the niche space (Czekanski-Moir & Rundell, 2019). However, some authors consider the number of lineages arising from an adaptive radiation as being of secondary importance to patterns of character disparity (Givnish, 2015). In aquatic angiosperms, the observed pattern is one of numerous adaptations to a new environmental niche but without a concomitant change in the diversification dynamics, ‘a radiation in slow motion’ (Barrett & Graham, 1997).

The low levels of diversity in aquatic angiosperms may be explained by three nonexclusive hypotheses. First, the ‘time-to-speciation’ hypothesis posits that the species richness of a given clade or geographic region is correlated with the amount of time available for speciation (Wallace, 1878; Stephens & Wiens, 2004; Mittelbach *et al.*, 2007). Thus, under this hypothesis, diversification in angiosperms proceeds at a constant rate and species accumulate gradually, or, in other words, species richness in a clade is expected to correlate with its age, that is the time elapsed since the ancestor of the clade arose by speciation. Second, aquatic clades may accumulate species at lower rates (diversification rate hypothesis), either through the effect of low speciation rates, high extinction rates or both. Third, aquatic lineages may have fewer species because of lower rates of transition from land to water than in the opposite direction (transition hypothesis). The last two hypotheses assume that diversification rates are not constant but vary across clades, and that a significant part of the heterogeneity in species richness between aquatic and terrestrial flowering plants is due to trait-dependence, that is the evolution of the lifestyle along the angiosperm phylogeny. So far, these three different evolutionary scenarios have been difficult to tease apart because of the lack of a solid, comprehensive phylogenetic and statistical framework to study them.

Here, we take advantage of the publication of a large angiosperm phylogeny, representing 79% of generic diversity (78 934 species) (Smith & Brown, 2018), and the development of trait- and time-dependent macroevolutionary models in a Bayesian framework (Maddison *et al.*, 2007; FitzJohn, 2012;

Beaulieu & O’Meara, 2016; Freyman & Höhna, 2018) to explore the following questions: (Q1) Was the invasion of aquatic environments concentrated in certain clades vs being randomly dispersed across the angiosperm TOL? (Q2) How many aquatic transitions took place in flowering plants? (Q3) Did the origin of extant aquatic lineages tend to be synchronized in geological time? (Q4) Was the colonization of aquatic habitats associated with decelerated diversification rates?

Materials and Methods

Data compilation

We use a fossil-calibrated molecular phylogeny of seed plants (Smith & Brown, 2018) that includes genetic data for 79 881 seed-plant species (their GBOTB tree) and 78 934 angiosperm species. This tree has considerably higher species sampling than other more recently published angiosperm timetrees; for example, Janssens *et al.* (2020) includes *c.* 31 000 species. Smith & Brown’s phylogeny was pruned at the genus level, keeping only the angiosperm lineages. The resulting tree contained as tips 10 360 angiosperm genera of the 13 164 described (Christenhusz & Byng, 2016). The decision to use a pruned phylogeny was based on the near completeness of the generic sampling (*i.e.* *c.* 79% genus diversity in angiosperms), and because species-level analyses, especially for state-dependent speciation extinction (SSE) models, are unfeasible to complete for a 78 934 species phylogeny, using current software implementing Bayesian phylogenetic inference (Höhna *et al.*, 2016). Also, by cutting the tree at the genus level, rather than random removal of a percentage of species, we ensured that representatives of all aquatic angiosperm lineages were included in the phylogeny. In other words, given the low proportion of aquatic species (< 1.5%; see the Results section), random species selection would probably lead to many aquatic genera being excluded from the analyses. Certainly, with our approach, we are disregarding those land–water transitions that occurred within genera, which might lead to underestimating the rate of transitions, or overestimating the age of some of those events. Conversely, using the species in Smith & Brown’s phylogeny as the evolutionary units of analysis could lead to the opposite result: overestimating transition rates within genera if species with the same character state as their sister-clades are not sampled, which is often the case in overdispersed, taxonomically guided taxon sampling (Reyes *et al.*, 2018). Nonetheless, most genera in our dataset are fully aquatic (178) or terrestrial (10 089), whereas the proportion of polymorphic genera (*i.e.* including both terrestrial and aquatic species) is relatively small (93, *c.* 0.8% of angiosperm generic diversity). Thus, we are confident that, even if absolute values for the diversification and transition rate parameters could not be estimated from our SSE analyses, we will be able to infer general patterns for the evolution of the aquatic lifestyle, reflecting evolutionary trends in the angiosperm TOL at high taxonomic levels (we further expand on this topic in the Discussion section).

Following previous studies (Sculthorpe, 1967; Cook, 1990, 1999; Barrett *et al.*, 1993), we consider aquatic plants in the

broadest sense, including all plants that occur in permanently or seasonally wet environments, such as plants defined as hydrophytes, limnophytes, macrophytes, amphiphytes, helophytes, amphibious, wetland or simple water plants. These life forms represent an evolutionary continuum, and it would thus be arbitrary to exclude species commonly found in aquatic habitats that do not complete their entire life cycle in water. Genera are coded as either terrestrial or aquatic following Cook (1990, 1999). The total number of species within each angiosperm genus is based on the Plant List (www.theplantlist.org), and the proportion of aquatic species in each genus is based on Cook (1990, 1999). For the analyses, we consider two alternative codings for the relatively few polymorphic genera in our dataset. First, we coded a genus as 'aquatic' if it comprises one or more submerged, emerged or seasonally submerged species, along with (or not) any additional terrestrial species. We consider this to be a conservative approach, as we wanted to test the hypothesis that the transition to aquatic habitats is associated with reduced diversification rates. In other words, by artificially increasing the number of aquatic genera, our analyses should be prone to higher Type II error rates (e.g. erroneously rejecting the hypothesis of low diversification rates). By contrast, in our alternative coding, we consider polymorphic genera, including both terrestrial and aquatic species, as 'terrestrial'. These analyses would be subject to a higher rate of Type I error or 'false positives', in comparison with the coding above.

Phylogenetic signal analyses

To test for the existence of phylogenetic signal in the aquatic lifestyle (i.e. the tendency for related species to resemble each other, more than they resemble species drawn at random from a phylogenetic tree), we used the δ statistics (Borges *et al.*, 2019) as implemented in the R package DIVERSITREE (FitzJohn, 2012). The higher the δ -value, the stronger is the phylogenetic signal. Since the value of δ could be arbitrarily large, significance was evaluated by comparing the inferred δ -value to a distribution of values obtained from 100 phylogenies where habitat trait is randomly assigned across tips.

Diversification analyses: constant-rate and trait-dependent models

To test whether the observed differences in diversity levels among aquatic and terrestrial angiosperm clades are explained solely by age, that is the time elapsed since the ancestor of the clade arose by speciation, we compared a constant-rate birth–death (BD) model assuming gradual accumulation of taxa in the phylogeny with SSE models (Beaulieu & O'Meara, 2016), in which transitions between the aquatic and terrestrial habitats were tied to the diversification dynamics in angiosperms. Additionally, we tested whether clade age (stem age of genera) could explain the observed species richness among aquatic angiosperms by performing phylogenetic generalized least squares (PGLS) analysis (Martins & Hansen, 1997) on the genus-level phylogeny in the R package CAPER (Orme *et al.*, 2013), with species numbers compiled as

above; a significant correlation between clade age and richness is interpreted as support for the time-to-speciation hypothesis.

We implemented the constant-rate BD and SSE models within a hierarchical Bayesian framework in the open software REVBayes (Höhna *et al.*, 2016). The advantage of this approach is the use of Markov chain Monte Carlo (MCMC) simulations and directed-acyclic-graph models to account for uncertainty in parameter estimates, and the possibility to statistically compare different models in terms of their strength of support by the data using Bayes Factor comparisons (Freyman & Höhna, 2018). We implemented the simpler binary state speciation–extinction (BiSSE) model, in which heterogeneity in diversity levels among angiosperms is tied to different diversification dynamics and/or asymmetry in transition rates between the two character states: aquatic vs terrestrial habitat. SSE models have been criticized because of the simple null hypothesis used for model comparison, which assumes no heterogeneity in clade diversification rates, and so can lead to elevated Type I error rates (Rabosky & Goldberg, 2015; Laenen *et al.*, 2016). To address this bias, we implemented the more realistic hidden state speciation and extinction (HiSSE) model (Beaulieu & O'Meara, 2016). HiSSE accounts for the existence of an unobservable character with 'hidden' states (state A, state B) that could impact diversification rates in addition to the observed trait of interest (state 0, terrestrial habit; state 1, aquatic habit). We used the complete Bayesian MCMC implementation of this model in REVBayes, with lognormal priors centered in the extant diversity for speciation and extinction rates, and exponential priors for transition rates in the focal character, habitat. We accounted for incomplete taxon sampling by providing a global sampling fraction, the *rho* parameter, as the ratio of the number of genera in our reconstructed phylogeny to the total genus-level diversity (0.79). We ran a 10 000 generation chain length, sampling every 1000 generations. Parameter mixing and stationarity were assessed using TRACER v.1.6 (Rambaut & Drummond, 2009) and the effective sampling size (ESS) criterion (ensuring values were > 200 for all parameters). After discarding 25% chain length as burn-in, we summarized ancestral character states as marginal posterior probabilities at the nodes in the phylogeny. We also employed a heuristic approximation to stochastic character mapping in REVBayes (Freyman & Höhna, 2018) to estimate the direction and timing of transition events between states along branches in the phylogeny, using 500 time slices.

To explore if our monomorphic coding of multistate genera could bias diversification and transition rate estimates, we ran the HiSSE models above using two coding strategies: polymorphic genera coded as terrestrial or as aquatic, and also quantified their respective level of support by the data using Bayes Factors. We decided not to use the geographic state-dependent speciation extinction model 'GeoSSE' (Goldberg *et al.*, 2011), which includes a speciation parameter (SAB) for polymorphic (widespread) states. This model is appropriate for modeling geographic distribution – a taxon can occupy two areas or geographic states simultaneously and speciate in allopatry – but this parameter is less realistic for a character trait such as lifestyle. A species is either aquatic or terrestrial (note that amphibious species were

coded as aquatic here), and therefore the polymorphic state in our analysis does not reflect a biological pattern but is a direct consequence of using genera as the tips in the phylogeny rather than species (i.e. if we had performed the analyses at the species level, no species would have been coded as polymorphic).

Both BiSSE and HiSSE implement trait-dependent diversification models. To allow for trait independence, we compared these models with a null model in which the diversification process is independent of the evolution of the observed character, but where there is still heterogeneity in diversification rates over the tree; that is, the rate of diversification is allowed to vary across branches but is not impacted by species habitats (Caetano *et al.*, 2018). Specifically, we ran a character-independent model with two hidden states (CID2 model), allowing diversification and transition rates to differ among hidden states but not within the focal character. We compared the level of support by the empirical data of the BD, BiSSE, HiSSE and CID2 models using Bayes Factor comparisons of their model marginal likelihood values; these were estimated via path sampling and stepping-stone sampling using parallel power posterior analyses. We ran 50 power posteriors, with a 500 generation chain length for each power posterior run. We compared marginal likelihood values using the likelihood ratio test, $2\log_e(H_0 - H_1)$, with values > 2 indicating positive support for one model over the other, and > 6 indicating strong positive support (Kass & Raftery, 1995).

Finally, to explore the susceptibility of our diversification inferences to the taxonomic level used to account for incomplete taxon sampling in the phylogeny, we repeated all the analyses above using a global sampling fraction reflecting the number of species in our phylogeny over the total species diversity in angiosperms ($\rho = 0.03$), that is assuming that tips represent species as evolutionary units instead of genera. All state-dependent analyses were performed on a cluster service provided by the CESGA supercomputing center.

Diversification analyses: episodic BD models

All diversification models above test for a signal of diversification rates varying across clades (i.e. clade-dependent heterogeneity), which may be trait-dependent (i.e. aquatic vs terrestrial as in BiSSE and HiSSE) or independent (CID2), but which assume diversification rates are constant within clades. To test for the possibility of diversification rates varying through time but affecting all clades simultaneously within a given lifestyle, we generated lineage-through-time (LTT) plots depicting the accumulation of extant lineages over time for aquatic and terrestrial angiosperms, separately. We used the function *drop.tip* from the R package APE (Paradis *et al.*, 2004) to prune the original phylogeny and generate subtrees comprising only current lineages with the same habitat type, terrestrial or aquatic. In addition, we generated ancestral-state-reconstructed 'ASR-LTT' plots depicting the accumulation of aquatic or terrestrial ancestral nodes over evolutionary time, as obtained from the marginal character-state reconstruction of the HiSSE model. This procedure ensures that our ASR-LTT plots include only lineages that were ancestrally reconstructed as exhibiting a given habitat type, aquatic or

terrestrial. Both extant and ancestral (ASR) LTT plots were analyzed using episodic BD (EBD) models (Stadler, 2011; Höhna, 2015) in order to estimate the time, magnitude and significance of discrete rate shifts in diversification (speciation minus extinction) and relative extinction or 'turnover' (speciation/extinction). To analyze the ASR-LTT plots, we used EBD models implemented in a maximum-likelihood framework in the R package TREEPAR (Stadler, 2011); we compared models with one or several discrete rate shifts against the constant BD model via likelihood-ratio tests. For the extant LTT plots, we used the Bayesian EBD model implemented in REVBayes: unlike TREEPAR, which implements a piece-wise constant process, rates in the REVBayes EBD model are autocorrelated between time intervals according to a locally adaptive process, namely a Horseshoe Markov random field allowing for nearly constant rates between time intervals interspersed with large jumps (Magee *et al.*, 2020). We ran analyses with 10 intervals (nine shifts) and compared them with a constant BD model via Bayes Factors.

Results

According to our estimates, aquatic genera have, on average, fewer species than terrestrial genera, 11 vs 22.7 species per genus, respectively (4467 aquatic species/407 genera; 292 015 terrestrial species/12 840 genera). We did not find phylogenetic signal of the aquatic lifestyle in the angiosperm phylogeny, with the observed δ value falling within the distribution of δ values generated from randomizing the focal trait in the phylogeny ($\delta = 27.9$; $P > 0.05$; Supporting Information Fig. S1). Similarly, we found no relationship between clade age and the log-transformed species richness across the full set of aquatic angiosperm genera (Fig. S2; F -statistic: 0.90; $P = 0.34$; $df = 267$; $\beta = 1.99$, where the regression coefficient β is the change in log-transformed diversity per million years).

For the genus-level phylogeny, Bayes Factor comparisons strongly rejected a constant-rate BD model in favor of the HiSSE model (Table 1). HiSSE was also favored over BiSSE with very strong support. Results from the HiSSE model supported higher speciation rates and lower extinction rates in terrestrial lineages

Table 1 Bayes Factor comparison of the marginal likelihood estimates for the constant BD, BiSSE, HiSSE and CID2 models.

Model	Marginal likelihood	
	ss	ps
Constant BD	-119 008.1	-119 008.2
BiSSE	-119 681.0	-119 679.7
HiSSE (polymorphic as terrestrial)	-117 970.2	-117 968.1
HiSSE (polymorphic as aquatic)	-118 289.1	-118 287.5
CID2	-118 291.2	-118 290.0

Marginal likelihood estimates are analyzed here using the global phylogeny of angiosperms with the genus fraction, and parallel power posterior analyses with path (ps) and stepping-stone sampling (ss). The model most strongly supported by the data among all five tested models is indicated in bold.

(state 1) compared to aquatic lineages (state 2), and these differences were maintained across the hidden states A and B (Fig. 1). Net diversification rates were thus estimated to be higher for terrestrial angiosperms than for aquatic genera. Estimated transition rates from land to water were lower than in the opposite direction, from water to land.

There was some overlap between the marginal posterior distributions of the extinction rate per habitat within hidden state B (Fig. 1). Nevertheless, pairwise comparisons of rate values between character states 1 and 2 within either hidden state A or B, and across the MCMC posterior set, generated a distribution of differences (e.g. $\lambda_{1A}-\lambda_{2A}$, $\mu_{1B}-\mu_{2B}$, $q_{1A}-q_{2A}$), in which the 95% credibility interval was each time greater or smaller than zero, vs overlapping zero (Fig. S3); this is interpreted as showing the existence of significant differences between speciation (λ), extinction (μ) and transition rates (q) for the focal states (aquatic vs terrestrial) in our angiosperm tree. The reconstruction of ancestral states over the maximum a posteriori (MAP) tree (Figs S4, S5), and the stochastic mapping analyses (Fig. 2) inferred the most recent common ancestor (MRCA) of all extant angiosperms as terrestrial, with the first transitions to water occurring early during angiosperm evolution.

We obtained similar results from the HiSSE analysis with the two alternative codings for polymorphic genera, either aquatic or terrestrial (Table 2), but rate posterior estimates were wider, and thus more uncertain, for the former (Fig. S6). In addition, the model with polymorphic genera coded as terrestrial (Fig. 1) received stronger Bayes Factor support (Kass & Raftery, 1995) than the model with them coded as aquatic (Table 1). Bayes Factor comparisons also indicated very strong support for the trait-dependent HiSSE terrestrial model over the character-independent CID2 model.

The analyses in which the sampling parameter reflected the species fraction of diversity in angiosperms showed similar results to the genus-level analyses in terms of net diversification and transition rates for the states of the focal character (Tables S1, S2; Fig. S7); the only exception was the extinction rate, which was estimated to be higher for terrestrial angiosperms than for aquatic lineages, unlike in the genus-level analysis. However, the main difference was found in the marginal character state reconstruction. The analysis assuming a species-level sampling reconstructed the MRCA of all extant angiosperms as an aquatic plant, as well as the MRCAs of all subsequent angiosperm orders (e.g. the ancestors of campanulids, lamiids, Cornales, Caryophyllales,

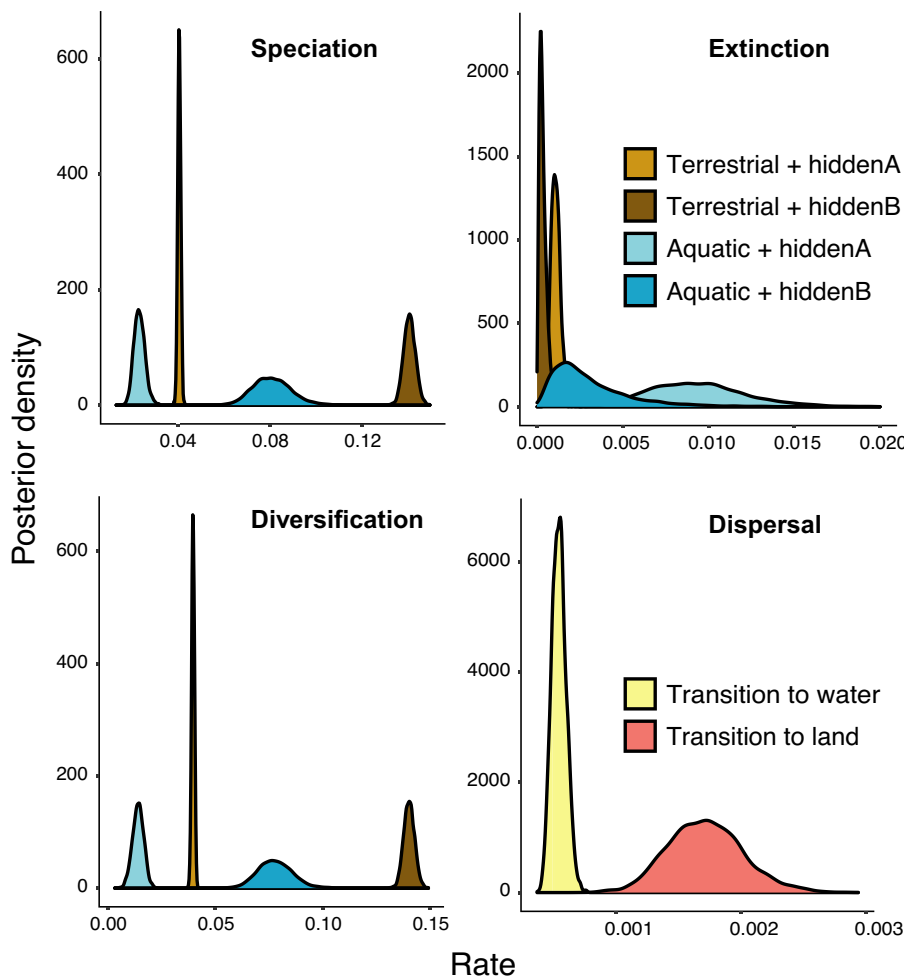


Fig. 1 Habitat-dependent diversification rates estimated with HiSSE under the genus-level sampling. Posterior densities of speciation (λ), extinction (μ), net-diversification ($\lambda - \mu$) and dispersal rates. Colors correspond to the posterior probabilities for a given state. Terrestrial species have higher speciation and lower extinction rates regardless of the hidden state considered.

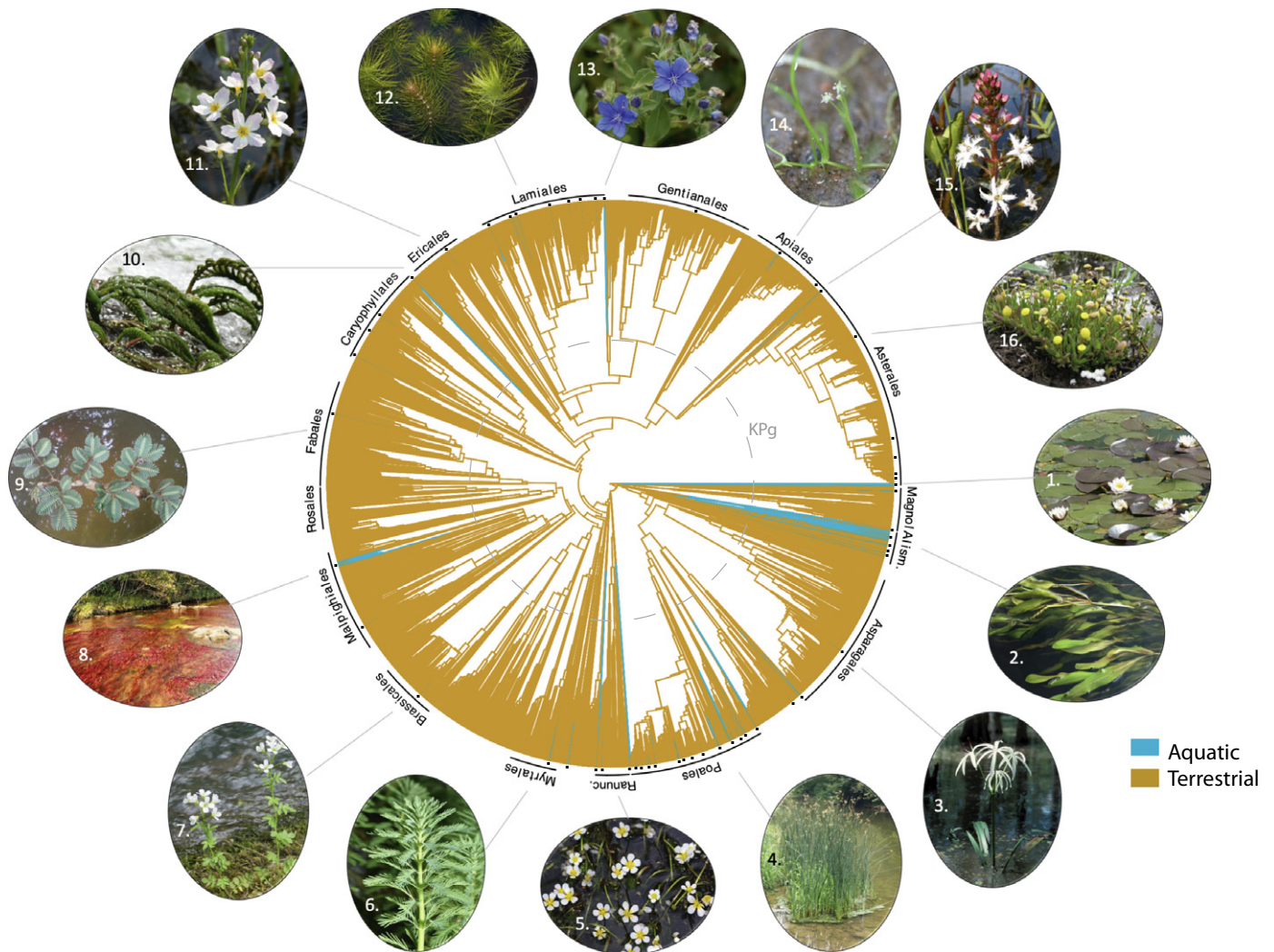


Fig. 2 Ancestral habitat reconstruction in angiosperms simulated under Bayesian stochastic character mapping in HiSSE using the genus-level sampling. Branch colors denote two different states: aquatic and terrestrial. Transitions between states are indicated by changes in color along the branches. Inset figures represent aquatic lineages of different angiosperm genera: 1. *Nymphaea* (Nymphaeaceae), 2. *Potamogeton* (Potamogetonaceae), 3. *Crinum* (Amaryllidaceae), 4. *Schoenoplectus* (Cyperaceae), 5. *Ranunculus* (Ranunculaceae), 6. *Myriophyllum* (Haloragaceae), 7. *Cardamine* (Brassicaceae), 8. *Macarenia* (Podostemaceae), 9. *Neptunia* (Fabaceae), 10. *Hydrostachys* (Hydrostachyaceae), 11. *Hottonia* (Primulaceae), 12. *Hydrotriche* (Scrophulariaceae), 13. *Hydrolea* (Hydroleaceae), 14. *Lilaopsis* (Apiaceae), 15. *Menyanthes* (Menyanthaceae), 16. *Cotula* (Asteraceae). Images from Wikimedia Commons, except number 14, from *Flora of the World*. Visualization of aquatic lineages on the angiosperm tree is limited by their low proportion. Dots were added on the edge of the tree to help with the visualization of aquatic lineages. KPg = Cretaceous–Paleogene boundary.

Table 2 Results from the HiSSE models using the global genus-level phylogeny of angiosperms when polymorphic genera (93 cases) are coded as either terrestrial or aquatic

HiSSE	Character	λ	μ	r	a	q
Polymorphic as terrestrial	1A = Terrestrial + A	0.0407	0.00110	0.03959	0.027076	0.00172 (21)
	2A = Aquatic + A	0.0234	0.00933	0.01407	0.398589	
	1B = Terrestrial + B	0.1400	0.00035	0.13965	0.002527	0.00055 (12)
	2B = Aquatic + B	0.0806	0.00310	0.07749	0.038486	
Polymorphic as aquatic	1A = Terrestrial + A	0.0404	0.00013	0.04026	0.003317	0.00133 (21)
	2A = Aquatic + A	0.0125	0.00171	0.01079	0.136800	
	1B = Terrestrial + B	0.1400	0.00009	0.13991	0.000633	0.00100 (12)
	2B = Aquatic + B	0.0436	0.00108	0.04252	0.024771	

Parameters in the model are speciation (λ) and extinction (μ) of terrestrial (1) and aquatic (2) lineages for the hidden characters A and B, as well as transitions (q) from water to land (21), and from land to water (12). Indirectly estimated net diversification ($r = \lambda - \mu$) and relative extinction rates ($a = \mu/\lambda$) are also provided.

Fabales, Malpighiales, etc., in addition to all magnoliids and monocots); the first transition to a terrestrial habitat was estimated to be as late as 50 Ma (Figs S8, S9). Inspection of MCMC traces revealed poor mixing of the diversification rate parameters, speciation and extinction, with very low ESS values (< 50 for some of the states), while transition parameters showed a better behavior. Parameter mixing did not improve after doubling the chain length (and analysis time, *c.* 2 months in a computer cluster), or running multiple chains starting from different random values. By contrast, the genus-level analysis showed good mixing behavior for all parameters, with ESS values > 200 for all states. Hence, we discarded the species-level analyses as unreliable, and discuss results below based on the genus-level analysis (see Discussion section).

The LTT plots based on the HiSSE marginal ancestral reconstruction (ASR-LTT plot with polymorphic genera coded as terrestrial) showed a steady accumulation of aquatic lineages over time, whereas the terrestrial ASR-LTT plot exhibited a pattern with an initial increase in diversity, followed by a slowdown in the rate of accumulation of terrestrial lineages, and a final upturn *c.* 10 Ma (Fig. 3a). This pattern is corroborated by TREEPAR, which infers a shift towards higher relative extinction rates at 70 Ma for ancestral terrestrial lineages, followed by a decrease in turnover at 10 Ma (all of them are significant, $P < 0.01$; Table S3); diversification rates were estimated to be low and relatively constant through time (Fig. S10), suggesting that rate shifts are mainly driven by extinction. For the aquatic ASR-LTT plot, TREEPAR supported a pattern of nearly constant diversification and turnover rates through time with only a small but significant shift in the rates of diversification and relative extinction at 10 Ma.

The extant LTT plot for the aquatic lineages (Fig. 3b) showed a very different pattern to the one obtained from the ancestral reconstruction (ASR-LTT plot): an initial rapid accumulation of lineages followed by asymptotic growth from 120 to 100 Ma onwards (Fig. 3a). This is corroborated by the autocorrelated EBD model in REV BAYES (Fig. S11; Table S4), showing a decrease in the net diversification rate after 120 Ma, coupled with an increase in the relative extinction rate. By contrast, the extant LTT plot for the terrestrial lineages showed a pattern very similar

to the one observed in the terrestrial ASR-LTT plot (Fig. 3a): an initial increase in terrestrial diversity, followed by a slowdown, and a final upturn at *c.* 10 Ma; the EBD model inferred higher relative extinction rates between 70 and 10 Ma, in line with TREEPAR inference based on the ancestral lineages.

Bayesian stochastic character mapping inferred 97 independent transition events from land to water, and 39 events from water to land (Fig. 4). Additionally, there were 44 reversal events involving more than one transition: 43 transitions from land to water followed by the reverse transition, from water to land, and one multi-change transition from water to land and back to water (Fig. S12). The timing and tempo of these events varied among transition types. Transitions to aquatic from a terrestrial habitat started earlier, nearly coincident with the origin of angiosperms in our phylogeny, and increased markedly between 40 and 10 Ma. Transitions from aquatic to terrestrial habitats are recorded later, with the first event of transition at *c.* 40 Ma and a peak at 20 Ma (Fig. 4). However, if we separate single-event transitions (from water to land) from multiple-event transitions (from land to water to land), the peaks are found at *c.* 10 and 25 Ma, respectively (Fig. S12).

Discussion

Why are there so few species among aquatic angiosperms?

Aquatic angiosperm lineages have been regarded as species-poor in comparison with their terrestrial relatives (Cook, 1990). Here, we update species numbers according to current diversity estimates (Christenhusz & Byng, 2016) and confirm the generality of this pattern. There are about *c.* 4500 aquatic species among angiosperms, in comparison to *c.* 291 000 terrestrial species described to date. This represents around 1.5% of the described angiosperm diversity, in line with previous estimates (1–2%) (Cook, 1999). Aquatic genera, including those polymorphic genera with terrestrial and aquatic species, also contain a lower number of species than their fully terrestrial counterparts on average (i.e. 11 vs 22.7 spp. per genus, respectively). This shows that there is a prevailing trend for aquatic lineages to be species-poor across angiosperms.

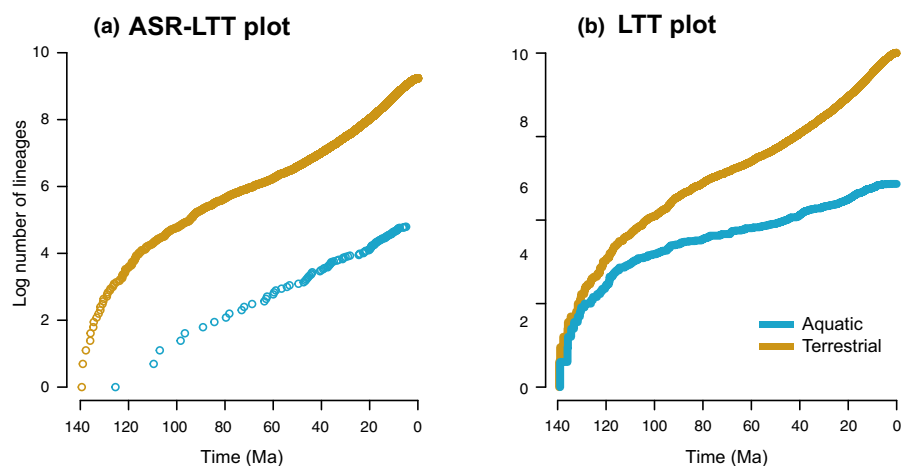


Fig. 3 Ancestral reconstruction lineage-through-time (ASR-LTT) plots for aquatic and terrestrial angiosperm lineages. (a) ASR-LTT plots are based on the reconstruction of ancestral habitat for phylogenetic nodes on HiSSE using the genus-level sampling. (b) Lineage-through-time (extant LTT) plots depicting the accumulation of extant lineages over time for aquatic and terrestrial angiosperms, separately. Ma, million years ago.

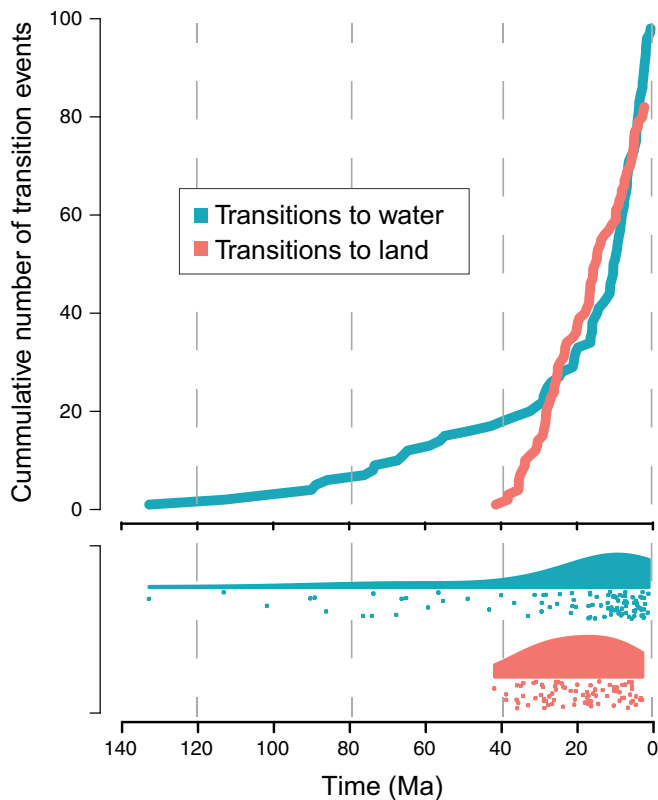


Fig. 4 Cumulative number of transition events from land to water and the reverse along the angiosperm phylogeny. Estimates derive from stochastic character mapping analyses in HiSSE under the genus-level sampling. Below, the raincloud plots show density distributions and raw counts (points) of each transition type. In this figure, single transitions from 'water to land' and multiple change transitions from 'land to water to land' were pooled, and similarly for transitions from 'land to water' and from 'water to land to water'. Dashed lines highlight 40 Ma (million years ago) intervals.

Our analyses reject the hypothesis of time as the sole explanation for differences in species richness between aquatic and terrestrial clades. First, a model of constant-rate diversification in angiosperms was rejected in favor of the trait-dependent HiSSE model (Table 1), indicating that gradual accumulation of species/genera over time is not sufficient to explain differences in diversity across clades. Second, given that a time-homogeneous model of constant-rate diversification is unrealistic for almost any empirical phylogeny (Caetano *et al.*, 2018), we examined how well supported the HiSSE model was in comparison with a different time-heterogeneous model which does not assume that rate heterogeneity is structured as an evolving trait. Again, the trait-independent CID2 model was rejected in favor of the trait-dependent HiSSE model by Bayes Factor comparisons (Table 1). Additionally, PGLS analysis found no significant correlation between age and species richness for the aquatic genera in our phylogeny, against the expectations of the time-to-speciation hypothesis (Fig. 2). Together, these results suggest that, even if time, or other unmeasured factors, could be influencing the diversification dynamics of angiosperms, a significant part of the observed heterogeneity in diversity levels among genera can be attributed to trait dependence, that is the evolution of the life-

history trait aquatic vs terrestrial along the phylogeny. Specifically, we argue that the observed heterogeneity can be explained by a strong asymmetry in the rate of transition events, and by different speciation and extinction rates in aquatic and terrestrial lineages, which also vary over time.

Despite the aquatic lifestyle being better represented among monocots than in other angiosperm lineages (Sculthorpe, 1967; Grace, 1993), the lack of phylogenetic signal found for this trait suggests that the transition to water is not concentrated in certain lineages (Fig. S1). Instead, the acquisition of the aquatic lifestyle occurred in multiple independent clades and at different times during angiosperm evolution (Figs 2, 4), supporting the idea that the aquatic environment was potentially accessible to many angiosperm lineages (and consistent with aquatic species being found in most angiosperm orders). Yet, despite the observed phylogenetic overdispersion of this trait, our results reveal that the transition to aquatic environments was relatively infrequent and often accompanied by reverse transitions to land. We estimate 97 independent transition events to the aquatic lifestyle (land to water) in a phylogeny of > 10 000 tips across nearly 140 million years (Myr) ($0.70 \text{ events Myr}^{-1}$), compared to 82 independent transitions back to land over *c.* 40 Ma ($2 \text{ events Myr}^{-1}$; Figs 1, 4). However, our results also suggest that some early diverging lineages that made the transition 'from land to water' were probably lost to extinction along the long branches of the angiosperm TOL. This is indicated by the multiple-event transitions 'from land to water and back to land' detected in our analysis (43 events), which peak earlier than the single-event transitions (Fig. S12) but did not leave recorded diversification events in the reconstructed phylogeny (some of these events might be accounted for by incomplete taxon sampling as well).

The reverse transition, from water to land, has been regarded as an unusual event in angiosperm evolution, which has been explained by the difficulty of reversing the major anatomical and physiological changes needed to adapt to water environments (Doyle, 2012), or by competition with other clades within nearly saturated terrestrial niches, a pattern observed in many other plant and animal clades (Vermeij & Dudley, 2000). Nevertheless, we inferred a rate of transition that is significantly higher from water to land than in the opposite direction (land to water), suggesting that the recolonization of land has been relatively frequent in angiosperms, at least in the past 40 Myr (Fig. 4). However, we cannot explain why no transitions from water to land occurred before this period. Although we inferred only a single multiple-event transition from water-to-land-to-water in our phylogeny (Fig. S12), we cannot discard the possibility that extinction has erased the signal, and, indeed, elevated extinction rates were inferred for the aquatic lifestyle in our SSE analyses (Fig. 1).

All of the above indicate that transitions to water started earlier but were apparently less frequent than the reverse transition over time (Fig. 4). In addition, angiosperm diversification rates were also lower in aquatic habitats (Fig. 1), with aquatic lineages presenting lower speciation and higher extinction rates than their terrestrial counterparts. Strong selective pressures exerted by aquatic environments are one explanation for the low

diversification rates detected here. Aquatic habitats represent stressful environments for plants. Despite not exhibiting the environmental extremes of terrestrial habitats, aquatic habitats are characterized by limited nutrient supply, low carbon availability and sometimes osmotic stress, all of which can limit photosynthesis and growth (Wetzel, 1988; Duarte *et al.*, 1994; Santamaría, 2002). Aquatic habitats also impose significant restrictions on sexual reproduction, by providing poorer conditions for seed production, germination and seedling establishment (Barrett *et al.*, 1993). This may induce selection for stress-tolerant taxa, which typically show broad tolerance ranges and high dispersal capabilities. Clonal growth and fragmentation may also be favored in these habitats (Barrett *et al.*, 1993; Van Groenendael *et al.*, 1996; Eckert *et al.*, 2016), with previous studies suggesting greater vulnerability of asexual lineages to extinction (Silvertown, 2008; Bromham *et al.*, 2016).

Area size could also explain the low diversification rates in aquatic habitats. The surface of aquatic habitats available to angiosperms is much reduced compared to that of terrestrial environments. Saltwater covers almost 68% of the planet, but freshwater habitats, where most aquatic angiosperm occur, cover only 0.02% (Gleick, 1993). Aquatic habitats are also quite uniform at regional scales (Sculthorpe, 1967; Cook, 1985; Les, 1988; Barrett *et al.*, 1993). Although there is considerable spatiotemporal variation in water availability, and aquatic habitats are heterogeneous at relatively small scales, this fine-scale environmental mosaicism tends to be repeated regionally (Santamaría, 2002), which might lead to loss of variation through selection, resulting in adaptation to a uniform environment. Together, these conditions probably limit the frequency of allopatric speciation between geographically isolated populations, and of *in situ* diversification, and may contribute to the maintenance of few species with broad ranges (Santamaría, 2002). Although these factors refer to water bodies today, they probably characterized ancient environments as well, which might explain the low diversification rates of aquatic angiosperm lineages characterizing their *c.* 100 Myr of evolution (Fig. 1).

Our ASR-LTT results provide further evidence for the differences in diversity levels between aquatic and terrestrial lineages. Whereas rates of diversification within the aquatic habitat were relatively low and constant throughout the Cretaceous and Cenozoic, the terrestrial ASR-LTT plot shows a pattern of rapid growth, followed by a slowdown at 70 Ma, and a final acceleration at 10 Ma (Figs 3a, S10; Table S3). Interestingly, analysis of the extant LTT plots provides a different picture: extant terrestrial and aquatic lineages exhibit a similar pattern of initial growth, resembling an adaptive radiation, followed by a slowdown in diversification; in fact, for some million years, in the Early Cretaceous, the number of aquatic and terrestrial lineages is shown to be equal (Figs 3b, S11). The observed differences in shape (and diversification dynamics) between the aquatic ASR and the extant LTT plot send out a note of caution about using LTTs based on pruned subtrees that include only extant lineages with a given character state. Our reconstructions (Figs 2, 4) suggest that many extant aquatic genera are the result of transitions from terrestrial ancestors occurring in recent times. However, in

connecting these sometimes distantly related genera, the aquatic phylogenetic subtree assumes that their MRCA was also aquatic, overestimating the age of the aquatic transitions, which will explain the extant pattern of initial rapid radiation (Fig. 3b). This artifact is not present in the ASR-LTT plot (Fig. 3a) because only the age of ancestral nodes that are reconstructed as aquatic is considered in the plotting of lineage accumulation over time.

Reconstruction of the ancestral angiosperm habit

Our analysis using genera as evolutionary units reconstructs the MRCA of all angiosperms as inhabiting terrestrial habitats in the Early Cretaceous, *c.* 140 Ma (Doyle, 2012). Transitions to an aquatic lifestyle occurred early during angiosperm evolution, starting with the branch leading to the common ancestor of Nymphaeales > 125 Ma (Figs 2, 4), and then followed by transitions within monocots (Alismatales, Acorales), Nymphaeales, Ceratophyllales, and some eudicot lineages with old crown ages, such as Cornales and Ranunculales – all these groups include families whose Cretaceous ancestors were adapted to aquatic environments in our analysis (Fig. 2). This reconstruction, with a terrestrial ancestor of flowering plants and an early transition to aquatic habitats, is in accordance with current angiosperm molecular phylogenetic evidence, indicating that several ANA-grade lineages (those that do not belong to mesangiosperms such as monocots, eudicots and magnoliids) include, or are solely, highly modified aquatic plants (Doyle, 2012; Iles *et al.*, 2014; Albert & Renner, 2020; Yang *et al.*, 2020; Zhang *et al.*, 2020). It is also consistent with the early appearance and abundance of freshwater angiosperms in the fossil record (Martín-Closas, 2003). For example, *Archeфраctus* represents one of the oldest, most complete angiosperm fossils (*c.* 125 Ma), and exhibits traits considered as special adaptations for the hydrophytic habitat (Sun *et al.*, 2002). Some authors have used this to suggest that the first angiosperms were indeed aquatic plants (Sun *et al.*, 2002; Gómez *et al.*, 2008; Gambaryan & Kuznetsov, 2021). Other authors argue that an aquatic ancestry is an anatomical impossibility (Doyle, 2012) because it would entail the loss of cambial activity (as observed in aquatic lineages such as Nymphaeales, *Ceratophyllum* or monocots), whereas lineages that are terrestrial and that arose very early in angiosperm evolution, such as *Amborella* and Austrobaileyales, have a relatively normal secondary xylem cylinder (Feild & Arens, 2007).

We infer the common ancestor of monocots, *c.* 130 Ma, to be a terrestrial plant. This result contradicts a recent study (Givnish *et al.*, 2018), which suggests that monocots descended from an aquatic ancestor based on the abundance of aquatic angiosperms among orders that branch off from the earliest divisions in monocot phylogeny (e.g. Acorales and Alismatales), and the aquatic habit of other ancient lineages such as *Ceratophyllum* (Ceratophyllaceae) or the order Ranunculales, the latter of which is sister to the rest of the eudicots (e.g. Chase, 2004). While our results do not support the monocots as being ancestrally aquatic, further research would be useful to clarify this, as we cannot discard the possibility that elevated rates of extinction in aquatic angiosperms (as evidenced by frequent multiple-event transitions and high

relative extinction rates; Figs 1, S12) have obscured the signal of an aquatic ancestor for this and other clades.

The conclusions above are based on analyses implementing the genus-level sampling strategy. Results from the analyses with tips representing species, that is using the considerably lower species-level sampling fraction, were very different: the MRCA of all angiosperms and of all subsequent nodes until 50 Ma were reconstructed as being aquatic (Figs S8, S9). While it could be possible to entertain the idea that the MRCA of angiosperms as a whole was aquatic (see above), the notion that the MRCA of a number of major 'derived' angiosperm clades was aquatic is difficult to sustain, based on multiple lines of evidence. We argue that this unexpected result is explained by the poor mixing and low EES values for the speciation and extinction parameters in the species-level analysis. This can be attributed to the highly incomplete phylogenetic sampling at the species level ($\rho = 0.03$), and the low representation of aquatic species in Smith & Brown's original phylogeny (which reflects the fact that they are not abundant among angiosperms; < 1.5% of species diversity). These composite sources of uncertainty were important factors precluding adequate mixing in the species-level analysis.

Analyzing macroevolutionary dynamics using genera rather than species is contentious, but has been attempted before under different statistical frameworks (Silvestro *et al.*, 2015; Hernández-Hernández & Wiens, 2020). Likelihood derivations used in the SSE and EBD models were designed for species-level phylogenies, with diversification interpreted as speciation minus species extinction (Nee *et al.*, 1994; Maddison *et al.*, 2007). In our genus-level analyses, diversification is equivalent to the rate of genus origination, and extinction to the rate of genus global extinction, so we are losing the part of diversification referring to the most recent time periods; that is, we are cutting the phylogeny and the diversification process at the origin time for all genera. However, we argue that this approach allows us to capture long-term dynamics in the evolution of the aquatic lifestyle in angiosperms with better accuracy than using the species-level fraction. This is because large-scale angiosperm phylogenies, such as that of Smith & Brown (2018) used here, typically try to maximize phylogenetic representation, using 'diversified', overdispersed taxon sampling; in contrast, the SSE and EBD models implemented in REV_{BAYES} assume random taxon sampling (Freyman & Höhna, 2018). Since the fraction of missing diversity is very small (*c.* 0.2) at the genus level, it is more likely that random sampling is more biased at the species level. Implementing clade-specific sampling fractions to better reflect the uneven representation of species in the original phylogeny (e.g. Rabosky *et al.*, 2014) would be a possible solution, but recent studies have shown that such a procedure may lead to incorrect likelihood estimates (Moore *et al.*, 2016; Beaulieu, 2020). Instead, we chose to use genera to approximate origination and extinction rates of evolutionary units, as well as to measure transition rates from/to the aquatic lifestyle. Future developments into algorithm efficiency for SSE and other birth–death models (e.g. Ronquist *et al.*, 2021) will hopefully allow us to analyze diversification dynamics on trees with hundred to thousands of species, such as the angiosperm TOL at the species level.




Acknowledgements

This study was funded through grant PID2020-120145GA-I00 to ASM and grant PID2019-108109GB-I00 to IS by the MCIN/AEI/10.13039/501100011033/ and EURFD 'A way to make Europe', the CAM Atracción de Talento program (2019-T1/AMB-12648) to ASM, and by a grant Juan de la Cierva Incorporación (IJCI-2017-32301) funded by MCIN/AEI/10.13039/501100011033 and 'ESF investing in your future'. SWG was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant. RC carried out this work during an undergraduate internship with ASM and IS supported by Universidad Complutense Madrid. We are grateful to the two reviewers, and especially to the Associate Editor Prof. Susana Magallón, for many insightful comments that helped improve the quality of our work.

Author contributions

ASM, IS and SWG designed research; ASM, IS and RC performed research; ASM and IS wrote the paper with the contributions of SWG.

ORCID

Sean W. Graham  <https://orcid.org/0000-0001-8209-5231>
Andrea S. Meseguer  <https://orcid.org/0000-0003-0743-404X>
Isabel Sanmartín  <https://orcid.org/0000-0001-6104-9658>

Data availability

The data, scripts (REV_{BAYES} and R), and output files supporting the findings of this study are available through https://github.com/asmeseguer/aquatic_transition.

References

- Albert VA, Renner T. 2020. Aquatic angiosperm ambiguities answered. *Nature Plants* 6: 181–183.
- Barrett SCH, Eckert CG, Husband BC. 1993. Evolutionary processes in aquatic plant populations. *Aquatic Botany* 44: 105–145.
- Barrett SCH, Graham SW. 1997. Adaptive radiation in the aquatic plant family Pontederiaceae: insights from phylogenetic analysis. In: Givnish TJ, Sytsma KJ, eds. *Molecular evolution and adaptive radiation*. Madison, WI, USA: Cambridge University Press, 225–258.
- Beaulieu JM. 2020. *The problem with clade-specific sampling fractions*. [WWW document] URL <https://rdrr.io/cran/hisse/f/inst/doc/Clade-specific-sampling.pdf>.
- Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* 65: 583–601.
- Borges R, Machado JP, Gomes C, Rocha AP, Antunes A. 2019. Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics* 35: 1862–1869.
- Bromham L, Hua X, Cardillo M. 2016. Detecting macroevolutionary self-destruction from phylogenies. *Systematic Biology* 65: 109–127.
- Caetano DS, O'Meara BC, Beaulieu JM. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72: 2308–2324.

- Chase MW. 2004. Monocot relationships: an overview. *American Journal of Botany* 91: 1645–1655.
- Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261: 201–217.
- Cook CDK. 1983. Aquatic plants endemic to Europe and the Mediterranean. *Botanische Jahrbücher* 103: 539–582.
- Cook CDK. 1985. Range extensions of aquatic vascular plant species. *Journal of Aquatic Plant Management* 23: 1–6.
- Cook CDK. 1990. *Aquatic plant book*. Amsterdam, the Netherlands: SPB Academic Publishing.
- Cook CDK. 1999. The number and kinds of embryo-bearing plants which have become aquatic: a survey. *Perspectives in Plant Ecology, Evolution and Systematics* 2: 79–102.
- Czekanski-Moir JE, Rundell RJ. 2019. The ecology of nonecological speciation and nonadaptive radiations. *Trends in Ecology & Evolution* 34: 400–415.
- Doyle JA. 2012. Molecular and fossil evidence on the origin of angiosperms. *Annual Review of Earth and Planetary Sciences* 40: 301–326.
- Duarte CM, Planas D, Penuelas J. 1994. Macrophytes, taking control of an ancestral home. In: Margalef R, ed. *Limnology now: a paradigm of planterary problems*. Amsterdam, the Netherlands: Elsevier, 59–79.
- Eckert CG, Dorken ME, Barrett SCH. 2016. Ecological and evolutionary consequences of sexual and clonal reproduction in aquatic plants. *Aquatic Botany* 135: 46–61.
- Feild T-S, Arens N-C. 2007. The ecophysiology of early angiosperms. *Plant, Cell & Environment* 30: 291–309.
- FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3: 1084–1092.
- Freyman W, Höhna S. 2018. The tempo of evolutionary decline in self-compatible plant lineages. *Systematic Biology* 68: 505–519.
- Gambaryan PP, Kuznetsov AN. 2021. Gnetum and Nymphaeaceans as models for a scenario of the origin of morphotype of flowering plants. *Biology Bulletin Reviews* 11: 237–253.
- Givnish TJ. 2015. Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist* 207: 297–303.
- Givnish TJ, Zuluaga A, Spalink D, Gomez MS, Lam VKY, Saarela JM, Sass C, Iles WJD, De Sousa DJL, Leebens-Mack J. 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. *American Journal of Botany* 105: 1888–1910.
- Gleick PH. 1993. *Water in crisis*. New York, NY, USA: Oxford University Press.
- Goldberg EE, Lancaster LT, Ree RH. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology* 60: 451–465.
- Gómez B, Daviero-Gómez V, Coiffard C. 2008. Comment vivaient les premières plantes à fleurs? *La Recherche* 419: 48–52.
- Grace JB. 1993. The adaptive significance of clonal reproduction in angiosperms: an aquatic perspective. *Aquatic Botany* 44: 159–180.
- Graham LE. 1993. *Origin of land plants*. New York, NY, USA: John Wiley & Sons Inc.
- Guiry MD. 2012. How many species of algae are there? *Journal of Phycology* 48: 1057–1063.
- Hernández-Hernández T, Wiens JJ. 2020. Why are there so many flowering plants? A multiscale analysis of plant diversification. *American Naturalist* 195: 948–963.
- Höhna S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *Journal of Theoretical Biology* 380: 321–331.
- Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. REVBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65: 726–736.
- Hutchinson GE. 1975. *A treatise on limnology: limnological botany*. New York, NY, USA: John Wiley & Sons.
- Iles WJD, Lee C, Sokoloff DD, Remizowa MV, Yadav SR, Barrett MD, Barrett RL, Macfarlane TD, Rudall PJ, Graham SW. 2014. Reconstructing the age and historical biogeography of the ancient flowering-plant family Hydatellaceae (Nymphaeales). *BMC Evolutionary Biology* 14: 1–10.
- Janssens S, Couvreur TLP, Mertens A, Dauby G, Dagallier L-P, Vanden Abeele S, Vandeloos F, Mascarello M, Beekman H, Sosef M *et al.* 2020. A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data Journal* 8: e39677.
- Jordan WC, Courtney MW, Neigel JE. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* 83: 430–439.
- Kass RE, Raftery AE. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.
- Knowlton N. 2000. Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* 420: 73–90.
- Laenen B, Machac A, Gradstein SR, Shaw B, Patiño J, Désamoré A, Goffinet B, Cox CJ, Shaw AJ, Vanderpoorten A. 2016. Increased diversification rates follow shifts to bisexuality in liverworts. *New Phytologist* 210: 1121–1129.
- Les DH. 1988. Breeding systems, population structure, and evolution in hydrophilous angiosperms. *Annals of the Missouri Botanical Garden* 75: 819–835.
- Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character’s effect on speciation and extinction. *Systematic Biology* 56: 701–710.
- Magee AF, Höhna S, Vasylyeva TI, Leaché AD, Minin VN. 2020. Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts. *PLoS Computational Biology* 16: e1007999.
- Martín-Closas C. 2003. The fossil record and evolution of freshwater plants: a review. *Geologica Acta* 1: 315–338.
- Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149: 646–667.
- Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA *et al.* 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.
- Moore BR, Höhna S, May MR, Rannala B, Huelsenbeck JP. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences, USA* 113: 9569–9574.
- Neer S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences* 344: 305–311.
- Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearce W. 2013. *The CAPER package: comparative analysis of phylogenetics and evolution in R*. R package v.5.
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.
- Philbrick CT, Les DH. 1996. Evolution of aquatic angiosperm reproductive systems. *BioScience* 46: 813–826.
- Rabosky DL, Donnellan SC, Grudler M, Lovette IJ. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology* 63: 610–627.
- Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64: 340–355.
- Rambaut A, Drummond AJ. 2009. *Tracer v.1.5, MCMC trace analysis package*. [WWW document] URL <http://tree.bio.ed.ac.uk/software/>.
- Reyes E, Nadot S, Von Balthazar M, Schönerberger J, Sauquet H. 2018. Testing the impact of morphological rate heterogeneity on ancestral state reconstruction of five floral traits in angiosperms. *Scientific Reports* 8: 1–12.
- Ronquist F, Kudlicka J, Senderov V, Borgström J, Lartillot N, Lundén D, Murray L, Schön TB, Broman D. 2021. Universal probabilistic programming offers a powerful approach to statistical phylogenetics. *Communications Biology* 4: 1–10.
- Santamaría L. 2002. Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecologica* 23: 137–154.
- Schluter D. 2000. *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.
- Sculthorpe CD. 1967. *The biology of aquatic vascular plants*. London, UK: Edward Arnold Ltd.

- Silvertown J. 2008. The evolutionary maintenance of sexual reproduction: evidence from the ecological distribution of asexual reproduction in clonal plants. *International Journal of Plant Sciences* 169: 157–168.
- Silvestro D, Cascales-Miñana B, Bacon CD, Antonelli A. 2015. Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytologist* 207: 1469–8137.
- Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.
- Sokoloff DD, Marques I, Macfarlane TD, Remizowa MV, Lam VK, Pellicer J, Hidalgo O, Rudall PJ, Graham SW. 2019. Cryptic species in an ancient flowering-plant lineage (Hydatellaceae, Nymphaeales) revealed by molecular and micromorphological data. *Taxon* 68: 1–19.
- Stadler T. 2011. Inferring speciation and extinction processes from extant species data. *Proceedings of the National Academy of Sciences, USA* 108: 16145–16146.
- Stephens PR, Wiens JJ. 2004. Convergence, divergence, and homogenization in the ecological structure of Emydid turtle communities: the effects of phylogeny and dispersal. *American Naturalist* 164: 244–254.
- Sun G, Ji Q, Dilcher DL, Zheng S, Nixon KC, Wang X. 2002. Archaefractaceae, a new basal angiosperm family. *Science* 296: 899–904.
- Van Groenendael JM, Klimeš L, Klimešová J, Hendriks RJJ. 1996. Comparative ecology of clonal plants. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351: 1331–1339.
- Vermeij GJ, Dudley R. 2000. Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biological Journal of the Linnean Society* 70: 541–554.
- Wallace AR. 1878. *Tropical nature, and other essays*. London, UK: Macmillan.
- Wetzel RG. 1988. Water as an environment for plant life. In: Symoens JJ, ed. *Vegetation of inland waters*. Dordrecht, the Netherlands: Kluwer Academic, 1–30.
- 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. *Nature Plants* 6: 215–222.
- Zeidler A, Schneider S, Jung C, Melchinger AE, Ditttrich P. 1994. The use of DNA fingerprinting in ecological studies of *Phragmites australis* (Cav.) Trin. ex Steudel. *Botanica Acta* 107: 237–242.
- Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X *et al.* 2020. The water lily genome and the early evolution of flowering plants. *Nature* 577: 79–84.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Phylogenetic signal of the aquatic habit trait, when the distribution of trait values is randomized along the tips of the phylogeny.

Fig. S2 Clade age and species richness correlation across aquatic angiosperms.

Fig. S3 Significant differences between diversification parameters of terrestrial and aquatic species with polymorphic genera coded as aquatic.

Fig. S4 Ancestral habitat reconstruction in HiSSE with polymorphic genera coded as ‘aquatic’ under the genus-level sampling fraction.

Fig. S5 Ancestral habitat reconstruction in HiSSE with polymorphic genera coded as ‘terrestrial’ under the genus-level sampling fraction.

Fig. S6 Habitat-dependent diversification rates estimated with HiSSE for the dataset with polymorphic genera coded as ‘aquatic’ under the genus-level sampling.

Fig. S7 Habitat-dependent diversification rates estimated with HiSSE for the dataset with polymorphic genera coded as ‘terrestrial’ under the species-level sampling.

Fig. S8 Ancestral habitat reconstruction in HiSSE with polymorphic genera coded as ‘aquatic’ under the species-level sampling.

Fig. S9 Ancestral habitat reconstruction in HiSSE with polymorphic genera coded as ‘terrestrial’ under the species-level sampling.

Fig. S10 ASR-LTT plots for aquatic and terrestrial angiosperms, and diversification rates through time estimated in TREEPAR.

Fig. S11 Extant LTT plots for aquatic and terrestrial angiosperms, and diversification rates through time estimated with episodic birth–death (EBD) model in REVBayes.

Fig. S12 Number and timing of transition events through time.

Table S1 Bayes Factor comparison for the BD, BiSSE, HiSSE and CID2 models with the sampling parameter reflecting the species fraction of diversity.

Table S2 HiSSE results under the species-level analyses.

Table S3 TREEPAR results.

Table S4 Comparison of marginal likelihood estimates for the constant (BD) and episodic (EBD) birth–death models in REVBayes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.