Freshwater diatoms diversify and turn over faster than marine in both planktonic and benthic habitats

Teofil Nakov[1,*], Jeremy M. Beaulieu[1], Andrew J. Alverson[1]

1 Department of Biological Sciences, University of Arkansas Fayetteville, 1 University of Arkansas, SCEN 601, Fayetteville, AR 72701

*Corresponding author: tnakov@uark.edu

Abstract

The marine–freshwater divide is a key determinant of the distribution and diversity of aquatic organisms. Many clades that span the salinity gradient are disproportionately more diverse in the younger, short-lived, and scarcer freshwater environments than they are in the marine realm. Differences in diversification rates between marine and freshwater lineages contribute to these imbalances. However, marine and freshwaters are not ecologically homogenous so assessing diversification across the salinity gradient together with interacting variables might provide important insights. Diatoms are an excellent system to study such interactions because they display a striking diversity imbalance across salinities, but also between planktonic and benthic habitats. We used multistate hidden state speciation and extinction models to test hypotheses regarding diversification and colonizations across the salinity gradient. Diatom lineages sharing a habitat diversified at substantially different rates linked to effects of unobserved variables accounted for by our models. However, despite this variation, a signal for state-dependent diversification was detectable. Both planktonic and benthic lineages diversified faster in freshwaters. Traversals of the salinity gradient were bidirectional, rejecting longstanding hypotheses about the permeability of the salinity barrier. Diatoms both speciate and go extinct faster in freshwater, suggesting an association between the rates of lineage and habitat turnover.

Keywords

Diatoms, Salinity gradient, Species turnover, State speciation and extinction

Introduction

Freshwater habitats comprise only about 2% of the biosphere, yet in lineages that span the salinity gradient the magnitude of freshwater diversity is often comparable, or surpasses, species richness in the marine realm, which covers 70% of Earth's surface (Appeltans et al., 2012; Grosberg et al., 2012; Miller & Wiens, 2017; Mora et al., 2011; Wiens, 2015). This is striking given that most such lineages originated in the oceans, and so—all else equal—have had more time to accumulate diversity within their ancestral marine environment. Disconnects between species richness and time for accumulation of diversity between marine and freshwaters are found across animals [e.g., flatworms, rotifers, molluscs, and ray-finned fish (Guinot & Cavin, 2015; Miller & Wiens, 2017; Wiens, 2015)] and specific tests about diversification across the salinity divide have revealed that in amniotes, for example, marine lineages have higher extinction rates (Miller & Wiens 2017, but see Vermeij et al. 2018), and in fish, freshwater lineages diversify faster (Bloom et al., 2013). Overall, there is not a clear consensus about the underlying mechanisms (Miller & Wiens, 2017; Vermeij et al., 2018) driving the patterns seen in animals, whereas plants and fungi are too depauperate in marine systems to be informative (Appeltans et al., 2012; Mora et al., 2011). Microbes, including protists, are ideal candidates for advancing our understanding of diversification across the marine–freshwater divide: they are diverse at both ends of the salinity gradient (Appeltans et al., 2012; Mora et al., 2011) and offer alternative perspectives due to their vastly different dispersal capabilities, population sizes, and physiological features compared to animals. Moreover, studies in other groups have yet to consider that the focal contrast (marine vs. freshwater) is almost certainly *not* the sole determinant of diversification dynamics, but rather, these patterns are likely the product of many potentially interacting aspects of the habitats and traits of different lineages (Beaulieu & O'Meara, 2016). Added layers of complexity might include, for example, latitudinal diversity gradients (Powell & Glazier, 2017) or subdivisions of marine and freshwater ecosystems into planktonic (suspended) and benthic (substrate-associated) habitats (Signor & Vermeij, 1994), which have long been recognized as important for the distribution and diversity of aquatic biota.

Our knowledge of microbial diversity lags far behind the extensive, and in many cases, near-complete, species catalogs of macro-organisms (Appeltans et al., 2012; Larsen et al., 2017; Mora et al., 2011). Despite this, patterns of imbalanced species richness across environmental gradients, including the marine–freshwater and plankton–benthos divides, are well known in bacteria, protists, and zooplankton (Hart et al., 2003; Logares et al., 2009; Lozupone & Knight, 2007; Rigby, 1997). In foraminifera and many zooplankton groups, planktonic lineages are often derived and appear to have originated through one or a few rare colonizations of the plankton by benthic ancestors, followed by *in situ* diversification (Darling et al., 2009; Hart et al., 2003; Rigby, 1997; Rigby & Milsom, 1996; Signor & Vermeij, 1994). Reversals to the benthos occur rarely, if ever at all (Rigby & Milsom, 1996). Similarly, most aquatic lineages are thought to be ancestrally marine and to have colonized freshwaters infrequently (Cavalier-Smith, 2009; Lee & Bell, 1999; Logares et al., 2009, 2007; Lozupone & Knight, 2007). Such transitions are considered challenging, as migrants need to mitigate osmotic stress linked to changes in cellular water potential, salt stress caused by increased uptake or loss of ions, and changes in the ratios of intracellular ions (Kirst, 1990, 1996). As a result, freshwater colonizations among protists are generally considered rare and phylogenetically deep events with broad evolutionary consequences (Cavalier-Smith, 2009; Logares et al., 2009; Mann, 1999).

Diatoms are a species-rich lineage of eukaryotic algae that play important roles in the global cycling of carbon and oxygen and form the base of many aquatic food webs (Armbrust, 2009; Bowler et al., 2010).

Diatoms span the marine–freshwater and plankton–benthos divides, though few data have been brought to bear on characterizing the directions, patterns, and consequences of transitions across these and other major environmental gradients. The history of plankton–benthos transitions in diatoms is just beginning to be resolved. Transitions between habitats are infrequent relative to changes between colonial and solitary growth forms, and colonial species are more likely to colonize the plankton (Nakov et al., 2015). Transitions between substrate-associated and suspended habitats are possibly linked to the mode of sexual reproduction. The numerous, flagellum-propelled male gametes of oogamous species might facilitate colonization of the plankton, whereas the few, aflagellated gametes of isogamous lineages might hinder survival and reproduction in dulute environments (Kooistra et al., 2009). With respect to salinity, most diatom genera are restricted to either marine or freshwaters (Round et al., 1990), and colonizations of freshwaters are considered infrequent and irreversible (Mann, 1999; Round & Sims, 1980). Notwithstanding, reconstructions of the timing and direction of traffic across the salinity gradient have revealed temporally spaced and bidirectional colonizations (Alverson, 2014; Alverson et al., 2007; Ruck et al., 2016). Importantly, however, previous studies have either focused on relatively small clades, making it difficult to derive broadly applicable conclusions (Beaulieu & O'Meara, 2018), or have not explored the interaction between rates of diversification and habitat transitions, a key component of testing character-dependent diversification (Maddison, 2006).

We characterize diatom evolution across the marine–freshwater and plankton–benthos divides using a large, time-calibrated phylogeny representative of the known breadth of extant diatom diversity. To account for variation in the rate of diversification and colonizations across environments, we employed a hidden state speciation and extinction modeling framework (HiSSE; Beaulieu & O'Meara 2016) modified for multiple characters. In doing so, we were able to jointly estimate rates of colonization and diversification across multiple environments and account for possible interactions between the focal characters (i.e., marine plankton, marine benthos, freshwater plankton, and freshwater benthos). We discuss the implications of our findings for diatom ancestry, diversification, and phylogenetic diversity across the ecologically important marine–freshwater and plankton–benthos divides.

Materials and Methods

Phylogeny and diversity data We used a time-calibrated phylogeny of 1,151 diatoms (Nakov et al., 2018)—the largest available phylogenetic tree of diatoms that covers their known phylogenetic and ecological breadth. We scored 1,132 out of the 1,151 species for two characters: "habitat" (plankton or benthos) and "salinity" (marine or freshwater) based on taxonomic compilations (Hasle & Syvertsen, 1996; Lange-Bertalot & Krammer, 2000; Lobban et al., 2012; Round et al., 1990; Witkowski et al., 2000) and previous work on character evolution (Alverson et al., 2007; Nakov et al., 2015; Ruck et al., 2016).

Given a total estimated diatom species richness of up to 100,000 species (Mann & Vanormelingen, 2013), this phylogeny, although the most taxonomically inclusive to date, still falls well short of the full diversity of diatoms. To account for this under-sampling in our diversification analyses, we estimated sampling fractions for each environment, e.g., the fraction of sampled marine species from the total number of marine species described to date. Data for salinity (marine–freshwater) were obtained from DiatomBase (Kociolek et al., 2017), which reports the total estimated number of species per genus and the total estimated number of marine species per genus, allowing for a straightforward calculation of the number of non-marine (freshwater) species. Similar data were not available for planktonic and benthic lineages, so we used the primary taxonomic literature (Hasle

& Syvertsen, 1996; Lange-Bertalot & Krammer, 2000; Lobban et al., 2012; Round et al., 1990; Witkowski et al., 2000) to record the number of species per genus occurring in the plankton or benthos. Based on these numbers, we then derived a per-genus fraction of planktonic species and multiplied the total estimated genus-level diversity by this fraction to obtain an estimate of the number of planktonic and benthic species per genus. We summed these numbers across all genera represented on our tree to approximate the diatom-wide estimates of species richness in each environment.

Analysis of binary characters Hypotheses about associations between lineage diversification and discretely coded traits or environments are often tested using state speciation and extinction models (-SSE; Beaulieu & O'Meara 2016; FitzJohn et al. 2009; Goldberg et al. 2011; Maddison et al. 2007). These models jointly estimate the rates of evolution of a focal character (i.e., transitions between states) and rates of diversification (i.e., speciation and extinction) and can therefore account for potential interactions between these two factors. In one of these models, the binary state speciation and extinction (BiSSE; FitzJohn et al. 2009; Maddison et al. 2007), each character state has its own speciation and extinction rates that are different from the parameters associated with the other state. This property provides a natural way to disentangle whether an imbalanced ratio of diversity between habitats reflects differences in diversification, or alternatively, asymmetry in rates of colonization between alternative habitats (Maddison, 2006; Maddison et al., 2007).

An assumption of the BiSSE model is that the rates of speciation, extinction, and transition are shared by all lineages occupying a particular environment (Beaulieu & O'Meara, 2016; Maddison et al., 2007). For example, with a BiSSE model we can test whether marine and freshwater diatoms diversified at different rates, but with the constraint that all marine and all freshwater lineages share the same rate. Whether this assumption is reasonable depends on the phylogeny and how the diversification regimes have been defined. In our case, with a phylogeny representing about 200 million years (My) of evolutionary history, known rate variation in diversification related to life history traits (Nakov et al., 2018), and broadly defined diversification regimes (e.g., tropical and polar oceans are grouped within the same "marine" regime), it is appropriate to assess and accommodate expected rate variation within our focal regimes.

Two approaches within the SSE family of models can be used to detect and account for heterogeneity in diversification on a phylogeny while co-estimating the transition rates between character states. They both use hidden (unobserved) states to implicitly account for the many potential factors, aside from the focal characters, that may have influenced diversification and character transition rates (Beaulieu & O'Meara, 2016; Caetano et al., 2018). First, hidden state speciation and extinction models (HiSSE; Beaulieu & O'Meara 2016) extend the BiSSE model by allowing two hidden states within each state of the focal character, but still assume that diversification is character-dependent. Second, character-independent models (CID) remove the link between diversification and the focal character entirely and associate diversification rates only to hidden (or unobserved) traits. By allowing rate variation independent of the focal character, the CID models represent non-trivial null models for testing hypotheses of state dependent speciation and extinction (Beaulieu & O'Meara, 2016; Caetano et al., 2018). In other words, for cases in which diversification rates vary across the phylogeny, parameter-rich SSE models might be favored even when the diversity patterns are not tied to the characters singled out for study (Beaulieu & O'Meara, 2016; Caetano et al., 2018; Rabosky & Goldberg, 2015). Evaluation of CID models, in turn, allows more realistic assessments of whether variation in diversification rates across the diatom phylogeny was related to habitat, salinity, or other unobserved characters.

We fit a total of five models for each binary character: **(1)** a "trivial null" model (number of parameters, k=3,

Fig. S1) with one parameter controlling the turnover rate (defined as speciation + extinction, which measures the rate of any event, birth or death, occurring per unit time) for both observed states, and one parameter controlling the rate of character transitions (e.g., marine $\rightarrow$ freshwater); **(2)** a BiSSE model (k=5, Fig. S1) with separate turnover rates estimated for lineages living in alternate environments and asymmetric transitions between environments (e.g., marine $\rightarrow$ freshwater $\neq$ freshwater $\rightarrow$ marine); **(3)** a HiSSE model (k=13) with separate turnover rates for each combination of observed and hidden states and eight transition rates for changes between states (e.g., marine hidden state A $\rightarrow$ marine hidden state B)—this model disallowed simultaneous transitions of the observed and hidden states (e.g., marine hidden state A $\rightarrow$ freshwater hidden state B; Fig. S1); **(4)** a CID2 model (k=6) with two turnover rate parameters similar to the BiSSE model, but unlinked from the observed character states; and **(5)** a CID4 model (k=8) with five diversification parameters matching the complexity of HiSSE (Fig. S1). In all cases, we kept the extinction fraction (defined as extinction/speciation, which measures the ratio of death and birth events per unit time) constant across states.

Analysis of multistate characters For the combined habitat + salinity character, we extended the HiSSE and CID model framework to handle a four-state case [i.e., marine-plankton (MP), marine-benthos (MB), freshwater-plankton (FP), and freshwater-benthos (FB)]. In this analysis, we again fit a "trivial null" model that assumed no variation in diversification and a multistate speciation and extinction model (MuSSE; FitzJohn 2012) that linked diversification rates to the four character states, but assumed no variation within regimes (Fig. S1). We then fit two multistate speciation and extinction models with hidden traits, MuHiSSE, and seven multistate character-independent models (MuCID) with increasing numbers of hidden traits (MuCID2 through MuCID8; Fig S1). The most complex MuCID8 model had eight parameters for turnover and one for the extinction fraction, matching the complexity of the MuHiSSE model, but with diversification parameters unlinked from the observed character states (Fig. S1). For both types of models, we allowed the transition rates between alternative environments to differ (e.g., MB $\rightarrow$ MP $\neq$ MP$\rightarrow$ MB), and again we disallowed dual transitions (e.g., MB $\rightarrow$ FP). The two MuHiSSE models differed in how they handle transitions between the hidden states. In the first one (MuHiSSE reduced), we assumed one rate for all transitions (e.g., MB hidden state A $\rightarrow$ MB hidden state B = FP hidden state A $\rightarrow$ FB hidden state B), while in the second model (MuHiSSE full) we allowed separate transition rates from hidden state A to B and hidden state B to A (e.g., MB hidden state A $\rightarrow$ MB hidden state B $\neq$ MB hidden state B $\rightarrow$ MB hidden state A; Fig S1). Finally, for the analyses of the combined habitat + salinity character we assessed two alternative implementations of the conditioning for survival of crown lineages, first the original `diversitree` and `HiSSE` implementation (Beaulieu & O'Meara, 2016; FitzJohn et al., 2009) (`HiSSE` setting `root.type="madfitz"`) and second the conditioning introduced by Herrera-Alsina et al. (2018) (`HiSSE` setting `root.type="herr_als"`).

In all cases, whether or not we allowed a separate estimate of turnover rate for each state, the extinction fraction was always fixed across observed and hidden states. It is important to note, however, that while extinction fraction is constrained among states, converting turnover and extinction fraction into speciation and extinction allows estimating separate birth and death rates for each state (Beaulieu & O'Meara, 2016). We also accounted for incomplete sampling by incorporating the sampling fractions described above directly into our diversification analyses (Beaulieu & O'Meara, 2016; FitzJohn et al., 2009). This approach assumes random sampling of each observed character state. For all models, we performed $\geq$ 50 maximum likelihood optimizations, each initiated from a distinct starting point. We also estimated the 2 log-likelihood unit support region around the maximum likelihood estimates (`HiSSE` functions `SupportRegion` and `SupportRegionMuHiSSE`), and

Table 1. Comparison of character-dependent and character-independent models for the binary coding of salinity and habitat. The models favored by AIC are in bold. Abbreviations: npd, number of diversification parameters (turnover + extinction fraction); npt, number of transition rate parameters; ωAICc, Akaike weight.

Character	Model	npd	npt	Log L	AICc	ΔAICc	ωAICc
Salinity	**HiSSE**	8	3	-5310.42	10647.16	0	0.999
Marine/Freshwater	CID4	8	8	-5354.43	10725	77.84	<0.001
	CID2	4	3	-5422.5	10857.07	209.91	<0.001
	BiSSE	4	2	-5484.45	10978.95	331.79	<0.001
	Null	2	1	-5591.24	11190.49	543.33	<0.001
Habitat	**CID4**	8	3	-5170.4	10356.93	0	0.999
Plankton/Benthos	HiSSE	8	8	-5189.03	10404.38	47.45	<0.001
	CID2	4	3	-5266.67	10545.42	188.5	<0.001
	BiSSE	4	2	-5394.85	10799.75	442.82	<0.001
	Null	2	1	-5563.33	11134.66	777.73	<0.001

reconstructed marginal ancestral states (`HiSSE` functions `MarginRecon` and `MarginReconMuHiSSE`). A vignette with further details regarding `MuHiSSE` analyses is available as part of the `hisse R` package. The data and scripts used here are available at Zenodo (DOI: 10.5281/zenodo.2558448).

Results and Discussion

Diatom diversification across the marine–freshwater and plankton–benthos divides

Diatom diversity data showed strong disparities in species richness across both the marine–freshwater and plankton–benthos contrasts (Kociolek et al., 2017; Mora et al., 2011). Approximately 70% of the species described to date are freshwater and 83% are benthic. These imbalanced diversity ratios suggested that diversification rates might be related to one or both of these characters. Indeed, BiSSE models that linked diversification and character transition rates to the alternate environments provided a much better fit than models that assumed one rate across the tree (Table 1). Freshwater and benthic lineages, as expected given the diversity ratios, had waiting times (inverse of the rate) of speciation + extinction (i.e., net turnover) that were 2.6 and 4.3 times faster, respectively, than their marine and planktonic counterparts. Similarly, net diversification (speciation − extinction) was 2.5 times faster in freshwaters than marine lineages, and 4.3 times faster in benthic than planktonic diatoms. The ratio of extinction to speciation (the extinction fraction) was high in both cases—0.85 for the marine–freshwater and 0.92 for the plankton–benthos contrast. Both of these environmental gradients, according to the BiSSE models, appear to have been important in shaping the diversity of diatoms.

To assess variation in diversification and transition rates within regimes, we also evaluated character-dependent (HiSSE) and character-independent (CID) models that included hidden states, i.e., factors other than our focal characters that might have influenced diversification. For both characters, hidden state models provided a significantly improved fit relative to the BiSSE models, highlighting substantial variation in the rates of diversification within, and transitions between, regimes (Table 1). The HiSSE model was favored for salinity, while CID4 was best for habitat, in both cases with Akaike weights of $\omega = 0.999$ (Table 1). The HiSSE model for salinity showed that in both hidden states, freshwater lineages speciated, but also went extinct, faster than marine lineages, resulting in faster overall species turnover in freshwater (Table 2). The difference between speciation and extinction also favored freshwater lineages with waiting times 17–39% shorter than in

Table 2. Parameter estimates and 2 log-likelihood unit support region (in brackets) for the best model (hidden-state speciation and extinction, HiSSE) for salinity (marine vs. freshwater). Rates of transition, speciation, extinction, diversification and turnover were inverted to waiting times in millions of years. The numbers next to character states indicate the hidden states.

	Marine 1	Freshwater 1	Marine 2	Freshwater 2
Transition				
Marine 1	-	149 [111-178]	21 [16-22]	0
Freshwater 1	250 [208-322]	-	0	14 [11-16]
Marine 2	588 [400-714]	0	-	2000 [909-2000]
Freshwater 2	0	270 [169-303]	188 [178-256]	-
Diversification				
Speciation	3.4 [2.9-3.6]	2.1 [1.9-2.2]	15.9 [14.6-17.9]	13.2 [10.6-18]
Extinction	4.7 [3.8-5.2]	2.9 [2.6-3.1]	21.9 [19.5-25.6]	18.2 [15.2-24.2]
Net turnover	2.0 [1.6-2.1]	1.2 [1.1-1.3]	9.2 [8.4-10.5]	7.7 [6.3-10.3]
Net diversification	12.6 [10.8-13.4]	7.7 [6.6-8.7]	58.5 [51.3-65.4]	48.5 [35.7-71.4]
Extinction fraction	0.726 [0.698-0.755]			

marine lineages (Table 2). For habitat, under the CID4 model, waiting times to either a speciation or extinction (turnover) event across the four character-independent rate classes varied between nearly 53 and 0.4 My (Table 3).

Like any diversification analysis that accounts for unsampled diversity, our results rely on the assumption that the total (described and undescribed) species richness in our focal regimes is directly proportional to the diversity described from those habitats to date. However, in addition to differences in diversification rates, these lopsided ratios could also arise from unequal taxonomic effort or from nonstandard taxonomic practices (e.g., splitting vs. lumping) between different habitats. It is difficult to assess how these problems might affect our analyses, or analyses of any species rich group, without some non-arbitrary way to scale the diversity data or sampling fractions to account for possible over-splitting (e.g., in freshwater lineages) or potentially lower taxonomic research effort (e.g., in marine groups). Nevertheless, we repeated the marine–freshwater analysis with arbitrarily reduced marine sampling fraction, from the observed value of 14% down to 10%, 6%, and 2%. Irrespective of how low we set the marine sampling fraction, we found that HiSSE models were still overwhelmingly favored over character-independent CID4 models. With lowered sampling fractions the diversification and turnover rate estimates for marine lineages increased. However, the signal for faster diversification in freshwater, at least in one of the hidden states, was detected in all but the analysis with a marine sampling fraction of 2%. In the latter, marine lineages diversified faster than freshwater in both hidden states. Overall, these results indicated that, if the past nearly 250 years of taxonomic work on diatoms is completely biased and non-representative of diatom diversity as a whole, then some of our inferences might be affected. However, to our knowledge, there are no data that would allow us to set the sampling fractions in a way other than we have done here.

In addition to estimating the most likely parameter values for predefined regimes, the HiSSE and CID models also allow for reconstruction of the probabilities of the hidden states occupied by each lineage. For example, a lineage might be freshwater, but its hidden state might be less certain, with some probability of occupying either of the hidden states in the model. In such cases it is unclear which parameter estimate should be assigned to the lineage because the diversification and transition rates might differ substantially between the

Table 3. Parameter estimates and 2 log-likelihood unit support region (in brackets) for the best model (character-independent with four hidden states, CID4) for habitat (plankton vs. benthos). Rates of transition, speciation, extinction, diversification and turnover were inverted to waiting times in millions of years. The numbers next to character states indicate the hidden states.

Transition				
Plankton-to-Benthos	667 [588-769]			
Benthos-to-Plankton	909 [833-1111]			
Between hidden states	133 [114-169]			
Diversification	Hidden 1	Hidden 2	Hidden 3	Hidden 4
Speciation	104.2 [82-113.6]	0.8 [0.6-0.8]	13.6 [11.4-15.5]	3.7 [2.8-4.5]
Extinction	107.5 [84-116.3]	0.8 [0.7-0.8]	14.0 [11.7-15.9]	3.9 [2.9-4.6]
Net turnover	52.9 [41.5-57.8]	0.4 [0.3-0.4]	6.9 [5.8-7.9]	1.9 [1.4-2.3]
Net diversification	3333.3 [3333-5000]	26.8 [23.6-32.1]	476.2 [434-666]	129.9 [89.3-178.6]
Extinction fraction	0.971 [0.969-0.978]			

two (or more) hidden states. One way to account for this uncertainty is to average across parameter estimates, weighted by the probabilities of occupying alternative hidden states. This type of averaging of per-lineage diversification parameters, provides a measure of rate variation within regimes, because, instead of splitting species and their rates within a regime into two groups corresponding to the two hidden states, we can obtain a distribution of weighted averages of the parameters of interest. In other words, even when character-dependent diversification is supported, it is unlikely that all lineages assigned to a regime would have diversified at the same rate, and this variation will be reflected in the hidden state reconstructions. Averaged over hidden states, waiting time to turnover events (either speciation or extinction in My) ranged between 2.1 and 8.5 My (mean across species $\pm$ SD, 6.0$\pm$2.1 My) in marine and between 1.2 and 6.2 My (2.85$\pm$1.43 My) in freshwater environments (Fig. 1A). The ranges for turnover in planktonic (4.75$\pm$3.46 My) and benthic (4.94$\pm$3.48 My) species were overlapping, with a minimum of 0.5 My and maximum of 12.3 My (Fig. 1B). The variation in net diversification was proportional to turnover (Fig. 1A,B) and for both binary characters, these tests highlighted broad variation in turnover and net diversification rates within regimes (Fig. 1A,B).

Combined salinity + habitat diversification regimes

One way to further evaluate this variation is to combine salinity and habitat into a single four-state character: marine–plankton, marine–benthos, freshwater–plankton, and freshwater–benthos. With this approach, we can specifically test for differences between, for example, the plankton and benthos, while allowing for the possibility that planktonic lineages in fresh vs. saline waters also differ in diversification. As an example of why such an interaction might be appropriate, consider the age and longevity of freshwater lakes that support planktonic communities. The formation and lifespan of most lakes is linked to Earth's glaciation cycles, with only a few dozen long-lived tectonic basins whose formation predates the last interglacial (ca. 130–115 thousand years ago, Hampton et al. 2018; Mackay et al. 2010). As the majority of freshwater lakes are geologically recent and have relatively short lifespans, freshwater planktonic lineages might be under drastically different environmental pressures relative to the marine planktonic species from environments that have existed continuously since the formation of the oceans. An indication of these potential differences in diversification dynamics can be gleaned from the oftentimes drastic change in relative abundance or composition of planktonic

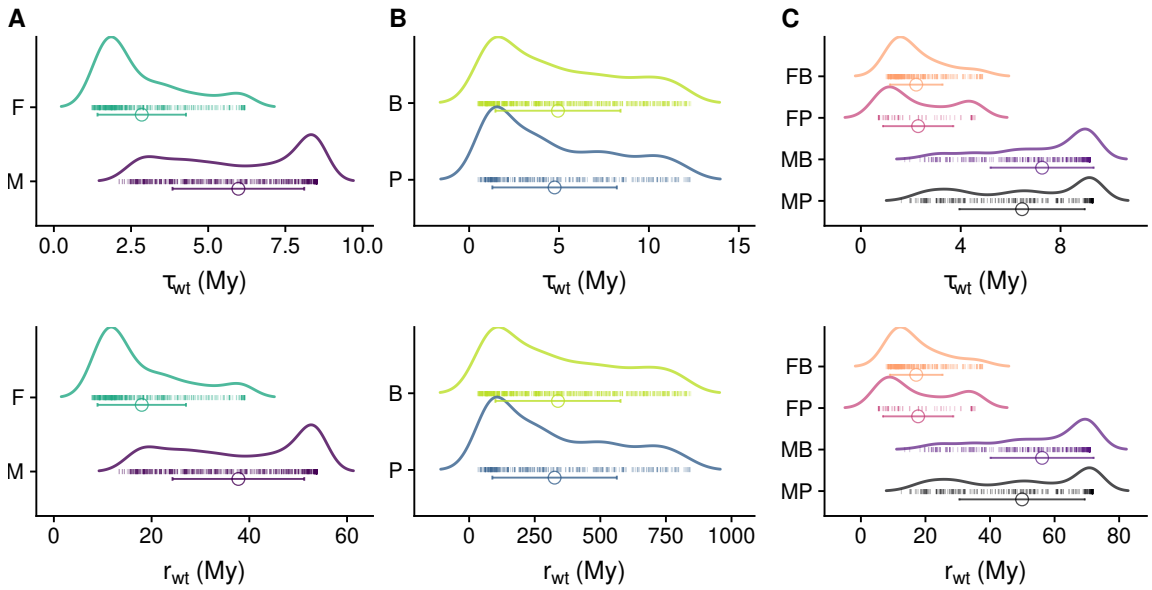

Figure 1. Distributions of tip estimates of net turnover (speciation + extinction) and net diversification (speciation − extinction) averaged across hidden states under HiSSE's original implementation of the conditioning for survival of crown lineages (Beaulieu & O'Meara, 2016; FitzJohn et al., 2009). (A) binary coding for marine vs. freshwater (HiSSE model), (B) binary coding for planktonic vs. benthic (CID4 model) and (C) multistate coding for the combination of salinity and habitat (MuHiSSE model). The circles and horizontal error bars show the mean ± standard deviation across tips of the phylogeny and the carpet points show the individual rates. The rates are inverted into waiting times (wt) in millions of years. Abbreviations: M, marine; F, freshwater; P, plankton; B, benthos.

communities in sedimentary records from ancient lakes that closely follow glaciation and drought cycles (Cohen et al., 2007; Edlund & Stoermer, 2000; Mackay et al., 2010; Stone et al., 2011).

Another example of the importance of the habitat + salinity interaction is the size and connectivity of marine vs. freshwater populations. Diatoms in marine environments have fewer physical barriers to migration and are generally expected to have broader geographic distributions and higher rates of gene flow (Palumbi, 1992, 1994), which together might result in lower speciation and extinction rates compared to their freshwater counterparts, which inhabit discontinuous and irregularly distributed habitats. Still, gene flow can be limited among marine diatom populations due to isolation by distance (Casteleyn et al., 2010) or ecology (Whittaker & Rynearson, 2017), sometimes even at small spatial scales (Godhe & Härnström, 2010; Rynearson & Armbrust, 2004; Sjöqvist et al., 2015), and this suggests that population subdivision might be even stronger in freshwater lineages (Evans et al., 2009; Vanormelingen et al., 2008). Taken together, these observations suggest that diversification or turnover proceeds at different paces in, for example, lineages of marine plankton vs. freshwater plankton.

To assess diversification rate differences across the four habitat + salinity regimes, we implemented a multistate variant of the HiSSE model (MuHiSSE, Caetano et al. 2018), with a separate turnover parameter for each of the four states and two hidden states within each combination. To guard against false-positive selection of these highly parameterized models, we designed and tested a set of multistate character-independent (MuCID) models with increasing numbers of diversification parameters, up to a MuCID model with eight hidden traits that matches the complexity of the MuHiSSE model [eight turnover and one extinction fraction that can be transformed into eight speciation and eight extinction rates (Beaulieu & O'Meara, 2016); `HiSSE` functions `ParameterTransform` or `ParameterTransformMuHiSSE`, Fig. S1].

Table 4. Comparison of character-dependent and independent models for the interaction between diversification and the combined salinity+habitat states. The favored model is in bold. Abbreviations: npd, number of diversification parameters (turnover + extinction fraction); npt, number of transition parameters; ωAICc Akaike weight.

Model	npd	npt	Log L	AICc	ΔAICc	ωAICc
MuHiSSE full	9	18	-5521.224	11096.45	0	0.999
MuHiSSE reduced	9	17	-5539.579	11131.16	34.71	<0.001
MuCID8 full	9	16	-5562.629	11175.26	78.81	<0.001
MuCID7 full	8	15	-5562.682	11171.36	74.91	<0.001
MuCID6 full	7	14	-5562.692	11167.38	70.93	<0.001
MuCID5 full	6	13	-5562.699	11163.4	66.95	<0.001
MuCID4 full	5	12	-5563.632	11161.26	64.81	<0.001
MuCID3 full	4	11	-5566.861	11163.72	11163.72	<0.001
MuCID2 full	3	10	-5572.266	11170.53	11170.53	<0.001

Data for the combined salinity + habitat states mirrored disparities in species richness observed for the individual binary traits. Marine plankton accounted for 11%, marine benthos 20%, freshwater plankton 6%, and freshwater benthos represent 63% of diatom species. Many marine planktonic lineages have much deeper fossil records than benthic lineages (Sims et al., 2006), and freshwater lineages are generally nested within, and younger than, their marine counterparts (Alverson, 2014; Ruck et al., 2016). A superficial look at these data suggests that younger lineages are disproportionately more diverse than older ones, perhaps due to differences in diversification dynamics. Indeed, under either of the alternative implementations for conditioning of survival of crown lineages, we found strong support for a MuHiSSE model (Akaike weight, ω=0.999) in which each of the four character states had separate diversification parameters and all transitions between states were free (Fig. S1, Table 4).

Under the MuHiSSE model, the maximum likelihood estimates and 2 log likelihood support regions indicated that marine lineages from the plankton and benthos diversified under similar rates in both hidden states (Table S1). For freshwater lineages, the diversification dynamics between the plankton and benthos differed, with shorter waiting times for net diversification and turnover in freshwaters, and non-overlapping 2 log likelihood support regions (Table S1). Averaged over hidden states, net turnover and net diversification in freshwater habitats were 2–3 times faster than in marine environments (Fig. 1C), irrespective of the type of conditioning for survival of crown lineages (Fig. S2). The relatively high global extinction fraction (ϵ = 0.77; Table S1), estimated here from extant species only, was similar to calculations made from Cenozoic marine fossils (ϵ = 0.75; Nakov et al., 2018).

The support for models that include hidden traits shows that, as would be expected, the focal characters alone (salinity and habitat) do not explain all of the variation in diversification rates across the phylogeny, even after partitioning the data into four groups. This further underscores the need to account for the multitude of environmental, ecological, and trait factors that were not explicitly considered, but whose cumulative effect on diversification might be comparable to the effects of salinity and habitat. To illustrate, we cam compare the differences in diversification between environments to the differences between hidden states within an environment. Marine planktonic lineages reconstructed in one of the hidden states had a waiting time to a speciation event of about 18.1 My. This was similar to marine benthic lineages in the same hidden state (18.1 My), but was several times shorter than marine planktonic lineages in the other hidden state (2.9 My; Fig. 2; Table S1). In other words, the alternative suspended vs. substrate-associated habitats conferred a smaller difference in diversification compared to differences conferred by unobserved factors approximated by the

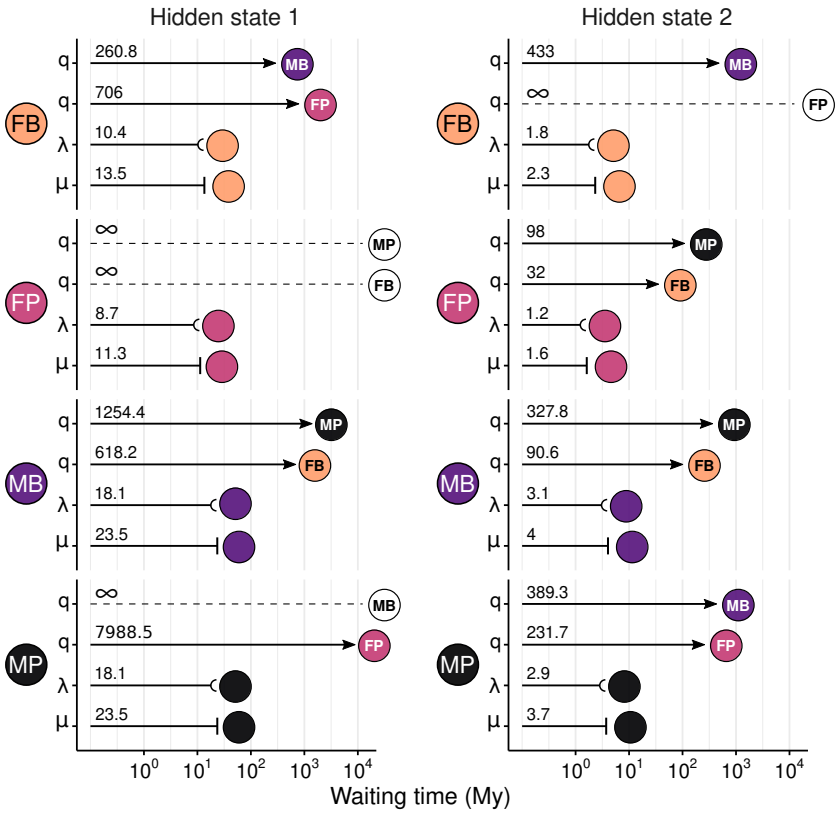

Figure 2. Waiting times (inverse of the rates) in million years (My) for speciation, extinction, and habitat + salinity transitions under the MuHiSSE model and original implementation of the conditioning for survival of crown lineages (Beaulieu & O'Meara, 2016; FitzJohn et al., 2009). The source environment is denoted with two-letter abbreviation on the y axis. For speciation (λ) and extinction (μ) the color of the end-point circle matches the source environment. For transition rates (q), the color and abbreviation at the end-point circle denote the destination of colonization. Numbers along the lines are the waiting times. Transitions rates estimated as indistinguishable from zero (and, consequently, infinity waiting times) are shown with dashed lines and open circles. Left and right side panels correspond to the two hidden states. Abbreviations: FB, freshwater–benthos; FP, freshwater–plankton; MB, marine–benthos; MP, marine–plankton.

hidden states (Fig. 2; Table S1). This is not to say that salinity and habitat are not associated with diversification patterns in diatoms, but the opposite—that their effects can be detected despite broad variation related to other, unobserved factors.

It is worth considering what some of these unmeasured factors might be. For example, differences in age, size, depth, and longevity between lakes of glacial, tectonic, and volcanic origin might be driving some of the variation within freshwaters. Likewise, the drastic contrast in conditions experienced by diatom communities in nutrient-rich marine polar waters vs. oligotrophic oceanic gyres could account for some of the hidden state components within marine lineages. For example, Crampton et al. (2016) found that peaks in species turnover of marine Antarctic diatoms coincided with cooling of southern latitudes and expansion of ground and sea ice. Five such pulses above the baseline turnover have been inferred over the past 15 My (Crampton et al., 2016), contributing to relatively fast turnover within Antarctic diatom communities. Similar cooling-driven fluctuations in turnover are less likely to have occurred in tropical ocean regions where temperature has been relatively constant over the past 15 My (Powell & Glazier, 2017). Therefore, it is possible that a portion of the variation captured by the hidden states might also be related to latitude. In this regard, diatoms are among the

few lineages with a reverse latitudinal diversity gradient, which is thought to be a product of asymmetric range expansion toward higher latitudes (Powell & Glazier, 2017).

Variation in rates of diversification within regimes could also stem from other traits that covary with diversity patterns in diatoms. For example, motility of vegetative cells, which allowed colonization of new types of habitats and changed the ways in which diatoms interact with each other and the environment, also appears to be associated with a faster rate of diversification (Nakov et al., 2018). A comparison between motile and non-motile lineages showed that the probability of occupying the second hidden state—where net turnover was much faster (Fig. 3, Table S1)—was higher in motile lineages (ANOVA, $p<0.001$) and suggested that motility might be an important factor not directly considered by our SSE models. It is important to reiterate that this *post hoc* exploration of potential hidden variables is not meant as a way to draw conclusions about other factors not directly tested here, but only as an illustration of the many other variables that may have contributed to diatom diversification (see e.g., Lewitus et al., 2018).

Diatom ancestry, absorbing habitats, and phylogenetic diversity

Ancestral state reconstructions supported a marine planktonic environment for the most recent common ancestor of diatoms, but they quickly colonized the marine benthos (see also Kooistra et al., 2007). The oldest extant benthic lineages were estimated to be about 180–175 My old (Fig. 3). Colonizations of freshwater did not occur until much later—the oldest extant lineages to gain foothold in the freshwater plankton date back to 120-130 Mya, some 60 My after the origin of diatoms (Fig. 3). Apart from a few isolated cases that have not left a substantial mark on present day diversity (e.g., *Ellerbeckia*, *Terpsinoe*, and *Pleurosira*), diatoms did not diversify greatly within the freshwater benthos until after the evolution of anisogamous sexual reproduction and axial cell symmetry in the pennate diatom lineage (Fig. 3). Marine benthic ancestry was inferred for the MRCA of the pennate clade, and within it, the clade of actively motile raphid diatoms (Fig. 3). Overall, although our reconstructions supported a planktonic ancestry for diatoms, it is clear that the marine benthos has been the center stage of diatom evolution, seeding diversity across a range of environments (e.g., the marine plankton and freshwater benthos) throughout their evolutionary history (Fig. 3; see also Kooistra et al., 2007).

A long-held view in diatom ecology and evolution is that colonizations of freshwater are infrequent and irreversible events (e.g., the Rubicon hypothesis of Mann 1999), suggesting that the modern freshwater flora has been assembled from a small set of phylogenetically diverse colonizers (rarity), and that once adapted to freshwaters, these lineages rarely, if ever, cross back to their ancestral marine habitat (irreversibility). The freshwater diatom flora is without question a composite of phylogenetically distant lineages that trace back to independent colonization events (Fig. 3), but the extent to which freshwater lineages have seeded diversity in the marine environment has not been tested at the level of all diatoms and in a framework that accounts for the interaction between rates of colonization and diversification.

The hypothesis of a unidirectional (marine-to-freshwater only) bridge across the salinity barrier implicates freshwaters as absorbing, meaning that marine colonists invade, adapt, and specialize to the point that transitions back to the ocean are unlikely. In modeling terms, this would translate as asymmetric transition rates, with the rates of transition from marine to fresh waters far exceeding the reverse. By simultaneously estimating the diversification and transition rates, our models provided a powerful test of this hypothesis, and accounting for rate heterogeneity within regimes allows for the possibility that salinity is a Rubicon in some, but not all, diatom lineages. Our analysis of the binary marine vs. freshwater data revealed that

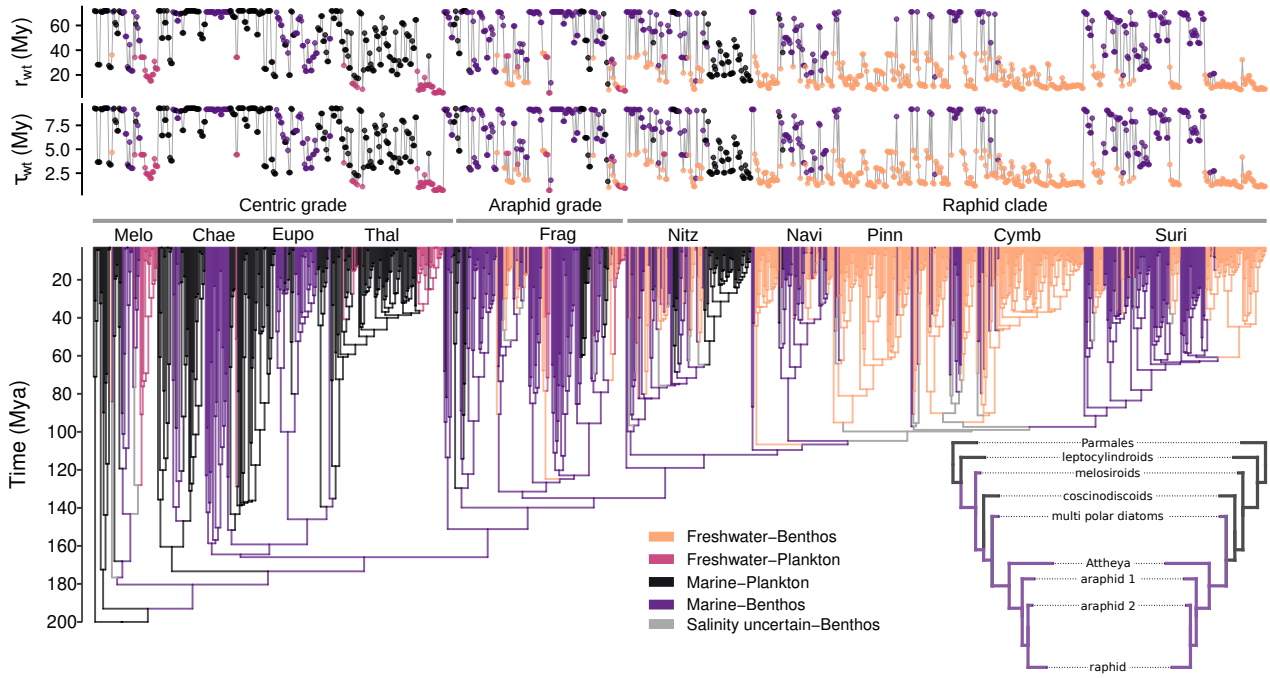

Figure 3. Ancestral state reconstructions of environments, and the tip estimates of net turnover (τ, speciation + extinction) and net diversification (*r*, speciation - extinction) rates inverted to waiting times (wt) in millions of years. The inset shows ancestral reconstructions along the backbone for two alternative conditioning for survival of crown lineages, following the original HiSSE implementation (left; Beaulieu & O'Meara, 2016; FitzJohn et al., 2009) and Herrera-Alsina et al. (2018) (right, see also Fig. S3). Both reconstructions support marine–planktonic ancestry of all diatoms and early colonization of the benthos. The high-turnover and high-diversification freshwater lineages (lower waiting times) are predominantly benthic within motile raphid pennate clade (e.g., Suri and Cymb) and predominantly planktonic within the non-motile centric and araphid grades (e.g., Frag, Thal, and Melo). Abbreviations: Mya, millions of years ago; Melo, Melosirales; Chae, Chaetocerotales; Eupo, Eupodiscales; Thal, Thalassiosirales; Frag, fragilaroids; Nitz, Bacillariales; Navi, naviculoids; Pinn, pinnulariids; Cymb, Cymbellales; Suri, Surirellales; Parm, Parmales; Lept, Leptocylindrales + Corethrales; Cosc, Coscinodiscales + Rhizosoleniales.

waiting times for marine-to-freshwater colonizations were indeed dependent on hidden factors. In some lineages, freshwater colonizations took about five times longer than the reverse, whereas in others the waiting time for marine-to-freshwater transitions was about half the waiting time of freshwater-to-marine transitions (Table 2). Under the MuHiSSE model, which further allowed for the possibility that marine-to-freshwater transitions differ between the plankton and benthos, there was also no evidence that marine-to-freshwater colonization were more frequent than the reverse freshwater-to-marine transitions (Fig. 3). It therefore appears that traditional notions about the nature of the salinity barrier in diatoms, specifically the irreversibility of freshwater colonizations, are largely unsupported by the data and apply only to a subset of diatom lineages in which recolonization of marine habitats is rarer but not impossible (Fig. 3).

Transition rates in the different hidden states also suggested that certain types of colonizations have drastically different probabilities. This is well illustrated by the contrasting rates of outflow from the marine and freshwater plankton between the first and second hidden states. For some lineages, the plankton represented an absorbing habitat, where they might have diversified, but transitions away from this habitat were extremely rare or impossible (Fig. 3). Yet, in other lineages, outflow from the freshwater plankton was among the most common type of transition (waiting times of 97 and 32 My; Fig. 3). These transition rates, combined with

the broad variation and bimodal distribution of turnover in freshwater plankton (Fig. 1C), imply that for some lineages the freshwater plankton is an absorbing state, where following colonizations, lineages diversify but do not emigrate, but for others, the freshwater plankton is a transient state, with a retention time much shorter than any other state. Moreover, when we consider the somewhat short-lived nature of freshwater planktonic habitats over deep geologic timescales, these findings suggest an explanation regarding the very low fraction of diatom diversity found in the freshwater plankton and the relatively narrow phylogenetic breadth of these communities. The few lineages that have diversified substantially in the freshwater plankton are almost entirely restricted to it and have been largely sourced from the marine plankton (e.g., the genus *Aulacoseira* within marine planktonic Melosirales or cyclostephanoid diatoms within the ancestrally marine planktonic Thalassiosirales). In turn, for the vast majority of the remaining lineages, colonizations of freshwater plankton were sporadic and did not lead to substantial species diversification (Fig. 3). This is in contrast to the much more phylogenetically diverse marine plankton, which has been colonized repeatedly by many marine benthic lineages (Fig. 3; see also Kooistra et al. 2007), harbors a considerable diversity of araphid (e.g., *Asterionellopsis* and *Thalassiosinema*) and raphid diatoms (e.g., *Fragilariopsis* and *Pseudo-nitzschia*; Fig. 3), and as suggested by the rates of outflow from the freshwater plankton, is also recolonized by freshwater planktonic lineages (Fig. 3, Table S1). These observations, combined with the non-uniform recruitment to the freshwater plankton—several old and species-rich marine planktonic lineages have not crossed into freshwaters or remain disproportionately more diverse in saline waters (e.g., Rhizosoleniales, Coscinodiscales, and Chaetocerales; Fig. 3; Kooistra et al. 2007)—also suggest that marine planktonic diatom assemblages are phylogenetically and functionally more diverse compared freshwater planktonic communities.

Summary

Water covers more than 70% of our planet, and of this, salt water habitats account for more than 99%. Yet many lineages, even ones that originated in the ocean, are far more diverse in fresh waters (Appeltans et al., 2012; Bloom et al., 2013; Grosberg et al., 2012; Guinot & Cavin, 2015; Miller & Wiens, 2017; Mora et al., 2011; Wiens, 2015). How these patterns were shaped is a topic of active research in various animal groups. In diatoms—a species-rich lineage of algae important for global ecology and biogeochemistry— we found that the marine–freshwater diversity imbalance is maintained by faster turnover and faster net diversification in freshwaters. This pattern holds for both planktonic and benthic lineages, illustrating the importance of considering interacting factors. However, variation in rates of diversification within regimes exemplifies the importance of accounting for unmeasured variables with potentially large cumulative effects on diversification.

We also report that that unlike many other aquatic lineages (Darling et al., 2009; Hart et al., 2003; Rigby, 1997; Rigby & Milsom, 1996; Signor & Vermeij, 1994), diatoms were ancestrally planktonic, but colonized the benthos early on in their evolution, such that most of their diversity was subsequently seeded by repeated emigration from the marine benthos (see also Kooistra et al., 2007). Marine-to-freshwater transitions are reversible, challenging longstanding hypotheses about the nature of the salinity barrier in diatoms. The marine plankton was sourced from a diverse set of marine benthic, and to a lesser extent, freshwater planktonic lineages. However, freshwater planktonic habitats were colonized by a much narrower pool of mainly marine planktonic lineages, suggesting that the marine planktonic assemblages are phylogenetically and functionally more diverse than their freshwater counterparts.

More generally, our results show that diversity patterns in diatoms, and presumably other lineages,

are unlikely attributable to a few, commonly broadly defined regimes. Support for character-dependent diversification can, in some cases, be overshadowed by broad variation in diversification rates arising from unconsidered factors whose individual or aggregate effects might equal or surpass the influence of factors traditionally considered most important. We illustrate how considering multiple possibly interacting characters might provide a more refined understanding of diversification in old, widespread, and species-rich clades.

Author contributions

TN, JMB, AJA designed the study. TN collected the data. JMB developed novel analytical tools. TN and JMB performed analyses. TN and AJA drafted the paper. All authors edited the final draft and approved the submission.

Acknowledgments

We thank Elizabeth Ruck for discussions about diatom phylogeny and distributions, and Daniel Caetano and James Boyko for helpful conversations about modeling diversification and trait evolution. This work was supported by grants from the NSF (DEB-1353131 to AJA) and from the Simons Foundation (403249 to AJA).

References

Alverson AJ, 2014. Timing marine–freshwater transitions in the diatom order Thalassiosirales. *Paleobiology*, 40:91–101.

Alverson AJ, Jansen RK, & Theriot EC, 2007. Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.*, 45:193–210.

Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, Bamber R, Barber A, Bartsch I, Berta A, Błażewicz-Paszkowycz M, Bock P, Boxshall G, Boyko CB, Brandão SN, Bray RA, Bruce NL, Cairns SD, Chan TY, Cheng L, Collins AG, Cribb T, Curini-Galletti M, Dahdouh-Guebas F, Davie PJF, Dawson MN, De Clerck O, Decock W, De Grave S, de Voogd NJ, Domning DP, Emig CC, Erséus C, Eschmeyer W, Fauchald K, Fautin DG, Feist SW, Fransen CHJM, Furuya H, Garcia-Alvarez O, Gerken S, Gibson D, Gittenberger A, Gofas S, Gómez-Daglio L, Gordon DP, Guiry MD, Hernandez F, Hoeksema BW, Hopcroft RR, Jaume D, Kirk P, Koedam N, Koenemann S, Kolb JB, Kristensen RM, Kroh A, Lambert G, Lazarus DB, Lemaitre R, Longshaw M, Lowry J, Macpherson E, Madin LP, Mah C, Mapstone G, McLaughlin PA, Mees J, Meland K, Messing CG, Mills CE, Molodtsova TN, Mooi R, Neuhaus B, Ng PKL, Nielsen C, Norenburg J, Opresko DM, Osawa M, Paulay G, Perrin W, Pilger JF, Poore GCB, Pugh P, Read GB, Reimer JD, Rius M, Rocha RM, Saiz-Salinas JI, Scarabino V, Schierwater B, Schmidt-Rhaesa A, Schnabel KE, Schotte M, Schuchert P, Schwabe E, Segers H, Self-Sullivan C, Shenkar N, Siegel V, Sterrer W, Stöhr S, Swalla B, Tasker ML, Thuesen EV, Timm T, Todaro MA, Turon X, Tyler S, Uetz P, van der Land J, Vanhoorne B, van Ofwegen LP, van Soest RWM, Vanaverbeke J, Walker-Smith G, Walter TC, Warren A, Williams GC, Wilson SP, & Costello MJ, 2012. The magnitude of global marine species diversity. *Curr. Biol.*, 22:2189–2202.

Armbrust EV, 2009. The life of diatoms in the world's oceans. *Nature*, 459:185–192.

Beaulieu JM & O'Meara BC, 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.*, 65:583–601.

Beaulieu JM & O'Meara BC, 2018. Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. *Am. J. Bot.*, 105:417–432.

Bloom DD, Weir JT, Piller KR, & Lovejoy NR, 2013. Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (Atherinopsidae). *Evolution*, 67:2040–2057.

Bowler C, Vardi A, & Allen AE, 2010. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.*, 2:333–365.

Caetano DS, O'Meara BC, & Beaulieu JM, 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution*, 72:2308–2324.

Casteleyn G, Leliaert F, Backeljau T, Debeer AE, Kotaki Y, Rhodes L, Lundholm N, Sabbe K, & Vyverman W, 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc. Natl. Acad. Sci. U. S. A.*, 107:12 952–12 957.

Cavalier-Smith T, 2009. Megaphylogeny, cell body plans, adaptive zones: Causes and timing of eukaryote basal radiations. *J. Eukaryot. Microbiol.*, 56:26–33.

Cohen AS, Stone JR, Beuning KRM, Park LE, Reinthal PN, Dettman D, Scholz CA, Johnson TC, King JW, Talbot MR, Brown ET, & Ivory SJ, 2007. Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proc. Natl. Acad. Sci. U. S. A.*, 104:16 422–16 427.

Crampton JS, Cody RD, Levy R, Harwood D, McKay R, & Naish TR, 2016. Southern Ocean phytoplankton turnover in response to stepwise Antarctic cooling over the past 15 million years. *Proceedings of the National Academy of Sciences*, 113:6868–6873.

Darling KF, Thomas E, Kasemann SA, Seears HA, Smart CW, & Wade CM, 2009. Surviving mass extinction by bridging the benthic/planktic divide. *Proc. Natl. Acad. Sci. U. S. A.*, 106:12 629–12 633.

Edlund MB & Stoermer EF, 2000. A 200,000-year, high-resolution record of diatom productivity and community makeup from Lake Baikal shows high correspondence to the marine oxygen-isotope record of climate change. *Limnol. Oceanogr.*, 45:948–962.

Evans KM, Chepurnov VA, Sluiman HJ, Thomas SJ, Spears BM, & Mann DG, 2009. Highly differentiated populations of the freshwater diatom Sellaphora capitata suggest limited dispersal and opportunities for allopatric speciation. *Protist*, 160:386–396.

FitzJohn RG, 2012. Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.*, 3:1084–1092.

FitzJohn RG, Maddison WP, & Otto SP, 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.*, 58:595–611.

Godhe A & Härnström K, 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. *Mol. Ecol.*, 19:4478–4490.

Goldberg EE, Lancaster LT, & Ree RH, 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.*, 60:451–465.

Grosberg RK, Vermeij GJ, & Wainwright PC, 2012. Biodiversity in water and on land. *Curr. Biol.*, 22:R900–3.

Guinot G & Cavin L, 2015. Contrasting "Fish" diversity dynamics between marine and freshwater environments. *Curr. Biol.*, 25:2314–2318.

Hampton SE, McGowan S, Ozersky T, Virdis SGP, Vu TT, Spanbauer TL, Kraemer BM, Swann G, Mackay AW, Powers SM, Meyer MF, Labou SG, O'Reilly CM, DiCarlo M, Galloway AWE, & Fritz SC, 2018. Recent ecological change in ancient lakes. *Limnol. Oceanogr.*, 37:17.

Hart MB, Hylton MD, Oxford MJ, Price GD, Hudson W, & Smart CW, 2003. The search for the origin of the planktic Foraminifera. *J. Geol. Soc. London*, 160:341–343.

Hasle GR & Syvertsen EE, 1996. Marine diatoms. In: CR Tomas, ed., *Identifying marine phytoplankton*, pp. 5–385. Academic Press, San Diego.

Herrera-Alsina L, van Els P, & Etienne RS, 2018. Detecting the dependence of diversification on multiple traits from phylogenetic trees and trait data. *Systematic Biology*, p. syy057.

Kirst GO, 1990. Salinity tolerance of eukaryotic marine algae. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 41:21–53.

Kirst GO, 1996. Osmotic adjustment in phytoplankton and macroalgae. In: RP Kiene, PT Visscher, MD Keller, & GO Kirst, eds., *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds*, pp. 121–129. Springer US, Boston, MA.

Kociolek JP, Balasubramanian K, Blanco S, Coste M, Ector L, Liu Y, Kulikovskiy M, Lundholm N, Ludwig T, Potapova M, Rimet F, Sabbe K, Sala S, Sar E, Taylor J, Van de Vijver B, Wetzel CE, Williams DM, Witkowski A, & Witkowski J, 2017. DiatomBase. http://www.diatombase.org. Accessed: 2017-11-30.

Kooistra WHCF, Forlani G, & De Stefano M, 2009. Adaptations of araphid pennate diatoms to a planktonic existence. *Mar. Ecol.*, 30:1–15.

Kooistra WHCF, Gersonde R, Medlin LK, & Mann DG, 2007. CHAPTER 11 - The origin and evolution of the Diatoms: Their adaptation to a planktonic existence. In: PG Falkowski & AH Knoll, eds., *Evolution of primary producers in the sea*, pp. 207–249. Academic Press, Burlington.

Lange-Bertalot H & Krammer K, 2000. *Diatoms of Europe: Diatoms of the European inland waters and comparable habitats*. ARG Ganter.

Larsen BB, Miller EC, Rhodes MK, & Wiens JJ, 2017. Inordinate fondness multiplied and redistributed: The number of species on Earth and the new pie of life. *Q. Rev. Biol.*, 92:229–265.

Lee CE & Bell MA, 1999. Causes and consequences of recent freshwater invasions by saltwater animals. *Trends Ecol. Evol.*, 14:284–288.

Lewitus E, Bittner L, Malviya S, Bowler C, & Morlon H, 2018. Clade-specific diversification dynamics of marine diatoms since the Jurassic. *Nature Ecology & Evolution*, 2:1715.

Lobban CS, Schefter M, Jordan RW, Arai Y, Sasaki A, Theriot EC, Ashworth M, Ruck EC, & Pennesi C, 2012. Coral-reef diatoms (Bacillariophyta) from Guam: New records and preliminary checklist, with emphasis on epiphytic species from farmer-fish territories. *Micronesica*, 43:237–479.

Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, & Rengefors K, 2009. Infrequent marine–freshwater transitions in the microbial world. *Trends Microbiol.*, 17:414–422.

Logares R, Shalchian-Tabrizi K, Boltovskoy A, & Rengefors K, 2007. Extensive dinoflagellate phylogenies indicate infrequent marine-freshwater transitions. *Mol. Phylogenet. Evol.*, 45:887–903.

Lozupone CA & Knight R, 2007. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U. S. A.*, 104:11 436–11 440.

Mackay AW, Edlund MB, Khursevich G, & Others, 2010. Diatoms in ancient lakes. *The Diatoms: Applications for the environmental and earth sciences*, 2:209–228.

Maddison WP, 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution*, 60:1743–1746.

Maddison WP, Midford PE, & Otto SP, 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.*, 56:701–710.

Mann DG, 1999. Crossing the Rubicon: The effectiveness of the marine/freshwater interface as a barrier to the migration of diatom germplasm. In: *Proceedings of the 14th International Diatom Symposium, Koenigstein, 1999*, pp. 1–21.

Mann DG & Vanormelingen P, 2013. An inordinate fondness? The number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.*, 60:414–420.

Miller EC & Wiens JJ, 2017. Extinction and time help drive the marine-terrestrial biodiversity gradient: Is the ocean a deathtrap? *Ecol. Lett.*, 20:911–921.

Mora C, Tittensor DP, Adl S, Simpson AGB, & Worm B, 2011. How many species are there on Earth and in the ocean? *PLoS Biol.*, 9:e1001 127.

Nakov T, Ashworth M, & Theriot EC, 2015. Comparative analysis of the interaction between habitat and growth form in diatoms. *ISME J.*, 9:246–255.

Nakov T, Beaulieu JM, & Alverson AJ, 2018. Accelerated diversification is related to life history and locomotion in a hyperdiverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *New Phytol.*, 219:462–473.

Palumbi SR, 1992. Marine speciation on a small planet. *Trends Ecol. Evol.*, 7:114–118.

Palumbi SR, 1994. Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Syst.*, 25:547–572.

Powell MG & Glazier DS, 2017. Asymmetric geographic range expansion explains the latitudinal diversity gradients of four major taxa of marine plankton. *Paleobiology*, 43:196–208.

Rabosky DL & Goldberg EE, 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.*, 64:340–355.

Rigby S, 1997. A comparison of the colonization of the planktic realm and the land. *Lethaia*, 30:11–17.

Rigby S & Milsom C, 1996. Benthic origins of zooplankton: An environmentally determined macroevolutionary effect. *Geology*, 24:52–54.

Round FE, Crawford RM, & Mann DG, 1990. *Diatoms: Biology and morphology of the genera*. Cambridge University Press.

Round FE & Sims PA, 1980. The distribution of diatom genera in marine and freshwater environments and some evolutionary considerations. In: *Proceedings of the sixth symposium on recent and fossil diatoms*, pp. 301–320.

Ruck EC, Nakov T, Alverson AJ, & Theriot EC, 2016. Phylogeny, ecology, morphological evolution, and reclassification of the diatom orders Surirellales and Rhopalodiales. *Mol. Phylogenet. Evol.*, 103:155–171.

Rynearson TA & Armbrust EV, 2004. Genetic differentiation among populations of the planktonic marine diatom Ditylum brightwellii (Bacillariophyceae). *J. Phycol.*, 40:34–43.

Signor PW & Vermeij GJ, 1994. The plankton and the benthos: Origins and early history of an evolving relationship. *Paleobiology*, 20:297–319.

Sims PA, Mann DG, & Medlin LK, 2006. Evolution of the diatoms: Insights from fossil, biological and molecular data. *Phycologia*, 45:361–402.

Sjöqvist C, Godhe A, Jonsson PR, Sundqvist L, & Kremp A, 2015. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea–Baltic Sea salinity gradient. *Mol. Ecol.*, 24:2871–2885.

Stone JR, Westover KS, & Cohen AS, 2011. Late Pleistocene paleohydrography and diatom paleoecology of the central basin of Lake Malawi, Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 303:51–70.

Vanormelingen P, Chepurnov VA, Mann DG, Sabbe K, & Vyverman W, 2008. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of Eunotia bilunaris sensu lato (Bacillariophyta). *Protist*, 159:73–90.

Vermeij GJ, Grosberg RK, Marshall CR, & Motani R, 2018. The sea as deathtrap: Comment on a paper by Miller and Wiens. *Ecol. Lett.*, 21:938–939.

Whittaker KA & Rynearson TA, 2017. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proc. Natl. Acad. Sci. U. S. A.*, 114:2651–2656.

Wiens JJ, 2015. Faster diversification on land than sea helps explain global biodiversity patterns among habitats and animal phyla. *Ecol. Lett.*, 18:1234–1241.

Witkowski A, Lange-Bertalot H, Metzeltin D, Lange-Bertalot H, & Gantner Verlag K, 2000. Diatom flora of marine coasts I. Iconographia Diatomologica. Annotated Diatom Micrographs. Diversity-Taxonomy-Identiflcation. Vol. 7.

Supporting Information

Supplementary Figure 1

A. Binary hidden state speciation and extinction

Null

	M	F
M	–	1
F	1	–
τ	2	2
ε	3	3

BiSSE

	M	F
M	–	2
F	1	–
τ	3	4
ε	5	5

HiSSE

	M1	F1	M2	F2
M1	–	3	5	–
F1	1	–	–	7
M2	2	–	–	8
F2	–	4	6	–
τ	9	10	11	12
ε	13	13	13	13

CID2

	M	F	H1	H2
M	–	2		
F	1	–		
H1			–	3
H2			3	–
τ			4	5
ε			6	6

CID4

	M	F	H1	H2	H3	H4
M	–	2				
F	1	–				
H1			–	3	3	3
H2			3	–	3	3
H3			3	3	–	3
H4			3	3	3	–
τ			4	5	6	7
ε			8	8	8	8

B. Multistate hidden state speciation and extinction

Null

	MP	MB	FP	FB
MP	–	3	5	–
MB	1	–	–	7
FP	2	–	–	8
FB	–	4	6	–
τ	9	9	9	9
ε	10	10	10	10

MuSSE

	MP	MB	FP	FB
MP	–	3	5	–
MB	1	–	–	7
FP	2	–	–	8
FB	–	4	6	–
τ	9	10	11	12
ε	13	14	15	16

MuHiSSE

	MP1	MB1	FP1	FB1	MP2	MB2	FP2	FB2
MP1	–	3	5	–	18	–	–	–
MB1	1	–	–	7	–	18	–	–
FP1	2	–	–	8	–	–	18	–
FB1	–	4	6	–	–	–	–	18
MP2	17	–	–	–	–	11	13	–
MB2	–	17	–	–	9	–	–	15
FP2	–	–	17	–	10	–	–	16
FB2	–	–	–	17	–	12	14	–
τ	19	20	21	22	23	24	25	26
ε	27	27	27	27	27	27	27	27

MuCID2

	MP	MB	FP	FB	H1	H2
MP	–	3	5	–		
MB	1	–	–	7		
FP	2	–	–	8		
FB	–	4	6	–		
H1					–	9
H2					10	–
τ					11	12
ε					13	13

MuCID8

	MP	MB	FP	FB	H1	H2	H3	H4	H5	H6	H7	H8
MP	–	3	5	–								
MB	1	–	–	7								
FP	2	–	–	8								
FB	–	4	6	–								
H1 –H8					9	10	11	12	13	14	15	16
τ					17	18	19	20	21	22	23	24
ε					25	25	25	25	25	25	25	25

Figure S1. **Character-dependent and character-independent diversification models for binary (A) and multistate (B) coding of salinity and habitat. The top part (grey) depicts the transition rate matrix with row and column names representing observed character states (M=marine, F=freshwater and their combinations) or hidden (H) states. The numbers 1-8 (e.g., M1, M2, F1, F2) differentiate hidden states within a character state. The bottom part (green), labeled with τ and ϵ, shows the parameters for net turnover (speciation+extinction) and extinction fraction (extinction/speciation), respectively. Each τ and ϵ parameter is linked to the state denoted in the column name. Note that diversification parameters are associated to character states in SSE models, but not in CID models, where they are linked to hidden states. Numbers inside the table represent the indexes of model parameters. For constrained parameters, the same index occurs in multiple cells (e.g., ϵ was fixed in all models, so it has one parameter in each table). Models for habitat (plankton–benthos) were the same as the binary models for salinity (A).**

Supplementary Figure 2

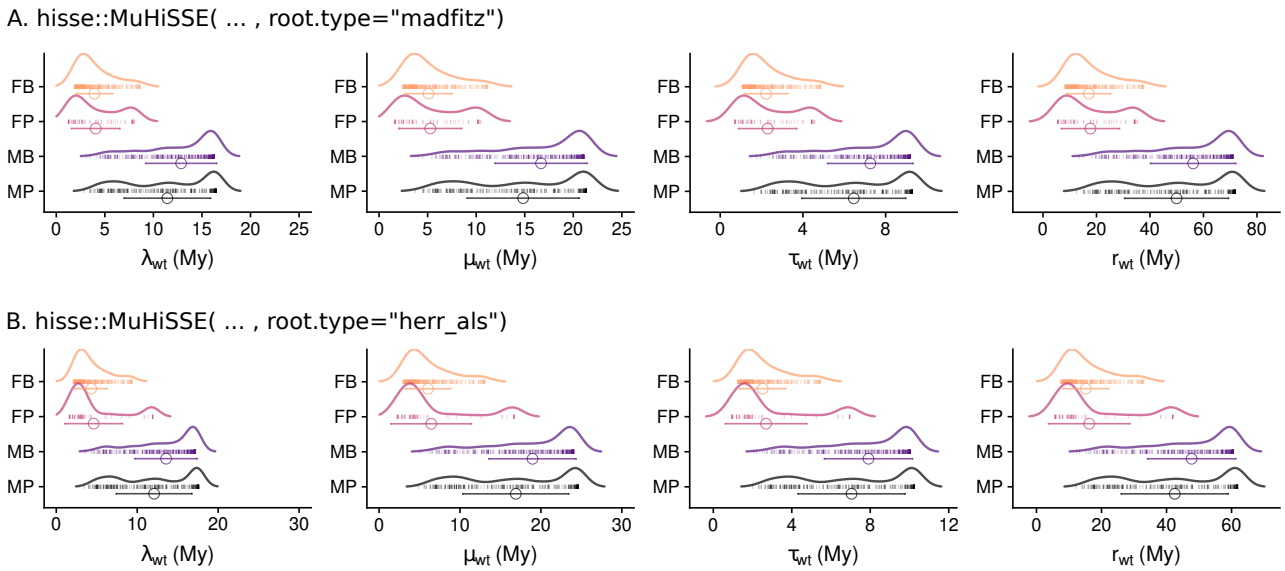

Figure S2. Distributions of tip-estimates of speciation (λ), extinction (μ), net turnover (τ, speciation + extinction) and net diversification (*r*, speciation - extinction) averaged across hidden states under the MuHiSSE model and two alternative implementations of the conditioning for survival of crown lineages (Beaulieu & O'Meara, 2016; FitzJohn et al., 2009; Herrera-Alsina et al., 2018). The circles and horizontal error bars show the mean $\pm$ standard deviation across tips of the phylogeny and the carpet points show the individual rates. The rates are inverted into waiting times (wt) in millions of years. Abbreviations: M, marine; F, freshwater; P, plankton; B, benthos.

Supplementary Figure 3

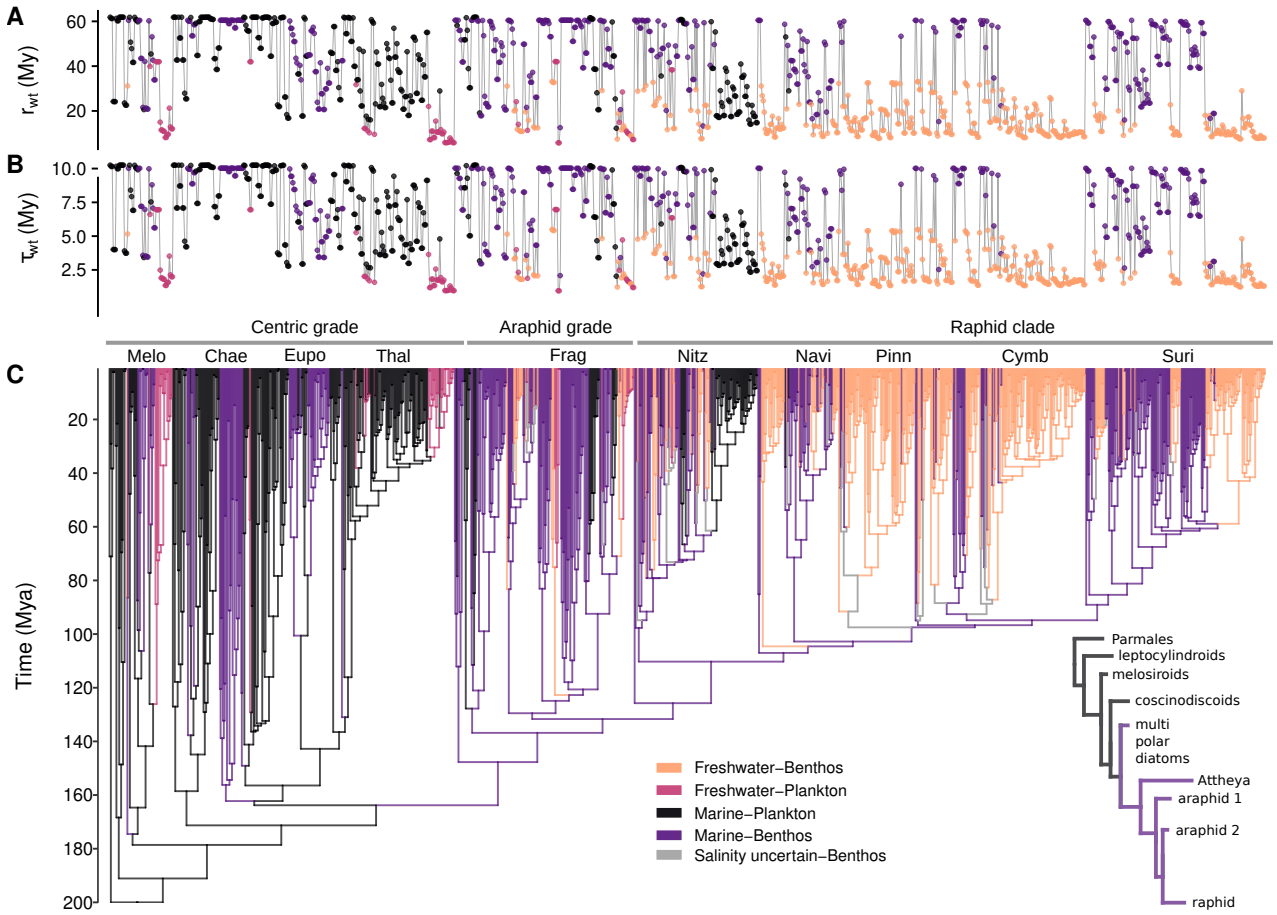

Figure S3. Ancestral state reconstructions of environments and the tip estimates of net turnover (τ, speciation + extinction) and net diversification (r, speciation - extinction) following Herrera-Alsina et al. (2018) implementation of the conditioning for survival of crown lineages. Rates were inverted to waiting times (wt) in millions of years. The inset shows ancestral reconstructions along the backbone supporting marine–planktonic ancestry of all diatoms and early colonization of the benthos. The high-turnover and high-diversification freshwater lineages (lower waiting times) are predominantly benthic among motile raphid pennate diatoms (e.g., Suri and Cymb) and predominantly planktonic within the non-motile centric and araphid grades (e.g., Frag, Thal, and Melo). Abbreviations: Mya, millions of years ago; Melo, Melosirales; Chae, Chaetocerotales; Eupo, Eupodiscales; Thal, Thalassiosirales; Frag, fragilaroids; Nitz, Bacillariales; Navi, naviculoids; Pinn, pinnulariids; Cymb, Cymbellales; Suri, Surirellales; Parm, Parmales; Lept, Leptocylindrales + Corethrales; Cosc, Coscinodiscales + Rhizosoleniales.

Supplementary table 1

Table S1. Parameter estimates and 2 log likelihood support region (in brackets) for the best model (hidden state miltistate speciation and extinction, MuHiSSE, with settings `hisse::MuHiSSE(..., root.type="madfitz")`) for the combined salinity + habitat character states [marine–plankton (MP), marine–benthos (MB), freshwater–plankton (FP) and freshwater–benthos (FB)]. Rates of transition, speciation, extinction, diversification and turnover were inverted to waiting times in millions of years. The numbers next to character states indicate the hidden states.

	MP1	MB1	FP1	FB1	MP2	MB2	FP2	FB2
Transition								
MP1	NA	1333 [1333-∞]	7988 [1698-9702]	0	357 [316-363]			0
MB1	1254 [981-1315]	NA	0	618 [572-646]		357 [316-363]	0	
FP1	∞ [1165-∞]	0	NA	∞ [978-∞]		0	357 [316-363]	
FB1	0	0.261 [244-268]	706 [664-767]	NA	0			357 [316-363]
MP2	15 [14-16]			0	NA	389 [369-404	232 [204-252]	0
MB2		15 [14-16]	0		328 [304-337]	NA	0	91 [87-96]
FP2		0	15 [14-16]		98 [91-104]	0	NA	32 [30-33]
FB2	0			15 [14-16]	0	433 [403-457]	∞ [1388-∞]	NA
Diversification								
Speciation	18.1 [16.4-19.4]	18.1 [16.8-19.2]	8.7 [8-9.5]	10.4 [9.6-11.1]	2.9 [2.7-3]	3.1 [3-3.3]	1.2 [1.2-1.3]	1.8 [1.7-1.8]
Extinction	23.5 [21.1-24.8]	23.5 [21.7-24.9]	11.3 [10.5-12.2]	13.5 [12.5-14.2]	3.7 [3.5-4]	4 [3.8-4.2]	1.6 [1.5-1.7]	2.3 [2.2-2.4]
Net turnover	10.2 [9.2-10.9]	10.2 [9.5-10.8]	4.9 [4.5-5.3]	5.9 [5.4-6.2]	1.6 [1.5-1.7]	1.8 [1.7-1.8]	0.7 [0.7-0.7]	1 [0.9-1]
Net diversification	79.2 [72.5-88]	79.1 [71.7-87.9]	37.9 [33.5-43.1]	45.4 [40.5-50.9]	12.6 [11.3-13.8]	13.6 [12.7-15.1]	5.4 [4.8-6.1]	7.8 [7.4-8.3]
Extinction fraction				0.7709 [0.7567-0.7853]				