

ORIGINAL ARTICLE

Comparative analysis of the interaction between habitat and growth form in diatoms

Teofil Nakov^{1,3}, Matt Ashworth^{2,3} and Edward C Theriot^{1,2}

¹Plant Biology Graduate Program, University of Texas at Austin, Austin, TX, USA and ²Section of Integrative Biology, University of Texas at Austin, Austin, TX, USA

We characterized the evolutionary history of growth form (solitary–colonial) and its interaction with species' habitat (planktonic–benthic) across a multi-gene phylogeny encompassing a broad sample of the order-level diversity of diatoms. We treated these characters broadly, modeling the evolution of aggregation of cells into a colony irrespective of the way aggregation is achieved, and relating the growth form to a broad concept of niche location: in the plankton or benthos. The results showed that habitat shifts are rare implying conservatism in niche location at the level of large clades. On the other hand, the evolutionary history of growth form is more dynamic with evolutionary rates that vary across the tree. Analyses of a possible interaction revealed that shifts in growth form are independent of habitat and that traversing between habitats does not hinge upon species' growth form. Our findings help to fill a gap in the understanding of diatom niche and growth form macroevolution and contribute toward a platform for the comparative study of the mechanisms underlying diatom species and functional diversity.

The ISME Journal (2015) 9, 246–255; doi:10.1038/ismej.2014.108; published online 1 July 2014

Introduction

Diatoms are an exceptionally diverse lineage of predominantly photoautotrophic heterokonts (Mann and Vanormelingen, 2013) responsible for substantial portions of the global primary production and atmospheric carbon removal (Nelson *et al.*, 1995; Hopkinson *et al.*, 2011). They have colonized the plankton and benthos, are frequently dominant in communities of lotic and lentic systems, and span the salinity barrier with substantial species diversity in each of these habitat types (Spaulding and Kociolek, 2000; Alverson *et al.*, 2007; Vyverman *et al.*, 2007). Their diversification across environments is a result of a combination of genetic, physiological and morphological factors (Falkowski *et al.*, 2004; Armbrust, 2009) shaped over an evolutionary history since the Mesozoic (Brown and Sorhannus, 2010).

One remarkable feature of diatoms is their extraordinary diversity in growth form. They range from simple spheroid unicells to complex three-dimensional colonies comprised of hundreds of cells and reaching macroscopic sizes. The mechanisms of colony construction are varied as well.

Diatoms form colonies through modified features of the silica cell wall, chitin threads and an array of extracellular mucilaginous secretions in the form of pads, stalks, tubes or sheets (Round *et al.*, 1990). This diversity in growth form has a functional role. The combination of growth form (for example, solitary or colonial) and habitat occupancy (for example, planktonic or benthic) approximates, albeit roughly, a diatom's ecological niche. Small-celled species that grow attached to a substrate via a mucilaginous pad, for instance, are early colonizers of benthic mats and are adapted for resistance to scouring from water currents (Hoagland *et al.*, 1982; Hoagland, 1983; McCormick and Stevenson, 1991; Johnson *et al.*, 1997). Long filamentous or branched colonies, on the other hand, tend to establish later in the succession, when the mat is crowded and cells improve access to nutrients and light by rising above the boundary layer (Hoagland *et al.*, 1982; Hoagland, 1983; McCormick and Stevenson, 1991). Growth form is similarly consequential in the plankton because species' sinking rate and vertical position in the water column are affected by colony morphology and symmetry (Padisák *et al.*, 2003; Reynolds, 2006). Thus, the amount of light and nutrients available to a cell living in a stratified environment is at least partially influenced by the ability to form colonies and their properties. Combining these considerations with the benefit of increased organism size as a strategy for defense against predation (Yokota and Sterner, 2010), colony formation emerges as a trait

Correspondence: T Nakov, Plant Biology Graduate Program, University of Texas at Austin, 1 University Station A6700 (Biolabs 214), Austin, TX 78712, USA.

E-mail: teofil@utexas.edu

³These authors contributed equally to this work.

Received 17 March 2014; revised 16 May 2014; accepted 29 May 2014; published online 1 July 2014

with wide-ranging consequences for life in the aquatic environment.

It seems plausible, therefore, to hypothesize that the combined influence of environmental factors and species interactions has guided lineages toward alternate growth forms in the strikingly different open water versus littoral zone habitats. Round *et al.*, for example, argued: ‘There is also no doubt that colonial organization has been subject to strong selection in particular habitats, in relation to attachment, light and nutrient capture ..., the control of sinking rate, etc.’ (1990, p. 29). The independent acquisition of colonial growth form, and indeed very similar colony morphology, in distantly related diatom lineages can be viewed as support for this assertion. However, despite the recognized importance of growth form and habitat occupancy, their evolutionary histories and interaction have seldom been investigated. Kooistra *et al.* (2007, 2009) identified lineages that transitioned to the plankton and discussed adaptations that may have accompanied such shifts. They also highlighted isogamous sexual reproduction as an obstacle for planktonic lifestyle in pennate diatoms (clade with bilaterally symmetrical cells) that has been successfully circumvented in few lineages (Kooistra *et al.*, 2009). Research in this area of diatom evolution, however, seems to have stalled and the evolutionary histories of habitat occupancy and growth form have not been evaluated in a modeling framework.

Recent efforts in reconstructing the diatom phylogeny are approaching nearly complete sampling of the major extant lineages of diatoms providing the opportunity to examine functional trait evolution in previously unattainable detail (Theriot *et al.*, 2009, 2010; Ashworth *et al.*, 2013). Moreover, advances in the methodology for modeling discrete traits that relax the assumption of rate constancy across a phylogeny allow evaluation of more realistic evolutionary scenarios (Beaulieu *et al.*, 2013). Here, we took advantage of these opportunities to study the evolutionary history of growth form (solitary–colonial) in relation to species’ habitat (planktonic–benthic) across a broadly sampled diatom phylogeny representative of 80% of order-level diversity.

Materials and methods

Character data and phylogenetic trees

The data set analyzed here consists of 281 diatom taxa capturing most major lineages of extant diatoms with representatives from ca. 80% of described orders. As outgroup we used the sister lineage to diatoms, *Bolidomonas* Guillou & Chrétiennot-Dinet. We coded each species for habitat (planktonic (0) or benthic (1)) and growth form (solitary (0) or colonial (1)) from personal observation of wild material from which the particular strain was isolated (growth form was not scored from cultures).

For cultures obtained from public collections, the characters were scored as reported in the primary literature (Supplementary Table S1). In cases where we were unable to unambiguously determine the character states, we repeated the analyses with alternative coding schemes (Supplementary Tables S2 and S3).

We reconstructed the phylogeny of these taxa using a concatenated alignment of three genes: the nuclear-encoded small ribosomal subunit rRNA (*nSSU*), the chloroplast-encoded Ribulose-1, 5-bisphosphate carboxylase/oxygenase (*rbcL*) and CP-43 chlorophyll *a*-binding protein (*psbC*). New taxa were added to complement previous data sets in areas of limited taxon sampling (Theriot *et al.*, 2010; Ashworth *et al.*, 2013). A total of 35 *nSSU*, 37 *rbcL* and 34 *psbC* sequences were newly generated (GenBank accessions: KJ577839–KJ577944). The new data were aligned as described in Theriot *et al.* (2010) and Ashworth *et al.* (2013).

The most likely tree topology was inferred from 1008 maximum likelihood (ML) optimizations each starting from a parsimony tree in RAXML v.7.4.2 (Stamatakis, 2006). Clade support values were assessed through 10^3 nonparametric bootstrap replicates using the rapid bootstrap algorithm (Stamatakis *et al.*, 2008). The phylogram with highest likelihood was converted to a relative-time chronogram with a root age of 100 time units using penalized likelihood as implemented in the R package ‘ape’ (Sanderson, 2002; Paradis *et al.*, 2004; R Development Core Team, 2013). To accommodate phylogenetic uncertainty, in addition to the ‘best tree’, downstream analyses were also performed with 100 trees sampled at random from the 1008 optimizations.

Individual traits

To model the evolution of habitat occupancy and growth form individually, we used constant- and variable-rate stochastic Markov models. These two classes of models differed based on the assumptions concerning the variation of transition probabilities between character states across the phylogeny (Beaulieu *et al.*, 2013). In the constant-rate models, transitions (‘forward’ = $0 \rightarrow 1$ and ‘backward’ = $1 \rightarrow 0$) are fixed across the entire phylogeny (Pagel, 1994). The variable-rate models, on the other hand, allow different portions of the phylogeny to have different forward and/or backward transition rates. This is achieved by creating separate rate classes for slow (S) and fast (F) transition probabilities accommodating the possibility that particular lineages can have accelerated or decelerated rates of evolution relative to other portions of the tree (Beaulieu *et al.*, 2013; R package ‘corHMM’). Any number of rate classes is possible. However, we restricted our analyses to models with two rate classes (S and F) due to the modestly sized data set and issues with parameter estimation from overly complex models. The terms

‘forward’ and ‘backward’ are used for convenience and do not imply transitions between ancestral and derived states.

We were interested in two types of models: those in which the forward and backward transitions are equally probable ($0 \rightarrow 1 = 1 \rightarrow 0$, ‘symmetric’) and those that relax this assumption ($0 \rightarrow 1 \neq 1 \rightarrow 0$, ‘asymmetric’). Thus, for the constant-rate class, we have two models referred to as symmetric (number of parameters, $k=1$) and asymmetric ($k=2$) (Figure 1a). The most complex variable-rate model considered had eight parameters corresponding to the transition rates between character states in different rate classes and the transitions between rate classes in alternate character states (Beaulieu *et al.*, 2013). For example, the gain of coloniality proceeds through two rate parameters: $0S \rightarrow 1S$ in the slow rate class and $0F \rightarrow 1F$ in the fast rate class (Figures 1b and c). Transitions between rate classes are modeled analogously with $0S \rightarrow 0F$ when the lineage is solitary and $1S \rightarrow 1F$ when the lineage is colonial (Figures 1b and c). We did not consider models where trait and rate class change simultaneously.

From this eight-parameter rate matrix, a number of simplified models can be constructed by removing or constraining parameters to equality. To maintain reasonable model complexity, we tested models in which the probabilities of change between rate classes were symmetrical (that is, $0S \rightarrow 0F = 0F \rightarrow 0S \neq 1S \rightarrow 1F = 1F \rightarrow 1S$; ‘rates-symmetric’ in Figures 1b and c) or equal across the entire phylogeny (that is, $0S \rightarrow 0F = 0F \rightarrow 0S = 1S \rightarrow 1F = 1F \rightarrow 1S$; ‘rates-equal’ in Figures 1b and c). The ‘rates-symmetric’ models assumed that slow \leftrightarrow fast transitions differed dependent on the state of the trait, while the ‘rates-equal’ models assumed that slow \leftrightarrow fast transitions are constant (Figures 1b and c). Focusing on testing the possibility of asymmetry in transition rates between character states within different rate classes, we compared models in which the forward and backward transitions were allowed to differ and models where these were constrained to equality (Figure 1c; compare ‘rates-equal, traits-symmetric’ to ‘rates-equal, traits-asymmetric’ models). We also considered variable-rate models that assessed the penalty of constraining the forward transitions to equality irrespective of the rate class while keeping

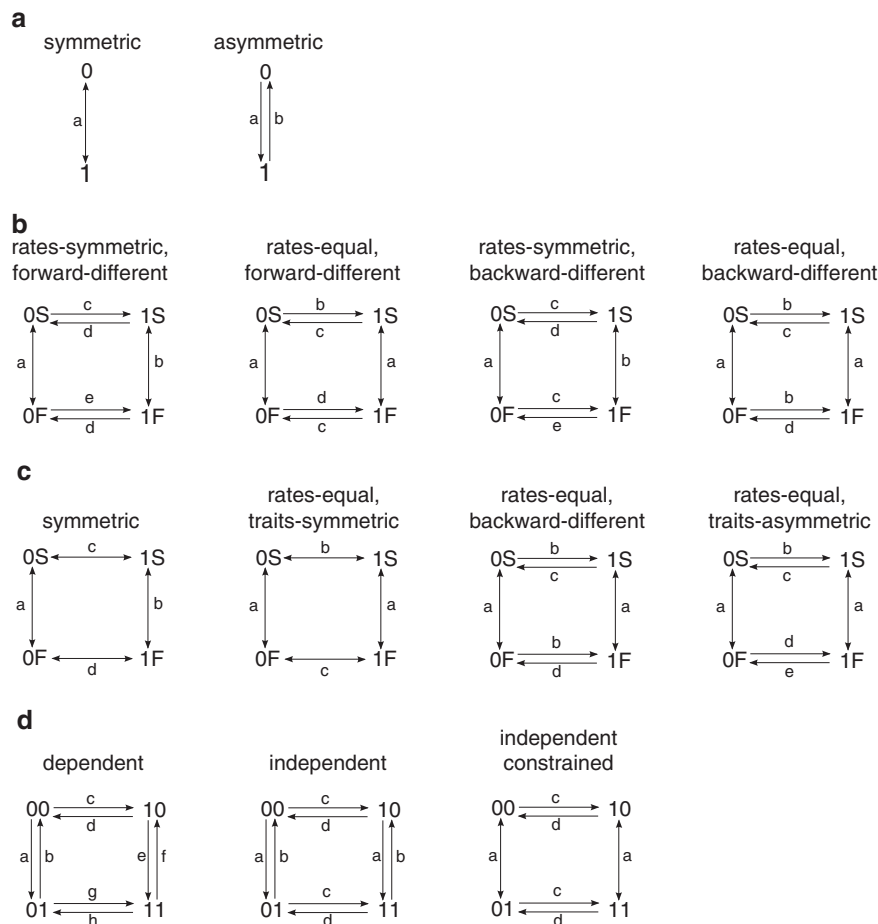


Figure 1 Diagrammatic representation of the considered models. (a) Constant-rate models. (b) Variable-rate models in the credible set for habitat occupancy. (c) Variable-rate models in the credible set for growth form. (d) Models for the combined habitat + growth form phenotypes.

the backward transitions different (that is, $0S \rightarrow 1S = 0F \rightarrow 1F$ and $1S \rightarrow 0S \neq 1F \rightarrow 0F$; ‘rates-symmetric (equal), backward-different’), and the reverse, constraining backward transitions to equality while keeping forward transitions different (that is, $0S \rightarrow 1S \neq 0F \rightarrow 1F$ and $1S \rightarrow 0S = 1F \rightarrow 0F$; ‘rates-symmetric (equal), forward-different’). These models were, in effect, testing the possibility that one type of transition in the trait (either the forward or backward) is constant across the phylogeny while the other varies. A total of 12 models constructed with the above reasoning were tested on the best phylogeny. A subset of four variable-rate models that fit the trait data best as well as the two constant-rate models were thereafter fitted to the sample of 100 ML trees. The set of six models are diagrammed in Figures 1a and b for habitat occupancy and Figure 1a and c for growth form.

Combined phenotypes

It is possible that growth form and habitat interact—colony formation might, for example, be favored in benthic species. The combination of two binary characters yields four combined phenotypes: planktonic + solitary (00), planktonic + colonial (01), benthic + solitary (10) and benthic + colonial (11). If growth form evolution depends on habitat, then the transition rate $00 \rightarrow 01$ is expected to be different than the transition rate $10 \rightarrow 11$ (Pagel, 1994; Pagel and Meade, 2006). The analogous situation for habitat is also of interest. Are transitions to the plankton dependent on growth form ($00 \rightarrow 10 \neq 01 \rightarrow 11$)? To test for the interaction, we fit two models: an independent model where the transitions $0 \rightarrow 1$ or $1 \rightarrow 0$ in one character were independent of the state of the other character ($k=4$; Figure 1d) and a dependent model where the probability of $0 \rightarrow 1$ or $1 \rightarrow 0$ in one character differed based on the state of the other character ($k=8$; Figure 1d). Preliminary analyses showed that a symmetric constant-rate model was favored for the evolution of habitat occupancy when viewed separately of growth form. Knowing this, we tested an independent model where planktonic \rightarrow benthic = benthic \rightarrow planktonic, but solitary \rightarrow colonial \neq colonial \rightarrow solitary (Figure 1d). All two-trait models were constant-rate because the current methodological framework is not extended to allow ‘hidden rates’ in two-trait models (Beaulieu *et al.*, 2013). As before, analyses were performed on the 100 phylogenies.

The analyses were performed in the R packages ‘corHMM’ and ‘ape’ (Paradis *et al.*, 2004; Beaulieu *et al.*, 2013) and character state transitions were calculated in Mesquite (Maddison and Maddison, 2011). Model selection was performed using the Akaike Information Criterion corrected for sample size (AICc). Importance of parameters was calculated as the sum of the average Akaike weights of all credible models that include a particular parameter (Burnham and Anderson, 2002).

Results

Trees and trait distribution

The phylogeny recovered all major groups identified in recent all-diatom trees (Theriot *et al.*, 2009, 2011) and is consistent with the current understanding of high-level relationships (Figure 2a, Supplementary Figure S1). For habitat occupancy, the ratio benthic:planktonic was about 2:1, and for growth form, solitary:colonial was about 1:1. In the grade of clades of non-pennate diatoms, phenotypes alternate fairly often and there appears to be no bias with respect to habitat or growth form (Figure 2a). Pennate diatoms were predominantly benthic (Figure 2a, Supplementary Figure S1 node 2) and among them the clade of actively motile raphid pennates was dominated by solitary forms (Figure 2a, Supplementary Figure S1 node 1).

Habitat

Five models, all within four $\Delta AICc$ units from the best and with a cumulative relative likelihood ω_i (Akaike weight) = 0.97, were within the credible set for the evolutionary history of habitat occupancy. The symmetric, constant-rate model performed best (Table 1; Figure 1a), followed closely by the asymmetric constant-rate model (mean $\Delta AICc = 0.82$; Table 1; Figure 1a). Given the cumulative weight of constant-rate models, variable-rate models cannot be ruled out as plausible for the evolution of habitat (Table 1). From the pool of tested variable-rate models, only those with symmetric or equal transitions between rate classes (slow \leftrightarrow fast) were included in the credible set (Table 1; Figures 1a and b). The penalty of reducing shifts between rate classes to one parameter was negligible (Table 1; ‘rates-equal’ in Figure 1b). These results are consistent with a scenario where transitions between the fast and slow rate class happen at similar rates across the entire phylogeny. Models where forward transitions (planktonic \rightarrow benthic) were allowed to vary, whereas backward (benthic \rightarrow planktonic) were kept equal, had a combined average $\omega_i = 0.38$ (Table 1; ‘forward-different’ in Figure 1b). The latter were better than models where backward, instead of forward transitions, were allowed to vary (Table 1; ‘backward-different’ in Figure 1b).

A parameter for a backward transition (benthos \rightarrow plankton) constant across the phylogeny, but different from the forward transition, was most important given the credible set of models (Table 2). There was little support for slow versus fast rate of backward transitions (Table 2). Parameters for slow versus fast forward transitions had weight, but their importance was not substantially higher than a constant forward rate parameter (Table 2).

Parsimony, maximum likelihood and stochastic character mapping on the best tree, the latter two conducted with the parameter estimates from the constant-rate symmetric model, agreed that a

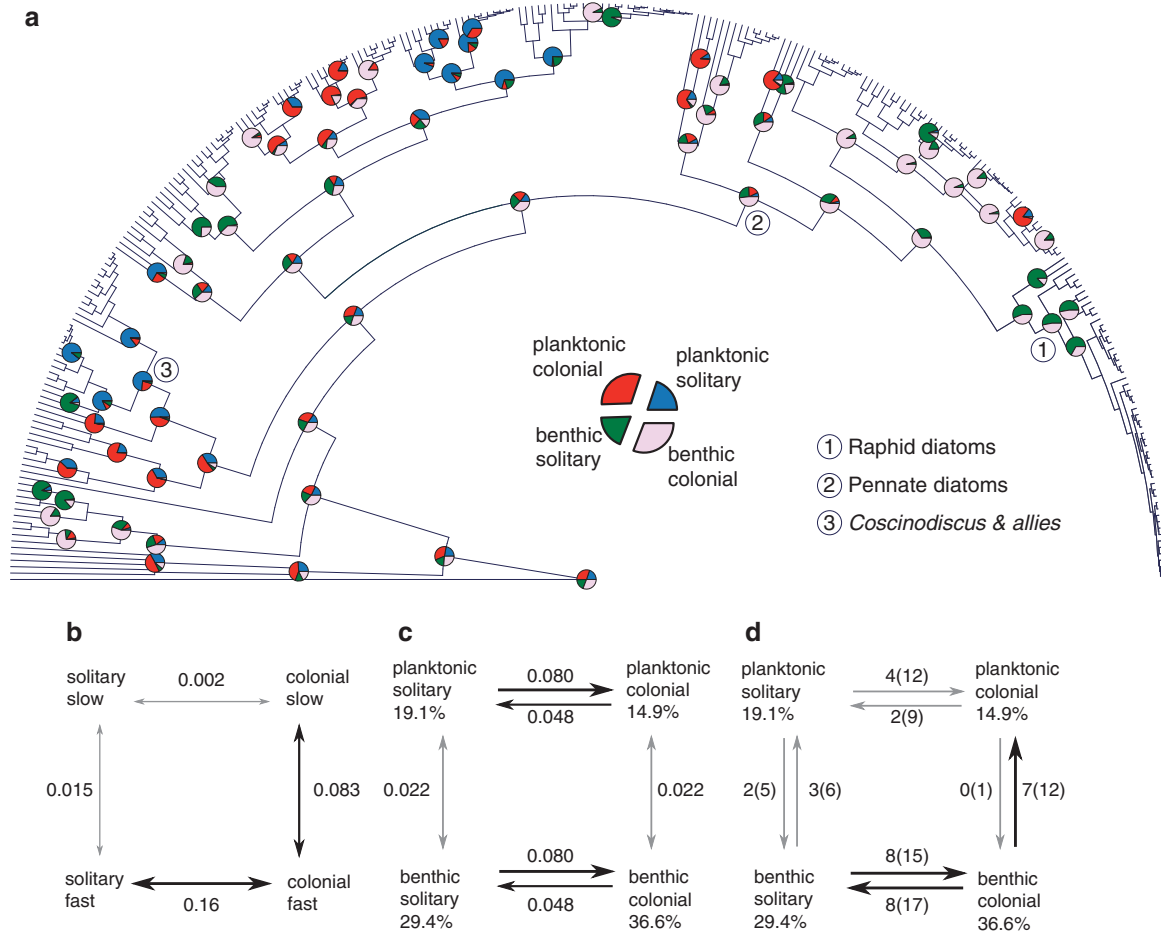


Figure 2 Evolution of the combined habitat + growth form phenotypes. (a) Ancestral states reconstruction of habitat and growth form based on 10^3 stochastic character histories under the favored independent constrained model. (b) Maximum likelihood estimates of the transition probabilities for the evolution of growth form (separate from habitat) under the symmetric, variable-rate model. (c) Maximum likelihood estimates of the transition probabilities for the combined four-state phenotypes under the independent constrained model of evolution. (d) Parsimony counts of state shifts for the combined four-state phenotypes.

Table 1 Constant-rate and variable-rate models for the evolution of habitat occupancy

Model	# Rate classes	Mean lnL	k	Mean AICc	Mean Δ AICc	Mean ω_i	Cumulative ω_i
Symmetric	1	-86.26	1	174.54	0.28	0.3	0.3
Asymmetric	1	-85.52	2	175.08	0.82	0.23	0.53
Rates-symmetric, forward-different	2	-82.54	5	175.29	1.03	0.23	0.76
Rates-equal, forward-different	2	-84.03	4	176.21	1.95	0.15	0.91
Rates-equal, backward-different	2	-84.78	4	177.71	3.45	0.06	0.97
Rates-symmetric, backward-different	2	-84.75	5	179.72	5.46	0.02	0.99

Abbreviation: AICc, Akaike Information Criterion corrected for sample size.

Models are ordered based on their average Akaike weights (ω_i) denoting the relative likelihood of each model. Refer to Figures 1a and b for model diagrams.

minimum of three plankton \rightarrow benthos transitions have happened along the diatom phylogeny. The maximum number of these transitions was estimated as high as 12 under parsimony and 24 under stochastic mapping. Benthos \rightarrow plankton transitions happened more frequently: a minimum of 9 (under ML) and 11 times (under parsimony and stochastic mapping) and a maximum of 20 and 33 times under parsimony and stochastic mapping, respectively.

Growth form

Four variable-rate models made up the credible set for the evolution of growth form (cumulative average $\omega_i = 0.98$; Table 3). A symmetric model with two rate classes performed best ($\omega_i = 0.47$; Table 3; 'symmetric' in Figure 1c). This is consistent with a scenario where the transition probabilities are either slow or fast, but in each rate class the probabilities of gain and loss of coloniality are equal. Restricting the

transitions between rate classes to one parameter did not incur a substantial cost in likelihood (average $\Delta\text{AICc} = 0.96$; Table 3; ‘rates-equal, traits-symmetric’ in Figure 1c). Relaxing the latter model to allow for asymmetric transition rates between growth forms or constraining forward transitions to equality irrespective of rate class did not offer a substantially better fit (Table 3; ‘rates-equal, backward-different’ and ‘rates-equal, traits-asymmetric’ in Figure 1c). In contrast to the results for habitat occupancy, constant-rate models performed poorly (average $\omega_i \leq 0.01$; Table 3).

Parameters for slow and fast, but symmetric, rates of solitary \leftrightarrow colonial transitions were of greatest importance for modeling the evolution of growth form (Table 2). Asymmetric transitions in either direction were less important (Table 2). Among these types of parameters, separate rate classes (slow versus fast) for loss of coloniality were more relevant (Table 2). Of the parameters implying asymmetry, but rate constancy across the tree, only the forward solitary \rightarrow colonial transition received some weight (Table 2). Based on the ΔAICc criterion, only the ‘symmetric’ and ‘rates-equal, traits-symmetric’ models fell within the credible set ($\Delta\text{AICc} \leq 4$; Table 3) trivializing any parameters not included in these models (Tables 2 and 3). Parameter estimates

Table 2 Relative importance of parameters (trait transition probabilities) based on the credible set of models for the evolution of habitat occupancy and growth form

Type of transition rate	Relative importance	
	Habitat occupancy	Growth form
Constant symmetric	0.30	0.00
Constant forward	0.29	0.10
Constant backward	0.61	0.00
Slow symmetric	NA	0.81
Fast symmetric	NA	0.81
Slow forward	0.38	0.07
Fast forward	0.38	0.07
Slow backward	0.06	0.17
Fast backward	0.06	0.17

Abbreviation: NA, not applicable.

The credible set was taken to include those models whose cumulative average Akaike weights reached or exceeded 0.95 (Tables 1 and 3). Refer to Figures 1a–c for model diagrams.

averaged over trees show that the solitary \leftrightarrow colonial transitions in the slow rate class ($= 0.0026$) are about 60 times slower than those in the fast rate class ($= 0.16$; Figures 2b and 3b). Shifts in rate class tend to proceed about 5.5 times faster on average in colony-forming lineages (Figures 2b and 3b).

Parsimony, maximum likelihood and stochastic character mapping on the best tree, the latter two conducted with the parameter estimates from the asymmetric constant-rate model, reconstructed a minimum of 12, 10 and 33, respectively, solitary \rightarrow colonial transitions across the diatom phylogeny. The maximum number of solitary \rightarrow colonial transitions was 20 under parsimony and 70 under the Bayesian stochastic mapping. Under parsimony, colonial \rightarrow solitary transitions were more frequent than the reverse with 21–29 total shifts. In contrast, under ML and stochastic mapping, the number of colonial \rightarrow solitary transitions was lower (five under ML and 24–58 under stochastic mapping).

Combined phenotypes

We tested for coordinated evolution between habitat and growth form by comparing a model that allowed transitions in one character to differ based on the state of the other character (for example, $00 \rightarrow 01 \neq 10 \rightarrow 11$) with models where such transitions were kept equal (for example, $00 \rightarrow 01 = 10 \rightarrow 11$). Models in which these two traits evolve independently were favored (combined average $\omega_i = 0.93$; Table 4; Figure 2b). Parameter estimates from the independent constrained model were similar to those estimated from the constant-rate models when traits were treated separately (Figure 2b).

Discussion

Variable pace of growth form evolution across the diatom phylogeny

Colonial growth form independently evolved in all major lineages of photoautotrophic eukaryotes (Niklas and Newman, 2013) and diatoms are no exception (Figure 2). Diatoms can aggregate into colonies through structures of the silica cell wall. In many cases, these are modifications of pre-existing features (for example, enlarged costae or heavily

Table 3 Constant-rate and variable-rate models for the evolution of growth form

Model	# Rate classes	Mean lnL	k	Mean AICc	Mean ΔAICc	Mean ω_i	Cumulative ω_i
Symmetric	2	−135.49	4	279.12	0.31	0.47	0.47
Rates-equal, traits-symmetric	2	−136.84	3	279.76	0.96	0.34	0.81
Rates-equal, backward-different	2	−138.29	4	284.72	5.92	0.10	0.91
Rates-equal, traits-asymmetric	2	−136.91	5	284.04	5.24	0.07	0.98
Asymmetric	1	−141.77	2	287.58	8.78	0.01	0.99
Symmetric	1	−143.99	1	289.99	11.19	<0.01	1.00

Abbreviation: AICc, Akaike Information Criterion corrected for sample size.

Models are ordered based on their average Akaike weights (ω_i) denoting the relative likelihood of each model. Refer to Figures 1a and c for model diagrams.

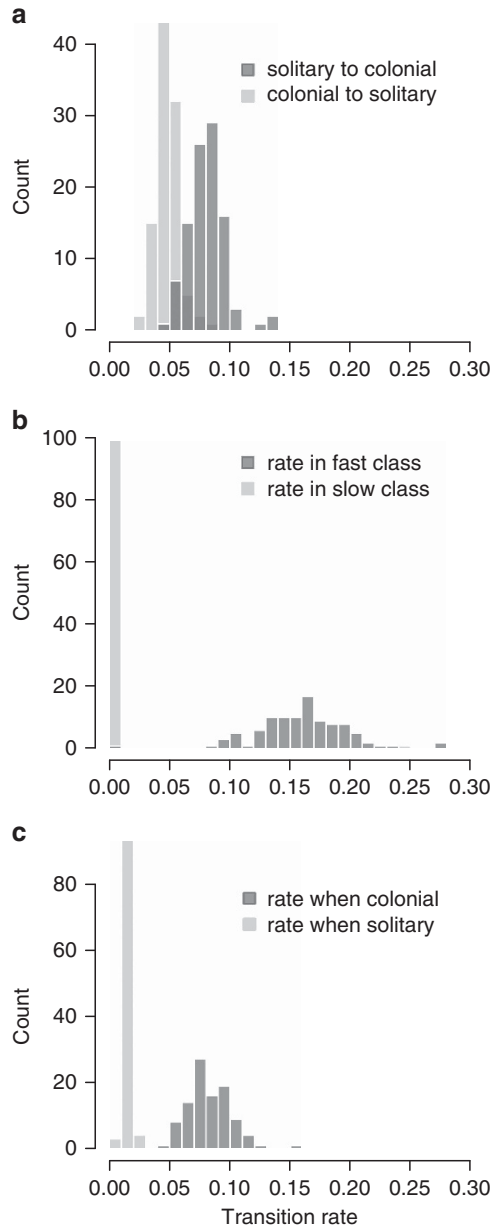


Figure 3 Transition probabilities for shifts between solitary and colonial growth form estimated from 100 phylogenies. (a) Under the asymmetric constant-rate model, the solitary→colonial transition is much faster than the reverse. (b) Transition probabilities for shifts between growth forms in the slow and fast rate class of the favored symmetric variable-rate model. (c) Transition probabilities for shifting between rate classes when lineages are solitary and colonial.

modified marginal strutted processes refashioned to serve in valve-to-valve interlocking) or rarely structures that seem specifically acquired for cell-to-cell attachment (for example, the periplekton of *Syndetocystis* Ralfs ex Greville). The other major mechanism is through extracellular mucilage production in form of pads, stalks, sheets or tubes. In many cases, solitary species already possess the ability to produce mucilage, commonly used for the attachment to benthic substrata or ‘cocooning’:

Table 4 Two-trait models for the evolution of the combined habitat + growth form phenotypes

Two-trait model	Mean lnL	k	Mean AICc	Mean $\Delta AICc$	Mean ω_i	Cumulative ω_i
Independent	-227.33	3	460.75	0.04	0.53	0.53
constrained						
Independent	-226.58	4	461.31	0.6	0.4	0.93
Dependent	-224.09	8	464.7	3.99	0.07	1

Abbreviation: AICc, Akaike Information Criterion corrected for sample size.

Models are ordered based on their average Akaike weights (ω_i) denoting the relative likelihood of each model. Refer to Figure 1d for model diagrams.

surrounding itself in a sheath of mucilage. Colony formation therefore might be a relatively simple process that requires the failure of cells to separate following mitosis and remain attached through elements of the cell wall or pre-existing mucilage formations. Perhaps due to this relative simplicity and the benefits of colonial lifestyle, colony formation has repeatedly evolved across the diatom tree. Estimates from different methods vary, but there seems to have been at least 10 and perhaps as much as a couple of dozen acquisitions of the colony-forming state (Figure 2). The solitary growth form predominates in the large, planktonic species of *Coscinodiscus* Ehrenberg (and allies) and the lineage of raphid pennate diatoms that have the ability of active movement (Figure 2a). Otherwise, gains of coloniality are dispersed across the phylogeny encompassing both planktonic and benthic species across an array of cell bauplānes (Figure 2a).

Under constant-rate models, there is support for asymmetry in the relative rates of transition between the solitary and colonial growth form: gains of coloniality are on average faster than losses (Figure 3a). This result can be interpreted as a tendency for the acquisition of the generally beneficial colonial state. When we considered the possibility of rate heterogeneity, however, there was no longer support for asymmetric transition probabilities (Table 3; Figures 2b and 3). Instead, the results argue for symmetric transitions between the states of growth form, but in separate rate classes. Thus, diatoms can be roughly divided into lineages in which the evolution of growth form can be considered stagnant and lineages in which this trait is labile, with frequent (60 times more probable) traversals between growth forms (Figures 2b and 3). Clades evolving in the slow or fast ‘regime’ are not restricted to a particular portion of the topology, but dispersed across the tree (Supplementary Figure S1). Apart from the coscinodiscoid lineage and the raphid pennates (Figure 2a, Supplementary Figure S1), which are estimated as exclusively slow- and fast-rate class, respectively, rate classes across the phylogeny alternate at the family or even genus level (Supplementary Figure S1). Asymmetric transition

probabilities within particular clades are certainly possible and a closer look at lineages where state shifts are most common might identify tendencies specific to particular groups.

Traversals between habitats are rare, but asymmetric in lineages with colonial growth form

For a photosynthetic unicell, the plankton and benthos are diametrically different environments. Differences in physico-chemical properties of the surrounding water, the availability of nutrients and light and the types of available microhabitats require different sets of morphological and physiological adaptations (Stevenson, 1997; Reynolds, 2006). Transitions between benthic and planktonic habitats therefore are accompanied by physiological and morphological adjustments streamlining the cells for existence in the respective habitat. A comparison of morphology of sister lineages differing in habitat occupancy revealed that benthos → plankton transitions tend to be accompanied by one or more of the following trait shifts: (i) increased morphological complexity of the cell, (ii) increase in cell size and (iii) transition to colonial growth form. These trait shifts might be related to adaptations for planktonic lifestyle (Kooistra *et al.*, 2007, 2009). Increased morphological complexity of the cell, achieved through the acquisition of various projections, spines and keels, might represent a mechanism-improved buoyancy. Departure from spherical cell shape increases form resistance—the difference in sinking velocity between a particle and a sphere with identical density and volume—and therefore decreases sinking velocity allowing cells to stay suspended longer (Padisák *et al.*, 2003). The morphology of a colony can have an effect on sinking velocity as well. Tubular, spiral or stellate arrangements that maintain colony symmetry exhibit reduced sinking relative to asymmetrical arrangements (Padisák *et al.*, 2003; Reynolds, 2006). Increase in cell size, a strategy that accompanied transitions to the plankton in some marine lineages (for example, the benthic *Odontella longicruris* (Greville) Hoban versus the larger planktonic *O. longicruris* var. *hyalina* (Schröder) Hoban), might be related to adaptation for higher capacity of nutrient storage given transition to the planktonic environment (Litchman *et al.*, 2009). The repeated appearance of these phenotypes coincident with the transition to planktonic habitats is suggestive of adaptations to life as a suspended particle. It is unlikely, however, that these traits are selected solely by the requirements of the planktonic environment. Large cell size, for example, is a strategy for defense against predators (Thingstad *et al.*, 2005; Verdy *et al.*, 2009; Yokota and Sterner, 2010) and appears to be favored in marine environments in general (Litchman *et al.*, 2009; Nakov *et al.*, 2014).

Estimates of the number of habitat traversals on the phylogeny offers several insights. First,

transitions to planktonic lifestyle from a benthic habitat occurred more frequently than the opposite. This result is consistent regardless of the method of inference (parsimony or model based). Second, the majority of transitions to the plankton happened in lineages that had already attained a colonial growth form (Figure 2c). Third, planktonic, colony-forming lineages rarely or never transition to benthic habitats (Figure 2c). Taken together, these observations indicate that lineages with solitary growth form traverse the habitat boundaries rarely, but in both directions. On the other hand, in the colonial state, transitions between habitats become highly asymmetric (under parsimony: 7–12 benthic → planktonic versus 0–1 planktonic → benthic). Planktonic colonial lineages can be viewed as somewhat of a ‘dead-end’ with respect to habitat traversals. For these species, transition to the benthos would be a two-step process involving, first, loss of colonial growth habit and, second, transition to the benthos (Figure 2c). Overall, the benthic colonial state seems most dynamic, as changes in either trait are more frequent compared with transitions to and from other states (Figure 2c).

Estimating character state shifts is sensitive to taxon sampling. In this data set, benthic species outnumber planktonic by factor of two, and among the colony formers this factor increases to 2.4. The bias in favor of benthic species could be problematic, but only if it incorrectly depicts the ratio of species richness observed in nature. Estimates of species numbers in diatoms are uncertain (Guiry 2012; Mann and Vanormelingen, 2013), but it is generally accepted that the benthos is more diverse than the plankton, especially in freshwater lakes (Mackay *et al.*, 2010). While we cannot ascertain that the ratio of colonial + benthic:colonial + planktonic species in our data set is a very accurate approximation of the diversity in nature, any potential bias in these data is likely in favor of the less numerous planktonic species as opposed to benthic taxa. Future studies will undoubtedly refine the findings reported here, but it is unlikely that these inferences are a result of a gross misrepresentation of the ratio of benthic:planktonic diversity.

Concluding remarks

We used a broadly sampled diatom phylogeny and flexible modeling framework to investigate the interaction between habitat occupancy, that is, where a species’ niche is, and growth form, that is, what is the species growth habit. We found that switches of habitat are rare and niche location, plankton or benthos, is conserved at the level of large clades. Transitions between growth forms are faster, but lineage specific, such that evolution is essentially stagnant in some clades, and dynamic in others. There is little interaction between these traits—switching niche location does not depend

on growth form and alternating between growth forms is not conditioned on habitat.

The approach taken here is somewhat simplistic in that the variety of microhabitats inhabited by diatoms and the diverse set of colony morphologies and aggregation mechanisms were reduced to binary characters. We were interested in capturing patterns at a very general level, that is, is cell aggregation in a colony, of any type, favored in planktonic versus benthic species? Thus, how colonies were constructed or whether the species were neritic or oceanic were not essential considerations for the present study. Fine-grained coding of microhabitats and colony features in lineages with a diversity of phenotypes allow posing specific questions that can refine the inferences made here. For example, what trajectories lead to and from the seemingly dead-end planktonic colonial state? Are mucilage- versus silica-derived cell-to-cell attachment mechanisms attained at similar rates? Has the acquisition of active motility slowed down the rate of growth form evolution? Understanding the evolution of these niche-related phenotypes would likely require multi-trait approaches that incorporate species cell size, morphology and physiological characteristics as well as investigation of the interplay between trait evolution and rates of species diversification.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work was partially funded by NSF EF 0629410 and Blumberg Fellowship to ECT. Computations were performed at the Texas Advanced Computing Center (<http://www.tacc.utexas.edu>).

References

- 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.
- Armbrust EV. (2009). The life of diatoms in the world's oceans. *Nature* **459**: 185–192.
- Ashworth MP, Nakov T, Theriot EC. (2013). Revisiting Ross and Sims (1971): toward a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae (Bacillariophyceae). *J Phycol* **49**: 1207–1222.
- Beaulieu JM, O'Meara BC, Donoghue MJ. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst Biol* **62**: 725–737.
- Brown JW, Sorhannus U. (2010). A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One* **5**: e12759.
- Burnham KP, Anderson DR. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer: New York, USA.
- Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O et al. (2004). The evolution of modern eukaryotic phytoplankton. *Science* **305**: 354–360.
- Guiry MD. (2012). How many species of algae are there? *J Phycol* **48**: 1057–1063.
- Hoagland KD. (1983). Short-term standing crop and diversity of periphytic diatoms in a eutrophic reservoir. *J Phycol* **19**: 30–38.
- Hoagland KD, Roemer SC, Rosowski JR. (1982). Colonization and Community structure of two periphyton assemblages, with emphasis on the diatoms (Bacillariophyceae). *Am J Bot* **69**: 188–213.
- Hopkinson BM, Dupont CL, Allen AE, Morel FMM. (2011). Efficiency of the CO₂-concentrating mechanism of diatoms. *Proc Natl Acad Sci USA* **108**: 3830–3837.
- Johnson RE, Tuchman NC, Peterson CG. (1997). Changes in the vertical microdistribution of diatoms within a developing periphyton mat. *J North Am Benthol Soc* **16**: 503–519.
- 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 RK, Medlin LG, Mann D. (2007). The origin and evolution of the diatoms: Their adaptation to a planktonic existence. In *Evolution of Primary Producers in the Sea*. Falkowski PG, Knoll AH (eds). Academic Press: Burlington, pp 207–249.
- Litchman E, Klausmeier CA, Yoshiyama K. (2009). Contrasting size evolution in marine and freshwater diatoms. *Proc Natl Acad Sci* **106**: 2665–2670.
- Mackay AW, Edlund MB, Khursevich G. (2010). Diatoms in ancient lakes. In *The Diatoms: Applications for the Environmental and Earth Sciences*. Smol JP, Stoermer EF (eds). Cambridge University Press: Cambridge, pp 209–228.
- Maddison WP, Maddison DR. (2011). Mesquite: a modular system for evolutionary analysis. <http://mesquiteproject.org>.
- Mann DG, Vanormelingen P. (2013). An inordinate fondness? the number, distributions, and origins of diatom species. *J Eukaryot Microbiol* **60**: 414–420.
- McCormick PV, Stevenson RJ. (1991). Mechanisms of benthic algal succession in lotic environments. *Ecology* **72**: 1835–1848.
- Nakov T, Theriot EC, Alverson AJ. (2014). Using phylogeny to model the evolution of cell size in marine and freshwater diatoms. *Limnol Oceanogr* **59**: 79–86.
- Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B. (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob Biogeochem Cycles* **9**: 359–372.
- Niklas KJ, Newman SA. (2013). The origins of multicellular organisms. *Evol Dev* **15**: 41–52.
- Padisák J, Soróczki-Pintér É, Reznér Z. (2003). Sinking properties of some phytoplankton shapes and the relation of form resistance to morphological diversity of plankton—an experimental study. *Hydrobiologia* **500**: 243–257.
- Pagel M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc R Soc B Biol Sci* **255**: 37–45.

- Pagel M, Meade A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am Nat* **167**: 808–825.
- Paradis E, Claude J, Strimmer K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- R Development Core Team (2013). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, URL <http://www.R-project.org/>.
- Reynolds CS. (2006). *The Ecology of Phytoplankton*. Cambridge University Press: Cambridge.
- Round FE, Crawford RM, Mann DG. (1990). *Diatoms: Biology and Morphology of the Genera*. Cambridge University Press: Cambridge.
- Sanderson MJ. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol* **19**: 101–109.
- Spaulding SA, Kociolek JP. (2000). Freshwater diatom biogeography. *Nova Hedwig* **71**: 223–242.
- Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**: 758–771.
- Stevenson RJ. (1997). An introduction to algal ecology in freshwater benthic habitats. In *algal Ecology: Freshwater Benthic Ecosystems*. Stevenson RJ, Bothwell ML, Lowe R (eds). Aquatic ecology series, Academic Press: San Diego.
- Theriot EC, Ashworth M, Ruck E, Nakov T, Jansen RK. (2010). A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecol Evol* **143**: 278–296.
- Theriot EC, Cannone JJ, Gutell RR, Alverson AJ. (2009). The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny. *Eur J Phycol* **44**: 277–290.
- Theriot EC, Ruck E, Ashworth M, Nakov T, Jansen RK. (2011). Status of the pursuit of the diatom phylogeny: are traditional views and new molecular paradigms really that different? *Diatom World* **119–142**.
- Thingstad TF, Øvreås L, Egge JK, Løvdaal T, Heldal M. (2005). Use of non-limiting substrates to increase size; a generic strategy to simultaneously optimize uptake and minimize predation in pelagic osmotrophs? *Ecol Lett* **8**: 675–682.
- Verdy A, Follows M, Flierl G. (2009). Optimal phytoplankton cell size in an allometric model. *Mar Ecol Prog Ser* **379**: 1–12.
- Vyverman W, Verleyen E, Sabbe K, Vanhoutte K, Sterken M, Hodgson DA *et al*. (2007). Historical processes constrain patterns in global diatom diversity. *Ecology* **88**: 1924–1931.
- Yokota K, Sterner RW. (2010). Trade-offs limiting the evolution of coloniality: ecological displacement rates used to measure small costs. *Proc R Soc B Biol Sci* **278**: 458–463.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)