

# Phylogenetic analyses reveal an increase in the speciation rate of raphid pennate diatoms in the Cretaceous

Alexandra Castro-Bugallo<sup>1,\*</sup>, Danny Rojas<sup>2,3</sup>, Sara Rocha<sup>4</sup>, Pedro Cermeño<sup>1</sup>

<sup>1</sup>Department of Marine Biology and Oceanography, Instituto de Ciencias del Mar, Consejo Superior de Investigaciones Científicas, E-08003, Barcelona, Spain

<sup>2</sup>Department of Biology and Centre for Environmental and Marine Studies, University of Aveiro, 3810-193 Aveiro, Portugal

<sup>3</sup>Department of Ecology and Evolution, Stony Brook University, 650 Life Sciences Building Stony Brook, NY 11794, USA

<sup>4</sup>Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310, Vigo, Spain

*\*Corresponding author: [alexandracbugallo@gmail.com](mailto:alexandracbugallo@gmail.com)*

Department of Marine Biology and Oceanography  
Instituto de Ciencias del Mar, Consejo Superior de Investigaciones Científicas,  
E-08003, Barcelona, Spain  
Tel: 0034 611411435

**Short title:** Cretaceous increase of marine diatom speciation

**Keywords:** Marine diatoms, raphid pennates, speciation rate, molecular phylogenies, Bayesian inference, heterothallism

# **Abstract**

The raphid pennates (order Bacillariales) are a diverse group of diatoms easily recognized by having a slit in the siliceous cell wall, called the raphe, with functions in cell motility. It has been hypothesized that this morphological innovation contributed to the evolutionary success of this relatively young but species-rich group of diatoms. However, owing to the incompleteness of the fossil record this hypothesis remains untested. Using the 18S ribosomal RNA gene, fossil calibrations, and Bayesian phylogenetic and diversification frameworks, we detect a shift in the speciation rate of marine raphid pennate diatoms in the Cretaceous, not detected in other diatom lineages nor previously recognized in the microfossil record. Our results suggest a positive link between the speciation of raphid pennate diatoms and the benefits derived from evolving motility skills, which could account for their outstanding present-day global diversity. The coincidence between the advent of the raphe and the increase in the speciation rate of raphid pennates supports the idea that simple morphological novelties can have important consequences on the evolutionary history of eukaryotic microorganisms.

# **Introduction**

Diatoms are silica-precipitating microalgae responsible for roughly one fifth of present day global primary production (i.e. ~20 Gt C per year), and contribute disproportionately to the maintenance of upper trophic levels and carbon sequestration through organic burial in sediments<sup>1,2</sup>. Diatoms are characterized by a peculiar diplontic life cycle involving gradual size reduction during asexual divisions followed by necessary size restitution via sexual reproduction<sup>3,4</sup>. Unlike the centric diatoms, which release flagellated microgametes to the extracellular medium, pennate diatoms normally produce non-flagellated amoeboid gametes of equal size in response to the pairing of vegetative cells of the opposite mating type. Another fundamental difference between centric and pennate diatoms lies in the mating system. Whereas the centrics are strictly homothallic and thus a single clone can produce both male and female compatible gametes, the pennates tend to be heterothallic, that is cross-fertilization must proceed from gametes produced by different clones or mating types<sup>5-7</sup>. The need for size restitution via sexual reproduction has confined pennate diatoms primarily into shallow benthic habitats, where sexual encounters among mating cells become more likely. Still some pennates have adapted successfully to planktonic habitats by maximizing sexual encounters during bloom events. The system of partner location works particularly well in raphid pennates, a lineage of pennate diatoms within the order Bacillariales, easily recognized by having a slit in the siliceous cell wall or frustule, the raphe, which confers the cells autonomous motility<sup>8,9</sup>. Indeed, in spite of their relatively short evolutionary history (i.e. raphid pennates are not known in the fossil record before the Palaeocene epoch), they are the most species-rich group of diatoms today<sup>9,10</sup>, which leads to the hypothesis that this morphological feature facilitated their diversification.

The dynamics of diatom diversity through time has been usually investigated from the analysis of the microfossil record<sup>11-16</sup>. Two major geological projects, the Ocean Drilling Program (ODP) and the Deep Sea Drilling Project (DSDP), now integrated into the International Ocean Discovery Program (IODP) have been instrumental to add new fossil data and unify taxonomic criteria into a global microfossil database with enormous potential for paleoecological

research<sup>17,18</sup>. However, the fossil record of marine diatoms is incomplete and severely biased toward recent times for several reasons<sup>13,19-21</sup>. First, silica structures are prone to dissolution at early stages of diagenesis and only a minor percentage of the silica frustules that accumulate in the sediments become eventually preserved. Second, silica recrystallizes under pressure, and as a consequence early diatoms are only preserved through unusual processes such as early carbonate cementation, pyritization or shallow burial. Third, sampling probability decreases with increasing geologic age owing to a deeper position of ancient sediments and the loss of oceanic crust at subduction zones. As a consequence, detailed taxonomic catalogues of fossil diatom assemblages are limited to unconsolidated sedimentary deposits dating back as much as the mid Cenozoic (~40 million years ago, Ma) and thus the evolutionary rates of marine diatoms prior to this time are virtually unknown<sup>20,21</sup>. These preservation biases call into question the suitability of the marine microfossil record for studying the macroevolutionary patterns of eukaryotic microorganisms<sup>9,19</sup>.

It is possible to estimate evolutionary rates (speciation and extinction) through time from molecular phylogenies based on extant species<sup>22</sup>. Distributions of time-calibrated phylogenetic trees are used to detect shifts in the speciation (and extinction) rate through time by simulating distinct evolutionary patterns and searching for the one(s) that best explain the dynamics of the clades within the phylogeny<sup>23-25</sup>. Here we infer a time-calibrated marine diatom phylogeny including available sequences of the 18S ribosomal RNA gene in GenBank and use a reversible jump Markov Chain Monte Carlo method<sup>26</sup> to detect speciation rate shifts and estimate rate values in major diatom lineages. The method takes into account biases associated with incomplete taxon sampling by incorporating missing lineages at the tree inference stage once provided sampling probabilities. Our objective is to explore the timing and patterns of marine diatom speciation, test for differences among groups, and identify potential causes for their ecological success in the modern oceans.

## Results

### *Sequence alignment and phylogenetic analyses*

An initial maximum length (without gaps) of 1645 base pairs (bp) resulted in a final alignment of 1157 bp after trimming for terminal regions. We detected that the trimmed sequences of *Thalassiosira pacifica*, *Nitzschia closterium*, *Chaetoceros neogracile*, *Porosira pseudodenticulata* and *Attheya septentrionalis* were equal to *Thalassiosira aestivalis*, *Cylindrotheca fusiformis*, *Chaetoceros gracilis*, *Porosira glacialis* and *Attheya longicornis*, respectively, so the first four sequences were deleted at the haplotypes collapsing step.

Maximum-likelihood (ML) and Bayesian inferences (BI) resulted in similar estimates of the phylogenetic tree, with congruent well-supported relationships (Figure S1, Figure S2). In these "unconstrained" ML/BI searches, most relevant groups obtained high support in accordance with previous phylogenetic analyses<sup>27,28</sup>. Rhizosoleniales, Coscinodiscales, Chaetocerotales, Thalassiosirales and Bacillariales (except *Attheya longicornis*) were highly supported both in Bayesian inference (Posterior probability, PP = 1) and ML (Bootstrap support, BS > 900). The sister-relationship between *Attheya longicornis* and the other Bacillariales included in the analyses was the lowest supported relationship (PP = 0.79; BS = 540), but this is in agreement

with previous studies showing the uncertain position of *Attheya longicornis* within diatoms<sup>29</sup>.  
The remaining Bacillariales received maximum support for defining a monophyletic clade.

Using the calibration points (Table S2), divergence time analyses estimated the root of the group at 193.32 Ma with a 95% highest posterior density (HPD = 121.98-274.74 Ma). This result is also consistent with other phylogenetic studies, which situate the origin of diatoms between the Triassic and lower Jurassic<sup>28,30</sup>. The mean age of the Thalassiosirales clade was 113 Ma (HPD = 69-155), and both Coscinodiscales and Bacillariales (including *Attheya longicornis*) were estimated to have originated in the Cretaceous, 115 Ma (HPD = 62-168) and 118 Ma (HPD = 62-180), respectively. Chaetocerotales and Rhizosoleniales seem to have a more recent origin, ~66 Ma (HPD = 56-86) and 91.5 Ma (HPD = 88-94), respectively (Figure 1, see also information available through the following Figshare DOI: [10.6084/m9.figshare.3795834](https://doi.org/10.6084/m9.figshare.3795834)).

### **Diversification analyses**

The effective sample size for the number of shifts and log-likelihood was superior to 2000 in both cases, indicating the appropriate sampling of parameters from the posterior. The 95% of the most credible rate shift sets were sampled with the Bayesian Analysis of Macroevolutionary Mixture (BAMM) from which twelve shift configurations were generated (see Material and Methods, Table S3). We found strong evidence for the one rate shift macroevolutionary model compared with the zero rate shift model (BF = 19.95). The rate shift was positioned at the base of the raphid pennate diatoms core within the order Bacillariales ~78-80 Ma (Figure 1). The two rate shift macroevolutionary model (BF = 5.03) was discarded because the overall best shift configuration positioned the second shift in the same location as the first one with less significant statistical support (Figure S3, Table S3). This speciation rate shift is the first one detected with non-fossil techniques and is in agreement with the expansion of raphid pennate diatoms during the Cenozoic era.

Phylorate plots showed that the maximum speciation rate of Bacillariales was as high as 0.2 species My<sup>-1</sup>, while the maximum for other diatom groups such as Thalassiosirales never exceeded 0.1 species My<sup>-1</sup> (Figure 2). All groups exhibited an increasing speciation rate through time. We found statistically significant differences in speciation rates between raphid pennate diatoms and centric diatoms ( $F = 191.96$ ,  $P = 0.0097$ , significance level = 0.05). The low speciation rate of Thalassiosirales throughout the Cenozoic era is conspicuous because this order is particularly successful in the modern oceans.

### **Discussion**

Using time-calibrated diatom phylogenies within a Bayesian framework we detected a shift in the speciation rate of raphid pennates 78-80 Ma. This speciation rate shift could not be previously recognized in the marine diatom fossil record, which is sparse for that time and strongly biased towards recent times<sup>13,18</sup>.

The fossil record situates the origin of raphid pennate diatoms some 70.6-55.8 Ma<sup>19</sup>, yet, our results (Figure 1) and previous molecular analyses push back their origin earlier in the

Cretaceous. Analyses of the 18S rRNA and ribulose biphosphate carboxylase/oxygenase large chain genes have shown that the raphid pennates comprise a monophyletic group of diatoms within the order Bacillariales<sup>27,28,30</sup>. Regarding to *Attheya longicornis*, its phylogenetic position is largely uncertain<sup>29</sup>, and although phylogenetic analyses seem to indicate that it shares an ancestor with the raphid pennates, it is clearly a distant relative and lacks a raphe. The inferred speciation rate shift was located in the Bacillariales clade, excluding *Attheya longicornis*, which further suggests that the most probable shift detected in our phylogeny corresponds to the evolutionary expansion of the raphid pennates.

The speciation rate shift reported here could result from i) a greater chance of sexual reproduction in raphid pennate diatoms with respect to other diatom lineages and ii) a tendency towards heterothallism in the pennate diatoms mating system, which may have contributed to erect hybridization barriers and reduce gene flow among sympatric populations.

Meiotic recombination in sexually-reproducing eukaryotes represents a first order cause of genotypic variability within mating populations and hence the frequency of sexual reproduction potentially accelerates the rate of species evolution<sup>31,32</sup>. Sexual reproduction is an obligate stage in the life cycle of most diatoms. However, the success of sexual reproduction may differ among diatom lineages depending on their sexual strategy and lifestyle<sup>9,33</sup>. For instance, in centric diatoms, which have adopted primarily a planktonic lifestyle, environmental cues induce the formation of eggs and flagellated sperm, which is released to the extracellular medium. A suite of concurrent requirements including appropriate environmental cues, the occurrence of sexualized cells and bloom-forming conditions that increase the chance of egg fertilization must be met to succeed<sup>34,35</sup>. Conversely, in pennate diatoms, the production of gametes is preceded by the pairing of vegetative cells of the opposite mating type<sup>33,36</sup>, a process that seems to be facilitated by the release of regulatory pheromones<sup>37,38</sup>. The system of partner location works especially well in raphid pennates, which use their motility skills to search actively for a partner and join their gametes<sup>9</sup>. Though there is little information regarding the frequency of sex in diatoms, previous reports indicate that, overall, pennate diatoms tend to exhibit shorter temporal lags between sexual events than centric diatoms<sup>7</sup>. For instance, the cell size threshold for inducing sexual reproduction in the diatom *Pseudo-nitzschia multiseriata* has been shown to be larger than previously thought<sup>39</sup>, expanding the window of opportunity for sexualization.

Sexual reproduction favors adaptation to new habitats and is particularly advantageous in rapidly changing environments, where some genetic variants might be wiped out by new conditions while others might be better adapted and thrive. In addition to increasing the chance of sexual recombination, gliding motility potentially increases resource use efficiency by conferring individuals the ability to seek for optimal light environments and nutrient-rich conditions<sup>40-42</sup>. We recognize that the outstanding current diversity of raphid pennate diatoms could be explained in part as a result of a finer niche differentiation in benthic habitats compared to more homogeneous planktonic ecosystems. However, this argument on its own is unable to explain the lower diversity of araphid pennates today despite their primarily benthic lifestyles and comparable evolutionary origins. Furthermore, our analysis shows that the lineages of raphid pennate diatoms adapted to planktonic lifestyles (e.g. *Pseudo-nitzschia*,

*Fragilaria*) also attained higher speciation rates than other planktonic lineages such as the Thalassiosirales, Chaetocerotales, Rhizosoleniales and Coscinodiscales, supporting the idea that the benefits derived from evolving motility skills (the raphe) played a role. The timing of the raphid pennate diatoms speciation rate shift suggests that changes in the extent of continental flooding along with concurrent increases in continental nutrient weathering fluxes<sup>43</sup> and sedimentary organic matter recycling<sup>44</sup> provided ideal conditions for habitat expansion since the mid-Cretaceous<sup>19</sup>.

The second critical control on evolutionary tempo deals with the fact that pennate diatoms show a tendency towards heterothallism<sup>35</sup>. It has been suggested that owing to their broad dispersal ranges and astronomical population numbers, microbial species cannot be geographically isolated<sup>45</sup>. Because geographic isolation is a necessary component of allopatric speciation models, global dispersal and the ensuing continuity of gene flow among sympatric populations are thought to lower the rate of speciation<sup>45</sup>. Ubiquitous dispersal is feasible in asexually reproducing microorganisms as long as global dispersal times do not exceed the pace of genetic variability. However, the mating system in heterothallic species is a fundamental control of syngamy, which increases the chance of geographic isolation and promotes the spatial structuring of genetic populations<sup>46</sup>. Our results support the idea that the advent of the raphe, a simple morphological feature involved in cell motility, facilitated sexual encounters among compatible mating types. The greater success of sexual reproduction in predominantly heterothallic taxa led to an increase in the speciation rate of raphid pennate diatoms. These results provide a feasible explanation for the outstanding current diversity of raphid pennates despite their relatively recent origin, and suggest that simple morphological novelties can have important consequences on the evolutionary history of eukaryotic microorganisms. While the signal seems strong in our data, it is important to recognize the limitations of single-gene-phylogenies,<sup>47,48</sup> and would be extremely valuable if future studies could employ more comprehensive sets of orthologs<sup>27,49,50</sup>, as these expand in databases, or their sequencing becomes more feasible, to keep testing hypotheses concerning timing and rates of diversification.

## Materials and Methods

### *Sequences and alignment*

All available sequences (102 sequences) of the 18S ribosomal RNA gene of major marine diatom orders (Thalassiosirales, Chaetocerotales, Rhizosoleniales, Coscinodiscales and Bacillariales) were downloaded from GenBank (Table S1). Sequences were aligned using MAFFT v7.058b<sup>51</sup> and the G-INS-I algorithm<sup>52</sup>, trimmed using GBlocks automatic parameters<sup>53</sup> and collapsed into haplotypes using ALTER<sup>54</sup> resulting in a final alignment with 97 sequences and 1157bp. The minimum number of sequences for defining a conserved position was 50 and 83 for a flanking position. The maximum number of contiguous non-conserved position was 8 while the minimum length of a block allowed after gap cleaning was 10. Alignments and haplotypes information are available through the following Figshare DOI: [10.6084/m9.figshare.3795834](https://doi.org/10.6084/m9.figshare.3795834).

## Phylogenetic analyses

To determine the best-fit model of sequence evolution for the dataset we used jModelTest v0.1.1<sup>55,56</sup> and the corrected Akaike Information Criterion (AICc)<sup>57</sup>. Maximum-likelihood (ML) and Bayesian (unconstrained) gene phylogenies were estimated using PhyML 3.0<sup>58</sup> and MrBayes v3.2.1<sup>59</sup>. ML was performed with 100 heuristic searches and its support accessed through 1000 bootstrap replicates. Two Bayesian inference runs of 10 million generations were performed, with default heating parameters, sampled each 1000<sup>th</sup> sample, and checked for convergence (of node's PP's) and congruence (of runs) using AWTY<sup>60</sup>. Both runs were summarized on a 50% majority-rule consensus. All details and outputs are available at DOI: [10.6084/m9.figshare.3795834](https://doi.org/10.6084/m9.figshare.3795834).

BEAST v2.1.3<sup>61,62</sup> was used for co-estimating divergence times and (constrained) gene phylogenies using relaxed molecular clock analyses. The nucleotide substitution model was implemented as estimated previously, with parameters to be co-estimated along the run. An uncorrelated lognormal model was used for the clock-rate and posterior estimates were obtained under both the Calibrated Yule and the Birth-Death models for the tree prior. Runs without data were also performed to evaluate prior and joint prior distributions.

Calibrations in the phylogeny were implemented either as minimum or “fixed” (interval) ages. Using data from the marine diatom fossil record, we constrained the minimum ages of several clades in the tree corresponding to the oldest unequivocal fossil belonging to that clade. In all cases these clades were highly supported in the unconstrained phylogenies (PP=1; BS>900/1000). The single exception was *Skeletonema grethae*, where specimens from the Atlantic and Pacific were not inferred as sister-taxa. Because the relationships within the clade they belong were largely unresolved (see Figure S1, S2), with no support either for them not being sister taxa, and the distance between them was very low (<0.3%), similar to the one between the other Atlantic-Pacific pair (*Thalassiosira weissflogii*), and to other Atlantic-Pacific taxa pairs which divergence is assumed to have been initiated at the final closure of the Isthmus of Panama<sup>63</sup>, we constrained this node and still included this calibration in our analyses.

The median age and the interval of probability were obtained from a lognormal distribution of the data (Table S2). Fixed age estimates were used for those clades whose divergence times occurred during a well-dated geologic event. In this case, the median and probability interval were calculated using a normal distribution (Table S2).

Two independent analyses (under each tree prior) were run for 400 million generations, sampling every 400,000. The resulting distributions of parameters were checked for convergence using Tracer v 1.5<sup>64</sup> ensuring effective sample size values (ESS) to be greater than 200. The posterior distributions of trees were summarized in a maximum clade credibility tree discarding 25% as burnin using median heights for node age estimates.

## Speciation rates and phylANOVA analyses

We used the Bayesian Analysis of Macroevolutionary Mixture (BAMM, [www.bamm-project.org](http://www.bamm-project.org)) to estimate marginal distributions of speciation and extinction rates for each branch

in a phylogenetic tree<sup>65</sup>. BAMM uses reversible jump Markov Chain Monte Carlo (rjMCMC) method to detect automatically rate shifts and sample distinct evolutionary dynamics (speciation and extinction) that best explain the whole diversification dynamics of the clade. The program is designed to work with datasets that contain large numbers of missing species. The method takes into account incomplete taxon sampling in phylogenetic trees by incorporating missing lineages at the tree inference stage once provided clade-specific sampling probabilities<sup>65,66</sup>.

We provided BAMM with the proportion of species sampled per genus (i.e., 1/number of species in genus). Priors were estimated with BAMMTools<sup>26</sup> using the function “setBAMMpriors”. We set a rjMCMC for  $11 \cdot 10^6$  generations and sampled every 1,000 generations. We used ESS values to assess the convergence of the run, considering values above 200 as indicative of good convergence. To visualize where in the tree the shifts occurred, we generated mean phylorate plots which represent the mean speciation rate sampled from the posterior at any point in time along any branch of the phylogenetic tree<sup>26</sup>. BAMM identifies a set of most credible rate shifts ordering them by posterior probability. Here, we selected twelve credible rate shift sets based on a Bayes Factor (BF) considering a BF value between 3 and 12 as positive evidence, BF > 12 as strong evidence and BF > 150 as very strong evidence<sup>67</sup>. All details and outputs are available at DOI: [10.6084/m9.figshare.3795834](https://doi.org/10.6084/m9.figshare.3795834).

Phylogenetic analyses of variances were used to test differences in speciation rates among raphid pennate diatoms and centric diatoms using the function phylANOVA in the R-package phytools<sup>68</sup>. phylANOVA compares results based on raw phylogenetic data to those based on phylogenetic model simulations to create an appropriate null distribution<sup>69</sup>.

## Acknowledgments

We thank R. Massana, F.M. Cornejo-Castillo and A. Roura for comments on the manuscript. This research was supported by grants 10PXIB312058PR from Xunta de Galicia and CTM2014- 54926-R from the Spanish Ministry of Economy and Competitiveness. P.C. was supported by Ramon y Cajal contract from the Spanish Ministry.

## Author contribution

ACB and PC conceived and designed the experiment, ACB selected and downloaded the sequences, ACB and SR performed the phylogenetic analysis, ACB and DR did the diversification rate analysis, ACB wrote the first draft of the manuscript and all authors contributed substantially to revisions.

## Competing financial interest

The authors declare no competing financial interest.

## References

- 1 Falkowski, P. G. *et al.* The Evolution of Modern Eukaryotic Phytoplankton. *Science* **305**, 354-360, doi:10.1126/science.1095964 (2004).

- 283 2 Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185-192,  
284 doi:10.1038/nature08057 (2009).
- 285 3 Round, F. E., Crawford, R. M. & Mann, D. G. *The diatoms Biology and Morphology of*  
286 *the Genera*. (Cambridge University Press, 1990).
- 287 4 Kaczmarek, I. *et al.* Proposals for a terminology for diatom sexual reproduction,  
288 auxospores and resting stages. *Diatom Research* **28**, 263-294,  
289 doi:10.1080/0269249X.2013.791344 (2013).
- 290 5 Drebes, G. in *Sexuality* (ed Werner D) 250-283 (University of California Press, 1977).
- 291 6 Mann, D. G. Patterns of sexual reproduction in diatoms. *Hydrobiologia* **269**, 11-20,  
292 doi:10.1007/bf00027999 (1993).
- 293 7 Hasle, G. R. & Syvertsen, E. E. in *Identifying Marine Diatoms and Dinoflagellates* (eds  
294 R.T. Carmelo *et al.*) 5-385 (Academic Press, 1996).
- 295 8 Lind, J. L. *et al.* Substratum adhesion and gliding in a diatom are mediated by  
296 extracellular proteoglycans. *Planta* **203**, 213-221, doi:10.1007/s004250050184 (1997).
- 297 9 Kooistra, W. H. C. F., Gersonde, R., Medlin, L. K. & Mann, D. G. in *Evolution of*  
298 *planktonic photoautotrophs* (eds P.G. Falkowski & A.H. Knoll) 207-249 (Academic  
299 Press, Inc., 2007).
- 300 10 Mann, D. G. & Droop, S. J. M. Biodiversity, biogeography and conservation of diatoms.  
301 *Hydrobiologia* **118**, 19-32, doi:10.1007/BF00010816 (1996).
- 302 11 Barron, J. A. Planktonic marine diatom record of the past 18 m.y.: Appearances and  
303 extinctions in the Pacific and Southern Oceans. *Diatom Research* **18**, 203-224 (2003).
- 304 12 Katz, M. E., Finkel, Z. V., Grzebyk, D., Knoll, A. H. & Falkowski, P. G. Evolutionary  
305 trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. *Annual*  
306 *Review of Ecology, Evolution, and Systematics* **35**, 523-556,  
307 doi:10.1146/annurev.ecolsys.35.112202.130137 (2004).
- 308 13 Rabosky, D. L. & Sorhannus, U. Diversity dynamics of marine planktonic diatoms  
309 across the Cenozoic. *Nature* **457**, 183-186, doi:10.1038/nature07435 (2009).
- 310 14 Lazarus, D., Barron, J., Renaudie, J., Diver, P. & Türke, A. Cenozoic Planktonic Marine  
311 Diatom Diversity and Correlation to Climate Change. *PLoS ONE* **9**, e84857,  
312 doi:10.1371/journal.pone.0084857 (2014).
- 313 15 Cermeño, P., Falkowski, P. G., Romero, O. E., Schaller, M. F. & Vallina, S. M.  
314 Continental erosion and the Cenozoic rise of marine diatoms. *Proceedings of the*  
315 *National Academy of Sciences* **112**, 4239-4244, doi:10.1073/pnas.1412883112 (2015).
- 316 16 Kotrc, B. & Knoll, A. H. A morphospace of planktonic marine diatoms. I. Two views of  
317 disparity through time. *Paleobiology* **41**, 45-67, doi:10.1017/pab.2014.4 (2015).
- 318 17 Lazarus, D. Neptune: a marine micropaleontology database. *Mathematical Geology* **26**,  
319 817-832, doi:10.1007/BF02083119 (1994).
- 320 18 Spencer-Cervato, C. Deep Sea Microfossil Record: Using the Neptune Database.  
321 *Palaeontologia Electronica* **2** (1999).
- 322 19 Sims, P. A., Mann, D. G. & Medlin, L. K. Evolution of the diatoms: insights from fossil,  
323 biological and molecular data. *Phycologia* **45**, 361-402, doi:10.2216/05-22.1 (2006).
- 324 20 Knoll, A. H., Summons, R. E., Waldbauer, J. R. & Zumberge, J. E. in *Evolution of*  
325 *Primary Producers in the Sea* (eds P.G. Falkowski & A.H. Knoll) 133-163 (Academic  
326 Press, 2007).
- 327 21 Cermeño, P. The geological story of marine diatoms and the last generation of fossil  
328 fuels. *Perspectives in Phycology*, **3**, 53-60, doi: 10.1127/pip/2016/0050 (2016).
- 329 22 Ricklefs, R. E. Estimating diversification rates from phylogenetic information. *Trends*  
330 *Ecol Evol* **22**, 601-610, doi:10.1016/j.tree.2007.06.013 (2007).
- 331 23 Alfaro, M. E. *et al.* Nine exceptional radiations plus high turnover explain species  
332 diversity in jawed vertebrates. *Proc Natl Acad Sci USA* **106**, 13410-13414,  
333 doi:10.1073/pnas.0811087106 (2009).

- 24 Silvestro, D., Schnitzler, J. & Zizka, G. A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology* **11**, 1-15, doi:10.1186/1471-2148-11-311 (2011).
- 25 Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proc Natl Acad Sci USA* **108**, 6187-6192, doi:10.1073/pnas.1016876108 (2011).
- 26 Rabosky, D. L. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *PLoS ONE* **9**, e89543, doi:10.1371/journal.pone.0089543 (2014).
- 27 Alverson, A. J. & Theriot, E. C. Comments on Recent Progress Toward Reconstructing the Diatom Phylogeny. *Journal of Nanoscience and Nanotechnology* **5**, 57-62, doi:10.1166/jnn.2005.007 (2005).
- 28 Sorhannus, U. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Marine Micropaleontology* **65**, 1-12, doi:http://dx.doi.org/10.1016/j.marmicro.2007.05.002 (2007).
- 29 Rampen, S. W. *et al.* Phylogenetic position of *Attheya longicornis* and *Attheya septentrionalis* (Bacillariophyta). *Journal of Phycology* **45**, 444-453, doi:10.1111/j.1529-8817.2009.00657.x (2009).
- 30 Kooistra, W. H. C. F. & Medlin, L. Evolution of the diatoms (Bacillariophyta) IV A reconstruction of their age from small subunit rRNA coding regions and fossil record. *Mol Phylogenet Evol* **6**, 391-407, doi:10.1006/mpev.1996.0088 (1996).
- 31 Maynard Smith, J. *The Evolution of Sex*. 236 (Cambridge University Press, 1978).
- 32 Colegrave, N. The evolutionary success of sex: Science & Society Series on Sex and Science. *EMBO Reports* **13**, 774-778, doi:10.1038/embo.2012.109 (2012).
- 33 von Dassow, P. & Montresor, M. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *Journal of Plankton Research* **33**, 3-12, doi:10.1093/plankt/fbq137 (2011).
- 34 Lewis, W. M. J. The Diatom Sex Clock and Its Evolutionary Significance. *The American Naturalist* **123**, 73-80, doi:doi:10.1086/284187 (1984).
- 35 Chepurnov, V. A., Mann, D. G., Sabbe, K. & Vyverman, W. in *International Review of Cytology* Vol. Volume 237 91-154 (Academic Press, 2004).
- 36 Chepurnov, V. A. *et al.* In search of new tractable diatoms for experimental biology. *BioEssays* **30**, 692-702, doi:10.1002/bies.20773 (2008).
- 37 Gillard, J. *et al.* Metabolomics Enables the Structure Elucidation of a Diatom Sex Pheromone. *Angewandte Chemie International Edition* **52**, 854-857, doi:10.1002/anie.201208175 (2013).
- 38 Frenkel, J., Vyverman, W. & Pohnert, G. Pheromone signaling during sexual reproduction in algae. *The Plant Journal* **79**, 632-644, doi:10.1111/tpj.12496 (2014).
- 39 Bates, S. S. & Davidovich, N. A. in *LIFEHAB Life histories of microalgal species causing harmful blooms* (eds E. Garcés *et al.*) 31-36 (European Commission. Research in Enclosed Seas series-12, 2002).
- 40 Cohn, S. A. & Weitzell, R. E. Ecological considerations of diatom cell motility. I. Characterization of motility and adhesion in four diatom species. *Journal of Phycology* **32**, 928-939, doi:10.1111/j.0022-3646.1996.00928.x (1996).
- 41 McLachlan, D. H., Brownlee, C., Taylor, A. R., Geider, R. J. & Underwood, G. J. C. Light-induced motile responses of the estuarine benthic diatoms *Navicula perminuta* and *Cylindrotheca closterium* (Bacillariophyceae). *Journal of Phycology* **45**, 592-599, doi:10.1111/j.1529-8817.2009.00681.x (2009).
- 42 Bondoc, K. G. V., Heuschele, J., Gillard, J., Vyverman, W. & Pohnert, G. Selective silicate-directed motility in diatoms. *Nat Commun* **7**:10540 doi:10.1038/ncomms10540 (2016).

- 43 McArthur, J. M., Howarth, R. J. & Bailey, T. R. Strontium Isotope Stratigraphy: LOWESS Version 3: Best Fit to the Marine Sr Isotope Curve for 0–509 Ma and Accompanying Look up Table for Deriving Numerical Age. *The Journal of Geology* **109**, 155-170, doi:10.1086/319243 (2001).
- 44 Cardenas, A. L. & Harries, P. J. Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nature Geosci* **3**, 430-434, doi:10.1038/ngeo869 (2010).
- 45 Finlay, B. J. Global Dispersal of Free-Living Microbial Eukaryote Species. *Science* **296**, 1061-1063, doi:10.1126/science.1070710 (2002).
- 46 Casteleyn, G. *et al.* Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proceedings of the National Academy of Sciences* **107**, 12952-12957, doi:10.1073/pnas.1001380107 (2010).
- 47 Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J. Mol. Evol.* **51**, 565–576, doi: 10.1093/molbev/msj011 (2000).
- 48 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179, doi:10.1038/nature14447 (2015).
- 49 David, L. A. & Alm, E. J. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93–96, doi:10.1038/nature09649 (2011).
- 50 Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048, doi:10.1038/nmicrobiol.2016.48 (2016).
- 51 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
- 52 Katoh, K., Kuma, K.-i., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511-518, doi:10.1093/nar/gki198 (2005).
- 53 Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* **17**, 540-552 (2000).
- 54 Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F. & Posada, D. ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research* **38**, W14-W18, doi:10.1093/nar/gkq321 (2010).
- 55 Guindon, S. & Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* **52**, 696-704, doi:10.1080/10635150390235520 (2003).
- 56 Posada, D. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* **25**, 1253-1256, doi:10.1093/molbev/msn083 (2008).
- 57 Posada, D. & Buckley, T. R. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology* **53**, 793-808, doi:10.1080/10635150490522304 (2004).
- 58 Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).
- 59 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574, doi:10.1093/bioinformatics/btg180 (2003).
- 60 Nylander, J. A. A., Wilgenbusch, J. C., Warren, D. L. & Swofford, D. L. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24**, 581-583, doi:10.1093/bioinformatics/btm388 (2008).

433 61 Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics  
434 and Dating with Confidence. *PLoS Biol* **4**, e88, doi:10.1371/journal.pbio.0040088  
435 (2006).

436 62 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling  
437 trees. *BMC Evol Biol* **7**, doi:10.1186/1471-2148-7-214 (2007).

438 63 Lessios, H. A. The Great American Schism: Divergence of Marine Organisms After the  
439 Rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and*  
440 *Systematics* **39**, 63-91, doi:10.1146/annurev.ecolsys.38.091206.095815 (2008).

441 64 Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. Tracer v1.6, Available from  
442 <http://beast.bio.ed.ac.uk/Tracer>. (2014).

443 65 Rabosky, D. L. *et al.* Rates of speciation and morphological evolution are correlated  
444 across the largest vertebrate radiation. *Nat Commun* **4**:1958, doi:10.1038/ncomms2958  
445 (2013).

446 66 Thomas, G. H. *et al.* PASTIS: an R package to facilitate phylogenetic assembly with  
447 soft taxonomic inferences. *Methods in Ecology and Evolution* **4**, 1011-1017,  
448 doi:10.1111/2041-210X.12117 (2013).

449 67 Raftery, A. E. Approximate Bayes factors and accounting for model uncertainty in  
450 generalised linear models. *Biometrika* **83**, 251-266, doi:10.1093/biomet/83.2.251  
451 (1996).

452 68 Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other  
453 things). *Methods in Ecology and Evolution* **3**, 217-223, doi:10.1111/j.2041-  
454 210X.2011.00169.x (2012).

455 69 Garland, T., Dickerman, A. W., Janis, C. M. & Jones, J. A. Phylogenetic Analysis of  
456 Covariance by Computer Simulation. *Systematic Biology* **42**, 265-292,  
457 doi:10.1093/sysbio/42.3.265 (1993).

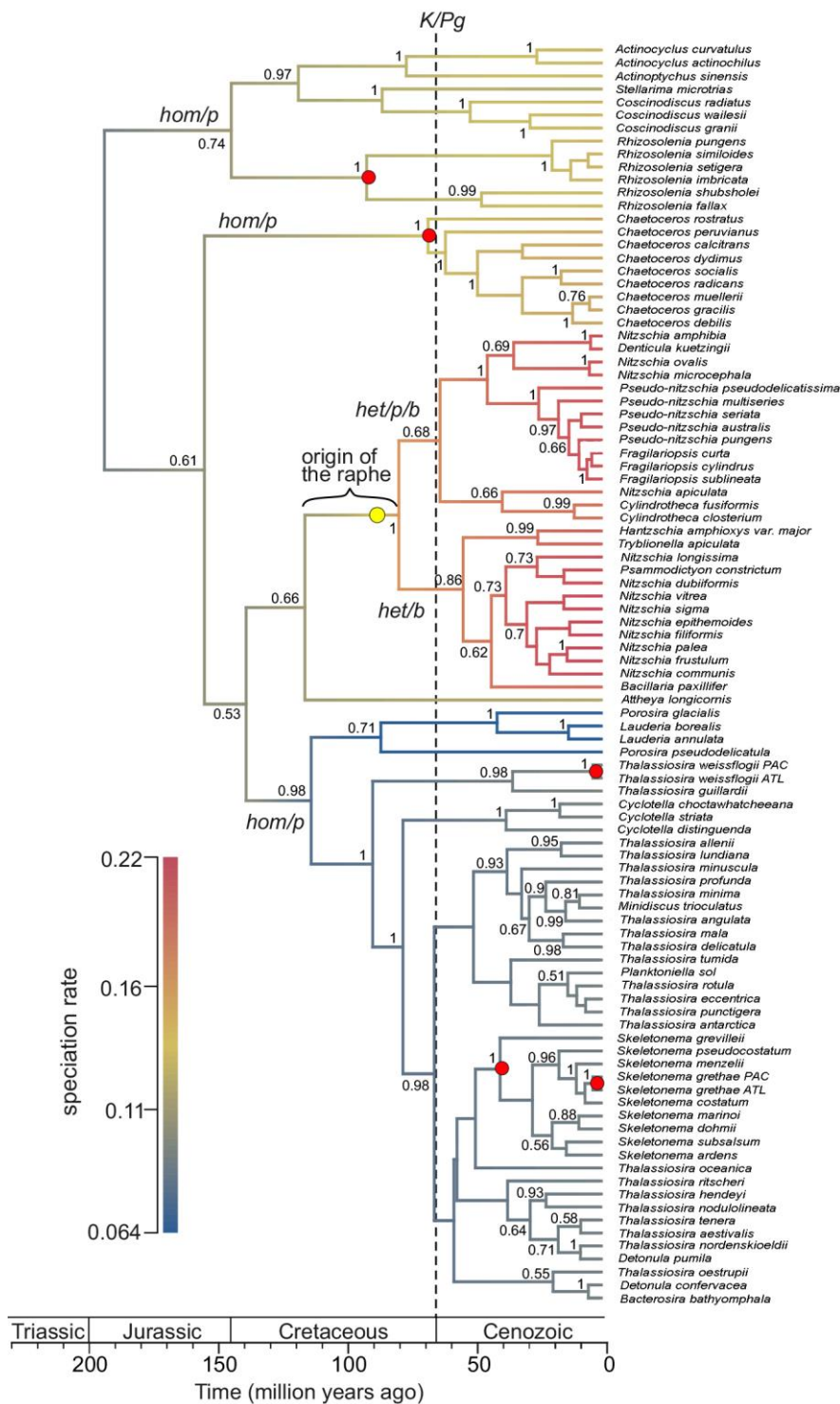
## Figure legends

**Figure 1.** Phylorate plot of the mean speciation rates of marine diatoms sampled from the posterior resulting from the BAMM output. Branches coloring (blue to red) indicates increasing speciation rates. The yellow dot marks the position of the speciation rate shift resulting from the overall best shift configuration. The red dots indicate the calibration points based on the sedimentary record (Table S2). Numbers at branches represent posterior probabilities (only values >0.5 are shown). The abbreviations over major clade branches denote the predominant sexual strategy and lifestyle within lineages: homothallism (hom), heterothallism (het), planktonic (p) and benthic (b). Also indicated is the time span for the presumed origin of the raphe.

**Figure 2.** Speciation rate through time plot of the different groups of marine diatoms compared with the evolutionary rate of the entire phylogeny (black line). Bacillariales are the primary responsible for the speciation rate shift detected in the phylogeny. Bacillariales plot includes all the species of the order except *Attheya longicornis*. Colored areas represent the quantiles of the posterior distribution. Pictures were taken by Isabel Gomes Teixeira.

## Figures

### Figure 1



**Figure 2**

