

1715

Title Clade-specific diversification dynamics of marine diatoms since the Jurassic

One Sentence Summary Phylogenetic analysis of metabarcoding data reveals effects of biotic and abiotic factors on the evolution of diatoms

Authors Eric Lewitus^{1,✧*}, Lucie Bittner², Shruti Malviya^{1,3}, Chris Bowler¹, H  l  ne Morlon¹

Affiliations ¹Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris 75005 Paris, France; ²Sorbonne Université, Univ Antilles, CNRS, Evolution Paris Seine - Institut de Biologie Paris Seine (EPS - IBPS), F-75005 Paris, France; ³Tata Institute of Fundamental Research, Bangalore, India; [◇]Current affiliation: US Military HIV Research Program, WRAIR, Silver Spring, Maryland, USA, and at the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland, USA

Email: elewitus@hivresearch.org

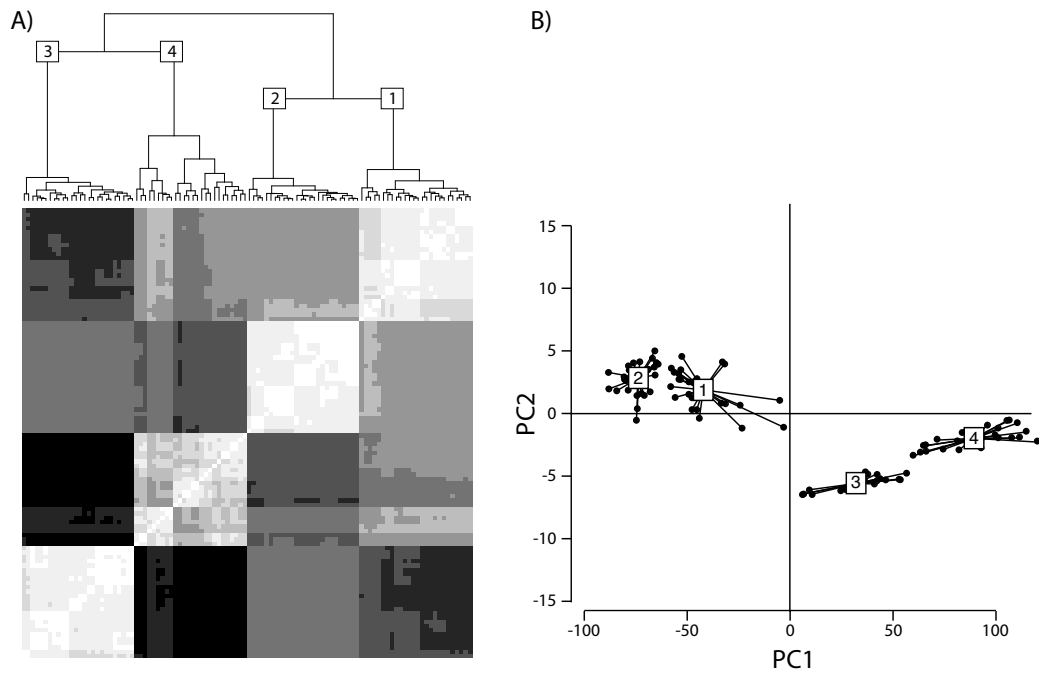
Supplementary Table 1: Ages and references for fossil and phylogenetic calibration points

taxonomic group	age (mya)	type	reference
rhizosolenids	91.5	fossil	Damste et al., Science, 2004
pinnularia	64	fossil	Souffreau et al., Mol. Phy. Evol., 2011
chaetoceros	60	fossil/phylogenetic	Ciesielski et al., Proc. Ocean Drill Program, 1991 Kooistra & Medlin, Mol. Phy. Evol., 1996
aulacoseira	48.6	fossil	Wolfe et al., PALAIOS, 1996
raphoneis	77.05	fossil	Tapia & Harwood, Micropaleontology, 2002
stellarima	96.55	fossil	Tapia & Harwood, Micropaleontology, 2002
skeletonema	77.5	fossil	Strelnikova, Nova hedwigia. Beihefte, 1975
thalassiosirales	88	fossil	Kooistra & Medlin, Mol. Phy. Evol., 1996
coscinodiscus	104	fossil	Girard et al., Geology, 2009 Sims et al., Phycologia, 2006
thalassionema	19.5	fossil	Sims et al., Phycologia, 2006
pennates	125	fossil	Sims et al., Phycologia, 2006; Sorhannus, Mar. Microp., 2007
bacillariophyceae	210-183	phylogenetic	Sorhannus, Mar. Microp., 2007 Katz et al., Ann. Rev. Ecol., Evol., and Syst., 2004

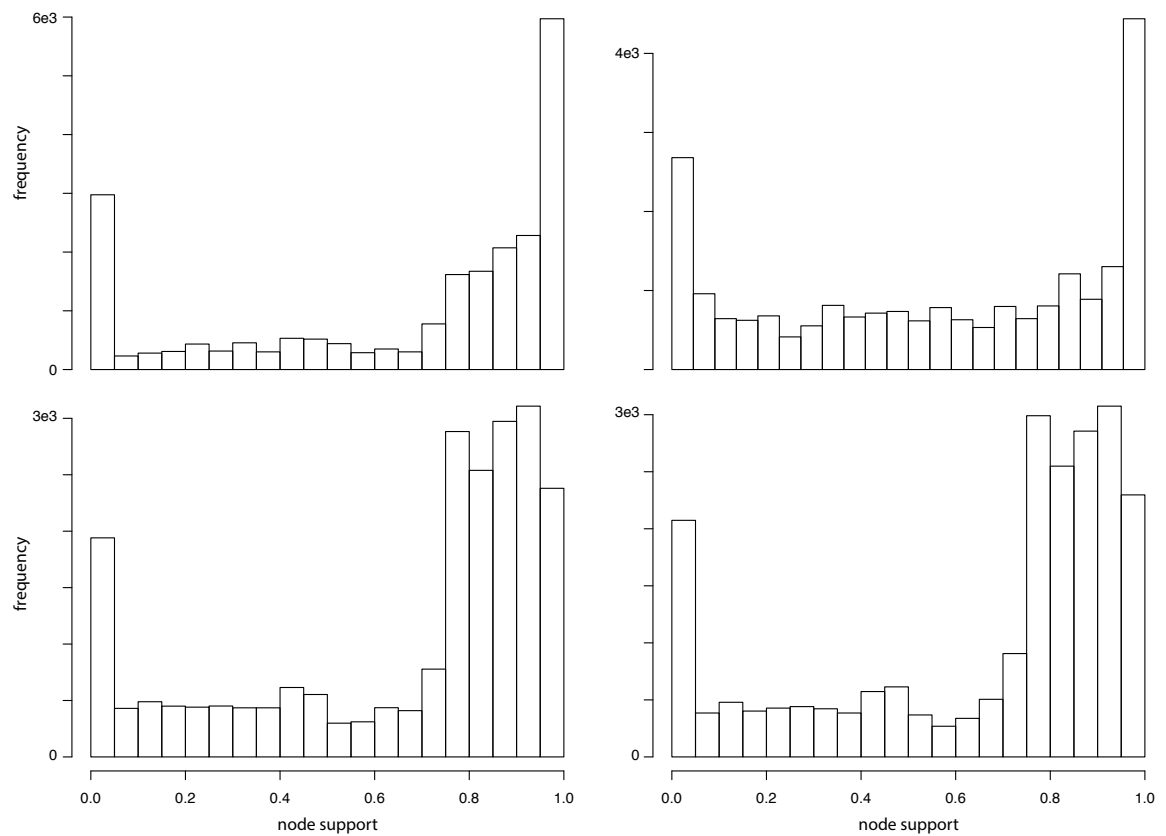
Supplementary Table 2: TreePar net diversification (ND) and relative extinction (RE) rates*

	MAFFT+RaXML	MAFFT+FastTree	CLUSTAL+RaXML	CLUSTAL+FastTree
pre-LE ND	0.022 (0.039,0.006)	0.026 (0.045,0.007)	0.006 (0.031,0.004)	0.009 (0.034,0.004)
post-LE ND	0.318 (0.138,0.436)	0.323 (0.143,0.51)	0.238 (0.126,0.357)	0.213 (0.140,0.479)
pre-LE RE	0.977 (0.961,0.991)	0.908 (0.899,0.992)	0.994 (0.932,0.998)	0.984 (0.903,0.998)
post-LE RE	0.0002 (0.0001,0.0003)	0.0002 (0.0001,0.0003)	0.0002 (0.0001,0.0003)	0.0002 (0.0001,0.0003)

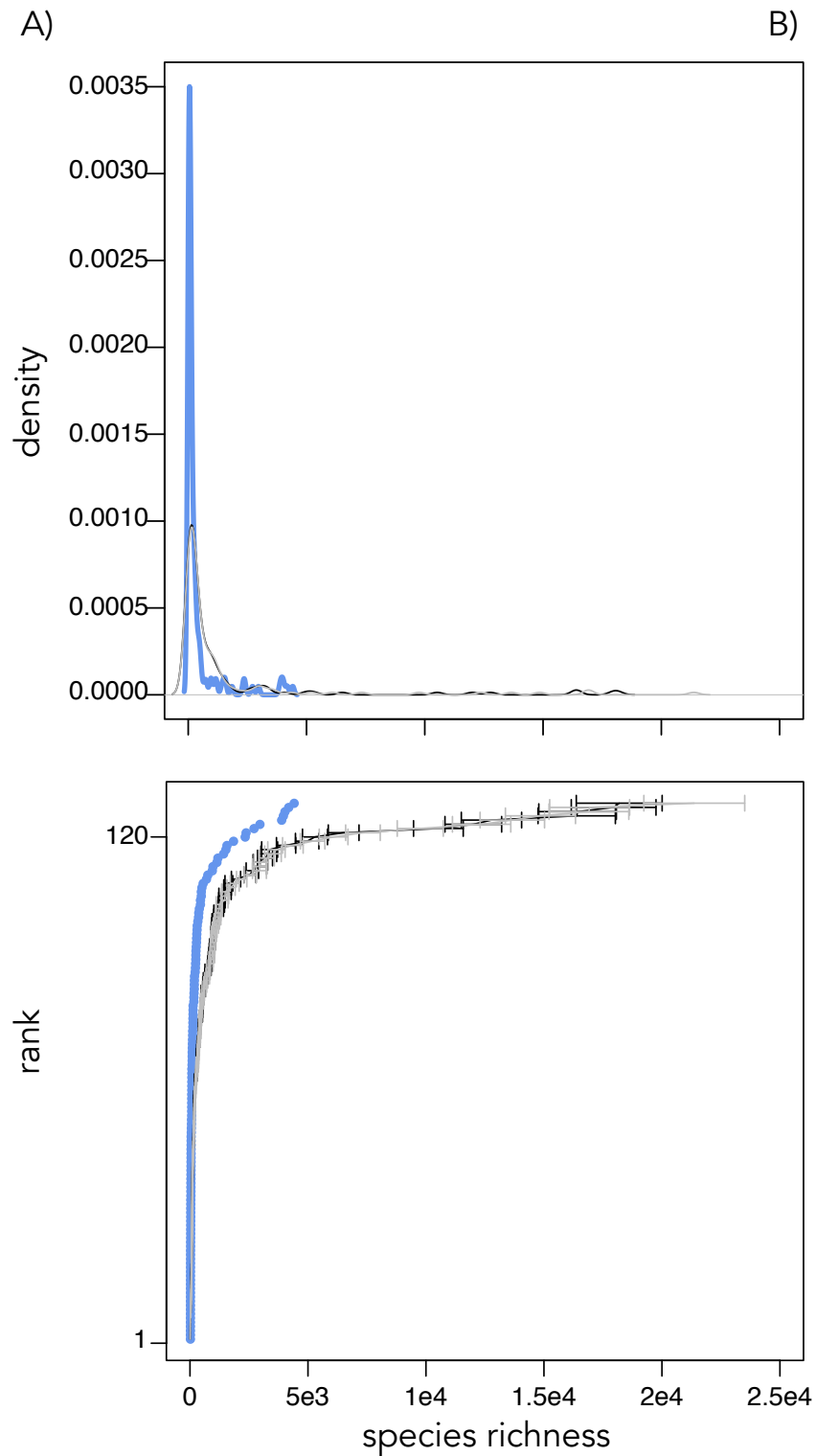
*rates are shown for median extant diversity estimates (lower extant diversity estimate, upper extant diversity estimate)



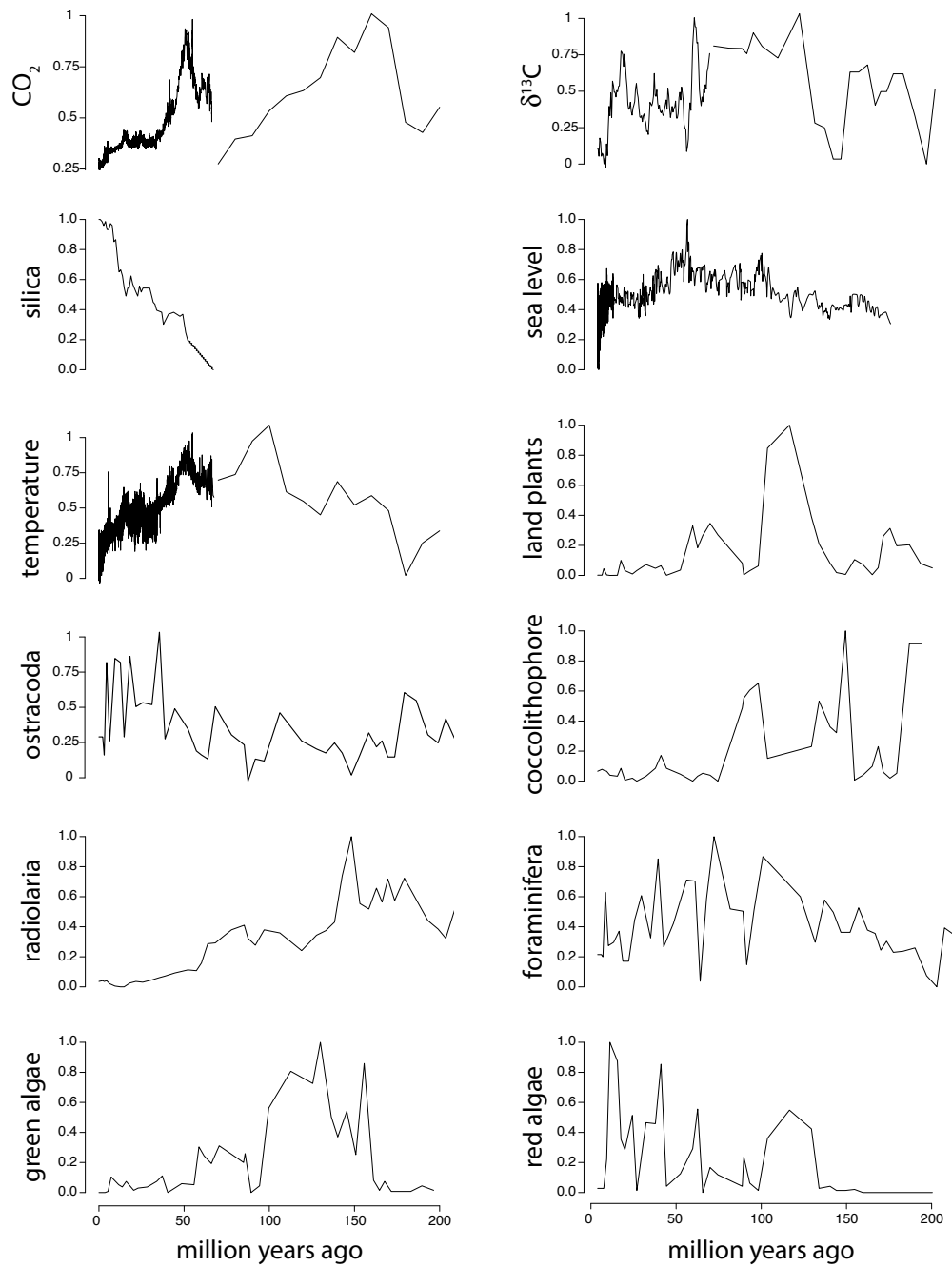
Supplemental Fig. 1: Comparison of phylogenies constructed with different PR2 backbones. (A) Hierarchical clustering of Jensen-Shannon index distances between spectral density profiles ? computed for phylogenies reconstructed using the (1) CLUSTAL+RaXML, (2) CLUSTAL+FastTree, (3) MAFFT+RaXML, and (4) MAFFT+FastTree backbones. Each phylogeny with each cluster was constructed using a different time-calibration scheme based on the 13 time-calibration points (Extended Data 3). (B) Partition clustering around medoids of phylogenies constructed from phylogenies based on Jensen-Shannon index distances as above. Principal components 1 and 2 explain 85% and 6% of the variability, respectively.



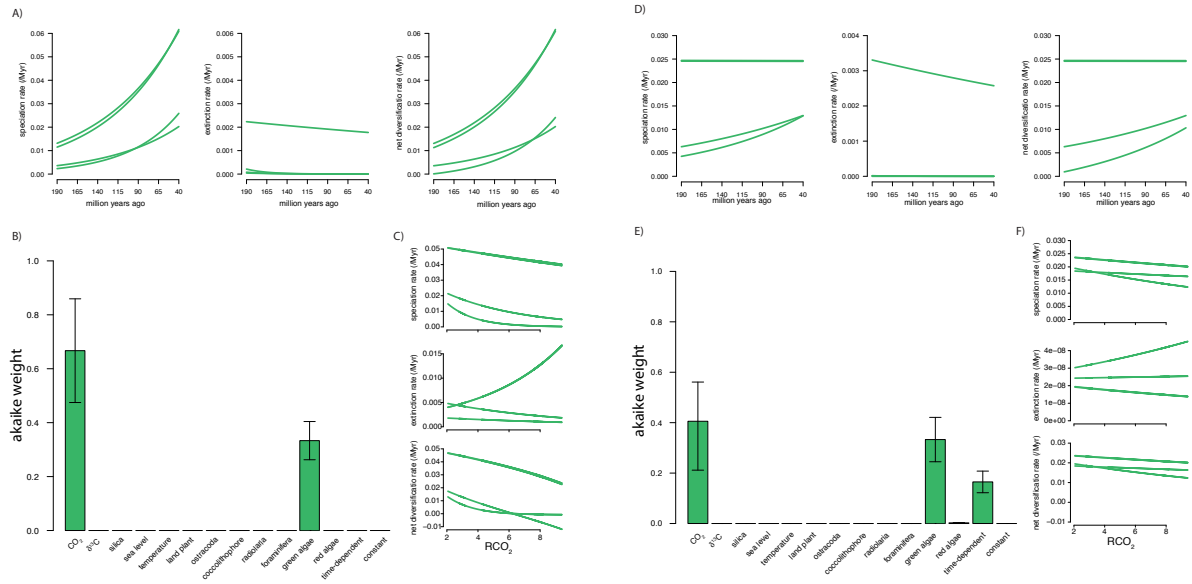
Supplemental Fig. 2: Frequency of node support for time-calibrated diatom OTU phylogenies. Average node support for all versions of the phylogeny (clockwise: MAFFT+RaXML, MAFFT+FastTree, CLUSTAL+RaXML, CLUSTAL+FastTree).



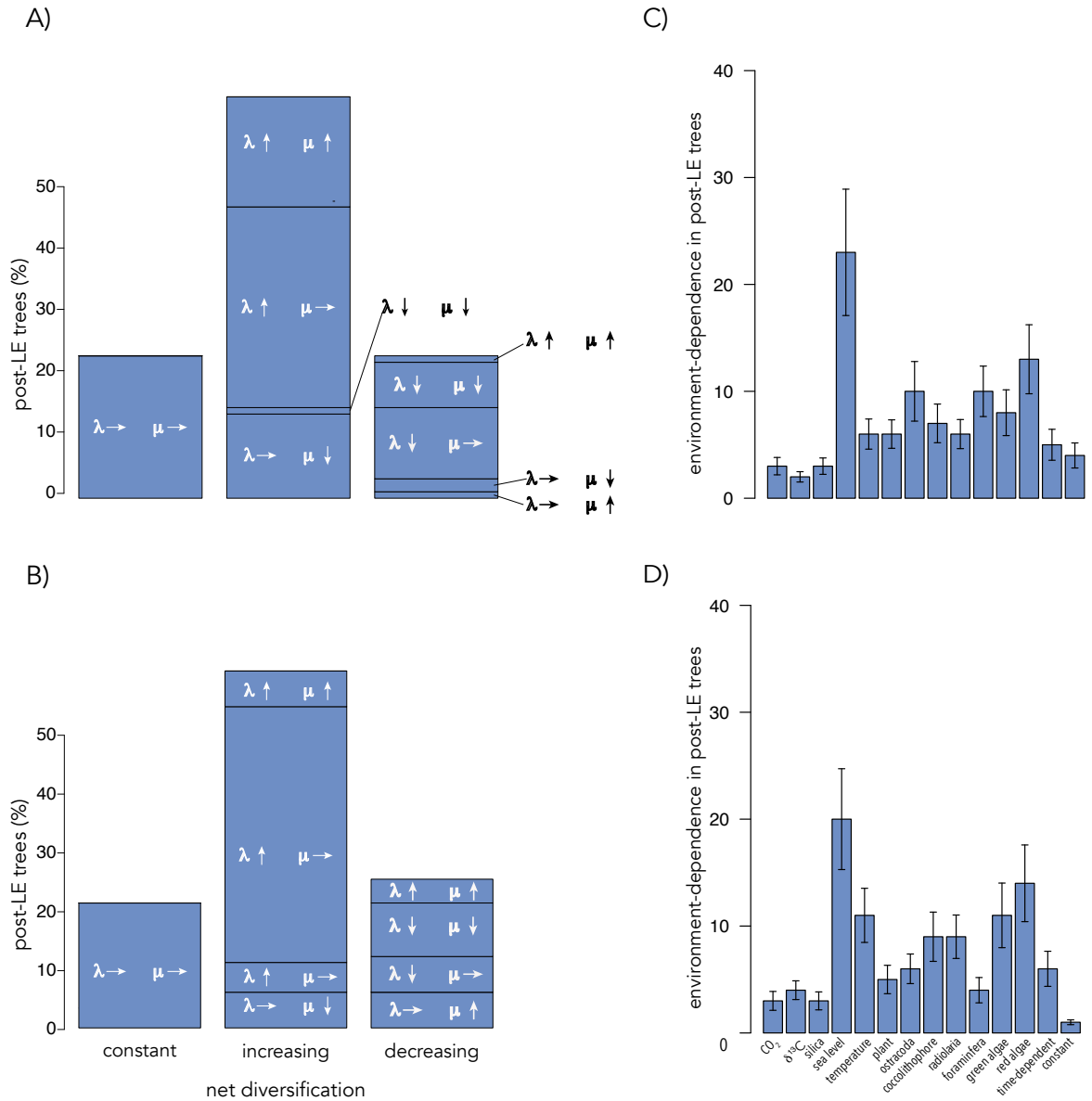
Supplemental Fig. 3: Extant diversity estimates for post-LE trees. (A) Density and (B) rank plot of actual (blue) and estimated extant diversity for post-LE trees using Bayesian fits to sampling abundances based on log-normal (black) and Sichel (gray) distributions. Errors bars in (B) represent 95% confidence intervals for each estimate.



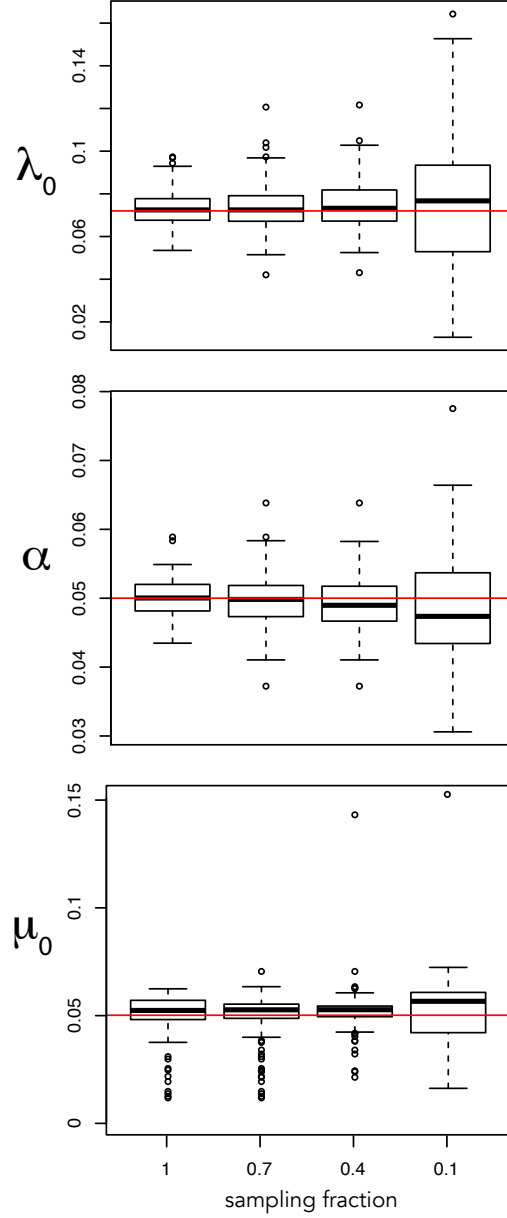
Supplemental Fig. 4: Plots of environmental curves used to estimate environment-dependent diversification. Each curve is scaled between 0 and 1. For CO_2 , $\delta^{13}C$, and temperature, different curves were used to analyse the pre- and post-LE trees. Each curve is shown in the range of time it was used to analyse.



Supplemental Fig. 5: Diatom diversification dynamics before the late Eocene with the lower (A-C) and upper (D-F) bounds of the 95% confidence interval estimates for extant diversity. (A,D) Time-dependent speciation, extinction, and net diversification rate for all four versions of the the pre-LE diatom OTU phylogeny. (B,E) Median Akaike weights for environment-dependent models fit to the pre-LE phylogenies. Error bars indicate the minimum and maximum support across all versions of the phylogeny. (C,F) Speciation, extinction, and net diversification rates as a function of RCO_2 (ratio of pCO_2 relative to the present) for each version of the phylogeny.



Supplemental Fig. 6: Diatom diversification dynamics from the late Eocene with the lower (A,C) and upper (B,D) bounds of the 95% confidence interval estimates for extant diversity. (A,B) Median percentage of post-LE trees supported by constant and time-dependent diversification models. Within each model, the speciation and extinction processes driving the net diversification are shown: up-arrow, increasing; down-arrow, decreasing; right-arrow, constant. (C,D) The median number of supported post-LE trees best fit by each environment-dependent, time-dependent, and constant model of diversification. Error bars in (C,D) indicate the minimum and maximum support across all versions of the phylogeny.



Supplemental Fig. 7: Parameter estimates for trees simulated under time-dependent birth-death processes. Trees were simulated at 50 million years, sliced at 15 million years from present. Parameter estimates recovered using an amended version of the RPANDA function `fit_bd` fit to trees jackknifed at 70%, 40%, and 10% of original diversity of the simulated tree. Red lines indicate the simulated parameter value.