

1 **Common species link global ecosystems to climate change**

2

3 Bjarte Hannisdal^{1*}, Kristian Agasøster Haaga¹, Trond Reitan², David Diego¹, Lee Hsiang

4 Liow^{2,3}

5 *¹Department of Earth Science, Centre for Geobiology, University of Bergen, PO Box 7803,*

6 *N-5020 Bergen, Norway.*

7 *²Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of*

8 *Oslo, PO Box 1066 Blindern, N-0316 Oslo, Norway.*

9 *³Natural History Museum, University of Oslo, PO Box 1172 Blindern, N-0318 Oslo, Norway*

10

11 *Corresponding author (bjarte.hannisdal@uib.no)

12 **Common species shape the world around us, and changes in their commonness signify**
13 **large-scale shifts in ecosystem structure and function¹⁻⁴. Dominant taxa drive**
14 **productivity and biogeochemical cycling, in direct interaction with abiotic components**
15 **of the Earth system^{3,4}. However, our understanding of the dynamic response of**
16 **ecosystems to global environmental changes in the past is limited by our ability to**
17 **robustly estimate fossil taxonomic richness^{5,6}, and by our neglect of the importance of**
18 **common species. To rectify this, we use observations of the most common and**
19 **widespread species to track global changes in their distribution in the deep geological**
20 **past. Our simple approach is robust to factors that bias richness estimators, including**
21 **widely used sampling-standardization methods⁵, which we show are highly sensitive to**
22 **variability in the species-abundance distribution. Causal analyses of common species**
23 **frequency in the deep-sea sedimentary record detect a lagged response in the ecological**
24 **prominence of planktonic foraminifera to oceanographic changes captured by deep-**
25 **ocean temperature records over the last 65 million years, encompassing one of Earth's**
26 **major climate transitions. Our results demonstrate that common species can act as**
27 **tracers of a past global ecosystem and its response to physical changes in Earth's**
28 **dynamic history.**

29 True species richness can be elusive even in well-studied ecosystems, because most
30 species are very rare, and relatively few species account for most of the total abundance^{1,2}.
31 For example, only ~1.4 % of the estimated tree species account for half of the biomass and
32 control the cycling of water, carbon, and nutrients in the Amazon forest³. Similarly, a recent
33 survey of eukaryotic diversity in the oceans found that ~0.24 % of the taxa accounted for half
34 of the total number of rDNA reads⁴.

35 Abundance and occupancy are typically positively correlated, with the more abundant
36 species being the more widespread^{4,7}. In the fossil record, species and higher taxa generally

37 have a humped temporal distribution of occurrences, being rare in the early and late stages of
38 their known stratigraphic range⁸⁻¹⁰.

39 Here we accommodate these ecological features by focussing only on species that are
40 common and widespread at any given time, using the Summed Common species Occurrence
41 Rate (SCOR), a very simple occurrence-based quantity that is sensitive to changes in total
42 abundance (Methods)¹¹. We apply SCOR to deep-sea sedimentary records of calcifying
43 plankton (coccolithophores and foraminifera) over the last 65 years to demonstrate how
44 relative changes in the distribution of common and widespread species were linked to climate
45 change on geological time scales.

46 First, we evaluate the sensitivity of SCOR and commonly used richness estimators to
47 potential biases using Poseidon, a simulation model of planktonic microfossil occurrences
48 (Fig 1; Methods and Supplementary Code). We target methods currently popular in
49 palaeobiology, and highlight the effects of two main factors: variability in the spatial
50 sampling completeness (Fig. 1b), and variability in the shape of the species rank-abundance
51 distribution (RAD; Fig. 1c).

52 Our simulations with Poseidon show that the SCOR estimate of relative changes in
53 total abundance is highly robust to variability in both spatial sampling and RAD shape (Fig.
54 1d). By definition, SCOR is immune to the loss of rare species, and decoupled from changes
55 in richness. As expected, the fidelity of raw sampled richness (S) decays rapidly with
56 increasing sampling variability, but shows little sensitivity to changes in the shape of the
57 RAD. Simple range-through richness (RT; assuming a species existed in all time bins
58 between its first and last occurrence) is relatively robust to both factors, indicating that the
59 level of sampling in Poseidon is sufficient to avoid severe edge effects. The Shannon entropy
60 H, which reflects both richness and evenness, is very sensitive to RAD shape variability,
61 ultimately tracking changes in evenness at the expense of changes in richness. Classical

62 rarefaction (CR) and shareholder quorum subsampling (SQS)⁵, being sampling-
63 standardization methods, are robust to the effect of spatial sampling variability on richness,
64 all else being equal. However, both CR and SQS are highly sensitive to changes in RAD
65 shape. As with Shannon H, increasing RAD variability causes CR and SQS to lose track of
66 richness and respond to changes in the shape parameter σ of the RAD instead. Note that the σ
67 values used in Poseidon generally correspond to high, moderately variable species evenness
68 (Supplementary Fig. 1). A third subsampling method, occurrences-squared weighted
69 (O2W)¹², shows overall poor agreement with true richness.

70 Turning to the rich deep-sea sedimentary record of the Cenozoic Era (0-65 million
71 years ago), we analysed global occurrences of the two most prominent groups of calcifying
72 plankton, coccolithophores and foraminifera, from the Neptune Sandbox Berlin (NSB)
73 database^{13,14} (Methods). In both groups, raw S generally increases along with the number of
74 boreholes representing the spatial sampling, while sampled evenness (J) decreases
75 (Supplementary Fig. 2a,b), as expected if improved sampling enhances the detection of rare
76 species (Fig. 1b,c). Sampling-standardized richness estimates (SQS) seem to remove the
77 sampling trend, but given the sensitivity of subsampling methods to RAD shape found in
78 Poseidon, we suspected an evenness signal in the SQS estimates. Indeed, SQS richness can
79 be reproduced by simply adding together the raw S and J curves (Supplementary Fig. 2c,d), a
80 relationship that emerges across NSB data and simulation runs (Fig. 2). This result implies
81 that changes in evenness are a major confounding factor for current sampling-standardized
82 richness estimators.

83 Focussing instead on common species, coccolithophores and foraminifera have
84 markedly different SCOR trajectories through the Cenozoic. On average, coccolithophores
85 have their highest SCOR values in the Eocene, followed by a decline in the Oligocene and a
86 resurgence in the late Miocene and Pliocene. Aspects of the coccolithophore SCOR pattern

87 have been linked to Cenozoic proxy records of atmospheric CO₂, suggesting that
88 coccolithophores could thrive in a high-CO₂ world^{11,15}. Since their rise in the Mesozoic,
89 coccolithophores shifted the dominant locus of carbonate burial from continental shelves to
90 the deep sea, providing a new mechanism for buffering ocean chemistry and atmospheric
91 CO₂ through carbonate compensation¹⁶. Oligocene cooling and CO₂ decline was
92 accompanied by a lowering of the carbonate compensation depth, which has been attributed
93 to changes in the supply of weathering products to the ocean¹⁷. The Oligocene reduction of
94 coccolithophore SCOR is opposite to that expected if SCOR were biased upward by
95 enhanced deep-sea preservation¹¹, and carbonate preservation trends cannot explain the
96 independent SCOR patterns in the two calcifying groups. Selective dissolution or taxonomic
97 preferences in sample processing may cause short-term volatility in SCOR, but only if
98 species presence or absence is random with respect to commonness (Supplementary Fig. 3).

99 Planktonic foraminifera SCOR was compared to Cenozoic deep-ocean temperature
100 (DOT) records¹⁸ (Methods; Supplementary Data Set). Although the net trends are inversely
101 related (foraminifera flourish as the world cools), shorter-term changes suggest positive co-
102 variation, including the Early Eocene climate optimum, Eocene cooling, as well as Miocene
103 and Pliocene optima (Fig. 3a). Geological proxy records are generally noisy mixtures of
104 signals representing multiple processes, derived from a sedimentary record that is itself an
105 active component of the Earth system. Any causal connection detected between proxy
106 records would necessarily be indirect with respect to the underlying processes of interest.
107 Nonetheless, the DOT record reflects a set of climate-related variables, including changes in
108 ocean thermohaline circulation, water mass structure, and nutrient dynamics, all considered
109 to be important abiotic controls on the long-term evolution of planktonic foraminifera^{6,19-21}.
110 Here we tested this drive-response hypothesis using three conceptually very different
111 methods for causal detection in time series (Methods): (1) Convergent cross mapping²², based

112 on the concept of state space reconstruction from time-delay embedding; (2) Information
113 transfer analysis^{23,24}, based on the concept of transfer entropy²⁵; and (3) Bayesian inference
114 of causal models based on linear stochastic differential equations^{26,27}.

115 Convergent cross mapping (CCM) from foraminifera SCOR to DOT peaks at a
116 negative lag, indicating that the SCOR signal carries a response to past changes captured in
117 the DOT record (Fig. 3b). The optimum lag is a single time bin, implying a causal delay of
118 0.5 million years (Myr) or less. CCM is also significant in the opposite direction but this is
119 stronger at positive lags (Fig. 3b), which are non-causal (future "drives" past). This result is
120 consistent with a unidirectional forcing where the dynamics of the response variable (SCOR)
121 is dominated by the driving variable (DOT), such that predictability flows both ways²⁸.
122 Information transfer (IT) analysis supports this inference: predictive information flow is
123 significant from past DOT to SCOR, although the optimal lag is shifted backward by one
124 time bin, implying a more protracted causal delay (Fig. 3c). In the opposite direction, IT
125 peaks at the corresponding positive (non-causal) lags, but is significantly weaker than in the
126 causal direction. Using a series of linear Stochastic Differential Equations (SDEs) to model
127 correlation and causality between the two records (Supplementary Fig. 4), we recover
128 relatively strong evidence that SCOR responds to changes in DOT, with a time lag of 0.33-
129 1.1 Myr, comparable to the CCM and IT analyses (Supplementary Tables 1, 2), although the
130 detailed nature of the causal relationship cannot be clearly resolved (Methods).

131 The congruence of these results strongly suggests that the ecological prominence of
132 planktonic foraminifera has evolved in response to past climatic and oceanographic changes
133 captured in the deep-ocean temperature proxy record. Furthermore, the inferred time delay
134 implies that the causal connection is highly indirect, involving climate changes propagating
135 through the Earth system to influence the commonness of foraminifera in the global plankton
136 on evolutionary time scales. In the modern global ocean, eukaryotic plankton richness

137 involves a vast number of parasite and symbiont species, highlighting the importance of
138 biotic interactions in driving diversification through trophic connectivity and complexity⁴.
139 Abiotic factors, such as differences in nutrient level among ocean basins, are more clearly
140 reflected in the relative abundance of the dominant species. A restructuring of water masses
141 and nutrient distributions is likely to cause a dramatic and discernible shift in the distribution
142 and abundance of many species, yet have a far less predictable impact on richness. Our
143 results imply that if such a fundamental ecosystem response were to leave a signature in the
144 fossil record, it would be far more evident in the robustly detectable distribution of the most
145 common species than in the indeterminate richness of rare species. Dominant groups also
146 reveal macroevolutionary trends in functional morphology otherwise obscured by rare taxa²⁹.
147 Given their critical importance to ecosystem functioning, common species provide a nexus
148 for understanding the role of an evolving biota in global environmental changes of the past.

149

150 **References**

- 151 1. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 254–283 (1948).
- 152 2. Connolly, S. R. *et al.* Commonness and rarity in the marine biosphere. *Proceedings of*
153 *the National Academy of Sciences* **111**, 8524–8529 (2014).
- 154 3. ter Steege, H. *et al.* Hyperdominance in the Amazonian Tree Flora. *Science* **342**,
155 1243092 (2013).
- 156 4. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**,
157 1261605 (2015).
- 158 5. Alroy, J. The Shifting Balance of Diversity Among Major Marine Animal Groups.
159 *Science* **329**, 1191–1194 (2010).
- 160 6. Ezard, T. H. G., Aze, T., Pearson, P. N. & Purvis, A. Interplay between changing
161 climate and species' ecology drives macroevolutionary dynamics. *Science* **332**, 349–

- 162 351 (2011).
- 163 7. Gaston, K. J. *et al.* Abundance–occupancy relationships. *Journal of Applied Ecology*
164 **37**, 39–59 (2000).
- 165 8. Foote, M. Symmetric waxing and waning of marine invertebrate genera. *Paleobiology*
166 **33**, 517–529 (2007).
- 167 9. Liow, L. H. & Stenseth, N. C. The rise and fall of species: implications for
168 macroevolutionary and macroecological studies. *Proceedings of the Royal Society B:*
169 *Biological Sciences* **274**, 2745–2752 (2007).
- 170 10. Liow, L. H., Skaug, H. J., Ergon, T. & Schweder, T. Global occurrence trajectories of
171 microfossils: environmental volatility and the rise and fall of individual species.
172 *Paleobiology* **36**, 224–252 (2010).
- 173 11. Hannisdal, B., Henderiks, J. & Liow, L. H. Long-term evolutionary and ecological
174 responses of calcifying phytoplankton to changes in atmospheric CO₂. *Glob Change*
175 *Biol* **18**, 3504–3516 (2012).
- 176 12. Alroy, J. New methods for quantifying macroevolutionary patterns and processes.
177 *Paleobiology* **26**, 707–733 (2000).
- 178 13. Spencer-Cervato, C. The Cenozoic deep sea microfossil record: explorations of the
179 DSDP/ODP sample set using the Neptune database. *Palaeontologia Electronica* **42**, 1–
180 268 (1999).
- 181 14. Lazarus, D. Neptune: A marine micropaleontology database. *Math Geol* **26**, 817–832
182 (1994).
- 183 15. Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C. E., Balch, W. M. & Guikema, S.
184 D. Multidecadal increase in North Atlantic coccolithophores and the potential role of
185 rising CO₂. *Science* **350**, 1533–1537 (2015).
- 186 16. Ridgwell, A. & Zeebe, R. E. The role of the global carbonate cycle in the regulation

- 187 and evolution of the Earth system. *Earth and Planetary Science Letters* **234**, 299–315
188 (2005).
- 189 17. Pälike, H. *et al.* A Cenozoic record of the equatorial Pacific carbonate compensation
190 depth. *Nature* **488**, 609–614 (2012).
- 191 18. Cramer, B. S., Miller, K. G., Barrett, P. J. & Wright, J. D. Late Cretaceous–Neogene
192 trends in deep ocean temperature and continental ice volume: Reconciling records of
193 benthic foraminiferal geochemistry ($\delta^{18}\text{O}$ and Mg/Ca) with sea level history. *J.*
194 *Geophys. Res.* **116**, C12023 (2011).
- 195 19. Cifelli, R. Radiation of Cenozoic Planktonic Foraminifera. *Syst Biol* **18**, 154–168
196 (1969).
- 197 20. Schmidt, D. N., Thierstein, H. R., Bollmann, J. & Schiebel, R. Abiotic forcing of
198 plankton evolution in the Cenozoic. *Science* **303**, 207–210 (2004).
- 199 21. Peters, S. E., Kelly, D. C. & Fraass, A. J. Oceanographic controls on the diversity and
200 extinction of planktonic foraminifera. *Nature* **493**, 398–401 (2013).
- 201 22. Sugihara, G. *et al.* Detecting Causality in Complex Ecosystems. *Science* **338**, 496–500
202 (2012).
- 203 23. Verdes, P. Assessing causality from multivariate time series. *Phys. Rev. E* **72**, 026222
204 (2005).
- 205 24. Hannisdal, B. Non-parametric inference of causal interactions from geological records.
206 *American Journal of Science* **311**, 315–334 (2011).
- 207 25. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
- 208 26. Reitan, T., Schweder, T. & Henderiks, J. Phenotypic evolution studied by layered
209 stochastic differential equations. *The Annals of Applied Statistics* **6**, 1531–1551 (2012).
- 210 27. Liow, L. H., Reitan, T. & Harnik, P. G. Ecological interactions on macroevolutionary
211 time scales: clams and brachiopods are more than ships that pass in the night. *Ecol*

212 *Letters* **18**, 1030–1039 (2015).

213 28. Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delayed
214 causal interactions using convergent cross mapping. *Sci. Rep.* **5**, 14750 (2015).

215 29. Jernvall, J. & Fortelius, M. Common mammals drive the evolutionary increase of
216 hypsodonty in the Neogene. *Nature* **417**, 538–540 (2002).

217

218 **Acknowledgments** We thank the contributors to the Neptune Sandbox Berlin database.

219 Thanks to J. Alroy for sharing his SQS 4.3 script. This project used DSDP/ODP samples
220 provided by the Integrated Ocean Drilling Program (IODP). IODP is sponsored by the U.S.

221 National Science Foundation (NSF) and participating countries under the management of

222 Joint Oceanographic Institutions (JOI), Inc. The authors were funded by the Norwegian

223 Research Council (231259 to B.H. and 235073 to L.H.L.) and the Bergen Research

224 Foundation (B.H. and K.A.H.).

225

226 **Author Contributions** B.H. and L.H.L. designed the study. B.H. ran the Poseidon

227 simulations, analysed data, performed the IT analysis, and wrote the paper. L.H.L. retrieved

228 the NSB data, calculated SCOR, and wrote the Poseidon code. K.A.H. and D.D. performed

229 the CCM analysis. T.R. performed the SDE analysis. All authors discussed the results and

230 commented on the manuscript.

231

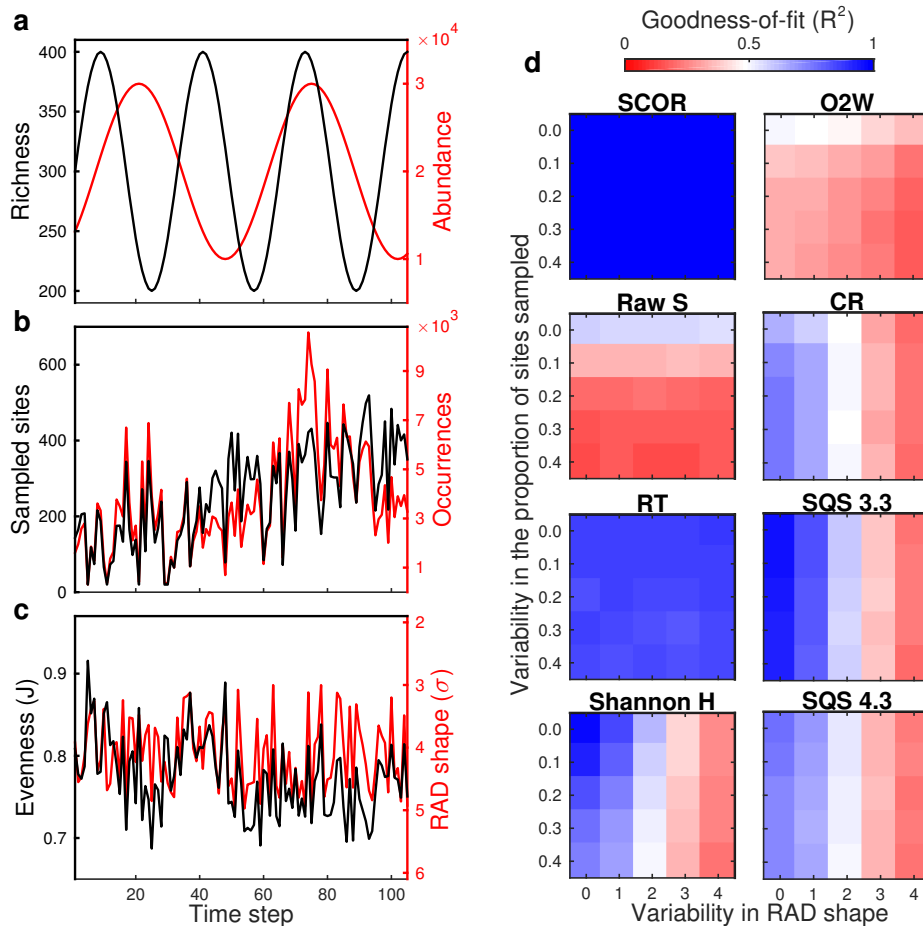
232 **Additional Information** Supplementary information is available in the online version of the

233 paper. Reprints and permissions information is available at www.nature.com/reprints.

234 Correspondence and requests for materials should be addressed to B.H.

235 (bjarte.hannisdal@uib.no).

236



237

238

239 **Figure 1 | Performance of SCOR and richness estimators in Poseidon model**

240 **experiments. a,** Simulated species richness and total abundance are decoupled. **b,** Sampled

241 species occurrences reflect abundance distorted by the trend and short-term variability in

242 sampled sites (in this example, variability = 0.1, corresponding to the standard deviation

243 around the mean trend). **c,** Sampled species evenness (Pielou's J) captures changes in the

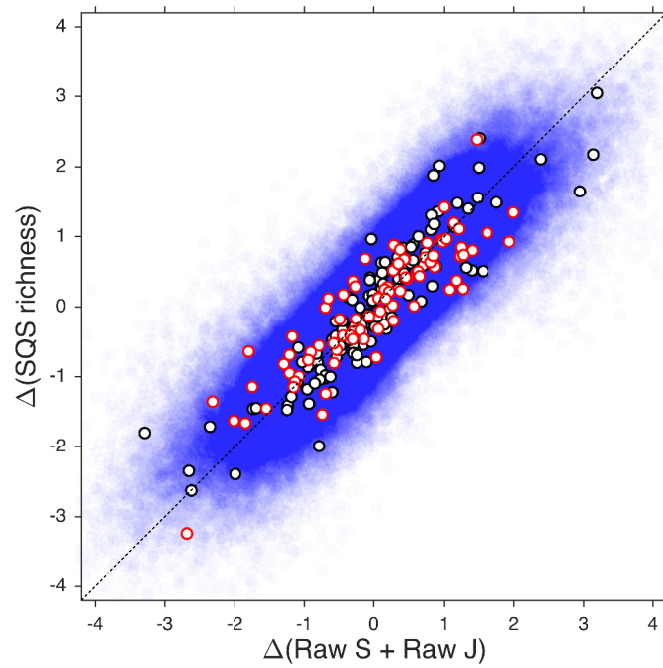
244 shape parameter σ of the RAD (in this example, variability = 2, corresponding to the range of

245 σ), superimposed on richness fluctuations and a net decrease caused by the sampling trend. **d,**

246 Sensitivity to sampling variability and RAD shape variability. Values are median goodness-

247 of-fit (R^2) of 50 model runs, comparing SCOR to true abundance, and richness estimates to

248 true richness. See text for abbreviations.



249

250 **Figure 2 | Empirical relationship between sampling-standardized richness and the sum**

251 **of raw richness and evenness.** Values are first differences of normalized time series of SQS

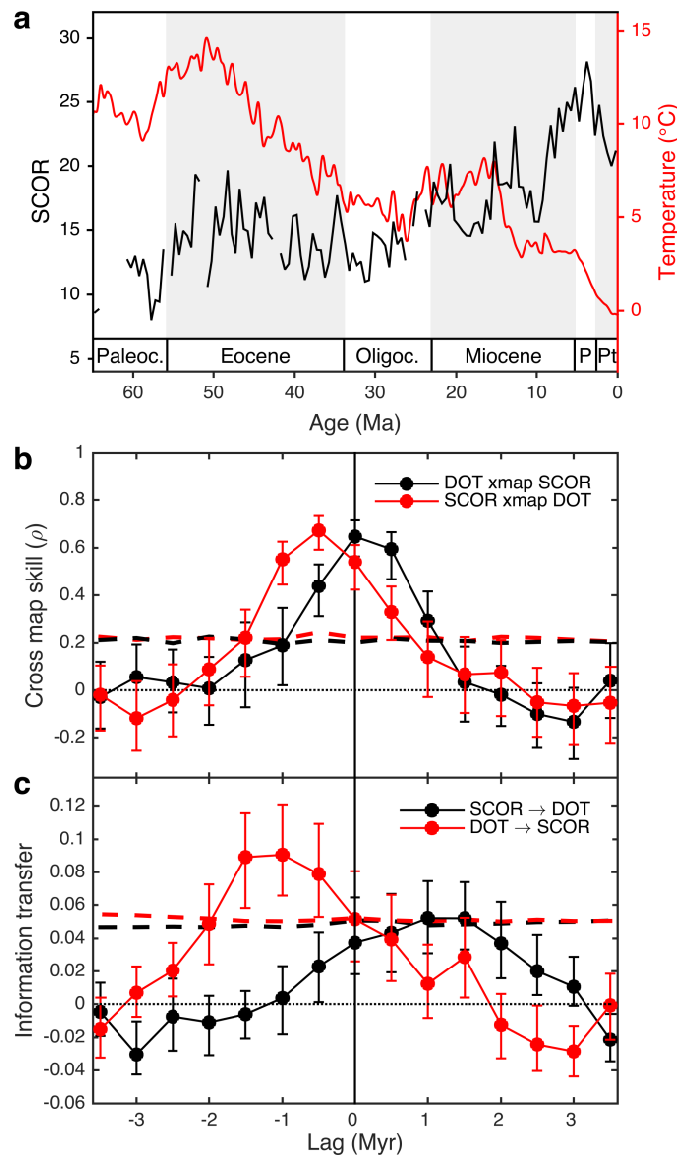
252 richness and of the sum of normalized raw richness (S) and evenness (Pielou's J). Data

253 include Cenozoic coccolithophores (black) and planktonic foraminifera (red) from the NSB

254 database (Supplementary Fig. 2), and all Poseidon model experiments (blue; $N = 262,500$).

255 Stippled line marks the 1:1 relationship.

256



257

258 **Figure 3 | Testing a causal link between planktonic foraminifera SCOR and Cenozoic**

259 **climate changes. a**, SCOR of planktonic foraminifera from the NSB database at 0.5 Myr

260 resolution, and DOT estimates¹⁸ at 0.1 Myr resolution. **b**, **c**, CCM skill (**b**) and IT (**c**) between

261 SCOR and DOT as a function of time lag. If past DOT drives SCOR, then SCOR *xmap* DOT,

262 while information flows DOT \rightarrow SCOR, at negative lags. Values are medians (dots) and 95

263 % ranges (whiskers) for 500 random subsamples of length 100, dashed lines are 95th

264 percentiles of 1,000 surrogates. All values are normalized to a surrogate mean of zero. Ma,

265 million years before present; Paleoc., Paleocene; Oligoc., Oligocene; P, Pliocene; Pt,

266 Pleistocene.

267 **METHODS**

268 **Data.** Microfossil occurrences were retrieved from the NSB database^{13,14} (accessed April 22,
269 2015). SCOR and richness estimates were calculated using 0.5 Myr time bins. For the
270 planktonic foraminifera, we compared the inferred times of species rise and fall in NSB¹⁰,
271 which encompass any period of potential commonness, to the species ranges in the
272 PlankRange database³⁰ (<http://palaeo.gly.bris.ac.uk/Data/plankrange.html>, accessed Aug. 24,
273 2014). After resolving most of the taxonomic discrepancies, ~82 % of the species have a rise-
274 fall interval that fits within the proposed range or is offset by < 2 Myr (a single time bin in
275 ref. 10). Of the remaining species, which could either not be matched taxonomically or have
276 a significant range offset, ~73 % are rare (peak occurrence frequency, as a proportion of all
277 sites with at least one species sampled, < 0.2), and none have a peak occurrence frequency >
278 0.4. Taxonomic or range errors in NSB are therefore unlikely to have a significant impact on
279 SCOR, which is sensitive only to the most widespread species. Quantification of this impact
280 awaits the public release of the updated PlankRange database³¹.

281 DOT estimates were obtained from ref. 18, using their > 9 Ma T_{d-SL} record (based on
282 subtracting New Jersey sea level records from a benthic $\delta^{18}O$ stack) and scaled $\delta^{18}O$ record
283 for the interval < 9 Ma, with their interpolation at 0.1 Myr resolution. Both SCOR and DOT
284 were tied to the GTS2004 time scale³².

285 **SCOR.** We treat the observation of a specific number of individuals as a Poisson-distributed
286 variable with parameter λ in each time bin. The probability of finding an individual of species
287 i in time bin j is then $p_{ij} = 1 - \exp(-\lambda_{ij})$, and thus $\lambda_{ij} = -\ln(1 - p_{ij})$. In practice, p_{ij} is estimated
288 as y_{ij}/n_j , where y_{ij} is the number of sites in which species i is observed at time bin j and n_j is
289 the number of sites in that time bin where at least one of the species included in the analysis
290 is observed. SCOR is the total density of a given set of m_j species in a time bin:

291
$$SCOR_j = \sum_{i=1}^{m_j} \lambda_{ij} \quad (1)$$

292 and we estimate its variance by the delta method³³:

293
$$Var(SCOR_j) = \sum_{i=1}^{m_j} \frac{p_{ij}}{(1 - p_{ij})n_j}. \quad (2)$$

294 SCOR is based on the observation that the more globally abundant a species is, the more
295 likely it is to occur at a greater number of sites⁴. As p approaches 1, the rate of increase in λ
296 grows rapidly, such that very widespread species have a much greater influence on SCOR
297 than restricted species. If a species occurs at all sites in a time bin, its λ for that time bin is
298 undefined.

299 SCOR is decoupled from species richness and relative abundance. If a species
300 becomes globally more abundant and widespread in a time interval, then its λ , and thus
301 SCOR, will increase. Even if all species became exactly equally more common in absolute
302 terms, with no change in relative abundance, their individual λ values will be higher and
303 SCOR will capture the proportional change in absolute abundance of the total species set.

304 **Poseidon simulations.** To evaluate the performance of SCOR relative to commonly used
305 diversity metrics, we designed a set of numerical experiments on the effects of temporal
306 variability in three factors: (1) spatial sampling completeness; (2) the shape of the species
307 rank-abundance distribution (RAD); and (3) the proportion of species lost randomly (with
308 respect to abundance).

309 Poseidon consists of 1,000 spatial grid cells and 105 time steps, where true species
310 richness and abundance are allowed to vary independently (Fig. 1a). In each time step, we
311 randomly assign an abundance value to each species, such that the entire community has a
312 log-normal rank-abundance distribution (RAD), the shape of which can be fixed or time-
313 varying. Species ranks are randomly reshuffled between time steps. We then randomly assign

314 a spatial grid cell (site) to an individual of a species, where it can potentially be preserved and
315 sampled.

316 Next, we sample only a proportion of the sites (spatial cells) such that this proportion
317 increases linearly from 0.1 to 0.4, representing a declining sampling coverage with age,
318 typical of deep-sea sedimentary records. Any short-term variability is thus superimposed on
319 this trend (e.g. Fig. 1b). Furthermore, a proportion of the remaining species can be randomly
320 removed (representing dissolution, selective picking, or other processes causing a species to
321 be absent in a time bin, regardless of its original abundance).

322 We then calculate raw S, RT, Shannon H, and three sampling-standardized richness
323 metrics widely used in palaeobiology (CR, O2W, and SQS). Although a generalized OXW
324 has been recommended for paleontological datasets³⁴, we use O2W here because our data
325 meet the assumptions of the latter¹². We used two versions of the shareholder quorum
326 subsampling (SQS) method^{5,35-37}: the SQS3.3 R script
327 (<http://bio.mq.edu.au/~jalroy/SQS.html>, downloaded Aug. 26, 2014), and the SQS4.3 perl
328 script, kindly provided by J. Alroy (Sept. 3, 2014). We did not modify the SQS codes but
329 wrote a function to format Poseidon output to species occurrence data for SQS4.3. For
330 SQS3.3, we input the list of sampled individuals (abundances). A quorum level of 0.6 was
331 used in all runs. Both SQS versions yielded very similar results when given the same type of
332 data (abundances or occurrences). For all subsampling methods (CR, O2W, and SQS), 100
333 iterations were used in obtaining each estimated richness value. Increasing the number of
334 iterations offered no discernible improvement. Shannon H was output from the SQS 3.3 R
335 script and used to calculate Pielou's J evenness. The goodness-of-fit between true and
336 estimated richness, and between true abundance and SCOR, was assessed by the coefficient
337 of determination (R^2) between time series. Poseidon R scripts are provided as Supplementary
338 Code.

339 **Time Series Analysis.** We used three different time series analysis methods to test for a
340 causal relationship between SCOR and DOT. Two of the methods are non-parametric, while
341 the third is based on linear models. To implement a time-displacement (lag) analysis for the
342 non-parametric methods, missing SCOR values were linearly interpolated, and the 0.1-Myr
343 resolution DOT record was bin-averaged on the SCOR time bins (0.5 Myr). Furthermore, the
344 non-parametric methods use surrogate time series to assess significance, which requires
345 detrending of the original time series. To avoid any bias from differences in non-stationarity
346 that are not reproduced by the surrogates, both records were detrended using a third-order
347 polynomial to satisfy a stationarity criterion³⁸, then normalized to zero mean and unit
348 standard deviation. For consistency, the model-based analysis was also performed on the pre-
349 processed data. However, the detrending may remove components of the variation relevant to
350 uncovering the parameters of underlying processes. We therefore repeated the model-based
351 analysis on untransformed data.

352 **CCM analysis.** CCM is a method for causal inference in nonlinear dynamical systems based
353 on the theory of state-space reconstruction²². Consider two time series X and Y consisting of
354 scalar observations of variables in a dynamical system. According to Takens's theorem³⁹, we
355 can construct a delay-coordinate embedding of the state space of the dynamical system into
356 an m -dimensional real space, by constructing the vectors $E_X = \{ (x(t_i), x(t_i - \tau), x(t_i - 2\tau), \dots,$
357 $x(t_i - (m\tau))) \}$, where $x(t_i)$ is the scalar value of the time series X at time t_i . That is, the vectors
358 in E_X are in one-to-one correspondence with the states of the dynamical system. If X and Y
359 are coupled variables of the same dynamical system (i.e. they are causally influencing each
360 other), this correspondence is also true for time series Y , and therefore E_X and E_Y are in one-
361 to-one correspondence with each other. CCM uses this result to predict scalar values of Y
362 from the coordinate-delay embedding of X and vice versa.

363 The CCM algorithm locates, for each scalar point P_i in the prediction set (subset of
364 time series Y), the contemporaneous state vector L_i in the library set (subset of state vectors in
365 the time-delay embedding constructed from time series X). Next, it finds L_i 's closest
366 neighbours and estimates a value for the predictee P_{i*} using simplex projection⁴⁰. CCM skill
367 is determined by the correlation (Pearson's ρ) between estimated P_{i*} and actual values of P_i .
368 With increasing library size, CCM skill is expected to increase and converge if the variables
369 are causally related. The notation " X xmap Y " refers to estimating $y(t_i)$ from corresponding
370 lagged-coordinate state versions of $x(t_i)$, which in a causal context is read as " Y is causally
371 influencing X ".

372 CCM analysis was performed using the rEDM software package⁴¹. We constructed
373 time-delay embeddings using embedding dimension $m = 2$ and delay time step $\tau = 1$. Cross
374 mapping was then performed using the entire time series as both prediction and library sets.
375 To avoid biased results, we used leave-one-out cross validation (i.e. the predictee P_i itself and
376 points in a time radius of E around P_i were excluded from the libraries, such that no points
377 sharing coordinates with P_i were used in the predictions; see refs. 41, 42).

378 If unidirectional forcing is sufficiently strong, the dynamics of the response variable
379 can become dominated by the driving variable. In this case, CCM may be significant in both
380 directions, and thus unable to distinguish unidirectional forcing from bidirectional causality.
381 To address this, we used the extended CCM approach²⁸, which repeats the cross mapping
382 using different time-displacements of the original time series. If there is a discernable lag
383 between cause and effect, then optimal cross map skill is expected to occur for negative time
384 lags in the direction(s) of true causality (past drives future). If true causality is unidirectional,
385 then any CCM skill in the non-causal direction is expected to peak for positive lags (future
386 "drives" past).

387 Extended CCM analysis of SCOR and DOT is reported as median cross map skill and
388 95 % ranges at different lags after drawing 500 samples with replacement from libraries of
389 size 100. Statistical significance is evaluated against a null distribution of CCM results for
390 1000 surrogate time series. For each lag, CCM skill is considered significant if it exceeds the
391 95th percentile of the surrogates. We verified the results using three different methods for
392 generating surrogate data: phase-randomized and amplitude-adjusted Fourier transform
393 (AAFT)⁴³, phase-randomized Fourier transform⁴⁴, and randomly shuffled surrogates. All
394 three methods indicate significant causality from DOT to SCOR, and we limit our results to
395 the AAFT method, which gave the most conservative significance estimates.

396 **IT analysis.** If two processes X and Y are independent, then a general Markov property will
397 hold²⁵:

$$398 \quad p(x_{n+1} | x_n^{(k)}, y_n^{(l)}) = p(x_{n+1} | x_n^{(k)}), \quad (3)$$

399 where $p(x_{n+1})$ is the transition probability to state $n+1$, and indices k and l are the dimensions
400 of vectors of past states. In the absence of information flow from Y to X , knowing the past l
401 states of Y has no influence on the transition probability of X beyond knowing the past k
402 states of X alone. Transfer entropy²⁵ quantifies the incorrectness of assuming independence
403 by means of a Kullback-Leibler divergence, a non-symmetric measure of the information lost
404 when the right hand side is used to approximate the left hand side of equation (3):

$$T_{Y \rightarrow X} = \sum_{x,y} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log \frac{p(x_{n+1} | x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} | x_n^{(k)})} \quad (4)$$

405 Transfer entropy is thus a non-symmetric measure of information flow, which has been
406 shown to be equivalent to a conditional mutual information⁴⁵, and equivalent to Granger
407 causality⁴⁶ for linear, Gaussian systems⁴⁷. We implement it in the modified IT form proposed
408 by Verdes²³, which has previously been applied to the analysis of geological records^{11,24,48,49}.

409 Here we expand on these earlier applications by repeating the IT analysis for different
410 time-displacements of the original time series, analogous to the extended CCM analysis
411 described above. Similar to CCM, predictive information flow may become symmetric if
412 unidirectional forcing and/or linear correlation is sufficiently strong. However, if there is a
413 discernable lag between cause and effect, then optimal information transfer is expected to
414 occur for negative time lags in the direction(s) of true causality (past \rightarrow future). If true
415 causality is unidirectional, then any information flow in the non-causal direction is expected
416 to peak for positive lags (future \rightarrow past). IT is a coarse-grained relative entropy measure,
417 which varies as a function of the data gridding resolution, summarized in a single IT value as
418 the area under the resulting curve²³. Lagged IT analysis of SCOR and DOT is reported as
419 median IT and 95 % ranges at different lags after drawing 500 random subsamples of size
420 100. IT is considered significant if it exceeds the 95th percentile of a null distribution of IT
421 results for 1,000 AAFT surrogate time series. Unlike CCM, this IT implementation does not
422 use time-delay embedding. Combined with coarse-graining of the data, this may help explain
423 the difference in optimal lag between IT and CCM (Fig. 3b, c), although more work is needed
424 to clarify this.

425 **Linear SDE analysis.** Given two time series representing two measured processes, linear
426 SDEs can be used to distinguish between correlation and Granger causality. Uni- and
427 bidirectional causation as well as hidden (unmeasured) processes can be modelled in the SDE
428 framework, expanding the space of possible connections^{26,27}. A basic linear SDE can be
429 written as

$$430 \quad dX = -\alpha_X (X - \mu_X) dt + \sigma_X dB_t^{(X)} \quad (5)$$

431 This describes a mean-reverting Ornstein-Uhlenbeck process (OUP) X , which contains a
432 systematic part (the dt term) and a stochastic part (the dB term). If the systematic part is
433 dropped ($\alpha_X = 0$), then equation (5) describes a Wiener process (WP, or random walk). The

434 OUP has expectation μ_X , stationary standard deviation $s_X = \sigma_X / \sqrt{2\alpha_X}$ and half-life

435 $t_{1/2,X} = \log(2) / \alpha_X$. To model a hidden process, we can write

$$\begin{aligned}
 436 \quad dY_1 &= -\alpha_{Y_1}(Y_1 - Y_2)dt + \sigma_{Y_1}dB_t^{(Y_1)} \\
 dY_2 &= -\alpha_{Y_2}(Y_2 - \mu_Y)dt + \sigma_{Y_2}dB_t^{(Y_2)}
 \end{aligned} \tag{6}$$

437 Here, the measured process Y_1 has a hidden process (or layer) Y_2 folded into its systematic
 438 part, such that Y_1 tracks Y_2 . Y_1 is similar to an OUP, but instead of fluctuating around a fixed
 439 expected value it fluctuates with a lagged response to the OUP Y_2 .

440 When modelling connections between processes, we use vector notation. A pure
 441 correlation between X and Y entails that the covariance matrix in front of the stochastic term
 442 $d\mathbf{B}$ will have off-diagonal elements. If there is a causal connection from Y_2 to X , for
 443 instance, the system takes the following form

444
 445

$$\begin{aligned}
 dX &= -\alpha_X(X - \mu_X - \beta_{Y_2 \rightarrow X}(Y_2 - \mu_X))dt + \sigma_X dB_t^{(X)} \\
 446 \quad dY_1 &= -\alpha_{Y_1}(Y_1 - Y_2)dt + \sigma_{Y_1} dB_t^{(Y_1)} \\
 dY_2 &= -\alpha_{Y_2}(Y_2 - \mu_Y)dt + \sigma_{Y_2} dB_t^{(Y_2)}
 \end{aligned} \tag{7}$$

447 where $\beta_{Y_2 \rightarrow X}$ describes the connection strength from Y_2 to X . Equation (7) describes a
 448 “common cause” situation, where Y_2 drives both X and Y_1 .

449 To analyse the SCOR and DOT records, we first characterized each time series
 450 separately, examining models with up to three layers (two hidden). In each model, the layers
 451 could be WP or OUP (including fully deterministic layers where $\sigma_i = 0$), excluding a one-
 452 layered WP, which prohibits incoming causal links. We also excluded internal feedback
 453 loops in multi-layer models because of numerical intractability. For both time series, μ was
 454 assigned a prior distribution $\mu_i \sim N(0,1)$, where i denotes the layer. All model parameters

455 were assigned normal priors, with 95 % prior probability ranges of $\sigma_i \in (0.01, 1.0)$ for the
456 stochastic term, $t_{1/2,i} \in (0.1My, 50My)$ for the half life, $\beta \in (-2, 2)$ for the causal connections,
457 and $\rho \in (-0.96, 0.96)$, for the logit-transformed correlation coefficients. We used MCMC
458 importance sampling to estimate Bayesian model likelihoods and calculate model
459 probabilities.

460 The best model for SCOR in isolation was a one-layered OUP, while a three-layered
461 model with a WP as the bottom driver was the best model for DOT in isolation. We then
462 investigated all 15 connection models between these two best models, including the null
463 hypothesis of no relationship (Supplementary Fig. 4). We allowed for causality from SCOR
464 to DOT because both proxy records ultimately derive from deep-sea carbonate sediments,
465 hence SCOR could in principle contain a signal of processes that have influenced DOT. The
466 null hypothesis was assigned 50 % prior probability, while 50 % was distributed evenly
467 among the 14 connection models. For model comparison, we used Jeffreys's scale to assess
468 the strength of evidence represented by the Bayes factor B , where $1 < B < \sqrt{10}$ is evidence
469 "barely worth mentioning", $\sqrt{10} < B < 10$ is "substantial evidence", $10 < B < 10^{3/2}$ is
470 "strong" evidence and $B > 10^{3/2}$ is "very strong" evidence⁵⁰.

471 The posterior probability of the null hypothesis was 11.2 % (Model 1; Supplementary
472 Fig. 4), hence the Bayes factor favouring a connection between SCOR and DOT is 7.9
473 (substantial evidence). The most probable model (Model 5; Supplementary Fig. 4) involves a
474 feedback loop, where the upper DOT process (DOT1) affects SCOR positively while SCOR
475 affects DOT1 negatively. All parameter estimates with credible intervals for the best model
476 are presented in Supplementary Table 1. In the second most (and almost equally) probable
477 model, (Model 12; Supplementary Fig. 4), SCOR affects the second DOT layer (DOT2)
478 instead.

479 With a half-life of ~ 0.5 Myr, SCOR responds to DOT processes on time scales
480 comparable to those inferred from the other analyses (Fig. 3b, c). In contrast, DOT processes
481 react very slowly to changes in SCOR (Supplementary Table 1), and because SCOR changes
482 rapidly, the response in DOT will be smoothed out. From equation 5, the effect of SCOR on
483 DOT1 is 0.07 Myr^{-1} , while the effect of DOT1 on the SCOR process is 0.6 Myr^{-1} . Thus, DOT
484 influences SCOR much more strongly per time unit than vice versa.

485 The third most likely model (Model 4; Supplementary Fig. 4) only had a causal
486 connection from DOT1 to SCOR, consistent with CCM and IT inferences. This model had a
487 posterior probability of 17.5 %, hence the evidence for a feedback loop is less than
488 “substantial” ($B = 2.6$). In summary, we find evidence for there being at least one connection
489 ($B = 7.9$); for the connections to be causal rather than correlative given that there are
490 connections ($B = 12.5$); and specifically for a causal connection from DOT to SCOR given
491 that there are connections ($B = 15.4$).

492 We then repeated the analysis on untransformed data (not detrended or normalized),
493 denoted uDOT and uSCOR. In this case, the best isolated model for both time series is a
494 three-layer model with a WP at the bottom. The Bayes factor favouring a connection over no
495 connection is 73, which is deemed “very strong evidence”. The best connection model
496 involves a feedback loop between the top layers uDOT1 and uSCOR1. However, there is a
497 very high probability for parameters enforcing cyclical behaviour, with a period of 1.5 Myr,
498 which is consistent with an internal feedback loop model as the best isolated model for
499 uDOT. Parameter estimates are shown in Supplementary Table 2.

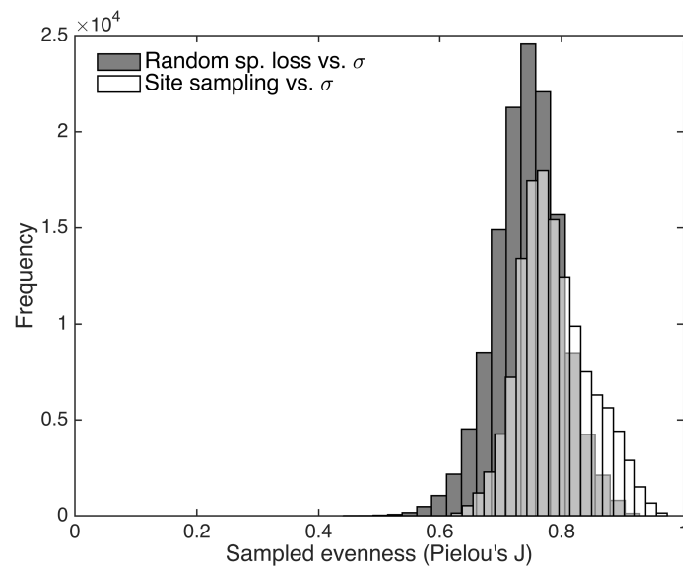
500

501 30. Stewart, D. & Pearson, P. N. PLANKRANGE: a database of planktonic foraminiferal
502 ranges. (2000). at <http://palaeo.gly.bris.ac.uk/Data/plankrange.html>

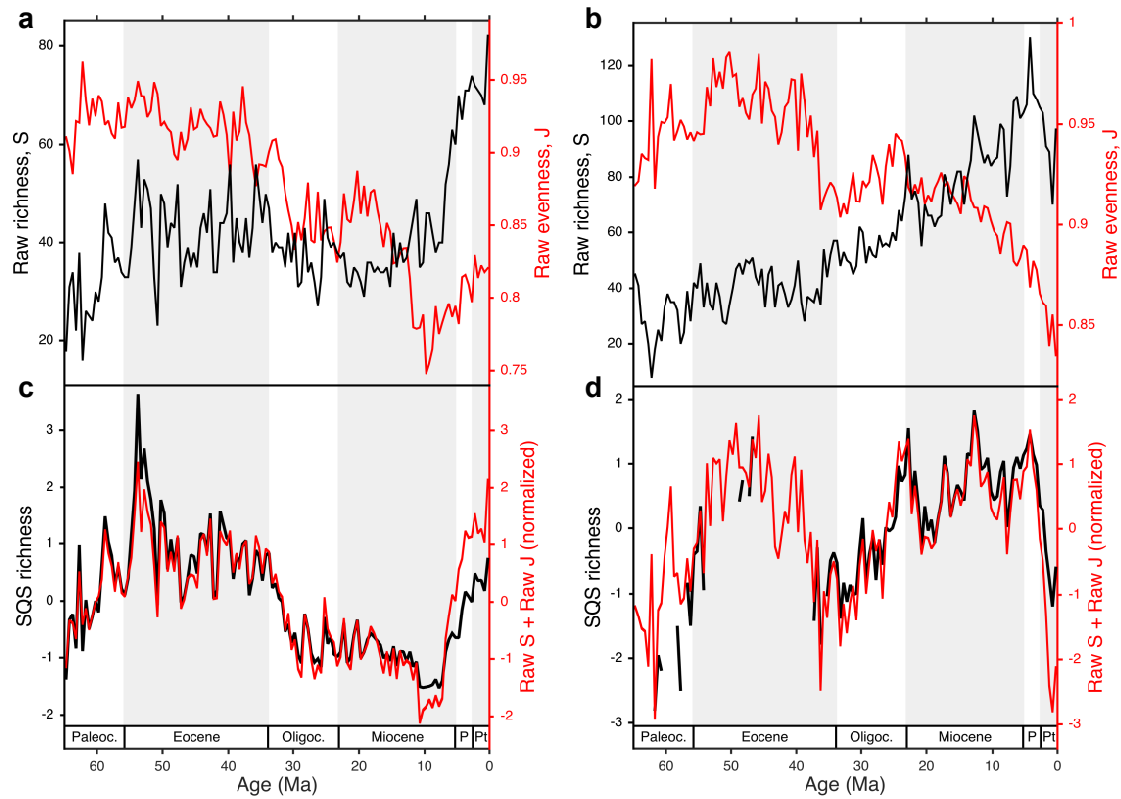
503 31. Aze, T. *et al.* A phylogeny of Cenozoic macroperforate planktonic foraminifera from

- 504 fossil data. *Biological Reviews* **86**, 900–927 (2011).
- 505 32. Gradstein, F. M., Ogg, J. G. & Smith, A. G. (eds.) *A Geologic Time Scale 2004*.
506 (Cambridge University Press 2004).
- 507 33. Casella, G. & Berger, R. L. *Statistical inference*. (Duxbury, 2002).
- 508 34. Bush, A. M., Markey, M. J. & Marshall, C. R. Removing bias from diversity curves:
509 the effects of spatially organized biodiversity on sampling-standardization.
510 *Paleobiology* **30**, 666–686 (2004)
- 511 35. Alroy, J. Accurate and precise estimates of origination and extinction rates.
512 *Paleobiology* **40**, 374–397 (2014).
- 513 36. Alroy, J. Fair sampling of taxonomic richness and unbiased estimation of origination
514 and extinction rates. *The Paleontological Society Papers* **16**, 55–80 (2010).
- 515 37. Alroy, J. Geographical, environmental and intrinsic biotic controls on Phanerozoic
516 marine diversification. *Palaeontology* **53**, 1211–1235 (2010).
- 517 38. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. Testing the null hypothesis
518 of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**, 159–
519 178 (1992).
- 520 39. Takens, F. in *Dynamical Systems and Turbulence* (eds. Rand, D. A. & Young, L. S.)
521 **898**, 366–381 (Springer, 1981).
- 522 40. Sugihara, G. & May, R. M. Nonlinear forecasting as a way of distinguishing chaos
523 from measurement error in time series. *Nature* **344**, 734–741 (1990).
- 524 41. Ye, H. *et al.* Reply to Luo *et al.*: Robustness of causal effects of galactic cosmic rays
525 on interannual variation in global temperature. *Proceedings of the National Academy*
526 *of Sciences of the United States of America* **112**, E4640–E4641 (2015).
- 527 42. Luo, M., Kantz, H., Lau, N.-C., Huang, W. & Zhou, Y. Questionable dynamical
528 evidence for causality between galactic cosmic rays and interannual variation in global

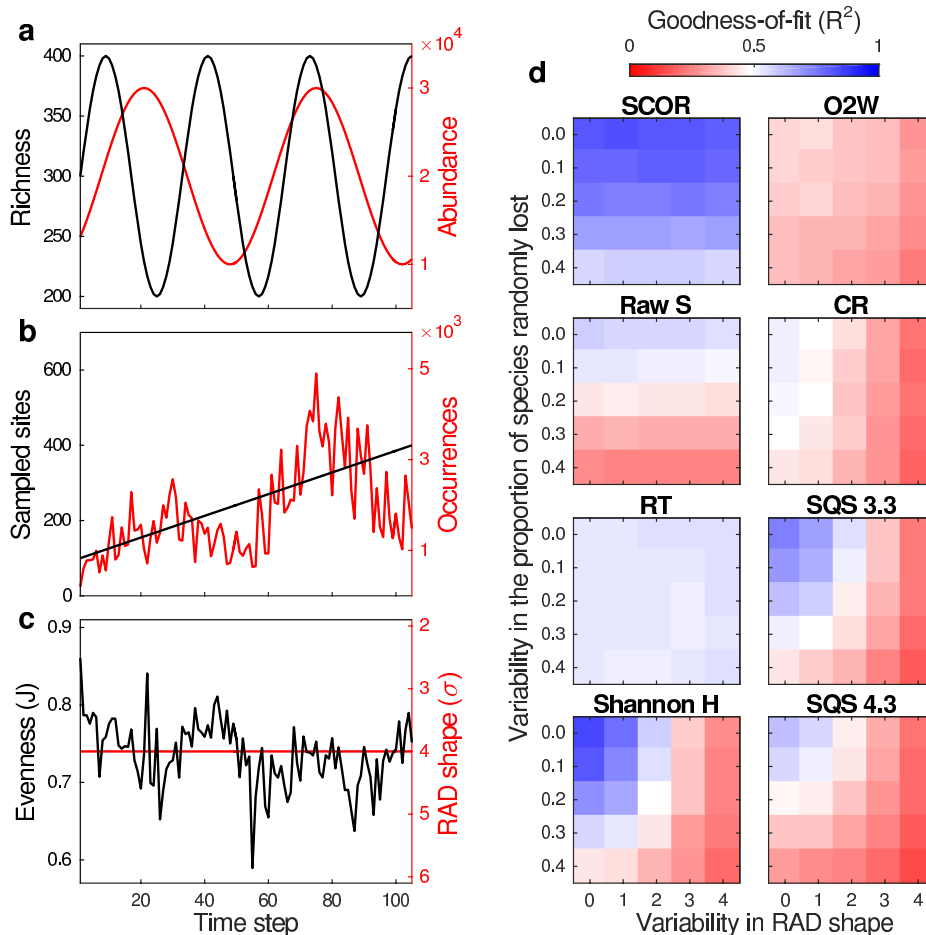
- 529 temperature. *Proceedings of the National Academy of Sciences of the United States of*
530 *America* **112**, E4638–E4639 (2015).
- 531 43. Theiler, J., Eubank, S., Longtin, A. & Galdrikian, B. Testing for nonlinearity in time
532 series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**, 77–94
533 (1992).
- 534 44. Ebisuzaki, W. A method to estimate the statistical significance of a correlation when
535 the data are serially correlated. *Journal of Climate* **10**, 2147–2153 (1997).
- 536 45. Paluš, M. & Vejmelka, M. Directionality of coupling from bivariate time series: How
537 to avoid false causalities and missed connections. *Phys. Rev. E* **75**, 056211 (2007).
- 538 46. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-
539 spectral Methods. *Econometrica: Journal of the Econometric Society* **37**, 424–438
540 (1969).
- 541 47. Barnett, L., Barrett, A. B. & Seth, A. K. Granger Causality and Transfer Entropy Are
542 Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **103**, 238701 (2009).
- 543 48. Hannisdal, B. & Peters, S. E. Phanerozoic Earth system evolution and marine
544 biodiversity. *Science* **334**, 1122–1124 (2011).
- 545 49. Dunhill, A. M., Hannisdal, B. & Benton, M. J. Disentangling rock record bias and
546 common-cause from redundancy in the British fossil record. *Nat Comms* **5**, 5818
547 (2014).
- 548 50. Jeffreys, H. *Theory of Probability*. (Oxford University Press, 1961).



Supplementary Figure 1 | Distribution of sampled evenness across all Poseidon experiments. Shaded histogram represents the model runs testing the sensitivity to variability in the proportion of species randomly removed, and variability in RAD shape parameter σ (Supplementary Fig. 3). Un-shaded histogram (note transparency in overlap) represents the model runs testing the sensitivity to variability in the proportion of sites sampled, and variability in σ (Fig. 1). The median evenness is 0.76, and 95 % of the values are in the range 0.65 - 0.90.

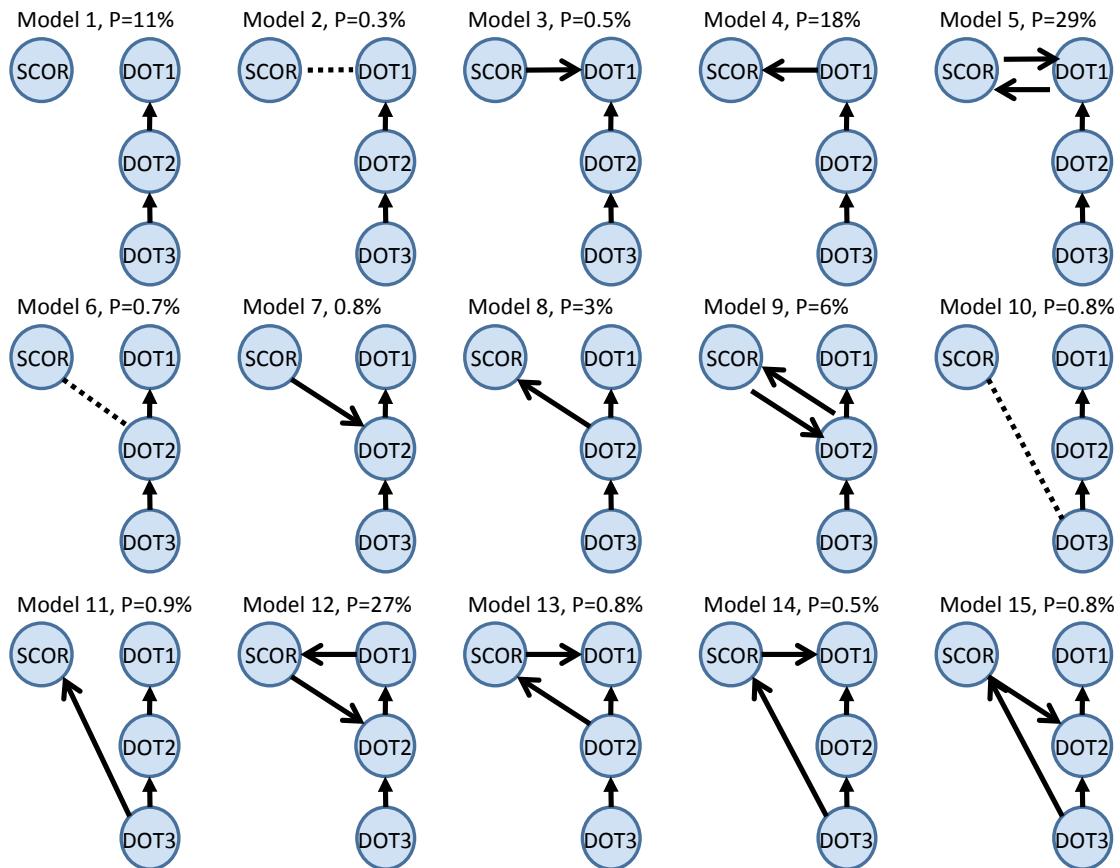


Supplementary Figure 2 | Sampling-standardized richness can be reproduced by the sum of raw richness and evenness. **a,b**, Raw sampled richness (S) and evenness (Pielou's J) of Cenozoic coccolithophores (**a**) and planktonic foraminifera (**b**) species from the NSB database. **c, d**, The sum of raw S and raw J superimposed on shareholder quorum subsampling (SQS) estimates of richness for coccolithophores (**c**) and foraminifera (**d**), all normalized to zero mean and unit standard deviation. SQS was calculated with a quorum level of 0.4 (higher quorum levels give nearly identical results but are less complete for the older part of the record). Ma, million years before present; Paleoc., Paleocene; Oligoc., Oligocene; P, Pliocene; Pt, Pleistocene.



Supplementary Figure 3 | Effect of random species loss in Poseidon model

experiments. a, Simulated richness and abundance as in Fig. 1a. **b**, Site sampling increases smoothly in all experiments. Instead, a proportion of the species is randomly removed in each time step, causing volatility in occurrences. No variability in the proportion lost means that 50 % are always removed. In this example, variability = 0.4, meaning that between 30 % and 70 % of species are lost. **c**, Even with a constant original RAD shape, random species loss, and variability in the proportion lost, generates volatility in sampled evenness (this example is an extreme case, see Supplementary Fig. 1). **d**, Sensitivity to variability in RAD shape and in the proportion of species lost. Values are median goodness-of-fit (R^2) of 50 model runs, comparing SCOR to true abundance, and richness estimates to true richness.



Supplementary Figure 4 | Schematic of all connection models between SCOR

and DOT in linear SDE analysis. The best model for SCOR in isolation is a one-layered OUP. The best model for DOT in isolation is a three-layered model with a WP as the bottom layer (DOT3). All models possible between these two best models are shown. Note that SCOR cannot drive DOT3 because DOT3 is a WP. Percentage values represent posterior model probabilities. Solid arrows represent casual connections pointing from driver to response. Dotted lines represent correlative relationships. See Methods for details.

Supplementary Table 1 | Parameter estimates for the most probable connection model between SCOR and DOT

Parameter	Estimate (posterior median)	95 % credible interval
$t_{1/2,SCOR1}$	0.53 Myr	(0.33, 1.1) Myr
s_{SCOR1}	0.94	(0.65, 1.1)
μ_{SCOR}	0.43	(-0.01, 0.94)
$t_{1/2,DOT1}$	8.0 Myr	(0.91, 31) Myr
s_{DOT1}	1.2	(0.37, 2.4)
$t_{1/2,SDOT2}$	16 Myr	(3.4, 166) Myr
s_{DOT2}	0.45	(0.03, 2.8)
σ_{DOT3}	0.027	(0.003, 0.22)
$\beta_{DOT1 \rightarrow SCOR1}$	0.45	(0.14, 0.76)
$\beta_{SCOR1 \rightarrow DOT1}$	-0.77	(-2.3, 0.5)

Parameter estimates for Model 5 (Supplementary Fig. 4), where $t_{1/2,i}$ are half-lives, s_i are the stationary standard deviations, μ_i are the expected values of OUP, σ_i are stochastic terms in WP. The bottom DOT layer (DOT3) is a WP. The interpretation of parameter values depend on other processes, e.g. the causal connection from DOT to SCOR will increase the total stationary standard deviation of the SCOR process. The reported s_i values represent what the process itself supplies in isolation. See Methods for details.

Supplementary Table 2 | Parameter estimates for the most probable connection model between uSCOR and uDOT

Parameter	Estimate (posterior median)	95 % credible interval
$t_{1/2,uSCOR1}$	0.91 Myr	(0.32, 1.6) Myr
S_{uSCOR1}	0.07	(0.01,0.35)
$t_{1/2,uSCOR2}$	2.4 Myr	(0.86, 68) Myr
S_{uSCOR2}	0.7	(0.02,21)
σ_{uSCOR3}	2.8	(0.04, 5.8)
$t_{1/2,uDOT1}$	0.24 Myr	(0.13, 0.33) Myr
$t_{1/2,uDOT2}$	0.50 Myr	(0.21, 1.5) Myr
σ_{uDOT3}	4.1	(1.9, 8.9)
$\beta_{uDOT1 \rightarrow uSCOR1}$	2.7	(-2.6, 3.8)
$\beta_{uSCOR1 \rightarrow uDOT1}$	-2.7	(-5.2, 1.9)

Symbols and interpretations as in Supplementary Table 1, but in this case both uDOT and uSCOR have 3-layer models with a WP at the bottom. See Methods for details.