Macroecological consequences of biotic and abiotic factors in marine ecosystems through time



Gwen Antell

St. John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

Supervised by:

Erin Saupe

Tim Coulson

Declaration

I declare that the work presented in this thesis is wholly my own work, except where otherwise stated. The views and opinions expressed herein are mine and not necessarily those of any other person or body unless so attributed.

Gwen Antell

Abstract

This dissertation statistically investigated putative biotic and abiotic drivers of ecological characteristics over the spatial scale of ocean basins and temporal scale of millennia to millions of years. For three fundamental hypotheses in evolutionary ecology, I compared patterns of natural experiments from the fossil record against expectations from theory. The first empirical study appraised the hypothesis that competition restricts geographic distributions at the species level, and thus taxa are more widespread when in the presence of fewer competitors, such as after mass extinction events. Contrary to the prediction of competitive exclusion, brachiopod and bivalve distributions were invariant to inferred biotic pressure, while habitat type and lineage age had detectable effects on range expansion or contraction on the scale of geologic stages across the Phanerozoic. The second study estimated the degree to which plankton modulate thermal preferences *vs.* migrate in response to glacial–interglacial climate fluctuations. Species conserved abiotic niches over the Middle–Late Pleistocene study interval, suggesting marine species will track preferred environments rather than adapt to Anthropogenic climate warming. The final study explored relative niche change within lineages *vs.* across an evolutionary tree of macroperforate planktonic foraminifera. The stark stasis observed within lineages could suggest niches evolve primarily during speciation, but any signal of such punctuated equilibrium was undetectable statistically with the small phylogenetic tree and large variance within and between species. Together, these results add perspective to longstanding debates about the scaling of population- to species-level ecology and micro- to macro-evolution.

There's a place you just can't reach unless you have
a dream too big to bear alone.

— Kubo Mitsurō

Acknowledgements

My supervisor Erin Saupe played equal parts initiator, sounding board, sponsor, critic, liaison, confidant, editor, cheerleader, and signatory during the past four years. Being the inaugural graduate student in any lab is a unique burden and privilege, and I have cherished the teamwork Erin and I shared as the Saupe lab grew.

I would not have come to Oxford for my DPhil except for Erin's persistent recruitment, and I might not have ventured into a palaeobiology doctorate at all save for the faith of many prior mentors, particularly Susan Butts and Jacques Gauthier at the Yale Peabody Museum.

Tim Coulson bravely jumped aboard my DPhil supervisory team sight unseen and has been a stalwart source of wisdom and wise-cracks ever since. I am endlessly grateful Tim adopted me into his lab and department.

Fulfilling the remainder of my dad joke quota, Roger Benson gave constant inspiration and readings to think about problems from different angles. Scattered through my research notebook are pages with Roger's suggestions, hastily scrawled after happy hour while still fresh in my memory.

PalaeoClub members took me under their wings at tea-time and pub-time, offering equal parts counsel and consolation. Thanks, in order of my first days in Oxford to my last: the cohort of 2018–2019 palaeo doctorates, Marie-Claire Koschowitz, Donald Davesne, Neil Brocklehurst, Ricardo Pérez-de la Fuente, Brooke Johnson, James Neenan, Ross Anderson, Roger Close, Isabel Fenton, Elsa Panciroli, and everyone from the Year of Zoom.

Officemates Tom Breithaupt, Allison Bryan, Brigit Tronrud, Duhita Naware, and Alice Turner turned our walk-in closet into a homey space held in confidence. The Eco-Evo Dynamics Lab, particularly Josie Mahonie, Tomos Potter, Aura Raulo, and Ash Sendell-Price, forged a joyous second departmental home when I needed a safe escape. Committee members Ed Love, Olivia Graves, Ria Dinsdale, and Adam Packer of the Middle Common Room at St John's College helped keep my life together during the year I was president, training for marathons, and re-building my first thesis publication (Chapter 2).

With love for other friends who celebrated my every step toward the end of my own rainbow: Tunrayo Adeleke-Larodo, Diana Avadanii, Diego Berdeja Suárez, Ellen Cliff, Em Dale, Laura Garmendia Sanchez, Sam Kim, Ricky Sengupta, and Jan Steinebrunner.

With apologies to my long-suffering family—Mum, Dad, Wes—who have not only tolerated but indeed actively enabled my love for all creatures weird and wonderful.

With gratitude to elders and ancestors not related by blood, in the order they changed my life: Patricia Wright, Pauli Murray, Audre Lorde, Eli Clare, Dr. Richard K. Bernstein, Sylvia Rivera, Marsha P. Johnson, David Jay, and countless others who have inhabited spaces between, both, and beyond. You toppled hierarchies and laws and literal walls to make my life and schooling possible. You not only showed me a path, but infinite paths. You continue to teach me that human innovation, resilience, and spirit know no bounds.

An historical note.

The last 22 months of this 50-month DPhil overlapped the global coronavirus crisis. I felt I owed my survival to the friends who brought me groceries before mask mandates in supermarkets, exchanged my medical sharps containers at the clinic, carried my placards to Black Lives Matter marches, met in parks for distanced walks, screened Shaun of the Dead and other appropriately apocalyptic films, and listened empathically to real and imagined scares. Thanks to all the scientists around the world who worked indefatigably against COVID-19, especially the clinical researchers and vaccine trial volunteers here at Oxford. I am proud to have made this city's DNA part of me, literally.

Whether during national lockdowns or the background extinction intervals in-between, I wrote most of the ensuing text from bedrooms, kitchen tables, or a makeshift home office. Jericho Coffee Traders roastery delivered beans to my door: many a draught yanked me through writing droughts. Regular meetings of the Clandestine Social Sciences Writing Community (the online, subject-inclusive offshoot of the official group) brought Chapter 1 to these pages. Ruby-Anne Birin, Michael Leadbetter (and Jannus), Meryem Kalayci, and Carolina de Mendonca Gueiros gave me especial welcome.

An account of pandemic postgraduate life would be incomplete without acknowledging that this disserting road is emptier now than when I started. Illnesses and traumas have taken many peers away, and I grieve our separation by time zones, suspended studies, or death. One friend who fled Oxford for the other side of the world recorded a voice message for me in the third UK lockdown. 'I'm so glad someone is pushing through,' they told me, 'and I'm glad that someone is you.'

—so this completed thesis is for you. All of you.

Gwen Antell
December 2021
Oxford, UK

"I was deeply stirred by the way that the marine creatures moved so freely; how the soft corals and sea invertebrates seemed to exist without physical boundaries, like warrior shape-shifters; the way the fish regally flaunted their colourful costumes. That's how I feel on the inside. In my soul, I'm that colourful; my sexuality, my gender—it's free-moving.

And then I met a tube-like structure from which emerged a fan of patterned feathers. I edged as close as I could to see inside the creature, but my sudden movement caused the feather structure to retreat. I was desperate to dive inside this wormhole with it, for this new world clearly had so much to teach me."

Amrou Al-Kadhi (2020), *Life as a Unicorn*, pp. 99–101

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction to marine ecosystem history **1**
1.1 Abstract . . . 2
1.2 Introduction . . . 2
1.3 History of marine ecosystems . . . 5
1.3.1 Ediacaran–Cambrian Radiation . . . 6
1.3.2 Great Ordovician Biodiversification Event . . . 9
1.3.3 Devonian Nekton Revolution . . . 10
1.3.4 Mesozoic Marine Revolution . . . 12
1.4 A framework for bottom-up ecosystem controls . . . 14
1.4.1 Nutrient flux . . . 17
1.4.2 Phytoplankton evolution . . . 19
1.4.3 Population abundance . . . 21
1.4.4 Consumer speciation . . . 24
1.4.5 Body mass and metabolism . . . 27
1.5 Other mechanisms of change . . . 29
1.6 Discussion . . . 30
1.7 Scope of the thesis . . . 33

2 Ecological release is undetectable in oceans at large scales **37**
2.1 Overview . . . 38
2.2 Results . . . 38
2.2.1 Time series analysis . . . 41
2.2.2 Biotic crises *vs.* background intervals . . . 42
2.2.3 Mixed-effects models . . . 42
2.3 Discussion . . . 47
2.4 Methods . . . 50
2.4.1 Fossil occurrences . . . 50
2.4.2 Spatial subsampling . . . 52
2.4.3 Statistical analyses . . . 53
2.4.4 Data and code availability . . . 54

3 Thermal niches are static throughout glacial–interglacial climate change 55
3.1 Abstract . . . 56
3.2 Introduction . . . 56
3.3 Results . . . 59
3.3.1 Time series correlation . . . 59
3.3.2 Niche lability *vs.* climate change magnitude . . . 61
3.3.3 Niche lability among extreme climate intervals . . . 61
3.4 Discussion . . . 63
3.5 Methods . . . 65
3.5.1 Palaeoclimate reconstructions . . . 65
3.5.2 Definition of 'available environment' . . . 66
3.5.3 Fossil occurrence data . . . 67
3.5.4 Spatiotemporal resolution . . . 69
3.5.5 Species attributes . . . 70
3.5.6 Kernel density estimation . . . 71
3.5.7 Niche dissimilarity measure . . . 73
3.5.8 Statistical analyses . . . 75
3.5.9 Data and code availability . . . 79

4 Rate and mode of niche evolution across a foraminifera phylogeny 80
4.1 Abstract . . . 81
4.2 Introduction . . . 81
4.3 Results . . . 83
4.3.1 Within-lineage evolution models . . . 85
4.3.2 Phylogenetic ecotype comparisons . . . 86
4.3.3 Across-lineage evolution models . . . 86
4.4 Discussion . . . 91
4.5 Methods . . . 94
4.5.1 Foraminifera data . . . 94
4.5.2 Within-lineage evolution models . . . 95
4.5.3 Phylogenetic ecotype comparisons . . . 97
4.5.4 Across-lineage evolution models . . . 98
4.5.5 Data and code availability . . . 107

5 General discussion and conclusions 108
5.1 Contribution to knowledge . . . 109
5.2 Limitations and opportunities . . . 112
5.2.1 Ancestral state reconstruction . . . 112
5.2.2 Multiple sources of autocorrelation . . . 114
5.3 Closing remarks on ecological and evolutionary scaling . . . 116

Appendices

A Supporting charts for Chapter 2 **119**
A.1 Figures . . . 119
A.2 Tables . . . 126
A.3 Data availability . . . 139

B Supporting methods for Chapter 2 **140**
B.1 Fossil occurrences compilation . . . 141
B.1.1 Global Phanerozoic dataset . . . 141
B.1.2 Northwest Atlantic Neogene–Quaternary dataset . . . 144
B.2 Standardisation of spatial coverage . . . 146
B.2.1 Further notes on proportional occupancy . . . 150
B.3 Assemblage framework . . . 153
B.3.1 Assemblage subsampling routine . . . 154
B.3.2 Time series analysis . . . 155
B.3.3 Sensitivity to subsampling parameters . . . 156
B.4 Taxon-tracking framework . . . 163
B.4.1 Environmental preferences . . . 164
B.4.2 Taxon-tracking subsampling routine . . . 166
B.4.3 Biotic crises *vs.* background intervals . . . 167
B.4.4 Mixed-effects models . . . 168
B.5 Power analysis . . . 169
B.5.1 Justification . . . 169
B.5.2 Simulation of ecological release . . . 170
B.5.3 Power to detect release . . . 171

C Supporting charts for Chapter 3 **173**
C.1 Figures . . . 173
C.2 Tables . . . 182

D Kernel density estimation methods for Chapter 3 **188**
D.1 Overview of KDE and 'kerneval' . . . 189
D.1.1 Why use kernels? . . . 189
D.1.2 How does it work? . . . 189
D.1.3 When should one use 'kerneval'? . . . 189
D.2 Mathematical framework . . . 190
D.3 Transformed density estimation . . . 190
D.4 Weighted density estimation . . . 191
D.4.1 Bootstrap bandwidth selection . . . 192

D.5 Examples 193
D.5.1 Length bias and inverse length bias 193
D.5.2 Polynomial bias function 197
D.6 Technical details 200
D.6.1 Approximating functions and integration 200
D.6.2 Boundary reflection 202
D.6.3 Mean integrated squared error 203

E Supporting charts for Chapter 4 **206**
E.1 Figures 206
E.2 Tables 212

References **221**

List of Figures

1.1 Components of marine ecosystems. 4
1.2 Continental configurations through time. 7
1.3 Bottom-up control pathways in marine ecosystems. 15

2.1 Regional species count *vs.* geographic range size through time. . . . 41
2.2 Species count and geographic range size from the Neogene–Quaternary Atlantic. 43
2.3 Geographic occupancy change during net species loss or gain. 44
2.4 Predicted change in geographic range size for boundary-crossing species. 45
2.5 Artistic reconstruction of Jurassic marine diversity-dependence from ecological release of geographic distributions. 48

3.1 Schematic of niche density estimation from palaeoclimate models. . 60
3.2 Correlation of niche dissimilarity and climate change magnitude. . . 62
3.3 Niche comparisons between glacial and interglacial extremes. 63
3.4 Reference for Hellinger's H distances between Normal distributions. 74

4.1 Phylogenetic tree of planktonic foraminifera. 84
4.2 Within-lineage evolution model weights. 87
4.3 Across-lineage evolution model weights. 90
4.4 Across-lineage evolution models fit on asynchronous tips. 91
4.5 Fossil sampling strategies for tree tips. 99
4.6 Example bootstrap distributions of likelihood ratios. 102
4.7 Example branch-length transformations. 104

A.1 Workflow of analysis in assemblage and taxon-tracking frameworks. 121
A.2 Simplified diagram of assemblage subsampling. 122
A.3 Simplified diagram of taxon-tracking subsampling. 123
A.4 Power analysis from simulated geographic expansion. 125

B.1 Illustration of the species–area effect on proportional occupancy. . . 148
B.2 Example of sampling area effect on range-size measurement. 151

C.1 Niche density distributions from example species and time steps. . . 174

C.2 Time series of mean temperature across sampling sites. 175
C.3 Schematic walk-though of the derivation of niche comparison bounds. 177
C.4 Correlation of niche dissimilarity and climate change magnitude within habitat depths. 178
C.5 Temperature time series at global *vs.* sampling sites. 179
C.6 Niche comparisons between glacial and interglacial extremes within depth habitats. 180
C.7 Successful detection of simulated niche lability. 181

D.1 Comparison of density curves from different estimators, $n = 50$. . . 199
D.2 Comparison of density curves from different estimators, $n = 500$. . . 200
D.3 Rectangular approximation of probability density integral. 201
D.4 Trapezoidal approximation of probability density integral. 202
D.5 Boundary reflection to improve density estimate on truncated data. 203

E.1 Time series of each species' occupied temperatures. 207
E.2 Time series of each species' occupied temperatures within depth habitats. 208
E.3 Within-lineage evolution model weights, within depth habitats. . . . 209
E.4 Across-lineage evolution model weights, within depth habitats. . . . 210
E.5 Across-lineage evolution model weights, with specified error. 211

List of Tables

1.1 Stepping-stones in bottom-up ecosystem control pathways. 16

A.1 Global Phanerozoic and Neogene–Quaternary time bins. 126
A.2 Subsampling parameter combinations for sensitivity analyses. . . . 128
A.3 Corrected time-series correlation coefficients. 129
A.4 Range-size shifts for each Phanerozoic interval. 131
A.5 Relative performance of non-trivial linear mixed-effects models. . . 133
A.6 Coefficient estimates and confidence intervals for 5 best-supported linear mixed-effects models. 136
A.7 Log-likelihood estimates and confidence intervals for global linear mixed-effects model. 138

C.1 Ecological attributes of foraminifera study species. 186
C.2 Age of glacial and interglacial intervals. 187

D.1 Performance of replicated bias correction method. 196
D.2 Original bias correction results from Barmi and Simoff (2000). . . . 196

E.1 Within-lineage estimates of trait mean and variance through time. . 213
E.2 Phylogenetic ANOVA coefficients of niche variance by ecotype. . . . 214
E.3 Across-lineage evolution model weights. 215
E.4 Across-lineage evolution model weights, with specified error. 217
E.5 Across-lineage estimates for *lambda*, *delta*, and *alpha*. 218

The historian of life takes not only knowledge of fossils but also a tremendous array of pertinent facts from other fields of earth sciences and of life sciences and weaves them all into an integral interpretation of what the world of life is like and how it came to be.

— George Gaylord Simpson (1949), *The Meaning of Evolution*

1

Introduction to marine ecosystem history

Contents

1.1 Abstract 2
1.2 Introduction 2
1.3 History of marine ecosystems 5
1.3.1 Ediacaran–Cambrian Radiation 6
1.3.2 Great Ordovician Biodiversification Event 9
1.3.3 Devonian Nekton Revolution 10
1.3.4 Mesozoic Marine Revolution 12
1.4 A framework for bottom-up ecosystem controls 14
1.4.1 Nutrient flux 17
1.4.2 Phytoplankton evolution 19
1.4.3 Population abundance 21
1.4.4 Consumer speciation 24
1.4.5 Body mass and metabolism 27
1.5 Other mechanisms of change 29
1.6 Discussion 30
1.7 Scope of the thesis 33

The literature review presented in this chapter has been published in Antell and Saupe (2021) (DOI: 10.1016/j.cub.2021.08.069). EES contributed structural advice and editing of the manuscript; GSA searched the literature, wrote the text, and made the charts. The present chapter constitutes an adaptation of the article to fit within the context of this thesis.

1.1 Abstract

Animal life originated in the oceans and evolved there for hundreds of millions of years before adapting to terrestrial environments. Today's oceans cover more than two-thirds of planet Earth and generate as much primary production as does land. The path from the first macrobiota communities to the modern menagerie of marine life involved parallel increases in terrestrial nutrient input, marine primary production, species' abundance, metabolic rates, ecotypic diversity, and taxonomic diversity. Bottom-up theories of ecosystem cascades arrange these changes in causal sequence. At the base of marine food webs, nutrient fluxes and atmosphere–ocean chemistry interact with phytoplankton to regulate production. First-order consumers (e.g. zooplankton) might propagate changes in quantity and quality of phytoplankton to changes in abundance and diversity of larger predators (e.g. nekton). However, many uncertainties remain about the mechanisms and effect size of bottom-up control, particularly as applied to the world's oceans and the entire history of animal life. Here, I distinguish and review modern and fossil evidence for hypothesised bottom-up pathways, and I assess the ramifications of these processes for four key intervals in marine ecosystems: the Ediacaran–Cambrian (635–485 million years ago [Ma]), Ordovician (485–444 Ma), Devonian (419–359 Ma), and Mesozoic (252–66 Ma). I advocate for clear articulation of mechanistic hypotheses to better understand causal relationships and proposed effects, combined with additional ecological experiments, palaeontological documentation, isotope geochemistry, and geophysical reconstructions. How small-scale ecological change transitions into large-scale evolutionary change remains an outstanding question for empirical and theoretical efforts to address.

1.2 Introduction

Identifying the factors that promote or limit biodiversity is key to understanding ecological and evolutionary change across timescales. Throughout modern biomes, predation pressure is cited as an important regulator of ecosystem structure,

determining the abundance, size, and behaviour of consumers at lower trophic tiers ('top-down control') (Myers, Baum, et al. 2007; Estes, Terborgh, et al. 2011). When a trophic level or size class is free of predators, a common assumption is that food availability limits the abundance of that level or class instead (Hairston, Smith, et al. 1960; Fretwell 1987; Sinclair, Mduma, et al. 2003). Ultimately, all food resources are limited by the rate of primary production, which in turn is strongly influenced by nutrient concentrations.

Research on modern ecosystems frequently has assumed that the environmental conditions governing nutrient pathways are static or homogeneous (the 'mean field approximation') (Hunter and Price 1992; Hunter 2001; Petersen, Kemp, et al. 2003). Where nutrient flux changes are noted, perturbations are often cyclic, with a phase of centuries at longest. Consequently, resource-driven ecological changes tend to be treated as reversible outcomes of brief environmental excursions. For instance, plankton productivity fluctuates partially in response to annual to decadal variation in polar climate (Behrenfeld 2014) and to decadal to centennial variation in the California Current System (Bringué, Pospelova, et al. 2014). Although anthropogenic nutrient redistribution is increasingly seen as a threat to biodiversity (Vitousek, Aber, et al. 1997; Bennett, Carpenter, et al. 2001), loss of apex predators remains a primary focus of research, conservation, and management in global-change biology (Schmitz 2005; Myers, Baum, et al. 2007; Estes, Terborgh, et al. 2011).

Over the more than 500 million years of animal evolution, however, the assumption of baseline environmental stasis is strongly violated. Widespread volcanism, planetary glaciation, and other dramatic perturbations to nutrient cycles lasted on the order of millions of years (Raymond and Metz 2004; Filippelli 2010). Ecological changes close in time to some of these geologic events were similarly momentous. On global, geologic scales, the role of primary productivity might shift from passively limiting local abundances to actively driving diversification of species and ecotypes. Such 'bottom-up controls' (Fig. 1.1) are hypothesised to have generated substantial evolutionary and ecological changes throughout Earth history (Bambach 1993; Martin 1996; Bambach 1999; Allmon and Martin 2014; Martin and Servais 2020).

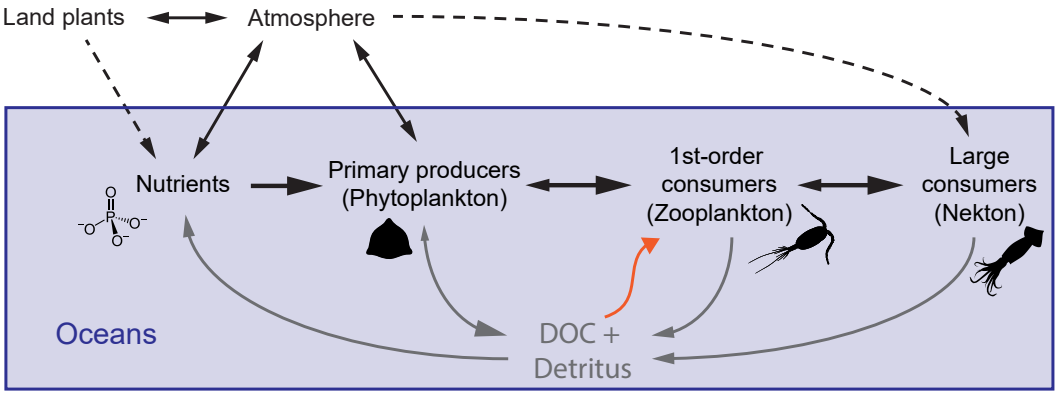


Figure 1.1: The abundance and diversity of marine primary producers, first-order consumers, and higher-order consumers depend on nutrient availability and atmospheric gas concentrations. Pelagic consumers can be coarsely divided into zooplankton and nekton; multiple tiers of consumers live on the benthos as well. All trophic levels leak organic carbon from food webs, in particulate form (detritus) or dissolved form (DOC). Substantial volumes of phytoplankton mass convert to DOC through cell death and viral lysis. The microbial loop (orange arrow) recycles some of this carbon to zooplankton via a chain of bacteria, nanoflagellates, and ciliate protozoans. Terrestrial ecosystems can affect marine ecosystems by eroding and weathering rock, altering the hydrologic cycle, and absorbing and respiring CO_2 and O_2 to the atmosphere. Thick arrows indicate the main candidate pathway for bottom-up ecosystem controls. Dashed arrows indicate pathways that may be weaker or of variable strength throughout Earth history.

I use the term 'ecological change' in the most inclusive sense, encompassing shifts in biomass, population density, relative abundances, species diversity, metabolic rates, frequency distribution of functional traits, and more. Changes in behaviour may also occur (Schmitz 2005; Laundré, Hernández, et al. 2010; Zhong, Li, et al. 2017), particularly as a top-down effect, but this outcome is unobservable in most of the fossil record. Ecologists distinguish between species-level cascades *vs.* community cascades, and abundance cascades *vs.* diversity cascades (Hunter and Price 1992; Polis 1999; Polis, Sears, et al. 2000; Dyer and Letourneau 2003). In contrast, most fossil studies aim to quantify only community or diversity responses, because it is harder to study individual species and absolute abundance from the rock record than in living systems.

Expansion of trophic tiering, acceleration of metabolic activity, and radiations of clades at many trophic levels are the outcomes of bottom-up controls discussed in the palaeontological literature most often (Bambach 1993; Martin 1996; Bambach 1999;

Allmon and Martin 2014; Martin and Servais 2020). Below, I briefly summarise examples of these and other ecological changes during four intervals of Earth history. I then synthesise theory and data from modern ecological studies as well as the fossil record to evaluate each proposed biological pathway of bottom-up control. Other recent reviews cover the geological record of putative bottom-up cascades in more detail, including comparing time series of family-level diversity and geochemical proxies for weathering and organic carbon production and burial (Martin and Quigg 2012; Allmon and Martin 2014; Martin and Servais 2020).

1.3 History of marine ecosystems

The ecological focus of the thesis centres on oceans because the history of marine ecosystems is substantially longer and better known than that of terrestrial ecosystems. Notable environmental events since the appearance of animals include multiple glaciations in the Neoproterozoic and Paleozoic (Raymond and Metz 2004; Hoffman, Abbot, et al. 2017), superplumes and volcanism in the Ordovician and Cretaceous (Barnes 2004; Lefebvre, Servais, et al. 2010; Sprain, Renne, et al. 2019), enhanced upwelling and runoff in the Jurassic (Giraud 2009), and land plant evolution in the Devonian and Cretaceous (Berner 1997; Algeo and Scheckler 1998; Boyce and Lee 2011). Marine ecosystem revolutions have been said to have occurred in each of four intervals overlapping these geologic events: the Ediacaran–Cambrian (635–485 Ma) (Erwin and Valentine 2013), Ordovician (485–444 Ma) (Servais and Harper 2018), Devonian (419–359 Ma) (Klug, Kroeger, et al. 2010), and Mesozoic (252–66 Ma) (Vermeij 1977). Biotic events during these intervals involved rapid speciation into previously unexplored or inaccessible areas of ecospace, such as the expansion of macroscopic life into the water column and the origin of metabolically-active predators throughout the Paleozoic (Bambach 1983; Bambach, Bush, et al. 2007; Sperling, Frieder, et al. 2013).

1.3.1 Ediacaran–Cambrian Radiation

Few intervals remain as contentious among geobiologists as the Neoproterozoic–Cambrian (1,000–485 Ma), a time that perhaps witnessed both glaciation and oxygenation on global scales, alongside the origin and radiation of animals. Earth scientists have established that the planet froze repeatedly, first in the Sturtian glaciation (720–660 Ma) and again in the Marinoan glaciation (650–635 Ma) (Hoffman, Abbot, et al. 2017). Unlike the Pleistocene ice ages of the last few million years, ice covered even the equatorial zone and sea surface in these 'snowball Earth' events. At least one more glaciation occurred, the Gaskiers (580 Ma), but this event may have been less temporally and geographically extensive than were the two preceding glaciations (Pu, Bowring, et al. 2016). Massive reconfiguration of continents overlapped the glaciation intervals. From 900–650 Ma the supercontinent Rodinia broke up; afterwards, the continents of East and West Gondwana collided into a new supercontinent, Pannotia (Fig. 1.2) (Squire, Campbell, et al. 2006; Scotese 2009; Campbell and Squire 2010).

Landmark evolutionary and ecological changes were interwoven with abiotic events in the Neoproterozoic. Molecular clocks place the evolution of animals before the first Neoproterozoic glaciation event, with metazoan–choanoflagellate divergence 1000–900 Ma and Metazoa origin *ca.* 800 Ma (reviewed in Sperling and Stockey 2018). Biomarker evidence indicates a pivotal turnover in marine primary production between the two snowball events: chlorophytes (green algae) and other primary endosymbiotic algae expanded from freshwater habitats to marine settings and displaced bacteria as the important primary producers in oceans (Brocks, Jarrett, et al. 2017). Eukaryotic algae have flourished as the main marine producers ever since, although component clades later turned over in abundance in the Phanerozoic ('Mesozoic Marine Revolution' below). The earliest known fossils of macroscopic animals (rangeomorphs) are dated at 574 Ma, placing them 5 Ma after the final, Ediacaran glaciation, the Gaskiers (Dunn, Liu, et al. 2021; Matthews, Liu, et al. 2021). Large multicellular organisms in the Ediacaran formed ecologically complex communities upon the seabed surface and set the

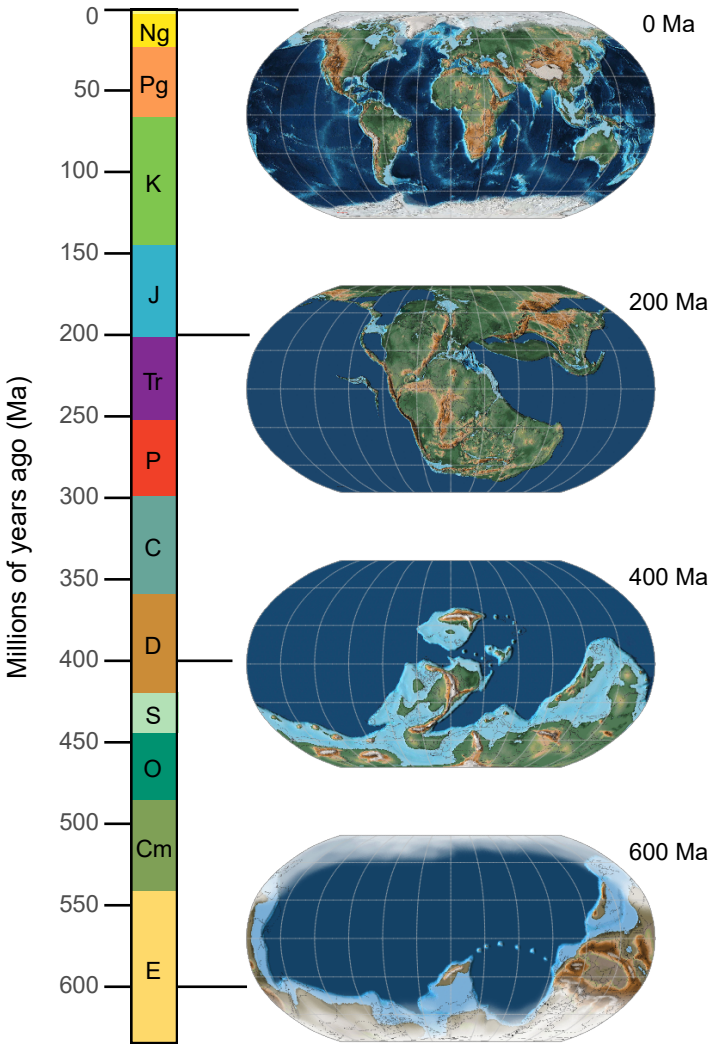


Figure 1.2: Reconstructions of continental positions, orientations, and topography are shown at 200-million-year intervals between the present day, Mesozoic, Paleozoic, and Neoproterozoic. Maps are plotted from PaleoAtlas data (Scotese 2016b; Scotese 2016a). E, Ediacaran; Cm, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene. Quaternary not labelled.

stage for unprecedented radiation of animal ecomorphology and clades (Schiffbauer, Huntley, et al. 2016; Droser, Tarhan, et al. 2017; Darroch, Laflamme, et al. 2018; Mitchell and Butterfield 2018). By the Early to Middle Cambrian, 40–60 million years later, body plans had evolved for nearly every animal group ranked as a phylum (Erwin, Laflamme, et al. 2011; Erwin and Valentine 2013), and trophic webs had attained the complexity of those in modern marine ecosystems (Dunne, Williams, et al. 2008). Note that molecular and body fossils set only minimum age

constraints, and it is possible eukaryotic algae, macrobiota, and metazoan crown groups proliferated earlier than currently documented.

The broadly similar timing of such dramatic abiotic and biotic events has prompted a slew of causal explanations, many of which assume spikes in oxygen or phosphorus mediated Earth and life system changes (Planavsky, Rouxel, et al. 2010; Mills and Canfield 2014; Xiao 2014; Laakso, Sperling, et al. 2020; Reinhard, Planavsky, et al. 2020). Unfortunately, the onset, duration, and magnitude of any Neoproterozoic Oxidation Event or phosphorus pulses are poorly constrained relative to glaciation events and the Ediacaran–Cambrian Radiation (ECR). One set of hypotheses posits glaciation (and deglaciation) released terrestrial nutrients such as phosphorus that fuelled phytoplankton production, and thence oxygenation (Canfield, Poulton, et al. 2007; Planavsky, Rouxel, et al. 2010; Martin and Quigg 2012); uplift of the Transgondwanan Supermountains might have shed nutrients into oceans with similar effect (Squire, Campbell, et al. 2006; Campbell and Squire 2010). Even if glacial or montane phosphorus input was transient, an associated sulphate increase could have jump-started the self-sustaining modern phosphorus cycle, recycling the nutrient for primary production (Laakso, Sperling, et al. 2020). A contrasting set of hypotheses reverses the order of causality and suggests plankton community restructuring caused a Neoproterozoic Oxidation Event and thence glaciation (Tziperman, Halevy, et al. 2011; Lenton, Boyle, et al. 2014), just as the Great Oxidation Event had preceded snowball glaciation *ca.* 2.4 billion years ago (Warke, Di Rocco, et al. 2020). Another biotic change that could have transformed ecosystems indirectly via subsequent abiotic change was the innovation of bioturbating metazoans. Bilaterians with a through-gut disrupted coverage of microbial mats that had sealed the Ediacaran seafloor, thereby allowing new, burrowing species to access the underlying sediment (Erwin and Valentine 2013; Erwin 2015c).

It is even possible environmental change was irrelevant to the origins of multicellularity and metazoans; Gene Regulatory Network innovations alone might have paced eukaryote evolution (Butterfield 2018 but see Erwin, Laflamme, et al.

2011). For such multifaceted changes as the ECR entailed, any single trigger may be insufficient as an explanation (Erwin, Laflamme, et al. 2011; Erwin and Valentine 2013; Schiffbauer, Huntley, et al. 2016; Sperling and Stockey 2018). A 'fire triangle' of oxygen, primary production (fuel), and favourable climate (temperature) together might have incited the ECR (Sperling and Stockey 2018).

1.3.2 Great Ordovician Biodiversification Event

Ecological and evolutionary radiations continued after the Cambrian. Diversifications in the Ordovician were so pronounced as to be named the Great Ordovician Biodiversification Event (GOBE). The GOBE entailed an explosion of marine diversity: species increased both in raw count and in occupation of feeding, motility, and water-column ecotypes (Bambach 1983; Bambach, Bush, et al. 2007; Mondal and Harries 2016). The radiation and adoption of pelagic lifestyles in graptoloids (filter-feeding, colonial pterobranch hemichordates) exemplifies this diversification in both taxonomic richness and ecospace (Servais, Perrier, et al. 2016). In addition to these evolutionary and ecological changes, the body size and therefore inferred metabolic demand of several clades, including brachiopods and crinoids, increased during the Ordovician (Finnegan and Droser 2008; Novack-Gottshall 2008). Despite similar patterns across clades, the GOBE was not one synchronous event. Rather, it occurred as a series of diversifications, first in planktonic fauna, then in benthic communities, and finally in reefs (Servais and Harper 2018). The first phase may have begun in the Cambrian with an increase in phytoplankton and zooplankton diversity (Servais, Perrier, et al. 2016). It is still inconclusive whether net primary production rose alongside producer diversity (Pohl, Harper, et al. 2018).

Nutrient enrichment may have instigated the many biotic changes associated with the GOBE (Botting and Muir 2008; Novack-Gottshall 2008; Servais, Lehnert, et al. 2008; Servais, Perrier, et al. 2016). Putative nutrient sources in the Ordovician include mountain building (Miller and Mao 1995); volcanic superplumes (Vermeij 1995; Barnes 2004; Lefebvre, Servais, et al. 2010); strong upwelling along the western coasts of Laurentia and Gondwana (Pohl, Harper, et al. 2018), possibly in response to

global cooling (Rasmussen, Ullmann, et al. 2016); trace metal availability associated with a pulse of oxygenation (Servais, Perrier, et al. 2016); and non-vascular plant expansion on land (Lenton, Crouch, et al. 2012; Porada, Lenton, et al. 2016; Servais and Harper 2018). An increase in detritivore abundance, diversity, and burrowing depth also could have facilitated bottom-up cascades in the Ordovician (Payne and Finnegan 2006; Martin and Servais 2020). By recycling organic matter from sediment and reclaiming detritus from the water column, deposit- and filter-feeding animals can increase nutrient availability to other marine trophic tiers even without an increase in nutrient delivery to oceans from land. This same contribution of burrowers is also suggested as a contributor to the ECR, as discussed above (Erwin and Valentine 2013; Erwin 2015c), but could have played a reprise role in the GOBE and later Paleozoic as burrowing modes continued to deepen (Mondal and Harries 2016; Tarhan 2018). Focussing entirely on biotic processes, some authors have argued for the general importance of competition in driving the GOBE (Harper, Zhan, et al. 2015). However, this idea is unsupported by several studies on specific taxa, which concluded Ordovician interactions were primarily non-competitive (Westrop and Adrain 1998; Finnegan and Droser 2008).

The GOBE was a complex and prolonged event that may have been a continuation of the ECR (Servais and Harper 2018). Cambrian–Ordovician reorganization of phytoplankton remains a strong candidate for enabling radiations at higher trophic tiers later in the GOBE (Servais, Lehnert, et al. 2008). Nevertheless, there are many additional putative causes for the GOBE, both biotic and abiotic, and a combination of factors might have been responsible for such varied evolutionary and ecological changes (Payne and Finnegan 2006; Harper, Zhan, et al. 2015; Pohl, Harper, et al. 2018; Servais and Harper 2018).

1.3.3 Devonian Nekton Revolution

The Ordovician ended in a mass extinction event, which severely curtailed species diversity but minimally pruned ecotypic diversity (McGhee Jr, Sheehan, et al. 2004; McGhee Jr, Sheehan, et al. 2012). Both types of diversity continued to

accumulate in many taxa throughout the later Paleozoic, until the late Devonian, when two slightly more disruptive biodiversity crises occurred at *ca.* 383 and 376 Ma (McGhee Jr, Sheehan, et al. 2004; Racki 2005; McGhee Jr, Sheehan, et al. 2012). One interpretation of Paleozoic ecosystem change suggests that Ordovician species expanded into pelagic habitats as passive drifters and filter feeders, whereas the fauna in the Devonian turned over to well-defended bottom-dwellers and active swimmers in the water column, including jawed fish (Signor and Brett 1984; Bambach 1999; Klug, Kroeger, et al. 2010). Both the processes and patterns of the so-called Devonian Nekton Revolution (DNR) have been contested, however. Chaetognaths, predaceous fish, and other nektonic groups originated before the Devonian, and some scholars have argued the mid-Paleozoic ecosystem change was too incremental to merit the title of revolution (Whalen and Briggs 2018). In recent network analyses and machine learning classifications, the ECR, GOBE, and Mesozoic stand out for evolutionary radiations while the Devonian does not (Cuthill, Guttenberg, et al. 2020; Rojas, Calatayud, et al. 2021). Assuming some amount of ecotypic and evolutionary turnover did occur in the Devonian, bottom-up mechanisms might have contributed to this restructuring.

Rather than glacial melts as in the ECR (Canfield, Poulton, et al. 2007), a top candidate source for nutrient enhancement during the Devonian is terrestrialisation. Land plants might have increased nutrient input through mechanical weathering of rock, acceleration of the hydrologic cycle, or mobilisation of phosphorus from soil acidification (Berner 1997; Algeo and Scheckler 1998; Dahl and Arens 2020; Sperling, Melchin, et al. 2021). Land plants are also proposed as drivers of the DNR through indirect effects on deep-water chemistry in the Devonian, with regional oxygenation or shifts from ferruginous to sulphidic anoxia (Dahl, Hammarlund, et al. 2010; Sperling, Melchin, et al. 2021). Burial of refractory plant material such as wood can increase partial pressure of oxygen in the atmosphere. Similar to proposals for the ECR, Devonian oxygen spikes might have enabled diversification of high-metabolism predators such as large fish (Dahl, Hammarlund, et al. 2010), although this link is disputed (Butterfield 2009). The net effect of terrestrial

vegetation is not straightforward, however, because botanic land cover also protected surface particles from blowing into the ocean, transformed river morphology from dispersed streams to meandering channels, and perhaps lowered the threshold of atmospheric CO_2 required to maintain mass balance of carbon and other elements in Earth's surface system (Gibling and Davies 2012; D'Antonio, Ibarra, et al. 2020; Dahl and Arens 2020).

As a standalone event, the DNR was a lacklustre revolution compared to the functional innovation and rapid speciation of the ECR and GOBE. In broader historical context, however, changes in the Devonian continued a Paleozoic trend of evolution in the water column and strengthened land plants as a lever on Earth surface processes that has persisted through the Mesozoic and Cenozoic.

1.3.4 Mesozoic Marine Revolution

The Paleozoic drama of the ECR, GOBE, and DNR—arguably a continuum of biotic changes rather than discrete events (Harper, Zhan, et al. 2015; Servais and Harper 2018; Whalen and Briggs 2018)—involved progressive expansion into trophic and water column tiers for the first time. In the Mesozoic (252–66 Ma), taxonomic and functional compositions of marine ecosystems turned over more than expanded (Erwin, Valentine, and Sepkoski Jr 1987). Classically, this Mesozoic Marine Revolution (MMR) is discussed in reference to diversification of active hunting modes in predators and of defence strategies in prey (Vermeij 1987). Crinoids evolved mobility at around the same time as their echinoid predators diversified and evolved stronger jaws (Baumiller, Salamon, et al. 2010), and gastropods evolved thicker, spinier shells (Vermeij 1977). Burrowing bivalves radiated, and although infaunalisation can be unrelated to predation pressure (Mondal and Harries 2016), hard-boring forms are linked to predation and diversified beginning in the Mesozoic (Morton 1990). These and other examples have been interpreted as top-down cascades, where grazers and carnivores exerted evolutionary pressure on lower trophic tiers. In addition to such top-down arms races, however, bottom-up mechanisms might have contributed to the MMR.

There are several possible sources of increased marine nutrient concentrations in the Mesozoic. Continental configuration may have played a role: over the course of the era, the supercontinent Pangaea split and the Tethys and then Atlantic ocean opened (Fig. 1.2). Submarine volcanism, long coastlines, wet conditions of interior seaways, and offshore upwelling from circulation changes during this interval could have contributed to enhanced nutrient runoff into oceans (Vermeij 1995; Giraud 2009; Allmon and Martin 2014). The radiation of angiosperms in the Cretaceous is also suggested to have increased terrestrial nutrient runoff (Bambach 1993; Martin 1996). However, the impacts of floral turnover on biogeochemical cycling are not straightforward. Although angiosperms increased the recycling rates of some nutrients (e.g. via annual leaf-fall), it remains unclear whether Mesozoic floral turnover increased the net volume of nutrient input to the marine reservoir—and even if angiosperms did drive a net nutrient loss from land to oceans, much of this terrestrial abundance change might have occurred after its supposed effects in the marine realm (Boyce and Lee 2011). Abiotic processes alone could have fuelled a surge in primary production across much of the Mesozoic (Allmon and Martin 2014), regardless of whether land plants played a smaller role in marine nutrient increases than during the Paleozoic.

Additional bottom-up theories for the MMR centre on a turnover in major phytoplankton taxa. Mixotrophic dinoflagellates radiated in the Jurassic, and their large cell size could have increased energy flow to top trophic tiers (Knoll and Follows 2016); in addition, diatoms radiated from the Cretaceous, and their efficient growth could have produced more biomass at the base of food webs (Knoll and Follows 2016). Dinoflagellates, diatoms, and other secondary and tertiary symbiotic algae primarily derive plastids from red algae and use chlorophyll *c* or rarely *b* as an accessory pigment to chlorophyll *a*. Green algae use chlorophyll *b* as an accessory to chlorophyll *a*. Differences in host cell phenotypes and plastid origins lead to differential uptake of nutrients and trace metals (Quigg, Finkel, et al. 2003). Chlorophyll *c* algae are denser than chlorophyll *b* algae in phosphorus, nitrogen, and the subset of trace metals that are more bioavailable in oxygenated waters, such as

cadmium and cobalt (Quigg, Finkel, et al. 2003). The Paleozoic–Mesozoic turnover of chlorophyll *b* to chlorophyll *c* algae therefore would have shifted the elemental stoichiometry of primary production, potentially increasing biomass quality at all trophic levels (Martin and Quigg 2012; Allmon and Martin 2014). Mesozoic oxygenation, evidenced by positive carbon isotope excursions (Hayes, Strauss, et al. 1999), would have rebalanced metal availability in oceans and might have ushered in the proliferation of chlorophyll *c* algae (Quigg, Finkel, et al. 2003; Martin and Quigg 2012). Ultimately, both nutrient addition and phytoplankton evolution could have contributed to the morphological and taxonomic diversity of animal consumers that increased during the Mesozoic.

1.4 A framework for bottom-up ecosystem controls

Increased biomass, abundance, taxonomic diversity, feeding and motility modes, and metabolic rate are all discussed as outcomes of bottom-up processes (Bambach 1993; Vermeij 1995; Martin 1996; Bush and Bambach 2011; Allmon and Martin 2014). Although it is possible for these changes to occur in lockstep, different mechanisms likely underpin such disparate outcomes. I identify eleven potential stepping-stones in bottom-up ecological pathways (Fig. 1.3, Table 1.1), where each stepping-stone can lead to multiple knock-on consequences at a higher trophic tier. This framework connects nutrient concentrations to primary producers to first-order consumers to higher-order consumers (Fig. 1.1). In the pelagic realm, these three trophic tiers correspond roughly to phytoplankton (primary producers), zooplankton (pelagic consumers with passive dispersal), and nekton (active swimmers such as large predators). Multiple tiers of trophic consumers live above, on, and within the seabed as well.

A tripartite division is a simplification of marine food webs. Many organisms fall between these categories, such as facultative heterotrophs (e.g. some planktonic foraminifera) and the nektoplankton (weak swimmers, e.g. Euphausiacea, krill). In these cases, multiple of the mechanisms discussed for individual trophic tiers may

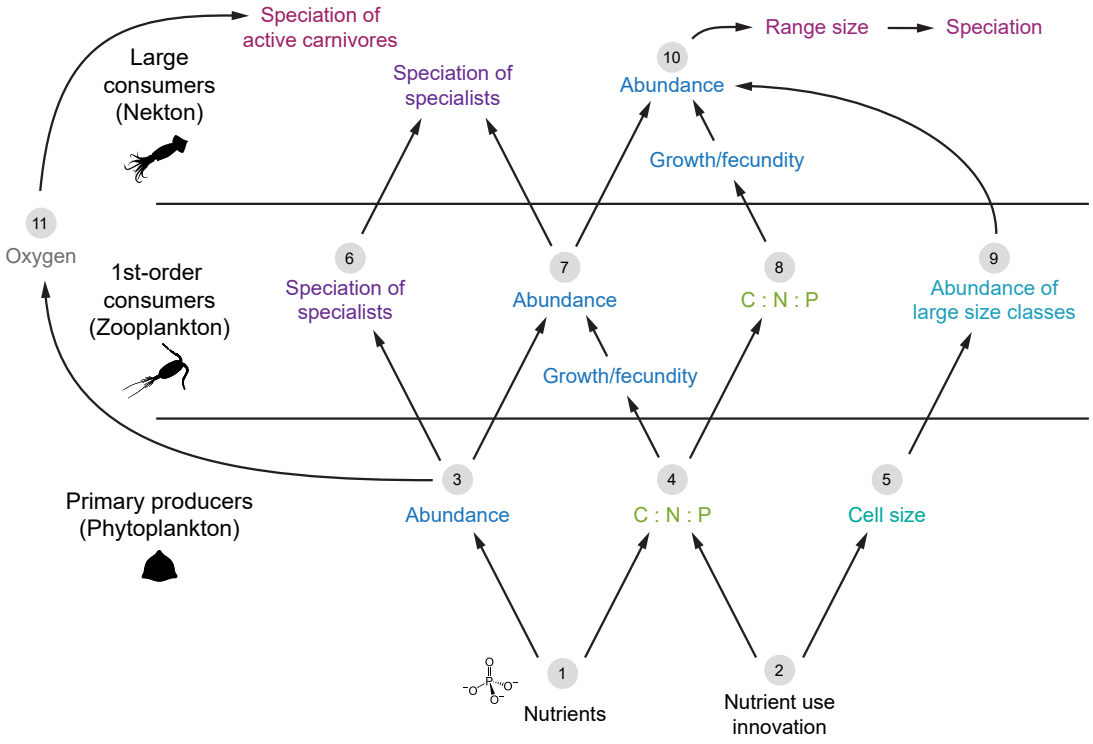


Figure 1.3: Ecological variables that affect successive trophic levels are numbered as stepping-stones in this simplified flowchart of bottom-up pathways. Initial triggers of bottom-up change include increased nutrients (stepping-stone #1) or evolutionary innovation in nutrient acquisition or metabolism (stepping-stone #2). Subsequently, these nutrient-related changes affect primary producers, first-order consumers, and higher-order consumers. Times in Earth history where each step is thought to have taken place are listed in Table 1.1. Top-down pathways are not shown but can co-regulate abundances (stepping-stone #3 and #7) and influence prey behaviour. Carbon, C; nitrogen, N; phosphorus, P.

apply. An additional influence of facultative heterotrophs is discussed in 'Phytoplankton evolution' below. Other organisms fall beyond the above categories altogether, such as chemoautotrophs (e.g. many bacteria and archaea) and detritivores (e.g. Holothuroidea, sea cucumbers). Chemoautotrophs may avoid dependence on marine food webs entirely, while detritivores can source organic material across every trophic tier (Fig. 1.1). Viruses and parasites also shape marine food webs (Rohwer and Thurber 2009; Roux, Brum, et al. 2016) but leave few traces in the fossil record.

Table 1.1: The bottom-up stepping-stones numbered in Fig. 1.3 are listed in order of increasing trophic level. Most effects are themselves the cause of subsequent changes, and the index numbers associated with these knock-on effects are included in parentheses. Time intervals where a given change is hypothesised to have occurred are listed at far right. Phosphorus, P; nitrogen, N; carbon, C; ↑, increase and ↓, decrease in count, concentration, or rate.

	Trophic level	**Change**	**Direct effect(s)**	**Occurrences**
1	Nutrients	↑ nutrient flux to oceans	↑ phytoplankton abundance (3) ↑ P and N relative to C in cells (4)	Cambrian, Ordovician, Devonian, Mesozoic
2	Phytoplankton	Key innovations in nutrient use and storage	↑ P and N relative to C in cells (4) ↑ cell size (5)	Cambrian, Mesozoic
3		↑ abundance	↑ zooplankton abundance (7) ↑ speciation of zooplankton specialists (6) ↑ O_2 (11)	Cambrian, Ordovician, Devonian, Mesozoic
4		↑ P and N relative to C in cells	↑ P and N relative to C in zooplankton (8) ↑ growth, ↓ generation time, ↑ abundance of zooplankton (7) ↑ fecundity, ↑ abundance of zooplankton (7)	Mesozoic
5		↑ cell size	↑ abundance of large zooplankton (9)	Neoproterozoic, Mesozoic
6	Zooplankton	↑ speciation of specialists	↑ speciation of nekton specialists	Cambrian, Ordovician, Devonian, Mesozoic

Continued on next page

Table 1.1 Continued from previous page

	Trophic level	Change	Direct effect(s)	Occurrences
7		↑ abundance (all size classes)	↑ speciation of nekton specialists	Cambrian, Ordovician, Devonian, Mesozoic
			↑ abundance of nekton (10)	
8		↑ P and N relative to C	↑ growth, ↓ generation time, ↑ abundance of nekton (10)	Mesozoic
			↑ fecundity, ↑ abundance of nekton (10)	
9		↑ abundance of large zooplankton	↑ abundance of nekton (10)	Mesozoic
10	Nekton	↑ abundance	↑ geographic range size, ↑ speciation	Devonian, Mesozoic
11	Atmosphere and oceans	↑ O_2	↑ speciation of high-metabolism carnivorous nekton	Cambrian, Devonian

1.4.1 Nutrient flux

Approximately half of net primary production today takes place in the marine realm (Field, Behrenfeld, et al. 1998). Nutrients limit or co-limit production in many parts of the oceans, as evidenced by blooms after nutrient addition (De Baar, Boyd, et al. 2005; Thingstad, Krom, et al. 2005; Boyd, Jickells, et al. 2007). This relationship is illustrated in Fig. 1.3 as the link between nutrients (stepping-stone #1) and phytoplankton abundance (stepping-stone #3). As mentioned in preceding sections, the greater availability of certain metals after oxygenation pulses is hypothesised to have contributed to the ECR, GOBE, and MMR (Quigg, Finkel, et al. 2003; Martin and Quigg 2012; Servais, Perrier, et al. 2016). There are many macronutrients and micronutrients, including essential elements such as iron, cobalt, and molybdenum. In modern oceans, nitrogen is well documented to limit primary production in many local regions on the scale of days to weeks—but ultimately, it is phosphorus

that limits marine production on large spatiotemporal scales (Tyrrell 1999; Sañudo-Wilhelmy, Kustka, et al. 2001; Moore, Mills, et al. 2013). Nitrogen concentrations dictate local population fluctuations about a long-term carrying capacity, while phosphorus concentrations set that capacity in many areas because there are fewer ways organisms can compensate for low phosphorus availability.

Marine organisms can source bioavailable nitrogen either from terrestrial runoff or from bacteria fixing atmospheric N_2 gas. However, runoff is the near-exclusive source of new marine phosphorus. Ash-fall and wind-blown dust contribute a smaller volume of phosphorus on the global scale, and volcanoes can enrich phosphorus concentrations locally (Benitez-Nelson 2000). Because terrestrial fluxes are so important to the phosphorus cycle, abiotic and biotic events on land have large potential to influence marine chemistry. Additionally, some changes that enhance terrestrial delivery of phosphorus to oceans could also enhance delivery of nitrogen and trace metals, further nourishing oceans. In the mid-20$^{\text{th}}$century, Helen Tappan wrote about this connection between terrestrial nutrients and marine productivity (Tappan 1968; Tappan 1986), thereby sparking much of the later inquiry into bottom-up controls across the Phanerozoic. As detailed above in the 'History of marine ecosystems' (section 1.3), developments on land relevant to oceans included multiple phases of mountain building, glacial scouring and melting, and expansion of plant root networks. Before the establishment of soils, runoff was unimpeded, so Precambrian and Paleozoic glacial and volcanic activity could have had an impact on oceans disproportionate to any such change observable today (Filippelli 2010). Therefore, nutrient addition, especially of phosphorus, was a potentially strong contributor to the ECR and GOBE.

Unfortunately, although the sources and sinks of nutrient cycles are known in qualitative detail, it is difficult even in the present day to quantify net flux of phosphorus and other nutrients to oceans globally (Benitez-Nelson 2000; Bennett, Carpenter, et al. 2001). Uncertainty stems in part from the complicated balance of burial *vs.* remineralisation on the seafloor (Ruttenberg and Berner 1993; Benitez-Nelson 2000) and from spatial heterogeneity in rates of weathering (Hartmann,

Moosdorf, et al. 2014) and of precipitation in rivers, estuaries, and shelves (Benitez-Nelson 2000; Sharples, Middelburg, et al. 2017). Recently, plankton have been found capable of recycling phosphorus through reduced forms, which may be an important reservoir in the marine phosphorus cycle; however, it remains unknown how environmental conditions affect biologic production of these reduced compounds (Van Mooy, Krupke, et al. 2015). The cumulative imprecision in quantifying marine phosphorus availability is even greater in deep time, challenging efforts to estimate even the direction of change (e.g. whether more or fewer nutrients reached the open ocean at a given time), let alone the magnitude of change (Allmon and Martin 2014). Substantial new geochemical research is needed to constrain the timing, duration, and location of nutrient addition to oceans both past and present.

1.4.2 Phytoplankton evolution

Even without an increase in nutrient flux to oceans, there are at least two ways to increase per-cell macronutrients in phytoplankton (stepping-stone #2). First, clades with denser macronutrient profiles could proliferate (stepping-stone #4). The relative amounts of phosphorus, nitrogen, and other elements are deeply conserved in algal lineages, and the clades with highest relative abundance today have lower carbon : nitrogen : phosphorus ratios (i.e. more phosphorus and nitrogen per unit dry mass) than clades that were most abundant during the Paleozoic (Quigg, Finkel, et al. 2003; Martin and Quigg 2012; Knoll and Follows 2016). This compositional turnover took place in the Mesozoic, when chlorophyll *a* and *c* algae such as diatoms replaced green algae such as prasinophytes in abundance ('Mesozoic Marine Revolution,' section 1.3.4 above). The richer nutrient profile of phytoplankton that resulted from algal radiations has cascading consequences for consumer fitness (discussed in 'Population abundance,' 1.4.3 below) and is suggested to have facilitated the ECR and MMR (Table 1.1) (Elser, Watts, et al. 2006; Martin and Quigg 2012).

Second, cell size distribution could shift rightward, with more cells at larger size fractions (stepping-stone #5). Generally, organism size increases at higher trophic levels, and so raising the size of primary producers at the base of food

chains might increase the carrying capacity of the largest zooplankton consumers (stepping-stone #9), effectively truncating the low-level consumer portion of food chains (Finkel 2007; Knoll and Follows 2016). Since energy transfer efficiency across trophic levels is low—usually less than 15% (Kemp and Boynton 2004; Dickman, Newell, et al. 2008)—bypassing low-level trophic tiers means a larger energy input to apex predators at the highest tiers (stepping-stone #10).

The size-based hypothesis is put forward as another bottom-up facilitator over marine ecosystem history, with initial cell size increases resulting from the evolution of mixotrophy in dinoflagellates during the Mesozoic and from fast-growing diatoms in the Cenozoic (Knoll and Follows 2016). Ecosystem models suggest that mixotrophy does increase cell size and trophic energy transfer efficiency (Ward and Follows 2016). At the same time, mesocosm experiments demonstrate that trophic energy transfer efficiency reciprocally affects food chain length: more efficient energy transfer can lengthen food chains, but the addition of higher-order consumers reduces the energy transfer efficiency of first-order consumption (Dickman, Newell, et al. 2008). It is difficult to directly estimate energy transfer efficiency in deep time, although trophic network properties derived from fossil and modern marine assemblages match closely (Dunne, Williams, et al. 2008; Roopnarine and Dineen 2018). The evolution of mixotrophy and of diatoms are milestones in marine history, but there are too many trade-offs in ocean food web dynamics to assign these events singular importance in driving post-Paleozoic ecosystem change.

Evolution in microbes can and does occur independently of environmental change or predation pressure. In tens of thousands of generations of *Escherichia coli* cultured under controlled laboratory conditions, growth rate and cell volume have increased steadily in every one of 12 replicate strains (Lenski, Wiser, et al. 2015). There are both adaptive and non-adaptive explanations proposed for this size increase, which is linked to changes in other physical properties and gene expression patterns (Philippe, Pelosi, et al. 2009). A lesson to remember when considering the history of microbial life is that environmental change is sufficient but not necessary for traits to evolve that have repercussions for higher trophic levels.

1.4.3 Population abundance

Abundance of phytoplankton ties intimately to that of zooplankton and higher-order consumers (stepping-stone #7 and #10). Understanding the bottom-up influence of nutrients *vs.* top-down influence of zooplankton on phytoplankton biomass is of immense interest to biological oceanography in the Recent, because marine primary producers are a major carbon reservoir in the biosphere with an uncertain future trajectory (Field, Behrenfeld, et al. 1998; Ridgwell and Zeebe 2005; Behrenfeld 2014). Since consumers represent a minor fraction of marine carbon, ocean and climate scientists sometimes lump together predator consumption, viral lysis, and other forms of phytoplankton losses under a single parameter (Behrenfeld 2014). However, to evaluate phytoplankton biomass increases as a candidate cause of ecological change in zooplankton and nekton throughout geological time, palaeoecologists need estimates on the specific amount of phytoplankton consumed by higher trophic levels. Mesocosm and ocean fertilisation experiments can address this question.

Experiments on living organisms demonstrate that zooplankton fed nutrient-rich phytoplankton become dense in nutrients themselves and pass these nutrients up to nekton (stepping-stone #8); conversely, in some cases nekton can become limited by the same nutrient as are phytoplankton (Malzahn, Aberle, et al. 2007; Dickman, Newell, et al. 2008). Animals have a partial ability to buffer this effect ('relaxed homeostasis'), particularly to raise their nutrient consumption when fed low-nutrient diets. For instance, *Daphnia* forage with greater fervour when fed phosphorus-poor diets to burn off carbon while increasing phosphorus intake (Plath and Boersma 2001). However, there may be fewer physiological mechanisms to mitigate diets with above-average nutrient density: phosphorus far in excess of levels to which consumers are adapted decreases growth and survival rates (Elser, Schampel, et al. 2005; Boersma and Elser 2006). Study of elemental stoichiometry as a control on growth rate and body condition, and thence generation time and fecundity, has developed into its own ecological subfield, with applications from aquaculture to palaeoecology (Sterner and Elser 2003; Elser, Watts, et al. 2006; Malzahn, Aberle, et al. 2007; Dickman, Newell, et al. 2008).

An experimental example of nutrient addition to marine food webs was an artificial phosphorus pulse made in an ultraoligotrophic patch of the Mediterranean (Thingstad, Krom, et al. 2005). This study stands out because it is one of few ocean fertilisation experiments that monitored zooplankton population responses (De Baar, Boyd, et al. 2005; Boyd, Jickells, et al. 2007). As expected from bottom-up theories, copepods used the extra dietary phosphorus to produce substantially more eggs (Thingstad, Krom, et al. 2005). However, the route by which phosphorus reached zooplankton was unexpected. Rather than stimulating primary production, the phosphorus fertilisation caused a decline in chlorophyll concentration and rise in bacterial production. The added phosphorus might have bypassed phytoplankton and instead travelled primarily through heterotrophic bacteria, to flagellates, to ciliates, to copepods (the microbial loop, Fig. 1.1). In addition or instead, both heterotrophic bacteria and phytoplankton could have stored the surplus phosphorus, increasing the quality rather than quantity of primary producer cells. This experiment holds several lessons for historians of life. First, the study demonstrates that the uses to which cells put nutrients are not always so straightforward as linear conversion of elements to biomass. Second, the results give greater importance to the quality than quantity of primary production, which might favour bottom-up hypotheses of phytoplankton evolution discussed in the previous section (1.4.2).

Nutrient addition experiments in a freshwater system led to similar conclusions of food quality determining ecosystem trajectories in deep time (Elser, Watts, et al. 2006). Researchers measured growth rate and survival of snail grazers along a gradient of phosphorus supplementation in a stromatolitic system, one of the closest modern analogues to Cambrian shallow ecosystems (Elser, Schampel, et al. 2005; Elser, Watts, et al. 2006). Snails appeared limited by low nutrient density but not by mass of primary production. This result was interpreted as support for the hypothesis that a phosphorus flux might have enabled Cambrian metazoan diversification (Planavsky, Rouxel, et al. 2010; Martin and Quigg 2012).

Short-term experiments can reveal complex pathways of energy flow, but the local community responses on the scale of weeks do not necessarily translate to

species-level effects on the scale of millennia or longer. Authors of nutrient addition experiments themselves have been quick to point out the dangers of spatiotemporal extrapolation (Thingstad, Krom, et al. 2005; Elser, Watts, et al. 2006). In the Mediterranean, for instance, the results documented might be specific to seasonal conditions of aeolian nutrient pulses, which cause phosphorus to be limiting for bacteria (Thingstad, Krom, et al. 2005). There is a clear need for nutrient addition experiments on intergenerational scales, to approximate steady-state population responses rather than individual plasticity in reaction to transient perturbations. In one such longer-term study, researchers transplanted guppy populations between predation-controlled and food-limited habitats (Reznick, Bassar, et al. 2019); they found lags in population evolution that would have been overlooked by a shorter study but were consistent with theory for life-history trait evolution.

Even if more laboratory and fertiliser experiments scale up in time, they will still be limited in space. There is a gap between the small spatial scale of manipulation and the larger scale at which both bottom-up and top-down effects can manifest in ecosystems (Estes, Terborgh, et al. 2011). Limnologists have long been aware that a phytoplanker response to predation in a mescocosm may be opposite to that in a whole lake, for instance (Carpenter and Kitchell 1988). Nevertheless, it is possible to generate results in mesocosms that concur with those found in the broader ecosystems they replicate, in both terrestrial and marine settings. The vital qualities for experimentation are (i) identification of fundamental mechanisms (e.g. herbivore behaviour response to predators), which can be scale-independent; (ii) careful design of enclosures (e.g. ensure relative dimensions do not curtail fish growth); and (iii) measurement in both field sites and mesocosms to corroborate results (Petersen, Kemp, et al. 2003; Schmitz 2005; Zhong, Li, et al. 2017). In fossil studies it is also important to recognise the role of both temporal variability and spatial variability—a shift from evaluating 'diversity through time' towards 'diversity through time and space' (Benson, Butler, et al. 2021).

1.4.4 Consumer speciation

There are both direct and indirect ways elevated productivity could increase speciation rate at higher trophic tiers. Directly, when a population increases in abundance it might expand geographically, become isolated across dispersal barriers, and undergo allopatric speciation (stepping-stone #10). Indirectly, an increase in the biomass within a trophic level could make it viable for new consumer species to specialise on resource types from that level (stepping-stone #6). At the core of both ideas is the question: how do ecological changes at the individual and population level translate into evolutionary changes within and between lineages?

Consumer abundance responds to primary production quality and quantity, as discussed in the previous section (1.4.3). In turn, abundance is strongly correlated with geographic range size, and so range expansion is a plausible outcome of population growth—although exceptions may occur if a species is limited by habitat availability or if niche breadth is ultimately the common cause of both abundance and geographic range size (Gaston and Lawton 1990; Gaston, Blackburn, et al. 1997). Since the number of peripheral populations and the chance of intersecting a geographic dispersal barrier both increase with range size, some biologists postulate a positive relationship between distribution area and speciation probability (Rosenzweig 1995; Maurer and Nott 2001). However, traits that promote broad distributions (e.g. fast dispersal rates) could also make it harder for populations to become isolated; therefore, the distribution–speciation relationship could be negative (Hansen 1980; Jablonski 1986). Empirical work in marine animals supports a negative distribution–speciation relationship (Jablonski and Roy 2003) yet finds weak concurrence between dispersal rate and range size (Gaines, Lester, et al. 2009). Accommodating these conflicting ideas and data, some authors propose a unimodal relationship between geographic range size and probability of successful speciation: species of intermediate distribution may split the most (Gaston 2003; Allmon and Martin 2014). In summary, population growth could cause geographic expansion, but a larger range area might either increase or decrease speciation rate, perhaps depending on the absolute range area.

The abundance–distribution–speciation relationship (the 'speciation cycle') has been put forward as a major mechanism linking primary production to diversifications across the history of life, including the revolutions described above (Allmon and Martin 2014). The merit of this idea is its explicit elaboration of every causal step, which may allow robust empirical tests in future. A challenge to the importance of the speciation cycle is the difference in timescale between its cause and effect. Phytoplankton and zooplankton abundances respond to environmental change on the scale of weeks to months, and all plankton, including planktonic larvae of benthic organisms, can ride currents across the planet on the scale of years. In contrast, the timescale of evolutionary radiations has been clocked at millions of years or more; for instance, the successive diversifications of planktonic, benthic, and reefal species in the GOBE spanned 40 million years (Servais and Harper 2018). Although abundance-driven speciation may be a weaker explanation of this prolonged change than previously thought, an increase in primary production could still have facilitated past diversifications via opening ecospace.

Palaeobiologists have identified a notable increase in marine ecospace occupation throughout the half a billion years since animals evolved, e.g. burrowing deeper into the sediment and swimming higher into the water column (Bambach, Bush, et al. 2007; Bush and Bambach 2011). One explanation for this pattern is that an increase in the number or size of prey at a trophic level made it possible for specialised predators to evolve at the trophic level above. The 'more specialization' hypothesis posits generally that specialists evolve whenever a category of food resource attains sufficient energetic density to sustain them (Abrams 1995; DeAngelis 1995; Srivastava and Lawton 1998). Resource expansion thereby can support the existence of a new species in communities, but the actual process of speciation could be due to any number of selective pressures or chance events, e.g. sexual selection, character displacement, or the emergence of a geographic barrier.

Adaptive radiation and niche construction are two areas of theory that seek to explain why species diversify into areas of niche space. Each idea helps structure thinking about causal processes behind speciation but is difficult to demonstrate

empirically. Evolutionary biologists invoke adaptive radiations frequently to explain rapid tree-branching, yet definitively eliminating alternative explanations for historical branching events can prove difficult or even insurmountable (Erwin 2015b; Wagner 2017). The strength of adaptive radiation literature lies in linking macroevolutionary patterns to evolutionary developmental biology. In recent formulations, ecological opportunity is rarely the first step in adaptive radiations but instead the last, lagging behind the initial steps of (i) potentiating or permissive mutations, (ii) actualisation or appearance of a novel trait, and (iii) adaptive refinement of that trait (Blount, Borland, et al. 2008; Erwin 2015b; Wagner 2017). In contrast to adaptive radiation, niche construction theory links population genetics to ecosystem changes as simultaneous processes, arguably as important as natural selection (Laland, Odling-Smee, et al. 1999; Laland and Sterelny 2006; Scott-Phillips, Laland, et al. 2014).

The salient feature of niche construction theory is that when organisms change their environment, they also change the evolutionary landscape of selective pressures on their own species and others. Niche construction theory applies equally well when ecosystem change occurs in food chains as in the abiotic environment (Laland, Odling-Smee, et al. 1999) and has also been applied in archeology to describe human modification of the natural environment, e.g. the origins of agriculture (Spengler III 2021). Niche construction is effectively synonymous with ecosystem engineering but is used in contexts that emphasise the evolutionary ramifications of ecosystem change. Detractors argue that although niche construction undoubtedly occurs, it fails to generate either specific, testable predictions or explanatory power beyond what classical theory can provide (Laland and Sterelny 2006; Scott-Phillips, Laland, et al. 2014; Spengler III 2021).

Labels of adaptive radiation and niche construction have been applied to many radiations including those of the ECR and GOBE. However, critics say ascribing the enormity of morphological changes in so many Cambrian taxa to adaptive radiation 'stretches the bounds of that term beyond recognition' (Erwin 2015b, *p.* 5). In contrast, Neoproterozoic oxygenation can be considered ecosystem engineering

that could easily influence evolutionary pressures across marine clades (Erwin and Valentine 2013; Erwin 2015a). Discussions of niche construction for the GOBE tend to be more implicit than for the ECR, but commentary is often couched in this interpretation (e.g. Botting and Muir 2008). Before evolutionary biologists can test any hypothesis for a given radiation, the assumptions and causal steps involved must be elaborated—explanations should be explicit. Ultimately, the relative contributions of adaptive radiation, niche construction, species selection, sexual selection, and other evolutionary processes may vary with every clade and historical context.

1.4.5 Body mass and metabolism

Alongside intervals of diversification, palaeontologists have suggested body mass and metabolic rate have risen through time (Bambach 1993; Bush and Bambach 2011; Allmon and Martin 2014). These latter traits are grouped under the category of energetics for their tight correlation. Palaeontologists infer changes in ecosystem energetics in deep time by measuring fossil body size, which can be substituted into metabolic rate scaling equations calibrated on modern descendants. This approach has demonstrated that after many animal clades suddenly appeared in the earliest Cambrian deposits, their body mass seems to have remained equivalent through the period (Novack-Gottshall 2008), at least in minimum and maximum size (Heim, Knope, et al. 2015). During the GOBE, however, there were increases in body size, abundance, and species count in brachiopods—a major component of the benthic fauna (Finnegan and Droser 2008; Novack-Gottshall 2008; Novack-Gottshall and Lanier 2008; Payne, Heim, et al. 2014; Zhang, Augustin, et al. 2015). Trilobite species that originated or survived into the end of the period tended to be larger than their predecessors and may have occupied a higher trophic level (Finnegan and Droser 2008). Brachiopods, trilobites, and crinoids continued to grow in body mass until at least the Early Devonian (Novack-Gottshall 2008; Payne, Heim, et al. 2014; Zhang, Augustin, et al. 2015). Across the Mesozoic, gastropod metabolic rates are estimated to have increased by 150% (Finnegan, McClain, et al. 2011).

Gastropods have been a prominent taxon in discussions of the MMR because of their vulnerability to crushing predators and their evolution of shell defences during this era (Vermeij 1977). Drilling gastropods in particular have grown in size consistently from the Ordovician to the Recent (Klompmaker, Kowalewski, et al. 2017).

Although many clades studied for their size evolution over the Phanerozoic are primarily filter-feeders and detritivores, predators such as many fish and marine mammals have also increased in mean body size through time (Heim, Knope, et al. 2015). Many large chordates are active predators, and their size is again theorised to be affected by phytoplankton production, because hunting incurs a high metabolic cost. The proportion of carnivorous taxa decreases in progressively lower-oxygen habitats today, and therefore ubiquitous anoxia or suboxia may have precluded active hunters prior to the Neoproterozoic Oxidation Event (Sperling, Frieder, et al. 2013). Primary producers are strongly responsible for regulating O_2 and CO_2 concentrations in the atmosphere on geologic timescales (stepping-stone #11), and so aerobic facilitation of apex predators could be considered another kind of bottom-up control. In addition to the ECR, the middle Paleozoic has been suggested as another interval where oxygenation facilitated evolution of the huge, metabolically-active animals of the time (Dahl, Hammarlund, et al. 2010; Heim, Knope, et al. 2015), although oxygen limitation after the Cambrian is more contentious.

Increased primary production has been proposed as a direct cause of energetic escalations across taxa in the Paleozoic and Mesozoic regardless of oxygen levels (Bambach 1993; Bambach 1999; Novack-Gottshall 2008; Finnegan, McClain, et al. 2011). The biological mechanisms underlying this link remain elusive, however. Why would greater density of resources lead consistently to larger body mass instead of to faster growth rate and earlier reproduction, for instance? Of course, it is possible for all these outcomes to co-occur: in living bivalves, faster growth leads to both faster maturation and larger adult size across a latitudinal gradient of temperature and food availability (Moss, Ivany, et al. 2016). Either earlier reproduction or greater fecundity associated with larger adult size would then

increase population abundance as well (stepping-stones #7 and #10), with further ecosystem consequences as discussed in previous sections.

Energetic effects of phosphorus, oxygen, and primary production volume should be considered in concert because they might have increased in lockstep. Low primary production may be essential to maintaining low atmospheric oxygen concentration, e.g. during low-phosphate spans of the Proterozoic (Derry 2015; Laakso and Schrag 2019). In modern marine ecosystems, independent limitation of oxygen and food are both associated with small body size and non-carnivorous feeding modes, and therefore it may not yet be possible to differentiate the relative importance of tightly correlated oxygen and primary production amounts for consumer body size and metabolic rate increases across life history (Sperling and Stockey 2018).

The evolution of growth rate, body size, fecundity, and other life history traits involves many trade-offs, including the balance of competing for resources *vs.* surviving predation (Reznick, Bassar, et al. 2019). The history of body size evolution is particularly troublesome to disentangle because size has so many correlates and causes, both biotic and abiotic (Smith, Payne, et al. 2016). Nevertheless, evidence suggests the orders-of-magnitude increase in Phanerozoic animal size reflects change in community composition rather than parallel evolution to larger size within many lineages (Novack-Gottshall and Lanier 2008; Heim, Knope, et al. 2015). In other words, radiation seems to occur preferentially in clades with larger ancestors. To understand this macroevolutionary pattern will require testing frameworks such as adaptive radiation and niche construction theory (section 1.4.4).

1.5 Other mechanisms of change

Many types of biological processes operate in parallel and interactively to generate ecological and evolutionary diversity. Bottom-up processes alone cannot explain fully the variety and richness of ecosystem history. Top-down controls, sexual selection, global-scale ecosystem engineering, Earth surface processes, extraterrestrial impacts, and other mechanisms should be considered as well when interpreting past ecological change.

In modern ecosystems top-down and bottom-up controls are known to interact with each other strongly and synergistically (Worm, Lotze, et al. 2002), although some communities are exceptions (Dyer and Letourneau 2003). For instance, to predict phytoplankton blooms over months, unravelling the balance of top-down *vs.* bottom-up control is cited as a key area for further research (Behrenfeld 2014). On geologic timescales over which species evolve, a compounding interaction between top-down and bottom-up control could take place: once bottom-up processes increase predator abundance, diversity, or energetics, arms races with prey species could ensue (Vermeij 1987). In this way, bottom-up processes could stimulate top-down processes that then unfold concurrently.

In addition to interactions among ecological processes, reciprocal relationships occur between life and Earth systems. Sections above note repeated instances of volcanic, tectonic, and other geologic events leading to sustained ecological change (Vermeij 1995; Rasmussen, Ullmann, et al. 2016). I have also discussed examples of biotic events affecting geologic processes, such as oxygenic photosynthesisers changing ocean chemistry and land plants accelerating weathering (Dahl and Arens 2020). Biological and Earth processes have mutually buffered and enabled each other in myriad other ways since life arose; these couplings occur across scales of space, time, and biological organisation (Laland and Sterelny 2006; Erwin 2015a; Ochoa-Hueso, Delgado-Baquerizo, et al. 2021). Perhaps nowhere are these interactions more apparent than in tracing the modern carbon cycle, which encompasses feedbacks between primary production, biomass, respiration, detritus, soil, atmospheric composition, global temperature, ice volume, calcium carbonate concentration, sediment burial, rock formation, weathering, wildfires, and now human emissions (Ridgwell and Zeebe 2005).

1.6 Discussion

The history of animal life in the ocean has involved parallel increases in marine nutrients, primary production, abundance, diversity, niche space occupancy, and metabolic rate. Rather than a monotonic, lockstep increase, these changes have

varied in pace through time and relative to each other. Notable intervals of concerted change were the ECR, GOBE, DNR, and MMR. These events share similarities but also differ in both putative causes and outcomes (Table 1.1). An increased rate of nutrient delivery from land is theorised for all four events, leading to increased primary production in oceans. However, the source, magnitude, and duration of primary production increase are specific to a given time interval.

Empirical knowledge of the ECR remains insufficient to state the extent to which radiation in animal diversity, trophic tiering, and water column occupancy were due to nutrient addition, oxygenation, both, or neither. Multiple interpretations of available data are plausible. Many lines of study can contribute to disentangling Ediacaran–Cambrian events, including geochemical and geophysical research on Neoproterozoic glaciations, palaeomagnetic reconstructions of continent position and configuration, and taxonomy and systematics of Ediacaran and Cambrian faunas. Physiological experiments on living organisms would complement these Earth science approaches, particularly with regard to the synergistic stresses of the metazoan fire triangle—low oxygen, restricted primary production and food supply, and warming climate (Sperling and Stockey 2018).

The GOBE and DNR continued filling the water column and sediment depths after the Cambrian. The Ordovician was a time of especially rapid diversification of species and ecotypes (Bambach 1983; Bambach, Bush, et al. 2007; Mondal and Harries 2016). Infaunalisation during the Paleozoic may have been key to recirculating more nutrients back into food webs (Payne and Finnegan 2006; Martin and Servais 2020), although a succession of volcanism, mountain building, land plant evolution, and glaciation also could have input nutrients to oceans (Miller and Mao 1995; Vermeij 1995; Algeo and Scheckler 1998; Raymond and Metz 2004). I have focussed on increases in functional and taxonomic diversity, but it is also interesting to consider the end-Ordovician and late Devonian mass extinctions with respect to primary production. Phytoplankton booms might facilitate population growth and speciation at higher trophic levels, yet nutrient addition and carbon burial beyond a threshold could tip oceans into a eutrophic state, drive subsurface

anoxia, and cool the global climate into icehouse conditions (Algeo and Scheckler 1998; Racki 2005; Dahl and Arens 2020). Hence, for the middle Paleozoic, nutrients and primary production might explain extinctions more than they do diversification.

Many lines of evidence place the MMR as a strong example of bottom-up ecological cascades, independent of strong top-down controls that probably occurred synchronously. Nutrient pulses related to continental fission, and possibly angiosperm evolution, are candidate triggers of phytoplankton biomass increase (Allmon and Martin 2014). The evolution and proliferation of chlorophyll *a* and *c* algae groups, such as dinoflagellates and diatoms, could have increased per-cell nutritional content (Martin and Quigg 2012; Knoll and Follows 2016). Based on results from ecological experiments, algal quality changes might exert stronger leverage on ecosystem complexity than do quantity changes, but as yet there is insufficient evidence to make this claim for the Mesozoic.

Strong theoretical and empirical links exist between nutrient concentrations and phytoplankton production. However, it becomes more difficult to predict the importance of bottom-up mechanisms at higher trophic levels. Nutrient addition experiments demonstrate the feasibility of cascades in biomass quantity and stoichiometry, from phytoplankton through to nekton (Dickman, Newell, et al. 2008), but there are many caveats to extrapolating the results of such studies in space and time. The data gap between micro scales (days to years, mesocosms to ocean patches) and macro scales (millennia to millions of years, ocean basins to biosphere) remains a black box for mechanisms of diversification, particularly in evolutionary diversifications interpreted as adaptive radiations or involving body size increases (Erwin 2015b). Future studies should work to close this gap from both sides. In living systems, insight into intermediate scales is attainable by extending the duration of nutrient addition experiments to multiple generations and all trophic levels, including secondary-or-higher predators like fish (Dickman, Newell, et al. 2008; Reznick, Bassar, et al. 2019). Using the fossil record, studies can examine the history of basins with high sedimentation rates and anoxic seabeds,

where it is possible to reconstruct historical pelagic ecology on decadal to centennial scales (Bringué, Pospelova, et al. 2014).

Aquatic ecologists led the way in quantifying the relative importance of bottom-up and top-down controls, and terrestrial ecologists eventually adopted this view of co-regulation (Hunter and Price 1992; Hunter 2001). In considering the whole of animal history, it is important to continue this holistic line of thinking. Unfortunately, it is no small feat to tease apart the relative contributions of bottom-up and top-down controls to speciation events and abundance changes that took place up to hundreds of millions of years ago. Further complicating inferences, any specialisation in predators that involved adaptations such as enhanced sensory perception and agility could have contributed to (and been amplified by) sexual selection processes (Bush, Hunt, et al. 2016). One aid to cut through such generalisations is to articulate mechanistic hypotheses rather than treat bottom-up processes as a monolith. The more that scholars specify the pathways they invoke, for instance by referencing particular stepping-stones itemised here (Fig. 1.3), and embed discussions in the historical context of specific past environments (e.g. Table 1.1), the more fruitful ensuing hypothesis-testing and interpretation will be.

Mechanistic investigations of ecosystem history hold deep import for human understanding of how our world came to be. If future studies confirm that nutrient fluxes or phytoplankton stoichiometry changes incited more than one marine revolution, then ecological diversification may be more deterministic than previously thought (Erwin 2015c). However, regardless of the strength of bottom-up controls, there remains scope to invoke other causal and stochastic Earth and life processes in shaping ecosystems through time.

1.7 Scope of the thesis

The ensuing chapters test hypotheses of how biotic and abiotic factors structure marine ecosystems, from primary producers to detritivores, from the pelagic realm to the benthos, and from millennia to hundreds of millions of years. I quantified ecological responses primarily in terms of species' geographic and environmental

occupancy, because species' distributions shape the biodiversity map of the planet, control extinction rates to a large degree (Harnik, Simpson, et al. 2012; Saupe, Qiao, et al. 2015), and might influence origination rates (Maurer and Nott 2001; Gaston 2003; Jablonski and Roy 2003). Distributions are an emergent property—a trait measurable only at the level of populations, species, or higher taxa rather than from an individual. Therefore, the thesis analyses big datasets of spatiotemporal fossil occurrences rather than describing physical material from any particular time, region, or taxon. The study system in each chapter was selected for its power to estimate the effect size of a given mechanism of ecosystem regulation.

Chapter 2 addresses the degree to which biotic competition restricts species distributions, drawing on global occurrences of brachiopods and bivalves across the Phanerozoic. Like other benthic animals, adults of these groups compete for space on the seafloor. The fossil record of brachiopods and bivalves has received particular study across regions and time periods, including focus on biodiversity, body size, and metabolic rates through time (section 1.4.5), and thereby provides dense information about the histories of competitor diversity and geographic distributions. The assumption that species competitively exclude each other from space and other resources underpins major theories in ecology and evolution, including community saturation and diversity-dependent diversification (Rabosky 2013; Aguilée, Gascuel, et al. 2018). However, competitive exclusion and its inverse, 'ecological release' after competitor removal, have rarely been tested in the marine realm or at scales larger than populations. My three independent analyses found that ecological release was undetectable in marine seafloor communities at the level of species and geologic stages: the number of potential competitors in a region failed to explain how widespread species were within that region. Furthermore, although ecological release predicts species to expand after biotic crises, survivors of extinction intervals did not invade the space of fallen competitors. These results suggest non-competitive factors, such as the availability of favoured environmental conditions, control species' distributions at coarse scales.

Chapter 3 examines explicitly the importance of environmental conditions for species' distributions. This study focussed on the last seven ice-age cycles (700 ka), an interval of dramatic fluctuations in climate that are well-constrained by ice core data. I quantified marine species' thermal occupancy in response to these Earth system oscillations at 8-ka resolution—a temporal scale fine enough to narrow the gap between ecological and palaeontological data discussed in the previous section. To achieve such unprecedented temporal resolution on this timescale, I used the fossil record of planktonic foraminifera, which international ocean drilling programs have sampled repeatedly and extensively across the globe. Planktonic foraminifera species occupy a range of trophic positions between primary producers and the zooplankton, including facultative autotrophs that host photosynthetic symbionts as well as obligate carnivores (Aze, Ezard, et al. 2011; Schiebel and Hemleben 2017).

I estimated the degree to which foraminifera species migrated to track preferred conditions or shifted their niche distributions to dwell in novel temperatures experienced *in situ* during climate change. Abiotic niche lability is a key parameter to estimate because it dictates the rate at which species must move to reach suitable conditions and the probability of extinction if there are few suitable areas remaining (Wiens and Graham 2005; Lavergne, Evans, et al. 2013). Three parallel statistical analyses indicated foraminifera niche time series were uncoupled from the global temperature curve. Niches were equally similar regardless of the magnitude of global temperature change, even between extreme contrasts of glacial and interglacial intervals. My findings of niche stability suggest warming during coming centuries could redistribute or reduce populations of foraminifera and potentially other marine species with pelagic life stages. As detailed in sections above, these ecological changes at the level of calcifying plankton will cascade to affect carbon sequestration and higher-order consumer dynamics.

Although my research indicated abiotic niche stasis for substantial lengths of species' lifetimes, foraminifera niches must evolve at some point, since species have diverged in their temperature preferences. Chapter 4 explores niche evolution across a phylogenetic tree, estimating the mode and rate of change within and between

lineages. Quantifying how much niche change accrues during species' lifetimes *vs.* at speciation events is an outstanding problem in evolutionary biology and one that fossil data from planktonic foraminifera are well-positioned to address. Trait evolution models within lineages confirmed occupied temperatures as static through time, rather than changing randomly, directionally, or in tandem with global climate. Phylogenetic models could not distinguish competing evolutionary hypotheses although might demonstrate model robustness to the age of training data supplied.

There are many open debates in macroecology and macroevolution (McGill, Chase, et al. 2019), and this doctoral research raises as many questions as it resolves. In Chapter 5 I review the contributions of the thesis. The themes of hierarchical scaling from community ecology to macroecology, disjunction between microevolution and macroevolution, and importance of quantitative approaches recur and offer direction for further investigation. Palaeobiology in isolation will never solve all the remaining puzzles in these areas; unearthing the history of life will require a plurality of theories and empirical contributions from many disciplines.

What are we missing in trying to read this world by the inappropriate scale of our small bodies and minuscule lifetimes?

— Stephen Jay Gould (1998), *Phil. Trans. R. Soc. London B*, p. 314

2

Ecological release is undetectable in oceans at large scales

Contents

2.1 Overview 38
2.2 Results 38
2.2.1 Time series analysis 41
2.2.2 Biotic crises *vs.* background intervals 42
2.2.3 Mixed-effects models 42
2.3 Discussion 47
2.4 Methods 50
2.4.1 Fossil occurrences 50
2.4.2 Spatial subsampling 52
2.4.3 Statistical analyses 53
2.4.4 Data and code availability 54

The content of this chapter and Appendices A and B has been published in Antell, Kiessling, et al. (2020) (DOI: 10.1016/j.cub.2019.10.065). The following text and charts constitute a minor reformatting of the accepted manuscript to fit within the context of this thesis. EES formulated the idea; GSA and MA curated the data; GSA, EES, and WK designed the analysis; GSA programmed the code, generated the results, and drafted the manuscript, which all authors critically edited and approved.

2.1 Overview

Fundamental ecological and evolutionary theories such as community saturation and diversity-dependent diversification assume that biotic competition restricts resource use, and thence limits realised niche breadth and geographic range size (Alroy, Aberhan, et al. 2008; Rabosky 2013; Aguilée, Gascuel, et al. 2018). This principle is called competitive exclusion. The corollary, ecological release, posits that after competitors disappear from a region, species that were previously excluded will invade. Hundreds of field experiments have demonstrated ecological release in living populations, but few of these studies were conducted in marine environments, and almost no work extended beyond 10 years and 1,000 km^2 (Schoener 1983; Sexton, McIntyre, et al. 2009). In limited investigation of marine taxa at larger spatiotemporal scales, macroecologists and palaeobiologists have observed little evidence of competitive exclusion (Benton 1996; Myers and Lieberman 2011; Klompmaker and Finnegan 2018; Rillo, Sugawara, et al. 2019). Here, I quantified spatial trends in the rich and densely-sampled fossil history of brachiopods and bivalves, while accounting for inconsistent sampling coverage through time using a new method of spatial standardisation. The number of potential competitors in a region did not explain the geographic distribution of constituent species or genera. Furthermore, although ecological release predicts species to expand after extinction events, survivors of intervals with net species loss expanded as rarely as species in other intervals. Regression model estimates indicated different spatial responses of brachiopods and bivalves, and of habitat specialists and generalists, but no effect from changes in number of potential competitors. Biotic competition may control the distribution of populations, but on larger spatiotemporal scales, non-competitive factors may have driven biogeographic patterns of brachiopods and bivalves.

2.2 Results

I performed three sets of analyses to test species' geographic responses to inferred competitive pressure (Fig. A.1). The first approach tested the degree to which mean

geographic range size of species in regional marine assemblages varied inversely with number of species in those regions. This analysis investigated trends across all taxa of bivalves and brachiopods that might have interacted in a given region, an 'assemblage approach' that produced continuous time series of central tendencies. I also analysed data in a 'taxon-tracking approach': each datum represented a single species that was traced through successive stages. I tested whether species that survived intervals of net species loss expanded in distribution compared to species from intervals of background extinction rates, and I modelled each species' change in geographic distribution through time as a function of environmental affinity, clade membership, and change in the number of species in a region.

All analyses focussed on marine bivalves and brachiopods—diverse and well-sampled clades that have figured prominently in discussion of how competition influences macroevolution (e.g. Gould and Calloway 1980). Each of the three analyses was applied to rasterised occurrences of brachiopod and bivalve species compiled from the global Paleobiology Database (Alroy, Aberhan, et al. 2008). I repeated the assemblage-based time series analysis on an independent, finer-resolution dataset of bivalve species in the northwest Atlantic. Members of Bivalvia and Brachiopoda overlap in body size, trophic position, and preferences for substrates and habitats. Therefore, species from these groups are candidate competitors, especially for substrate availability, even though the clades differ physiologically and diverged before the Early Cambrian (Payne, Heim, et al. 2014; Carlson 2016). The study does not assume that all co-existing species competed simultaneously for space or other resources, or that competition was stronger between than within clades. Rather, because the global dataset includes species that occupied a variety of habitats, I expect that subsets of species competed in the living community, regardless of whether these competitors can be identified individually.

I converted occurrences to presence or absence values for grid cells of standard area. My novel subsampling procedure controlled both total sampling area and dispersion of sampling sites through time (Figs. A.2 and A.3; Appendix B). Although spatial subsampling compresses the range of observed values in a similar way to

classical rarefaction, power analysis indicated a low likelihood of type II error (Fig. A.4). Geographic range size was calculated here both as the number of equal-area grid cells in which a species was observed ('occupancy') and as the length of the minimum spanning tree connecting occupied grid cell centroids (Close, Benson, et al. 2017). Occupancy directly reflects area on a map, whereas tree length is a continuous variable that accounts for dispersion.

I parameterised competitive pressure in terms of the number of species in an assemblage, to consider the cumulative pressure from many species on any focal species. Taxon count has been treated as a proxy of competitive pressure in recent theoretical work (Herrera-Alsina, Pigot, et al. 2018) and in previous empirical work on brachiopods and bivalves (Gould and Calloway 1980; Miller and Sepkoski Jr 1988) and other taxa (Laube, Graham, et al. 2013). Co-occurrence frequencies and temporal replacement, alternative metrics for competitive pressure often used for fossil data, have power only to detect extreme, pairwise cases of exclusion (Benton 1996; Myers and Lieberman 2011; Klompmaker and Finnegan 2018). Common metrics used in modern ecology, such as changes in abundance or experimental removal of competitor populations (examples in Sexton, McIntyre, et al. 2009), are difficult or impossible to apply to the fossil record.

The reliability of species count as a metric of competitive pressure could depend on the degree of niche overlap in a study ecosystem, a property that may have differed between Paleozoic and post-Paleozoic marine fauna (Wagner, Kosnik, et al. 2006). However, analyses herein considered many time bins and spatial regions so as to determine dominant patterns across the Phanerozoic; I also replicated analysis on fine-scale spatiotemporal data in the short and ecologically complex interval of the Neogene–Quaternary. Ultimately, the particular biotic parameter to which species respond most strongly (e.g. abundance of the closest competitor *vs.* total number of competitors) remains unknown.

2.2.1 Time series analysis

I calculated the correlation between replicate time series of species count and geographic range size in an assemblage drawn at random from each time bin. If ecological release occurred on the spatiotemporal scale of analysis, one would expect a negative correlation between time series of species count and geographic range size. Instead, the estimated relationship was too close to null to consider the signal to be either positive or negative (Fig. 2.1; Table A.3A). Plots of cross-correlation functions at lags of one to five time bins tested for delayed ecological responses, but the patterns were congruent with the simultaneous comparisons (i.e. lag zero).

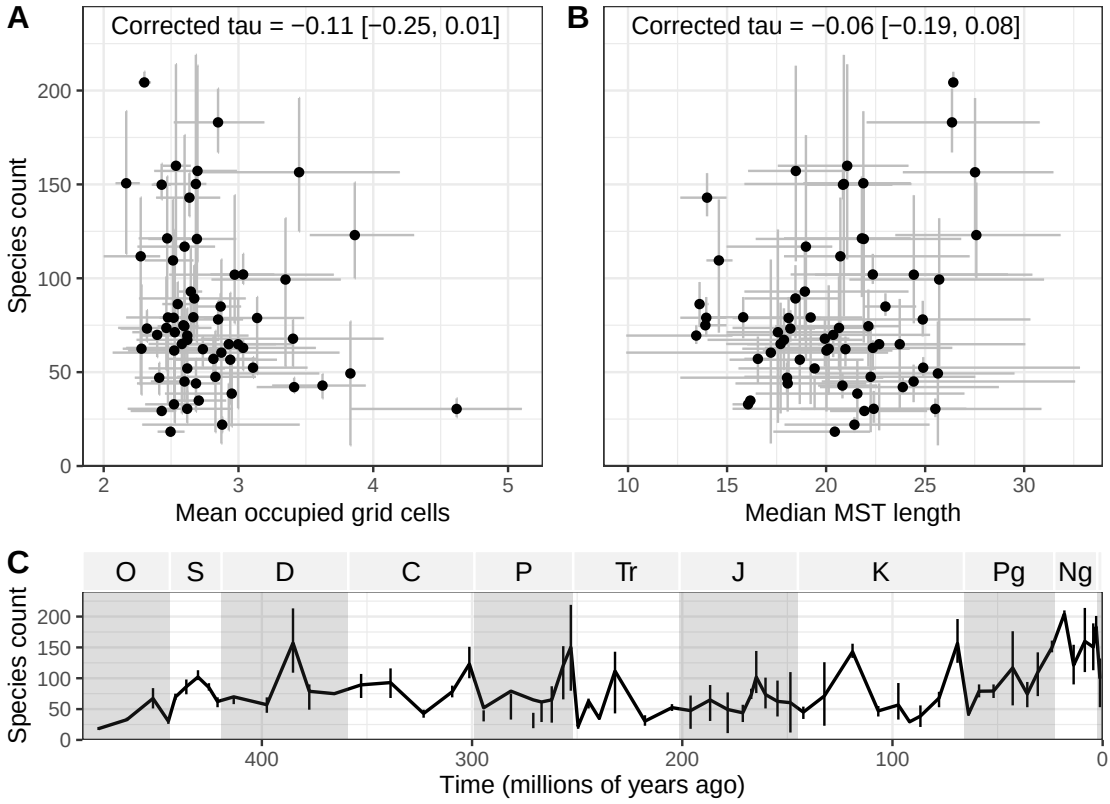


Figure 2.1: Scatterplots indicate the relationship between species count and **(A)** mean number of occupied grid cells or **(B)** median square-root length of tree connecting occupied cells (MST length) from 63 time bins (Table A.1). Error bars mark the interquartile range. Correlation coefficients are corrected for autoregression in the predictor time series. The slight tendency toward negative values is attributable to the sampling bias of equal magnitude (Table A.3). **(C)** Mean species count values among regional subsamples of global data are plotted across the 485-million-year study interval. Geological periods: O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene. Quaternary not labelled.

Results from the finer-grain northwest Atlantic dataset also concurred (Fig. 2.2; Table A.3H), as did results from sensitivity analyses—for instance where global data were subsampled more dispersedly or within smaller regions, treated with classical or coverage-based richness rarefaction, or analysed at the level of genera (Tables A2 and A.3B–G; Appendix B.3.3).

2.2.2 Biotic crises *vs.* background intervals

I tracked individual species across stage boundaries to test whether taxa expanded their distributions after biotic crises in comparison to periods of background extinction dynamics. Geographic behaviour varied considerably both among species within each interval and in the mean species' trend among intervals; however, the central tendency of species' geographic range size changes was stasis (Table A.4). Expansion and contraction in species' geographic ranges were distributed evenly across intervals of increases and decreases in species count, regardless of geographic range metric (Fig. 2.3). This result is incongruent with a hypothesis of ecological release, which would have predicted species that survived the extinction of competitors to expand into newly-vacated geographic space.

2.2.3 Mixed-effects models

Change in species count did not predict change in geographic range size of species tracked through time. The variable for change in species count, *delta* in Table A.5, was inconsistently selected as a predictor in well-supported linear mixed effects models (16 of 29 non-trivial models for geographic range size calculated as occupied grid cell count, and 18 of 34 non-trivial models for range size as minimum spanning tree length). In models that did include species count as a predictor variable, the coefficient estimates for *delta* were small in magnitude and positive in direction (Fig. 2.4, Tables A.6 and A.7). The direction of effect predicted by competition restricting whole species' distributions would instead be negative, i.e. a decrease in range size associated with an increase in potential competitors. Thus, regression analysis agreed with the other two analyses in that it did not provide evidence for ecological release.

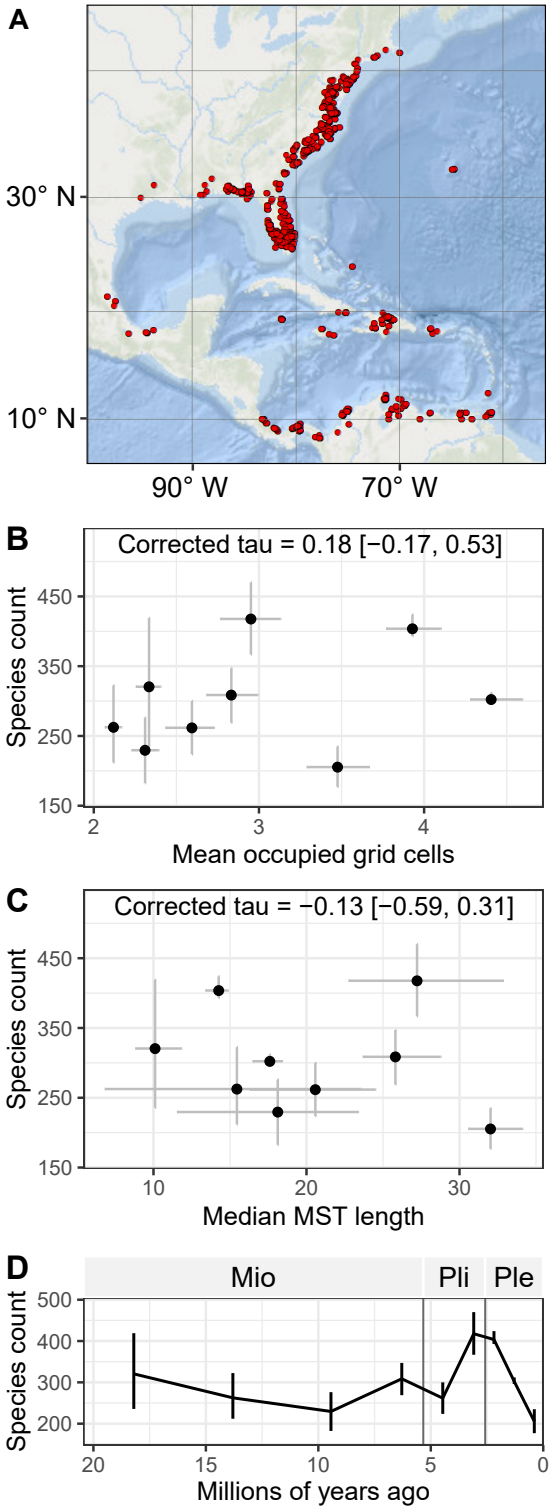


Figure 2.2: **(A)** Northwest Atlantic bivalve specimen localities are plotted at modern-day coordinates. Scatterplots between species count and **(B)** mean number of occupied grid cells or **(C)** median square-root length of tree connecting occupied cells (MST length) are plotted as in Fig. 2.1. Correlation coefficients for every pair of time series are listed in Table A.3. Data ranged from the Burdigalian (Early Miocene) through the Pleistocene (Table A.1B). Geological epochs: Mio, Miocene; Pli, Pliocene; Ple, Pleistocene.

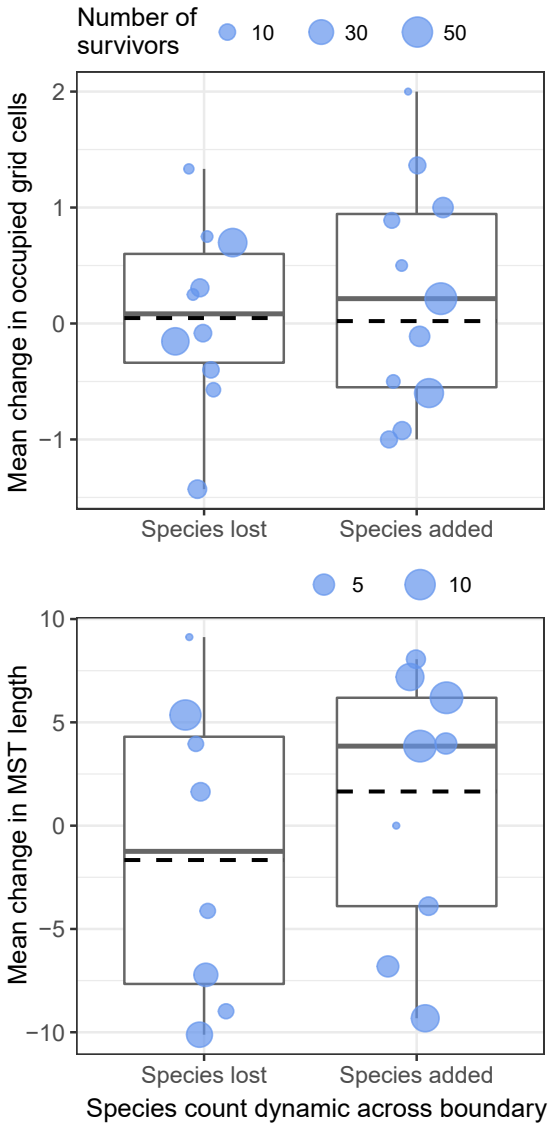


Figure 2.3: Point size is proportional to the number of species sampled on both sides of a given boundary. Phanerozoic intervals were included only when successive bins contained ≥ 10 cells of the same habitat type inhabited by multiple surviving species (Table A.4). **(A)** Species' occurrences were similarly dispersed (weighted Wilcoxon p-value $= 0.79$, t-value $= -0.27$, $df = 19$) between intervals when species count in a region decreased ($n = 10$) and increased ($n = 11$). The unweighted means (solid horizontal lines) aligned closely with n-weighted means (dashed lines, expansion of 0.05 grid cells per species for intervals of species count decrease *vs.* expansion of 0.02 grid cells per species for intervals of species count increase). **(B)** Results were congruent when geographic distributions were measured using minimum spanning tree length: geographic trends were equivalent (p-value $= 0.31$, t-value $= 1.1$, $df = 15$) between intervals of species count decrease (average reduction of 1.66 km per species, $n = 8$) and increase (average expansion of 1.65 km per species, $n = 9$). When newly-originated and nearly-extinct species were evaluated along with mid-duration species, strong outliers appeared but the general trends were still consistent with those of mid-duration species alone as reported in Table A.4.

In contrast to change in species count, a strongly influential predictor of species' distribution dynamics was the point in time in a lineage (e.g. mid existence) at which geographic ranges were measured. Evidence for the strength of this variable includes the fact that it was the only predictor to appear in every non-trivial model for occupancy and 32 of 34 models for minimum spanning tree length (Table A.5). Species expanded after appearing in the fossil record and contracted prior to disappearing from the record (Fig. 2.4, Tables A.6 and A.7). Hence, these results corroborate previous descriptions of waxing and waning through time (Foote, Crampton, et al. 2007; Liow and Stenseth 2007; Harnik, Simpson, et al. 2012).

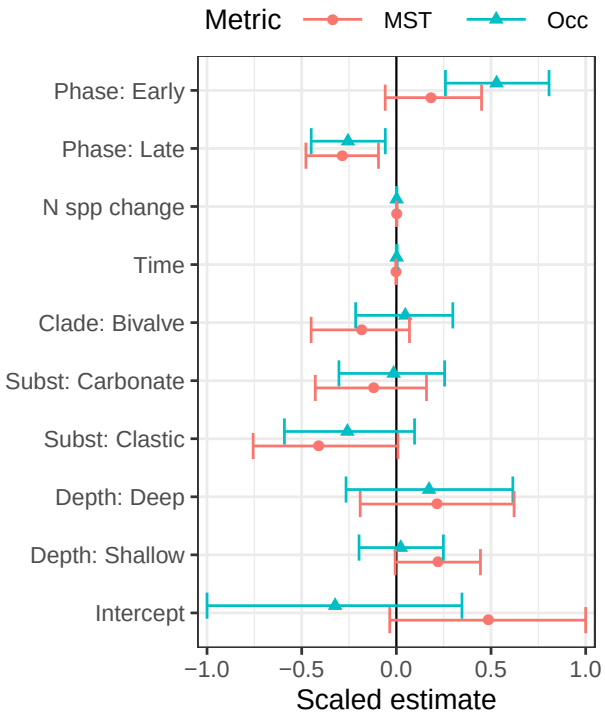


Figure 2.4: A mixed-effects model estimated a null effect from change in species count, contrary to expectations of ecological release. Estimates were relative to mid-existence (*vs.* early or late) brachiopod species (*vs.* bivalve) with no preference for habitat (*vs.* preference for substrate or water depth). Substrate preference (subst.) was either carbonate (e.g. reef) or siliciclastic (clastic, e.g. sand). Error bars = 95% confidence intervals from profiling. Data for each species was standardised within subsamples of grid cells that contained suitable substrate and water depth. Colour indicates whether a model was based on occupancy (Occ, $n = 708$) or square-root minimum spanning tree length (MST, $n = 239$); coefficients for the two models were scaled for comparison. Unscaled estimates are given in Table table:a-global-lmm. Models based on subsets of predictor variables are described in Tables A.5 and A.6.

The clarity of regression predictions varied depending on whether geographic range size was measured in terms of occupied grid cells or minimum spanning tree length. When range size was measured in grid cells, expansion of newly-originated species was approximately twice as large as contraction of nearly-extinct species (Table A.6A), an asymmetry found previously (Liow and Stenseth 2007; Harnik, Simpson, et al. 2012). In contrast, the magnitude of spanning tree expansion early in a species' duration was smaller than the magnitude of later contraction (Table A.6B). In addition, specialists in either of the two categories of substrate lithology tended to shrink in grid cell area compared to generalists when range size was measured in terms of grid cells. This is evident in the negative estimates for coefficients corresponding to carbonate or siliciclastic lithology preferences (Table A.6A). Bathymetric preference did not appear as an important predictor of occupancy change in models with the most weight (Table A.6A). However, when range size was measured in terms of spanning tree length, preference for water depth replaced preference for lithology as an important predictor in the best-supported models (Table A.6B). Surprisingly, specialists in either shallow or deep water tended to expand in spanning length compared to bathymetric generalists (Table A.6B). The confidence intervals related to water depth preferences are wide, however, since there were few specialists for which the number of grid cells with suitable habitat was large enough to perform subsampling (89 shallow-water species but only 18 deep-water species, compared to 131 generalists).

Of the variance not attributed to the predictors described above (i.e. the fixed effects), the time bin of observation explained some of the differences between species' responses. Time bin was included as a random effect *a priori* in all models and accounted for 24% of variance for occupied grid cells and 27% of variance for spanning tree length within fixed-effect treatment groups. The variance terms indicate a limited but non-negligible similarity in geographic responses among species of the same time interval.

2.3 Discussion

Debate has raged for centuries (Carlson 2016) on whether bivalves displaced brachiopods competitively (Miller and Sepkoski Jr 1988; Liow, Reitan, et al. 2015) or diversified independently (Gould and Calloway 1980; Kiessling and Aberhan 2007a; Payne, Heim, et al. 2014). The present study did not intend to answer this question explicitly. However, results from the mixed-effects models provide information that bears on the issue tangentially. Unexpectedly, the model predicted brachiopods, not bivalves, to expand more in spatial distribution when suitable habitat was present. This result seems particularly anomalous with a hypothesis of competitive displacement. Moreover, many species of both clades were associated with specific habitats, and the mixed-effects model predicted differential expansion for each environmental affinity. Together, these results are consistent with hypotheses that abiotic environmental conditions mediated shifts in dominance among benthic macrofauna. For instance, water chemistry perturbations in offshore shelf settings during the Permo-Triassic transition might have preferentially extirpated brachiopod species and prompted different diversity dynamics in the two clades (Gould and Calloway 1980; Clapham and Bottjer 2007). Alternatively or additionally, brachiopods might have specialized on substrates during the post-Paleozoic more than did bivalves. Siliciclastic habitat specialists (e.g. sandy-bottom dwellers) seem to have experienced a dampening of geographic range expansion (Fig. 2.4) and origination rates (Kiessling and Aberhan 2007a), for instance.

Evolutionary theorists beginning with Darwin have assumed that biotic competition at the finest spatiotemporal scales could impose a stabilising feedback on diversity at the largest scales (Reznick and Ricklefs 2009). Geographic range size has been a foremost candidate for facilitating this process (Rabosky 2013). Mechanistically, death of competitors during mass extinction could release surviving species from refugia (Sidor, Vilhena, et al. 2013; Button, Lloyd, et al. 2017). As large geographic ranges increase the probability of survival (Harnik, Simpson, et al. 2012; Saupe, Qiao, et al. 2015), albeit to a lesser degree in the midst of a mass extinction event (Kiessling and Aberhan 2007b), species richness could rebound in the interval

after an extinction pulse (assuming no reduction in origination rates) (Fig. 2.5). With many species in competition, geographic distributions might shrink. Since restricted spatial distributions come with a high risk of extinction, richness might drop back to an equilibrium level (assuming no increase in origination rates) (Fig. 2.5).

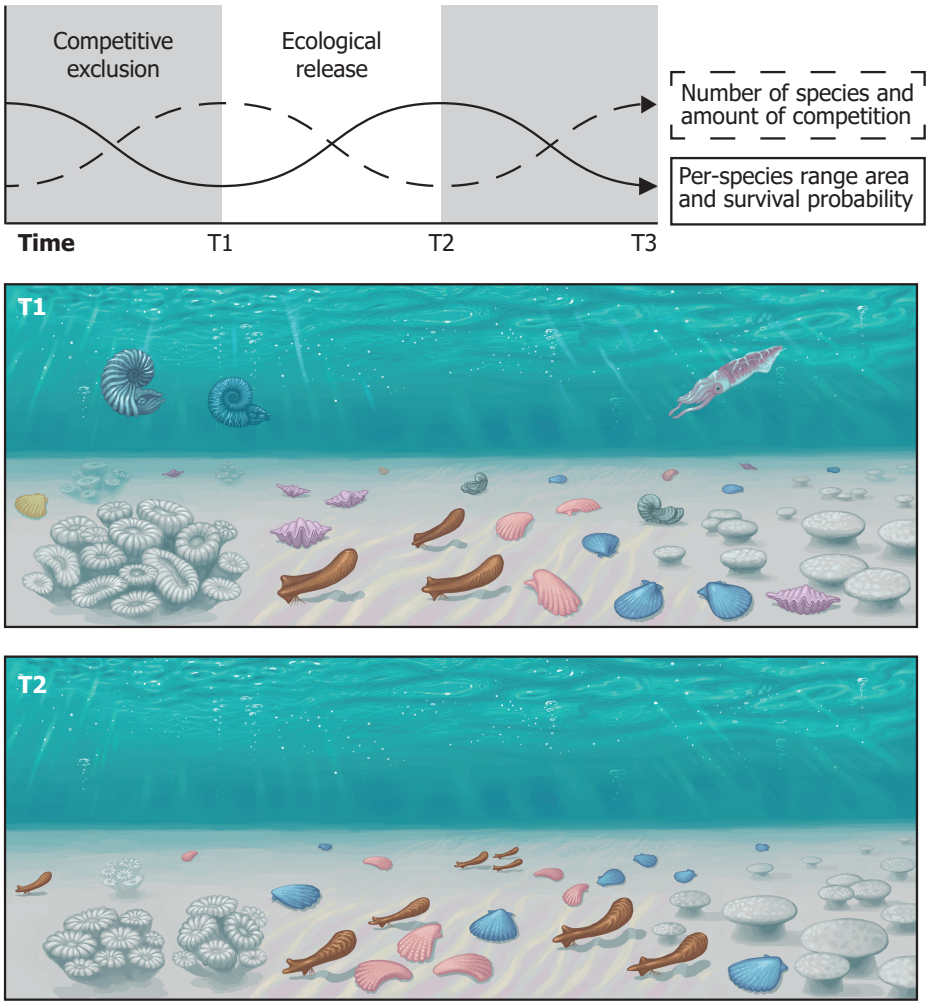


Figure 2.5: The hypothesis of ecological release posits that after a decrease in the number of competitors for a species (time T1 to time T2), individuals of that species will invade areas of lessened resource use. Since range expansion would tend to buffer species from extinction, the number of species in competition might accumulate until species excluded each other again (T3). Thus, geographic ranges could mediate diversity-dependence (Rabosky 2013; Aguilée, Gascuel, et al. 2018). In this artistic reconstruction of hypothesised ecological release (middle and bottom), victims of the Early Jurassic extinctions included several bivalves (e.g. *Gryphaea gigantea* and *Pseudopecten equivalvis*) and brachiopods (e.g. *Tetrarhynchia tetrahedra*), as well as nektonic species. Survivor species (e.g. *Eopecten velatus*, *Gervillaria hartmanni*, and *Oxytoma inequivalvis*) might have invaded vacant areas of seafloor. Original illustration by Andrew Orkney.

The scale of my analysis is appropriate to search for such a stabilising feedback between the number of species and the size of their geographic ranges, since the median temporal resolution of the datasets (2–5 million years) matches the timespan on which diversity has responded to certain mass extinction events (Smith and Jackson 2009; Sidor, Vilhena, et al. 2013; Darroch and Wagner 2015). A mechanism for diversity-dependence based on ecological release, however, appears unfounded in marine bivalves and brachiopods. An alternative explanation for diversity-dependence, if present, is that strong incumbency could reduce the persistence or spread of nascent species (Wagner, Kosnik, et al. 2006; Rabosky 2013; Ranjard, Welch, et al. 2014).

My assessment of geographic occurrences across all oceans through 485 million years indicates that competitive exclusion and ecological release have not dominated as macroecological patterns in the marine benthos at the level of species. This result contrasts with evidence from population-level studies (Schoener 1983; Sexton, McIntyre, et al. 2009). One explanation for these differing patterns is that fine-scale spatiotemporal heterogeneity might induce asynchronous dynamics among populations of a species. Populations at one edge of a geographic range might contract while those elsewhere expand, such that the cumulative evolutionary pressure over all populations would be null (Ricklefs 2004; Eldredge, Thompson, et al. 2005; Schindler, Armstrong, et al. 2015). Environmental filtering and other abiotic processes could emerge as the dominant determinants of species' distribution areas at coarse scales (Willis and Whittaker 2002; Myers and Lieberman 2011; Peterson, Soberón, et al. 2011; Wiens 2011; Götzenberger, de Bello, et al. 2012). Brachiopod and bivalve species interact with members of many other clades (Miller and Sepkoski Jr 1988), and antagonistic interactions with predators or parasites also could affect macroevolution in our focal clades (Stanley 2008). The scale-dependent effect of within-guild biotic interactions investigated here could extend to these other negative interactions, as discussed elsewhere in relation to the Red Queen model (Benton 2009; Strotz, Simoes, et al. 2018). Studies at the meso-scale of 10^3–10^5 years,

in regional habitat zones, are a clear next step to pinpoint the scale at which abiotic processes dominate species' geographic patterns to the exclusion of biotic controls.

Characteristics of species, communities, or environments could alter the effect size of competitive interactions. The conclusion that factors other than competition control most species' geographic range sizes might be generalisable to other marine taxa, for which geographic shifts are more predictably tied to environmental conditions such as temperature than are terrestrial taxa (Benton 1996; Myers and Lieberman 2011; Sunday, Bates, et al. 2012). Terrestrial organisms that disperse long distances, such as many plants (Götzenberger, de Bello, et al. 2012; Wisz, Pottier, et al. 2013), may share macroecological patterns with marine taxa. Meanwhile, in terrestrial organisms with limited dispersal capacity, such as many animals, competition might have controlled large-scale spatiotemporal distributions to a larger degree (Laube, Graham, et al. 2013; Pigot and Tobias 2013; Sidor, Vilhena, et al. 2013; Wisz, Pottier, et al. 2013; Ranjard, Welch, et al. 2014; Dunhill and Wills 2015; Button, Lloyd, et al. 2017). Biotic interactions also might exert more influence in communities with simple interaction networks, or in environments subject to widespread, homogenising disturbances.

2.4 Methods

2.4.1 Fossil occurrences

Global Phanerozoic dataset

Bivalve and brachiopod occurrences across the Phanerozoic were downloaded from the Paleobiology Database (PBDB; `paleobiodb.org`) in July 2018. Download and vetting criteria are described in Appendix B.1.1. The dataset included 57,621 rasterised occurrences of 16,995 species. I binned data at the finest temporal resolution (stages or clustered stages; Table A.1A) that retained sufficient occurrences in every time bin to calculate the variables of interest. Spatial sampling in the Cambrian was too sparse to achieve sub-epoch resolution, so the Phanerozoic dataset was truncated to Ordovician–Pleistocene records. Time bins were < 20 Ma in duration (Table A.1A); the median was 5.1 Ma (Q1 = 3.6 Ma; Q3 = 7.7

Ma). I coded environmental conditions of each specimen record along two axes: substrate (carbonate, siliciclastic, or both) and bathymetry (deep, shallow, or both), according to the definitions used by Nürnberg and Aberhan (2013). After grading the environmental state for every occurrence record, I inferred the environmental preferences of each species. For each axis, a species with $\geq 80\%$ of occurrence records classed as only one category was considered a specialist for that environmental condition; otherwise the species was treated as a generalist along the focal axis. For further details, please refer to Appendix B.4.1.

I converted occurrence coordinates to presence or absence values for raster grid cells. To ensure that all cells contained the same area, an initial 1° global grid was transformed to the Eckert VI projection. I eliminated occurrence records for any species observed in exactly one grid cell from a single time bin; this step reduced the influence of Lagerstätten and the incidence of invalid species identifications.

Northwest Atlantic Neogene–Quaternary dataset

The northwest Atlantic dataset included 12,035 occurrences of 1,856 bivalve species within the area of 6–42° North latitude and 50–100° West longitude. Records were compiled from the Integrated Digitized Biocollections database (iDigBio; `idigbio.org`). As with the global dataset, I converted occurrences to presence–absence values in equal-area raster cells, but with twice as fine a resolution (a 0.5° initial grid). Time bins for the northwest Atlantic dataset (Table A.1B) were shorter than those used in the global dataset. All bins were < 5 Ma in duration, with a median of 1.8 Ma (Q1 = 1.0 Ma; Q3 = 4.4 Ma). Sampling in the Aquitanian was so sparse in comparison to the other intervals that the dataset was truncated to Burdigalian–Late Pleistocene records only. Many iDigBio records lacked precise age information, either as a numeric estimate or geologic stage name. I therefore researched the chronostratigraphy of fossiliferous units in the northwest Atlantic to determine the minimum and maximum age estimates for 485 formations and members (Dataset S2), expanding the correlation chart of Saupe, Hendricks, et al. (2014). For further details, please refer to Appendix B.1.2.

2.4.2 Spatial subsampling

To account for the heterogeneity of preserved fossil occurrences, I devised a subsampling method that controls jointly the spatial extent and dispersion of sample localities. Each iteration picks a random occurrence point as a centre around which to circumscribe a buffer region with a predetermined radius; the procedure then draws a given number of equal-area grid cells within the circular buffer (Figs. A.2 and A.3). These selected cells form a subsample, within which geographic range size and species count are calculated. I assessed the performance of the subsampling routine by calculating a proxy of spatial coverage, here called sampling aggregation; I evaluated the explanatory power of this variable in time series analysis.

I tested the sensitivity of analyses to each parameter of the subsampling routine, including the radius and number of cells selected (Appendix B.3.3). For the analysis results presented in this chapter, I set a 1,500 km radius to confine each subsample in the global data, which agrees well with previous quantitative thresholds (Close, Benson, et al. 2017) and the common approach of constraining the dispersion of a sample to one continent or ocean basin. Consequently, the circle that constrained a random sample covered 7×10^6 km^2, an area approximately equivalent to the size of Australia. Within this region, each subsample iteration drew 12 cells without replacement, within the range that others have set when sampling grid cells (Darroch and Wagner 2015). Appendix B.3.1 and B.4.2 give step-by-step descriptions of assemblage-based and taxon-tracking subsampling procedures. Although any individual subsample may have captured only some of the fossil species present in the global record at the time, all available data contributed to analysis because the regional subsample procedure was replicated 500 times in each time bin. Appendix B.2 discusses further considerations of spatial standardisation, such as discussion of proportional occupancy metrics. Appendix B.5 and Fig. A.4 provide power analysis of the data compression caused by the subsampling procedure.

2.4.3 Statistical analyses

Time series analysis

For each iteration of subsampling, I calculated the correlations between predictor time series (species count or sampling aggregation) and response time series (occupancy or minimum spanning tree length). I fit an autoregressive error structure of order 1 to each predictor time series and then estimated correlation strength (non-parametric Kendall's *tau* coefficient) between the residuals and each response time series. The bounds of the confidence interval were the 2.5% and 97.5% quantiles of *tau* values obtained across subsampling iterations. Appendix A, section 4.1.2, contains additional details.

Biotic crises *vs.* background intervals

For every species that was observed in two or more consecutive time bins, I quantified the change in the survivor's range size between the earlier and later time bin. I compared mean change in range size among survivors ('range-size shift') of two sets of time intervals: intervals when species count increased and intervals when species count decreased. Phanerozoic intervals were included only when successive bins contained ≥ 10 raster cells of the same habitat type inhabited by multiple surviving species. To classify each interval as associated with either a net gain or loss in species number, I calculated whether the median change in species count experienced by survivors across a given boundary was positive or negative. More species crossed certain stage boundaries than others (Table A.4), which resulted in unequal variances among intervals. I therefore weighted each range-size shift by the number of survivors that contributed to that shift and compared the difference in the two groups with a two-sided Mann-Whitney (Wilcoxon rank-sum) test.

Mixed-effects models

I explored the nature of the relationship between change in species count and change in geographic range size with a generalised additive mixed-effects model to determine that a linear fit was appropriate. I then built a global linear mixed-effects

model (all predictor variables included) and also compared the 64 models with every permutation of the fixed effect variables: change in species count (integer), clade (categorical: Brachiopoda or Bivalvia), point in lineage duration (categorical: Early, Middle, or Late), time bin age (continuous and positive), lithological preference (categorical: carbonate, siliciclastic, or both), and bathymetric preference (categorical: shallow, deep, or both). The identity of the time interval associated with each survivor was included as a random effect in every model. I included this random effect because species from the same time bin were expected to pseudoreplicate each other due to shared environmental conditions and spatial sampling structure. I fit models based on restricted maximum likelihood for estimation purposes and based on maximum likelihood for testing purposes. I estimated Akaike's Information Criterion (AIC) to determine the subset of models that received non-trivial weight (ΔAIC < 10; for derivation and discussion of weights, consult Burnham and Anderson (2010)). Models are listed in order of relative performance in Table A.5. For fixed effects of these supported models, I recorded point estimates and the 95% confidence interval bounds from the likelihood profile (Table A.6).

2.4.4 Data and code availability

All generated data and custom software developed for this project are archived at Zenodo (DOI: 10.5281/zenodo.3491853) and maintained at `https://github.com/GwenAntell/EcoRelease`.

Each foraminifer is a tiny window into the past, so tiny that window can fit in the eye of a needle.

— Tracy Aze

3

Thermal niches are static throughout glacial–interglacial climate change

Contents

3.1 Abstract 56
3.2 Introduction 56
3.3 Results 59
3.3.1 Time series correlation 59
3.3.2 Niche lability *vs.* climate change magnitude 61
3.3.3 Niche lability among extreme climate intervals 61
3.4 Discussion 63
3.5 Methods 65
3.5.1 Palaeoclimate reconstructions 65
3.5.2 Definition of 'available environment' 66
3.5.3 Fossil occurrence data 67
3.5.4 Spatiotemporal resolution 69
3.5.5 Species attributes 70
3.5.6 Kernel density estimation 71
3.5.7 Niche dissimilarity measure 73
3.5.8 Statistical analyses 75
3.5.9 Data and code availability 79

The content of this chapter and Appendix C has been published in Antell, Fenton, et al. (2021) (DOI: 10.1073/pnas.2017105118) and is reformatted here from the accepted version. GSA and EES designed research; ISF compiled fossil data; PJV ran climate models; GSA analysed data, generated results, and wrote the paper.

3.1 Abstract

Abiotic niche lability reduces extinction risk by allowing species to adapt to changing environmental conditions *in situ*. In contrast, species with static niches must keep pace with the velocity of climate change as they track suitable habitat. The rate and frequency of niche lability have been studied on human timescales (months to decades) and geological timescales (millions of years), but lability on intermediate timescales (millennia) remains largely uninvestigated. I quantified abiotic niche lability at 8-ka resolution across the last 700 ka of glacial–interglacial climate fluctuations, using the exceptionally well-known fossil record of planktonic foraminifera coupled with Atmosphere–Ocean Generalised Circulation Model reconstructions of palaeoclimate. I tracked foraminiferal niches through time along the univariate axis of mean annual temperature, measured both at the sea surface and at species' depth habitats. Species' temperature preferences were uncoupled from the global temperature regime, undermining a hypothesis of local adaptation to changing environmental conditions. Furthermore, intraspecific niches were equally similar through time, regardless of climate change magnitude on short timescales (8 ka) and across contrasts of glacial and interglacial extremes. Together, these results suggest that warming and ocean acidification over the next hundreds to thousands of years could redistribute and reduce populations of foraminifera and other calcifying plankton, which are primary components of marine food-webs and biogeochemical cycles.

3.2 Introduction

Abiotic niche dynamics determine patterns of community composition over space and regulate trajectories of diversity over time (Wiens and Graham 2005). Both niche lability (Gómez-Rodríguez, Baselga, et al. 2015; Castro-Insua, Gómez-Rodríguez, et al. 2018) and conservatism (Wiens and Graham 2005; Pyron, Costa, et al. 2015) have been proposed to spur speciation, and abiotic niche lability has been associated with ecological invasions (Stigall 2014; Velasco, Martinez-Meyer, et al. 2018; Folk, Stubbs,

et al. 2019) and with reduced risk of extinction during times of climate change (Lavergne, Evans, et al. 2013). Thus, a deeper understanding of species' propensity for niche stasis *vs.* lability could improve predictions of biodiversity restructuring in response to anthropogenic climate change (Munday, Warner, et al. 2013).

Stasis in species' abiotic niches through time has been documented in empirical research, but most such studies have been limited to ecological niche modelling on decadal scales (reviewed in Peterson 2011) or palaeoecological examination on 10^6- to 10^7-year scales (Holland and Zaffos 2011; Saupe, Hendricks, et al. 2014; Stigall 2014). Since empirical rates of niche change are scarce and difficult to acquire, many studies merely assume that niche evolution occurs at a constant rate along branches of a phylogeny (Gómez-Rodríguez, Baselga, et al. 2015; Castro-Insua, Gómez-Rodríguez, et al. 2018; Velasco, Martinez-Meyer, et al. 2018; Folk, Stubbs, et al. 2019). Niche dynamics at intermediate timescales of centuries to millennia are particularly poorly documented (Peterson 2011), and studies at this meso-scale have been restricted to terrestrial systems (e.g. Banks, d'Errico, Peterson, Kageyama, Sima, et al. 2008; Pearman, Randin, et al. 2008; Record, Fitzpatrick, et al. 2013) or to comparisons between the present day and the single historical time step of the Last Glacial Maximum, ~21 ka ago (Martínez-Meyer, Townsend Peterson, et al. 2004; Martínez-Meyer and Peterson 2006; Banks, d'Errico, Peterson, Kageyama, and Colombeau 2008; Waltari and Guralnick 2009; Waterson, Edgar, et al. 2017). Quantifying the rate and relative frequency of niche change in marine species over timescales of 10^2–10^5 years is important, however, because species will adapt or go extinct in response to anthropogenic ocean changes over this timescale (Hutchins and Fu 2017).

In this chapter, I investigate climatic niche lability from the rich sedimentary archive of global planktonic foraminifera across the last 700 ka of glacial–interglacial cycles, at 8-ka resolution. Planktonic foraminifera (Protista) construct 'shells' (tests) of calcite, thereby sequestering carbon and recording an isotopic signature of past ocean conditions. Tests readily accumulate over large expanses of the seafloor. Consequently, the fossil record of foraminifera—arguably 'the best fossil record on Earth' (Kucera 2007)—affords an exceptionally high-resolution view into

past species distributions. This detailed record fuels studies of biostratigraphy, palaeoclimatology, and palaeoecology (Kucera, Weinelt, et al. 2005; Kucera 2007; Ezard, Aze, et al. 2011; Morard, Darling, et al. 2015; Waterson, Edgar, et al. 2017). Moreover, the complete species diversity of planktonic foraminifera has been described for the Plio-Pleistocene, with good agreement between morphological and molecular phylogenies (Kucera 2007; Aze, Ezard, et al. 2011; Morard, Darling, et al. 2015; Schiebel and Hemleben 2017). Although some have speculated that foraminifera competitively exclude each other (Ezard, Aze, et al. 2011), recent work found that planktonic foraminifera species seldom restrict each other's distributions (Rillo, Sugawara, et al. 2019). Presumably, therefore, species occupy the full envelope of existing environmental conditions within their tolerance limits, and geographic distributions are determined almost entirely by physical ocean conditions.

I developed three analyses to investigate the degree of abiotic niche lability in foraminifera relative to climate change. All methods examined the univariate niche axis of temperature, which is the single most important explanatory variable with regard to geographic distributions of foraminifera (Bijma, Faber, et al. 1990; Rutherford, D'Hondt, et al. 1999; Fenton, Pearson, et al. 2016; Waterson, Edgar, et al. 2017; Richter, Watteaux, et al. 2019) and is a climate-related stressor and extinction driver for diverse marine fauna across timescales (Reddin, Nätscher, et al. 2020; Mathes, van Dijk, et al. 2021). The adaptive potential of thermal niches has been taken as a key determinant of global community structure and genetic connectance in plankton (Ward, Cael, et al. 2021). Primary productivity and other environmental variables, however, may also structure abiotic niches of plankton (Brombacher, Wilson, et al. 2018). The suite of analyses herein quantified whether and by how much planktonic foraminiferal niches shifted along a temperature axis. First, I correlated time series of species' thermal optima with global temperature to determine whether species tracked suitable habitat or experienced environmental fluctuations *in situ*. Then, I quantified species' niche dissimilarity between pairs of time bins, first by tracking niches across bin boundaries and then by contrasting niches at climatic extremes of glacial maxima and interglacial thermal peaks.

3.3 Results

Occurrence data were derived from a global dataset of Cenozoic planktonic foraminifera records (Fenton, Woodhouse, et al. 2021). I converted fossil coordinates into presence–absence values within 1.25° raster grid cells at 8-ka intervals over the past 700 ka (i.e. $n = 88$ time bins), as detailed in sections 3.5.3 and 3.5.4, 'Fossil occurrence data' and 'Spatiotemporal resolution'. Mean annual temperature (MAT) was extracted from a coupled Atmosphere–Ocean Generalised Circulation Model at every cell containing an occurrence for each time interval (Fig. 3.1; section 3.5.1, 'Palaeoclimate reconstructions'). Occurrences from ocean drilling cores covered all oceans and latitudinal bands. I repeated all analyses first using temperature extracted at the sea surface and then using temperatures extracted within each species' preferred depth layer (a 'depth-habitat approach,' defined in section 3.5.5 and Table C.1). The final dataset contained 24 species and 42,233 unique age-cell-species combinations.

Thermal niches were estimated from occurrences using kernel density estimation (KDE; Figs. 3.1 and C.1). KDE generates a continuous occurrence probability distribution for an organism along an environmental or resource gradient, based on a finite sample of presence points (e.g. Broennimann, Fitzpatrick, et al. 2012; Godsoe and Case 2015). I programmed KDE in a way that standardised for differential sampling of environmental conditions through time (section 3.5.6 and Appendix D). Thermal optima (preferred temperatures) were defined as the MAT associated with the probability function maximum for a species in a time bin (Fig. 3.1). Niche dissimilarity was defined as the Hellinger's H distance between two probability distributions; identical distributions give $H = 0$ and non-overlapping distributions give $H = 1$ (Fig. 3.4). Table C.1 lists the depth habitat, thermal optimum, and mean H for each species.

3.3.1 Time series correlation

For the first test of niche lability, I correlated concurrent time series of preferred and available temperatures to quantify the degree of climate tracking on short (8-ka) timescales (Methods section 3.5.8, 'Time series correlation'). A species'

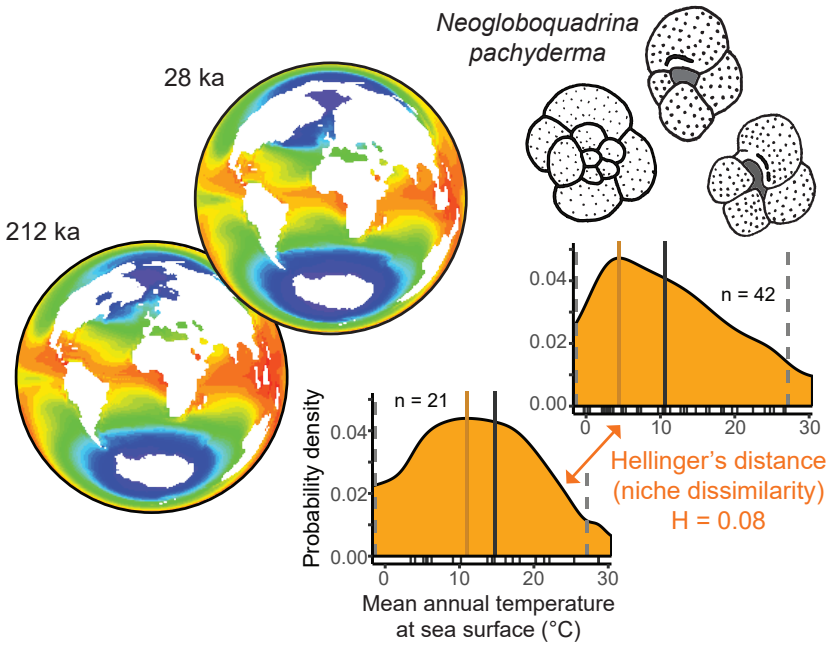


Figure 3.1: This study combined Atmosphere–Ocean Generalised Circulation Model temperature reconstructions (sea surface estimates at 28 ka and 212 ka shown) with fossil occurrences of planktonic foraminifera (polar species *Neogloboquadrina pachyderma* illustrated at top). Temperatures at occurrence sites are marked with ticks below the x-axis of the plots; these values are the basis for kernel density estimates of thermal niches (orange). Vertical black lines mark the mean of occupied temperatures for a sample. Dark orange lines mark optimal or 'preferred' temperatures, inferred as the temperature associated with the maximum of the probability density function. Hellinger's H distance quantifies dissimilarity between a pair of probability density functions, from 0 (identical distributions) to 1 (no overlap). In the figured example, the H distance between *N. pachyderma* niches at the 28 ka glacial and 212 ka interglacial time steps is 0.08. H distances were calculated over a standard temperature interval (bounded by dashed gray lines at −1.4 and 27.1 °C) for all species and time bins.

preferred temperature was defined as the MAT associated with the largest estimated probability density from KDE in a time bin (Fig. 3.1). Available environment was summarised as average MAT across global marine points (section 3.5.2, 'Definition of "available environment"'). I considered separately any discontinuous time series for a focal species, resulting in 32 time series from the 24 study taxa.

If species adapted to changing environmental conditions *in situ*, one would expect a positive correlation between preferred and available global temperatures. If instead niches remained static and species migrated to track preferred thermal conditions, one would expect a negligible correlation, and results were consistent

with this scenario. The mean correlation (Pearson's r coefficient) between species' thermal optima and available temperatures was 0.062 with standard deviation of 0.26. Results were nearly identical under the depth-habitat approach ($r = 0.031 \pm 0.23$).

3.3.2 Niche lability *vs.* climate change magnitude

In addition to correlating standing time series of thermal preferences and global temperature (above), I also compared bin-to-bin changes in these two variables. This second examination of niche lability quantified changes in the entire niche distribution from one time step to the next, rather than tracking a single point estimate of a niche. For each species, I compared thermal niche estimates at consecutive time bins using Hellinger's H distance (Fig. 3.1; section 3.5.7, 'Niche dissimilarity measure') and then averaged H distances among species that crossed each time bin boundary. The process of building, standardising, and comparing niche distributions is presented in sections 3.5.6 and 3.5.7 and Figs. C.2 and C.3. As a correlate to the average species' niche shift across each boundary, I calculated the magnitude of change in mean global temperature from one time step to the next.

If species shifted, broadened, or otherwise changed their niches in response to climatic perturbation, one would expect larger H distances with larger climate change magnitude. Instead, mean H appeared to vary independently of global change, under both the sea-surface and depth-habitat approaches (Figs. 3.2 and C.4). Niches were similar (H was small) between every pair of intervals (Figs. 3.2 and C.4).

3.3.3 Niche lability among extreme climate intervals

In addition to tracking thermal niche change across consecutive time steps, described in the previous section, I compared niches at glacial minima and maxima (Figs. 3.3A and C.5, Table C.2). For every possible pair of different (i) glacial, (ii) interglacial, or (iii) glacial–interglacial time bins, I calculated mean H distance among species sampled in both bins (Figs. 3.1 and C.1). If niches adapted to available environmental conditions, one would expect more niche divergence (larger H distances) in the glacial–interglacial category than in either of the two control

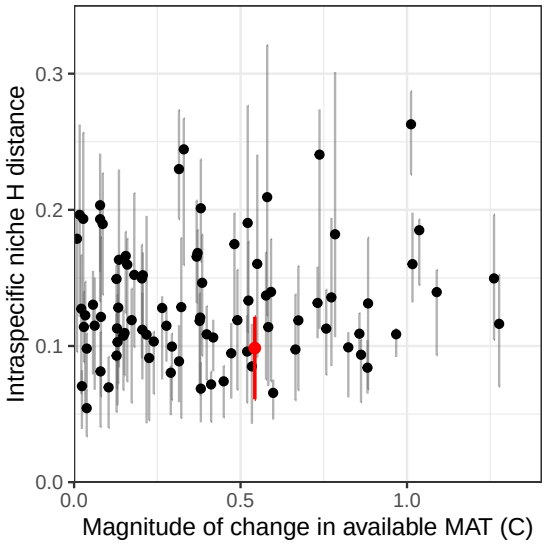


Figure 3.2: Niche dissimilarity across each time bin boundary (mean Hellinger's H among boundary-crossing species) was independent of absolute differences in available temperature (surface Mean Annual Temperature averaged among global marine sites) across each boundary. Corrected for temporal autocorrelation, Pearson's r correlation coefficient was 0.18, with 95% confidence interval [-0.03, 0.37] (uncorrected $r = 0.081$, 95% CI = [-0.13, 0.29]). The most recent interval (12—4 ka ago) is plotted in red. Grey bars denote interquartile ranges. $H = 0$ for identical distributions.

categories. By comparing niche lability at extreme endpoints of climate-change fluctuations, this test could reveal lability differences that might be too subtle to detect in the preceding analysis, which evaluated the narrow gradient of change in available environment from one time bin to the next.

Niches were no more different between intervals of climatic extremes than they were in intervals of similar climate (Figs. 3.3B and C.6A), supporting the preceding results that suggested thermal niches of planktonic foraminifera have been static. Sampling in the youngest interglacial and oldest glacial interval was anomalous, and these bins were leverage points in the analysis (section 3.5.8, 'Niche lability among extreme climate intervals' for details). When only the middle six glacial and interglacial intervals were considered, H distances were smaller (niche estimates were more similar) and less variable, but the pattern across categories remained the same (Figs. 3.3C and C.6B). Furthermore, I tested the analytical power to detect signals of niche change by simulating fossil occurrences under a hypothesis of local niche adaptation within unchanging geographic distributions (section 3.5.8, 'Test for type

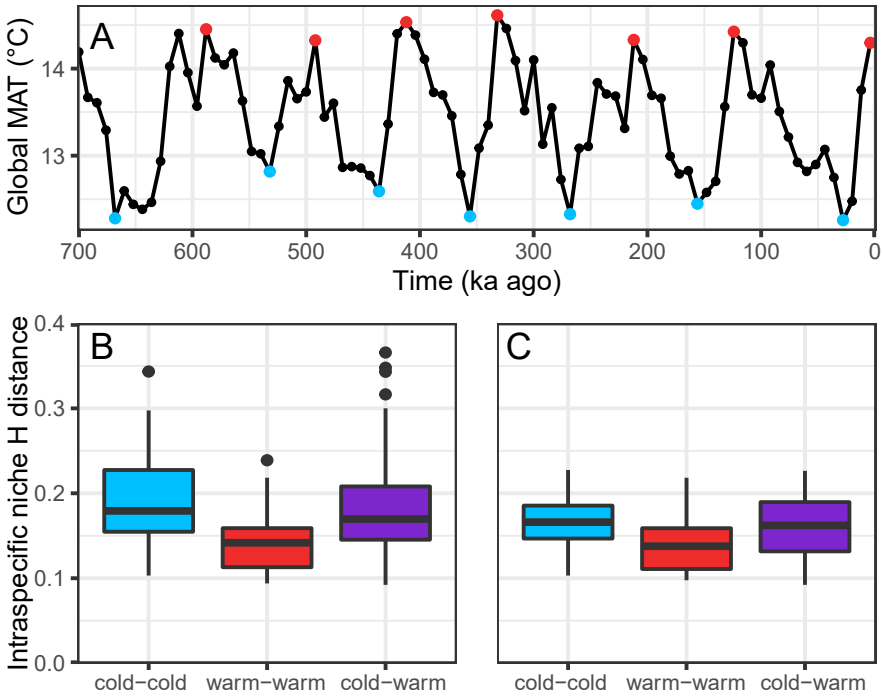


Figure 3.3: **(A)** Time series of mean annual temperature (MAT) averaged over a standardised global grid. Blue points identify glacial maxima, and red points highlight interglacial warm peaks (ages listed in Table C.2). **(B, C)** Hellinger's H distances were equivalent for niches between pairs of glacial maxima ('cold–cold'), pairs of interglacial peaks ('warm–warm'), and pairs of glacial and interglacial extremes ('cold–warm'). Boxplots depict the variation in mean species' H among all pairs of time bins in each category. In **C**, the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were omitted from comparisons due to anomalous sampling in these intervals, as described in section 3.5.8, 'Niche lability among extreme climate intervals'; note the smaller variance in **C** than **B**.

II error'). A signal of adaptation was detected in all 100 simulations (Fig. C.7), which adds confidence that niche lability would have been detected if present and strong.

3.4 Discussion

My analyses found prevalent stability of thermal niches in planktonic foraminifera across 700 ka. The independence of preferred and available temperature time series indicated that niche lability was not tied to climate cycles at the scale of 10^3 years. Similarly, analyses of intraspecific niche dissimilarity through time indicated a decoupling of niche lability and climate change magnitude, both on short intervals (8 ka) and across glacial–interglacial extremes. Previous studies have argued that

foraminifera do (Irwin, Finkel, et al. 2015) and do not (Waterson, Edgar, et al. 2017) adapt to environmental change on millennial scales, but the results above, from a much larger number of time intervals, suggest planktonic foraminifera track suitable habitats rather than changing their thermal preferences.

Niche stasis could result from failure to adapt at range margins, where individuals would be most likely to encounter novel combinations of abiotic and biotic conditions (Donelson, Sunday, et al. 2019). One hypothesis for limited adaptation is that dispersal of individuals from central populations swamps locally-adapted individuals in the gene pool at range margins (Bridle and Vines 2007; Polechová and Barton 2015). Alternatively, lack of emigrants to marginal populations could stifle genetic variation and thereby limit adaptation (Bridle and Vines 2007; Polechová and Barton 2015). Outcomes depend on dispersal ability relative to environmental gradient steepness, population density relative to carrying capacity, and selection strength relative to genetic drift (Bridle, Polechová, et al. 2010; Polechová and Barton 2015). The balance of each set of factors differs for marine ecosystems in comparison to their better-studied terrestrial counterparts (Donelson, Sunday, et al. 2019). Foraminifera, like other plankton and animals with pelagic larvae, disperse great distances on directional currents and attain enormous population sizes. Terrestrial organisms with large populations and long dispersal distances, such as some plants, may share the pattern of limited niche lability found here.

Abiotic niche stasis puts species at risk of population declines and extinction from climate-related stressors (Lavergne, Evans, et al. 2013). Survival hinges both on suitable conditions continuing to exist somewhere on Earth, and on populations establishing in those areas at least as quickly as populations elsewhere go extinct. Due to shallower environmental gradients over ocean expanses compared to on land, marine species must migrate quickly to keep pace with preferred conditions (Burrows, Schoeman, et al. 2011)—poleward shifts in living marine species have been clocked at nearly six times faster than those of terrestrial species (Lenoir, Bertrand, et al. 2020). For calcifying plankton, ocean stratification and acidification add to the stress of warming water temperature (Jackson 2008; Dutkiewicz, Morris,

et al. 2015). Declines and redistributions in this functional group will cascade to affect marine food-webs (Jackson 2008; Cahill, Aiello-Lammens, et al. 2013) and biogeochemical cycles (Hutchins and Fu 2017).

The specific result of thermal niche stability in foraminifera carries important consequences even if conclusions are not extrapolated to other clades. In particular, 'transfer functions' in palaeoclimatology assume that thermal niches of foraminifera remain constant. These functions convert microfossil census data to proxies of sea surface temperature, which serve as the basis for forcing and validating ocean circulation models (CLIMAP Project 1981; Kucera, Weinelt, et al. 2005). Hence, evidence of prolonged niche stability in foraminifera strengthens the theoretical basis for the tools used to understand climate dynamics and infer past Earth conditions.

3.5 Methods

3.5.1 Palaeoclimate reconstructions

The temperature at all 3-dimensional locations in prehistoric oceans is not directly observable but can be reconstructed using a coupled Atmosphere–Ocean Generalised Circulation Model (AOGCM). I obtained mean annual temperature (MAT) estimates from the UK Met Office Unified Model Hadley Centre Coupled Model (HadCM3), with the MOSES 2.1 land-surface model and TRIFFID dynamic global vegetation model (Cox 2001). The precise configuration was HadCM3b-M2.1aD (Valdes, Armstrong, et al. 2017). This AOGCM linked atmospheric circulation (at a resolution of 3.75°x2.5° and 19 levels) to ocean circulation (at 1.25°x1.25° horizontal resolution and 20 vertical ocean levels). The AOGCM included many environmental variables besides MAT, such as salinity, water age, and North–South and East–West current velocity. Primary productivity and terrestrial nutrient input were not modelled in the AOGCM because these variables are less accurate and more difficult to reconstruct globally, although they are important resource axes for foraminifera in some contexts (Brombacher, Wilson, et al. 2018).

A climate simulation was run at 4-ka intervals for the last 700 ka, initialised with climate estimates from similar HadCM3 simulations based on the last glacial–interglacial cycle (Davies-Barnard, Ridgwell, et al. 2017). Each simulation was run for 1,000 years: 900 years of spinup to allow equilibration, and a further 100 years that were averaged to produce the output. The ocean layers of interest (within the upper few hundred meters) required only ~500 years to equilibrate. Deeper waters reached equilibrium more slowly (on the scale of millennia) but impacted surface waters only slightly.

Each simulation at each time step was accorded time-specific boundary conditions for orbital position, greenhouse gas concentrations (carbon dioxide, methane, and nitrous oxide), and ice sheet extent and volume (and corresponding ocean salinity). The greenhouse gas forcing for each time step was linearly interpolated from ice core records. Ice sheet reconstructions were those of De Boer, Van de Wal, et al. (2013) and De Boer, Lourens, et al. (2014), which are not as well constrained as more recent reconstructions—e.g. ICE-6G (Peltier, Argus, et al. 2015) and GLAC-1D (Tarasov and Richard Peltier 2002; Tarasov, Dyke, et al. 2012; Briggs, Pollard, et al. 2014; Tarasov, Hughes, et al. n.d.) for the past 27 ka. However, comparisons of climates at the Last Glacial Maximum (21 ka ago) using these different ice sheets revealed relatively small changes except over the ice sheets themselves, and the DeBoer ice sheet reconstructions are the only available estimates of global ice sheets over the last 1 Ma.

3.5.2 Definition of 'available environment'

I defined 'available environment' as the set of ocean temperatures from a standard latitude–longitude grid at a given time step, as reconstructed by AOGCM. Analyses compared species' thermal niches to the mean or first difference of mean available environment. Hence, the way available conditions were defined could affect results. One potential issue is that averaging MAT of all AOGCM grid cells indiscriminately in a time step could falsely dampen climatic signals. During glacial intervals some cells were covered in ice and therefore would not contribute to MAT estimates,

bringing up global mean temperature; during warm intervals some of these polar cells would be open water, bringing down global mean temperature. To obviate this issue of ice-volume fluctuation, I laid a 10° grid of points on the Earth and identified the points that consistently recorded open-water conditions throughout the study interval. I then took the mean of MAT at these ocean points to generate a time series of global temperature, plotted in Fig. 3.3A. For the depth-habitat approach to analyses, I took separate global means at each of three ocean depth layers.

Available environment could be defined as the set of temperatures at all sampled occurrence points instead of at all marine points. Depending on the direction of the null hypothesis, one could argue that using sampled sites would lessen the chance of type I error. In practice, however, I found that replacing the global mean with the sample-site mean gave equivalent results. This is because the entire study interval was well sampled and the mean MAT among all sampled points closely tracked the global mean (Pearson's $r = 0.80$), although with slight discrepancies in the timing of some of the interglacial peaks (Fig. C.5). Therefore, results herein are derived from the global marine environment.

3.5.3 Fossil occurrence data

I drew on the most comprehensive dataset of Cenozoic foraminiferal occurrences to date, Triton (Fenton, Woodhouse, et al. 2021). I subset the dataset to records from the last 700 ka to match the temporal extent of AOGCM reconstructions. Most occurrence records (10^5) were assembled into Triton from datasets in the PANGAEA Open Access library (`https://pangaea.de`). The compilation also includes records (10^3) from the Deep Sea Drilling Project (1968–1983), the Ocean Drilling Program (1985–2004), the Integrated Ocean Drilling Program (2004–2013), and the International Ocean Discovery Program (2013–present); these were integrated from the Neptune database (`http://www.nsb-mfn-berlin.de`) or the program's online data repository. Additionally, the Triton dataset includes occurrences from ForCenS, which itself draws on four compilations and six datasets, including the compilations

of CLIMAP, the Brown University Foraminiferal Database, the ATL947 database, and MARGO (Siccha and Kucera 2017).

Taxonomy in the dataset is harmonised according to currently valid nomenclature (Fenton, Woodhouse, et al. 2021). Macro- and micro-perforate species are included, but benthic foraminifera are excluded. I excluded two species that originated within the last 700 ka and eight species that went extinct over that span, to ensure all study species were extant throughout the entire study interval.

All specimens were collected from sediment cores and core tops, with no live-collection samples such as from plankton tows. Even the youngest sediment samples are estimated to be centuries or millennia in age (Jonkers, Hillebrand, et al. 2019). Thus, all specimens reflect pre-industrial species distributions, which were unaffected by anthropogenic global warming. Palaeo-coordinates were reconstructed by rotating modern coordinates according to the plate model of Matthews, Maloney, et al. (2016), in accordance with Neptune methods.

I omitted records from sites at shallow depths, where planktonic foraminifera cannot establish viable populations (Darling, Kucera, et al. 2007). The exact minimum depth for reproduction is unknown and may vary with water conditions. Here, I excluded samples collected at < 100 m water depth under the assumption that any foraminifera at those sites drifted in from elsewhere. Deep-water sites could also be problematic because carbonate dissolution could preferentially destroy fossils of certain species. However, such preservational artefacts had negligible impact in previous studies of abiotic niche estimates in foraminifera (Waterson, Edgar, et al. 2017), and so I retained deep-water sites for this study.

I extracted MAT values from AOGCM raster cells containing foraminiferal occurrences. If the MAT value of a cell was NA (because the cell partially fell on ice or land), I averaged the values of the nine surrounding cells. Finally, I converted presence, absence, and count data to presence-only occurrences for individual species. Absence data were not used subsequently except to estimate distributions of sampling effort (i.e. sampled core locations as in Fig. C.3). Occurrences for a species in a time bin were included only if they totalled six or more, a minimum

cutoff for robust probability density estimation. Similarly, any species not observed to cross at least six consecutive time bin boundaries was excluded. These sample size constraints reduced the scope of the dataset from 58 to 24 planktonic species. Therefore, results might be more representative of species that are widespread and abundant than those that are restricted and rare.

3.5.4 Spatiotemporal resolution

I binned occurrence records into 88 intervals of 8 ka, from the recent sub-fossil record to 700 ka ago. The chosen bin length reflects the resolution of the AOGCM outputs (4 ka), the time-averaging of fossil assemblages and imprecision in fossil ages (up to a few thousand years [Martin 1999]), and the rate of climatic change (fluctuation from glacial minimum to interglacial maximum approximately every 50 ka). To minimise error in bin assignments due to age uncertainty, records were omitted if age estimates were derived from foraminiferal zones or had confidence intervals longer than 2 ka; remaining records could be assigned confidently to a single time bin based on the mean age estimate. Occurrences were paired with AOGCM data from the midpoint of time bins (e.g. climate was modelled at 12 ka ago for occurrences 16 ka to 8 ka in age).

I rasterised point coordinates to a 1.25° latitude–longitude grid and shortened the dataset to unique combinations of species, grid cells, and time bins. This spatial resolution matched the scale of the AOGCM reconstructions and is appropriate for the magnitude of error in occurrence coordinates. Spatial imprecision could occur from ocean currents displacing foraminiferal tests as the dead organisms sank to the seafloor. For sites at < 1 km depth or for large foraminifera, the distance between where the organism died and where it settled may be on the scale of a single grid cell (Gyldenfeldt, Carstens, et al. 2000; Qiu, Doglioli, et al. 2014). For depths of 2–3.5 km, a maximum displacement of 100–400 km is reasonable, corresponding to an error of up to 3 grid cells (Siegel and Deuser 1997; Waniek, Koeve, et al. 2000). Because MAT values are generally quite similar in adjacent cells (i.e. spatially

autocorrelated; Fig. 3.1), even the largest transport of dead foraminifera to the ocean floor would affect niche estimates undetectably.

3.5.5 Species attributes

Most living species of foraminifera are benthic, but the ~50 extant planktonic species are an abundant constituent of the zooplankton at the base of the marine food web. Until recently, all planktonic species were thought to reproduce exclusively sexually, with synchronous release of gametes at the mixed-layer depth of the water column (Schiebel and Hemleben 2017). Two research teams have now reported asexual reproduction in planktonic species, but the frequency of asexual reproduction in natural conditions is unknown (Davis, Livsey, et al. 2020; Takagi, Kurasawa, et al. 2020).

Foraminifera species inhabit specific and consistent vertical ranges and do not participate in vertical diel migration as do some other plankton, although these protists do migrate to the mixed layer depth to breed (Meilland, Siccha, Weinkauf, et al. 2019). Therefore, I performed each analysis two ways: using MAT at the sea surface to estimate niche differences, or using MAT extracted from each focal species' depth range (a 'depth-habitat approach'). In analyses based on the depth-habitat approach, each species' thermal occupancy was then compared against global ocean temperature from the corresponding depth layer. I searched the literature for quantitative information on each species' modern depth range to assign taxa to one of three depth habitats: surface (40 m in the AOGCM), surface-subsurface (78 m), or subsurface (164 m). Table C.1 lists the depth habitat and reference(s) for each species.

Habitat depth estimates are complicated by the fact that vertical structure of the thermocline, halocline, and photic gradient can shift with latitude, longitude, and time. For instance, a surface-dwelling species might live at 30 m in temperate zones but 10 m in equatorial zones. My analyses simplified this complexity by using a constant depth point to define each habitat zone. Another assumption of the depth-habitat approach is that species' depth ranges have been conserved

throughout the study interval. Since the goal of the study was to quantify niche lability, there is danger of theoretical tautology: assuming stability of depth niches to test the stability of thermal niches. Nonetheless, living depth relates intimately to fundamental aspects of foraminiferal autecology, such as whether a species is constrained to the photic zone by its symbionts, and so depth conservatism on the study timescale is probable. On one hand, the depth-habitat approach incorporates biological complexity into consideration, and so results might be more accurate than those based on sea surface data. On the other hand, due to the assumptions of the depth-habitat approach, results from sea surface data might be more precise than those from the depth-habitat approach. Ultimately, the differences between the two methods were small, and temperature values at the sea surface correlated tightly with those from habitat zones (Pearson's $r > 0.95$). Previous studies of foraminiferal ecology also recovered equivalent results whether analysis was based on data from the sea surface or at depth (Richter, Watteaux, et al. 2019).

3.5.6 Kernel density estimation

I used kernel density estimation (KDE) to construct niche distributions. KDE generates a continuous probability density function (PDF) from a finite set of observations along one or more independent axes. Unlike a probability mass function for discrete data, the PDF at a given point does not represent the probability of an event (e.g. species occurrence) at that point; rather, the integral of the PDF over a defined interval is the probability of an event occurring within those bounds. KDE is a cornerstone of nonparametric statistics with a robust mathematical foundation and applications in many fields, including ecology (Silverman 1986; Worton 1989; Horne, Garton, et al. 2007; Broennimann, Fitzpatrick, et al. 2012). KDE can be conducted in any number of dimensions, but every additional dimension requires exponential increases in sample size to maintain low mean square error (Silverman 1986). I chose to prioritise high accuracy of niche estimates (low error) and fine temporal resolution (dividing available data into many time bins) at the expense of a multivariate response (e.g. both MAT and salinity niche occupancy).

One of the difficulties of using KDE with ecological data is the prevalence of non-uniform sampling along environmental gradients. In these instances, one must account for the sampling distribution to make fair and accurate inferences about niches (Broennimann, Fitzpatrick, et al. 2012; Godsoe and Case 2015). This problem is often called 'length-bias' in the statistical literature, for the common case where the probability of observing an event scales linearly with the independent variable (e.g. the chance of identifying an animal increases directly with group size). When using KDE to construct abiotic niches, non-linear bias could arise from the unequal land cover of different biomes. For instance, because the curvature of the Earth and ocean–atmosphere circulation patterns lead to larger surface areas of temperate than polar habitats, the probability of observing a species in a temperate region of niche space may be higher than that of observing the species in a polar region.

I developed the **kerneval** package in R to construct niches for each species' occupied MAT in each time bin (version 0.1.2, available at `https://github.com/GwenAntell/kerneval`). This tool can correct for uneven sampling using the method of Barmi and Simonoff (2000). First, I estimated the PDF of sampling probability along the MAT axis, based on the MAT values from all sediment core sites in each time bin. Each core is expected to contain a nearly-continuous sedimentary record of ocean conditions and fauna, and so I interpolated sampling locations for cores in the middle of their temporal duration for which no records were otherwise present in the database (Fig. C.3A–D). These 'pseudo-absences' were used only to estimate the sampling density when standardising niche estimates; no species presence points were added. Next, the MAT axis for each species was transformed by the inverse of the cumulative distribution function of the sampling probability. I performed KDE for each species in the transformed space and back-transformed the resultant PDF to interpret the niche on the original axis scale (Appendix D).

Choice of bandwidth is a critical step in KDE (Silverman 1986), and so I explored densities with rule-of-thumb ('nrd0'), cross-validation ('CV'), and Sheather and Jones solve-the-equation ('SJ-ste', Sheather and Jones 1991) methods of bandwidth selection. The rule-of-thumb over-smoothed densities to the point of obscuring

differences between species and intervals, while CV generated unrealistically rough PDFs. I choose the SJ-ste method for its middle-ground behaviour, although this too could be criticised for under-smoothing densities somewhat (Fig. C.1).

3.5.7 Niche dissimilarity measure

I used Hellinger's H distance to quantify the similarity of niche distributions between time bin pairs (Figs. 3.1 and 3.4). This measure has a long history in the statistics literature, treats the two entities of comparison symmetrically, can be adapted for discrete or multidimensional data, and behaves well even when distributions are highly peaked or contain stretches of zero probability. H ranges from 0 to 1, with 0 indicating identical distributions; this is the opposite interpretation of Schoener's D, which is a similarity measure rather than a distance measure.

To compare PDFs in a fair way, the sampling universe (observation probability) of each distribution must be the same. In the context of niche overlap calculations, two biological entities must be able to access the same region of niche space for the niche comparisons to be fair (Peterson 2011). If the sampling universe is not shared, one might infer an entity was absent from a region of niche space where another entity was present when, in reality, the region was only available to the latter species. For instance, a species might occupy extremely cold regions during a glacial period but not during an interglacial period merely because there were no equally-cold locations on Earth during the interglacial. Without accounting for this difference in availability of environment, one would calculate a larger niche discrepancy than would be justified.

One could standardise niche space either prior to or during niche construction. In either case, one must determine the segment of the niche axis accessible to species in every time bin. The upper and lower limits of the comparison interval are listed in Fig. C.2 for every depth habitat and plotted as dashed lines on niche distributions in Figs. 3.1 and C.1. If I were to standardise niche space *a priori*, I would have had to discard any fossil occurrences associated with temperatures outside the interval of comparison (i.e. data points outside the dashed lines in Figs.

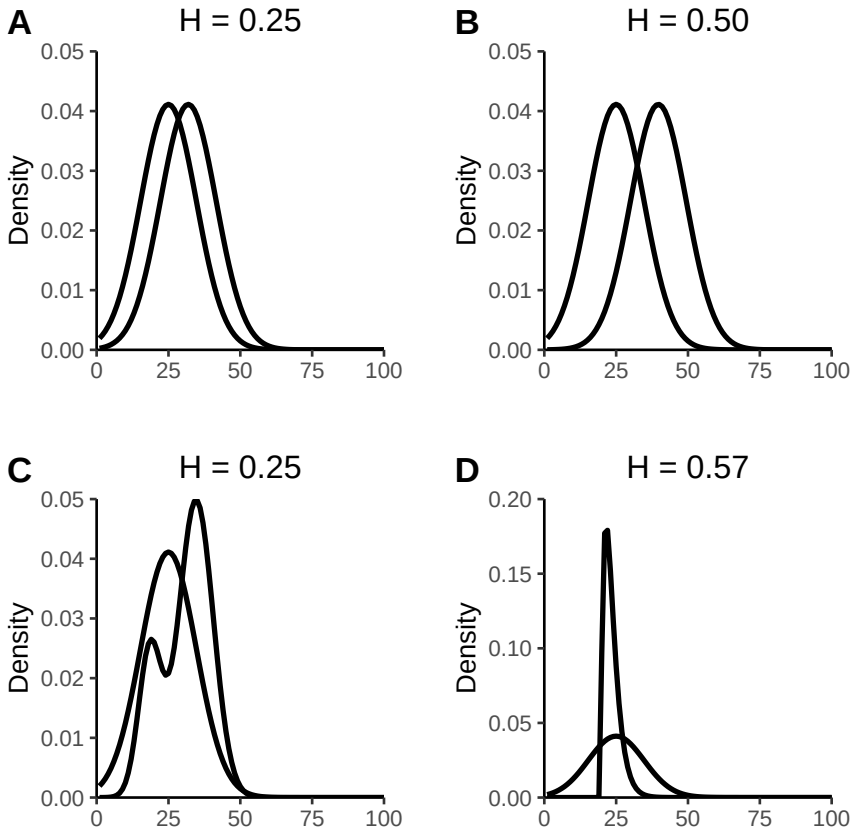


Figure 3.4: Examples of H distances between standard statistical distributions. H ranges from 0 (identical distributions) to 1 (non-overlapping distributions). In **A** and **B**, one of two identical Normal distributions is translated horizontally to give an H distance of 0.25 and 0.50, respectively. **(C)** The original Normal distribution is compared with a mixture of two Normal distributions, with H distance of 0.25. **(D)** The original Normal distribution is compared with a chi-squared distribution (4 degrees of freedom), with H distance of 0.57.

3.1 and C.1). This truncation would remove 19% of presences, almost all of which would exceed the upper limit. Instead, I used all presence points to estimate niche densities but then trimmed the density estimates to a standard MAT interval. I rescaled density functions over this interval so the probability integrated to unity.

Figure C.3 outlines the process of calculating the standard temperature interval for niche comparisons. As when estimating the sampling density for niche estimation (section 3.5.6, 'Kernel density estimation' above), I interpolated gaps in sediment core records with a 'range-through' approach (Fig. C.3A–D). I extracted MAT values from AOGCM cells corresponding to core sites in each time bin (Fig. C.3E).

I determined the sampled temperature range in each time bin and then defined the standard interval for niche comparisons as the temperature range accessible to species in all time bins (Fig. C.3F).

For some niche estimation problems, it may be desirable to reflect data across estimation bounds so that the density does not drop towards zero as it approaches those bounds (Appendix D.6.2). For example, if the range of sampled values is much narrower than the suspected true breadth of the distribution, then boundary reflection could improve KDE accuracy. I did not reflect boundaries during KDE, however, because the temperature bounds of niche comparison were close to the physiological tolerances of foraminifera. The lower bound for sea-surface data (−1.4 °C) approaches the freezing point of sea ice (−2 °C). The upper bound (27 °C) is slightly lower than the maximum thermal tolerance of foraminifera, 30–32 °C (Bijma, Faber, et al. 1990). However, under default KDE settings a distribution generally does not fall to zero exactly at each bound, but rather tapers towards zero at a point three times the bandwidth farther away. Thus, for my data, KDE distributions without boundary reflection match closely the laboratory evidence for foraminiferal thermal limits. In addition, 27 °C is the point at which the tight linear relationship between sea surface temperature and thermocline structure begins to break down. Comparing foraminiferal presences to global temperature in a linear way is justified only up to this point (Rutherford, D'Hondt, et al. 1999).

3.5.8 Statistical analyses

Time series correlation

The first analysis correlated time series of species' thermal optima and global temperature. I constructed separate time series for each species, imposing a minimum length requirement of seven time bins. If a species was sampled in multiple, discontinuous sets of at least seven consecutive bins, then each segment was analysed separately. For each time step in a species' series, I used KDE to define preferred (optimal) temperature as the MAT value associated with the PDF maximum. I then correlated each time series of thermal optima with concurrent

estimates of mean global MAT. I report the mean and standard deviation of correlation coefficients (Pearson's r) across all time series ($n = 32$). A strongly positive correlation coefficient could indicate niche shifts in adjustment to available environment ('Definition of "available environment",' section 3.5.2 above). In the depth-habitat version of time series analysis, MAT across global marine points was extracted from the same depth layer as was preferred by each species.

Niche lability *vs.* climate change magnitude

I calculated niche dissimilarity (Hellinger's H distance) for each species across each boundary it crossed. I then averaged H distances among all species that crossed a given boundary, for each of the 87 time bin boundaries. Ultimately, the goal was to relate these niche lability estimates to climate change magnitude, which I defined as the absolute value of the difference in mean global MAT in each pair of consecutive bins. If climate change caused species to alter their niches, one would expect larger niche differences with larger magnitude of environmental change.

I correlated preferred and available temperature both with and without removing temporal autocorrelation ('pre-whitening'). First I checked for autocorrelation: an augmented Dickey-Fuller test indicated the time series of climate change magnitudes was stationary but the series of mean H distances was not. Therefore, a first-order autoregressive time series model was fit to the H series; the estimated coefficient of autocorrelation was 0.49. In the case of pre-whitening the data to remove autocorrelation, I used the series of residuals from the autoregressive model in place of the original H values as a correlate to climate change magnitude. Temporal autocorrelation was negligible in the time series of H model residuals and the time series of climate change magnitude, and in the residuals of a linear model of the former on the latter. This last result confirmed that no further correction was needed to account for pseudoreplication in the correlation of interest.

Niche lability among extreme climate intervals

I calculated Hellinger's H distances not only across consecutive time bins (described in the previous section), but also in pairs of discontinuous time bins representing

glacial maxima and minima. I determined the timing of these climatic extremes based on the average global marine temperature curve (Figs. 3.3A and C.5, Table C.2). The timings derived from the AOGCM data agreed closely with interglacial onsets reported in the literature (Lisiecki and Raymo 2005). For each pair of time bins (e.g. the glacial maximum 28 ka ago and the interglacial peak 212 ka ago, shown in Fig. 3.1), I took the mean H distance among all species observed in both bins. I used boxplots to summarise variation among bin pairs in each of three climate categories: glacial–glacial, interglacial–interglacial, and glacial–interglacial intervals (Figs. 3.3 and C.6). The mean global temperature difference for time bin pairs in each category was 0.23 °C, 0.14 °C, and 1.99 °C, respectively.

I conducted a variation on the extreme climate comparisons analysis to test sensitivity to sample size deviations through time. One of the benefits of KDE is its invariance to slight deviations in sample size; however, the number of observations influences bandwidth calculation to a small degree. The estimate of a complicated distribution will be somewhat rougher with denser sampling than with sparse sampling. This responsiveness to sample size is considered a desirable statistical quality in theory, because it means that estimates will be perfect representations of distributions as sample size goes to infinity.

Unfortunately, in the empirical case of a comparison of foraminifera niches in the most recent interglacial (4 ka ago) against any older interval, the difference in sample size was up to two orders of magnitude: of the 42,233 occurrences, 19,711 (47%) were from the 4-ka-ago time step, 2,079 were from 12 ka ago, and the rest were spread nearly evenly across older time steps. The density estimates for species at 4 ka ago are noticeably rougher than in much older intervals (Fig. C.1), which could inflate the H distance between densities. In the 86 time bins older than 12 ka, the correlation between sample size and bandwidth estimate was negligible across density estimates (Pearson's $r = 0.040$, 90% CI $= [-0.0087, 0.089]$), and the sample size difference between any consecutive pair of time bins was small enough to be safely ignored in time series and bin-to-bin niche lability analyses. Among glacial intervals, the oldest (668 ka) was sampled least and therefore could influence

H distances with other bins. Given the sampling heterogeneities, I repeated the analysis of niche lability among climate extremes after excluding these two temporal endpoints of glacial–interglacial cycles from comparison (Figs. 3.3C and C.6B).

Test for type II error

I simulated fossil occurrences through time to test for the possibility that the empirical analyses were unable to detect niche changes that occurred (i.e. the possibility of falsely detecting negative signals of niche change, a type II error). In each simulation, species' occurrence sample sizes and the magnitude of climate change were kept the same as those in the empirical data. I used the geographic occurrences of species at 12 ka ago as a reference for species' 'true' distributions and assumed geographic ranges were static (and hence niches adapted to local environmental changes) through time. I chose the time step at 12 ka since global temperatures were close to the long-term mean (Fig. 3.3A) and sampling was dense throughout species' geographic distributions.

I generated occurrence data for species in glacial and interglacial intervals (listed in Table C.2), excluding the youngest and oldest extremes (4 ka and 668 ka) as in the niche lability analysis presented in Figs. 3.3C and C.6B. For each interval and focal species, I randomly drew the observed number of occurrences for that interval and species, sampling without replacement from the species' occurrences at 12 ka. Then, I extracted the sea surface MAT at the simulated presences based on AOGCM data for the relevant time step. I used the simulated data to run the 'Niche lability among extreme climate intervals' analysis as described in section 3.5.8 to determine whether mean H distance was larger between pairs of glacial and interglacial intervals than either (i) pairs of glacial intervals or (ii) pairs of interglacial intervals. The entire simulation and comparison procedure was repeated 100 times. Boxplot comparisons of results from the first five replicates are illustrated in Fig. C.7A–E.

The simulation framework considered scenarios of complete niche adaptation, but species could respond in many ways beyond a binary of either niche stasis (and geographic habitat tracking) or niche adaptation (and geographic stasis). For

instance, species could expand their niche breadth to *both* occupy the existing geographic range with new environmental conditions *and* track preferred habitat into new geographic regions. Species could also shrink in realised niche space by occupying only the subset of the original range that remained within tolerance limits. Either of these alternatives would manifest as a difference in the response of the leading and lagging geographic range edge. The simulations conducted here represent a scenario of 'lock-step' changes in range edges, the most common empirical pattern (Lenoir, Bertrand, et al. 2020), which allowed the test of whether strong and consistent lability would be detectable given the data and methods of the study.

3.5.9 Data and code availability

Analyses were conducted in an R statistical computing environment, version 4.0.3 (R Core Team 2020). All data are archived at Zenodo (DOI: 10.5281/zenodo.4658884). The **kerneval** software package can be installed from `https://github.com/GwenAntell/kerneval`. Scripts for data manipulation and analyses are available at `https://github.com/GwenAntell/ForamNiches`.

At one of her many public talks, [Lynn Margulis] asks the molecular biologists in the audience to name a single unambiguous example of the formation of a new species by the accumulation of mutations. Her challenge goes unmet.

— Michael Behe (2006), *Darwin's Black Box*, p. 26

4

Rate and mode of niche evolution across a foraminifera phylogeny

Contents

4.1 Abstract 81
4.2 Introduction 81
4.3 Results 83
4.3.1 Within-lineage evolution models 85
4.3.2 Phylogenetic ecotype comparisons 86
4.3.3 Across-lineage evolution models 86
4.4 Discussion 91
4.5 Methods 94
4.5.1 Foraminifera data 94
4.5.2 Within-lineage evolution models 95
4.5.3 Phylogenetic ecotype comparisons 97
4.5.4 Across-lineage evolution models 98
4.5.5 Data and code availability 107

The within-lineage trait evolution models and ecotype comparison analyses of this chapter have been published in Antell, Fenton, et al. (2021) (DOI: 10.1073/pnas.2017105118). The text and charts relating to these analyses appear here in edited form for inclusion with other content in this chapter.

4.1 Abstract

Abiotic niche evolution influences species' distributions, origination, and extinction. A key goal of macroevolution, the study of evolution above the population level, is quantifying rates of niche change to decipher its driving processes. Palaeontology's major contribution to this aim has been theorising and empirically demonstrating punctuated equilibrium, a mode of evolution in which traits change rapidly during speciation but experience no net change for the rest of a species' existence. The contentious history of punctuated equilibrium and divergent conclusions about niche evolution from empirical analyses of living species leave many unanswered questions about rate and mode of evolutionary change across phylogenetic trees. Combining palaeoclimate reconstructions and the fossil record of planktonic foraminifera, I fit a suite of trait evolution models to thermal niche means both within individual lineage time series and across a phylogeny of the clade. Niches remained largely static within lineages for the entire 700 ka study interval, consistent with a hypothesis of punctuated evolution. Evolutionary models including phylogenetic covariance between species lacked power to confirm a punctuational mode; even a non-phylogenetic model could explain observed differences in species' niche means to the same degree as complex evolutionary models, regardless of how fossil data were used to define niche means at tree tips. In addition, the minor variations in within-lineage niche lability could not be attributed to ecotypic differences between species, such as the presence of autotrophic symbionts. Distinguishing a signal of past evolutionary processes will require a more species-rich study clade with known relationships and fossil representatives, as well as conscientious model building, comparison, and interpretation.

4.2 Introduction

One of the most important questions at the interface of ecology and evolution is the degree to which niches are static, and the manner in which they change (McGill, Chase, et al. 2019). Macroevolutionary biologists seek to explain the

rate and mode of change in niches and other traits, in conjunction with those of speciation and extinction (Saupe and Myers 2021). As suggested by Chapter 2 and examined directly in Chapter 3, abiotic niches strongly control where species live and how likely they are to go extinct. Moreover, there is general agreement that abiotic niche evolution is more important than morphological evolution in accounting for diversification rates (Kozak and Wiens 2016; Seeholzer, Claramunt, et al. 2017). However, empirical studies have come to conflicting interpretations about the direction of causality in this relationship, mainly due to uncertainties around speciation processes. Niche evolution might cause speciation, and thence increase diversity (Kozak and Wiens 2010; Gómez-Rodríguez, Baselga, et al. 2015; Castro-Insua, Gómez-Rodríguez, et al. 2018). Alternatively, niche conservatism could promote speciation by stimulating allopatry (Wiens and Graham 2005; Pyron, Costa, et al. 2015), and diversification might precede an increase in rates of abiotic niche evolution (Folk, Stubbs, et al. 2019).

To infer historic rates of evolution from living species, evolutionary biologists must assume that rates (or their first or second derivative) are constant within lineages and that the phylogenetic relationships between lineages are known. Palaeobiologists, in contrast, have opportunity to measure trait change directly from the fossil record. Ironically, empiricists who dedicated themselves to fossil examination in hopes of confirming directional, adaptive change within lineages often found stasis in traits instead (Gould 1980; Dresow 2019). Out of this observation grew theory that most evolutionary change occurs during speciation, and once originated, species conserve their traits with minimal alteration. Gould and Eldredge took the spotlight for coining this phenomenon 'punctuated equilibrium,' although other authors had published related ideas before (Mayr 1989). Explicit descriptions of punctuated equilibrium entered the literature in the early 1970s, but it took half a decade for the ideas to mature into a strong articulation, in which the level of species was set as an additional tier atop a hierarchy of evolutionary processes (Eldredge and Gould 1972; Gould and Eldredge 1977; Dresow 2019). This bold framing sparked heated debate for a generation of scholars, and although critics

remain (Pennell, Eastman, et al. 2014), evolutionary biology is approaching 'an almost silent integration' of the major claims of punctuated equilibrium and the Modern Synthesis (Serrelli and Gontier 2015, p. *vi*).

While palaeobiologists tend to concentrate on morphological traits, it is possible nevertheless to estimate abiotic niche traits as well, by combining occurrence palaeocoordinates with environmental reconstructions (e.g. Saupe, Hendricks, et al. 2014; Stigall 2014). I followed this approach to estimate thermal niche occupancy in the study species from Chapter 3: planktonic foraminifera across the Middle–Late Pleistocene. The group makes for an interesting new case study because previous, morphological studies of planktonic foraminifera returned variable evidence for and against punctuated equilibrium—in contrast to benthic fauna where punctuation predominated (Jackson and Cheetham 1999). In this chapter I applied three sets of macroevolutionary analyses to fossil foraminifera to quantify: (i) rate and mode of abiotic niche change from time series within species' lineages ('Within-lineage evolution models'), (ii) variation in within-lineage niche lability between ecotypes ('Ecotype comparisons'), and (iii) rate and mode of niche evolution across a phylogeny, and the power and sensitivity of such estimation ('Across-lineage evolution models').

4.3 Results

The data encompassed microfossil occurrences from ocean sediment cores and mean annual temperature (MAT) reconstructions from Atmosphere–Ocean Generalised Circulation Models, binned at 8-ka resolution back to 700 ka ago. While some studies perform environmental niche modelling to derive transformed axes for niche evolution analysis (Title and Burns 2015; Salariato and Zuloaga 2017; Velasco, Martinez-Meyer, et al. 2018), most reconstruct niches from environmental values at occurrence points directly (Kozak and Wiens 2010; Quintero and Wiens 2013; Gómez-Rodríguez, Baselga, et al. 2015; Cooney, Seddon, et al. 2016; Lv, Xia, et al. 2016; Seeholzer, Claramunt, et al. 2017; Castro-Insua, Gómez-Rodríguez, et al. 2018; Folk, Stubbs, et al. 2019). I followed the majority in running within- and between-lineage models

on raw environmental values (MAT) averaged across sampling sites for a given species. To quantify niche lability for ecotype comparisons, I first constructed density estimates of entire niche distributions (Appendix D). The pairwise difference between densities was calculated as Hellinger's H distance, which ranges from zero (identical distributions) to one (no overlap). Evolutionary relationships within macroperforate foraminifera were assumed to follow the phylogenetic hypothesis of Aze, Ezard, et al. (2011). Twenty-two of 24 study species were included in the tree (Fig. 4.1).

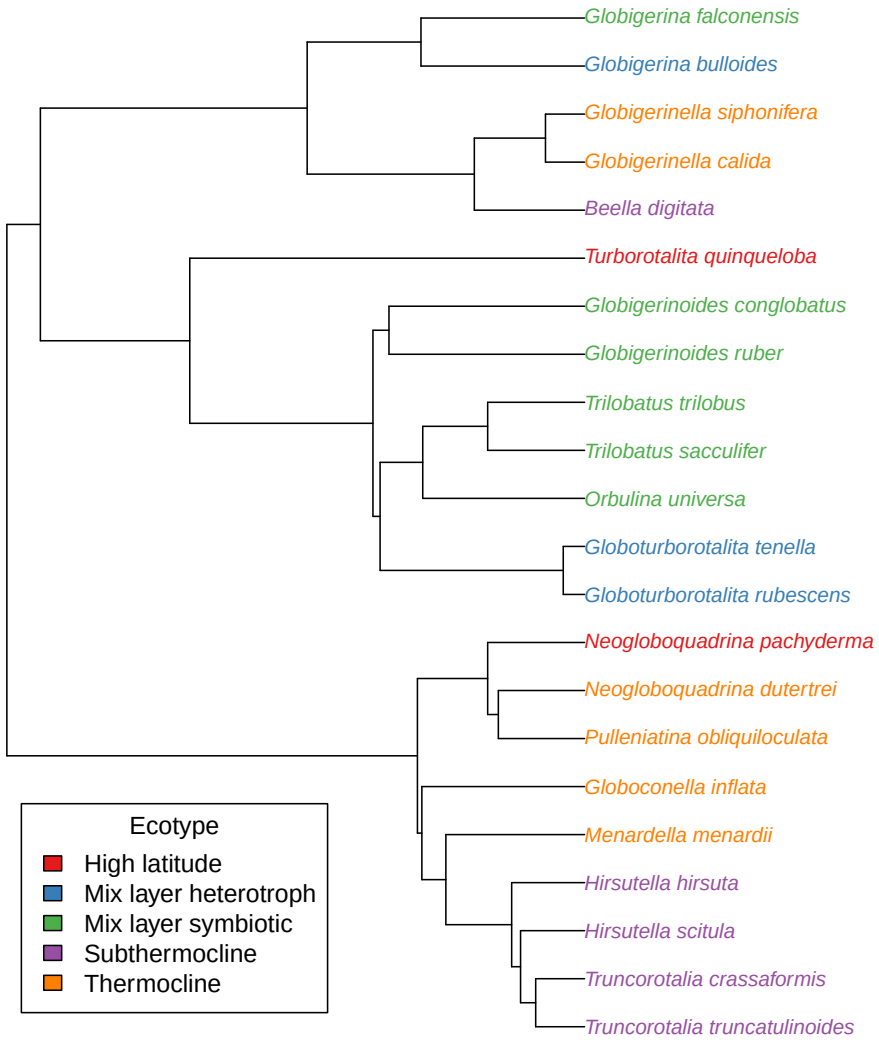


Figure 4.1: The 22 study species with phylogenetic data in the tree of Aze, Ezard, et al. (2011). Tips are colour-coded according to ecotype, assessed during phylogenetic analysis of variance. The age of the tree root is 64.97 Ma.

4.3.1 Within-lineage evolution models

I fit five trait evolution models to time series of occupied temperature for each species to quantify relative support for interpreting each series as (i) strict stasis, (ii) stasis *sensu lato* (white noise), (iii) an unbiased random walk, (iv) a biased (directional) random walk, or (v) tracking of mean global temperature (Methods section 4.5.2, 'Within-lineage evolution models' for details). Support for either of the stasis models would suggest niche stability (or white-noise fluctuation about a static mean) within lineages, support for either of the random walk models would indicate niche lability independent of climate, and support for covariate-tracking would indicate niche shifts that matched global trends in environmental conditions. Historically, trait evolution models have been applied to phenotypic data but were adapted here to environmental occupancy data as a new way to characterise abiotic niche change. I split fossil occurrence data into 32 time series of mean MAT within species (Figs. E.1 and E.2).

Stasis models fit all species' time series better than random walk or covariate-tracking models: strict stasis was best supported for 10 series and general stasis for 22 series, using temperatures extracted at the sea surface (Fig. 4.2). Support for static trait models confirms qualitative inspection of time series (Figs. E.1 and E.2). Results were similar under the depth-habitat approach to MAT estimation (Fig. E.3; approach defined in section 3.5.5): strict stasis models received the most support for 9 series, general stasis for 22 series, and unbiased random walk for 1 series (*Pulleniatina obliquiloculata*).

Evolutionary models parameterise the variance in a trait mean through time as a parameter called *omega*. In strict stasis models, the trait mean is invariant through time (i.e. *omega* is zero). In general stasis models, *omega* represents the amplitude of 'white-noise' biological fluctuation; in practice *omega* might also include some variation from heterogeneous fossil sampling. *omega* estimates from general stasis best-fit models ranged from 0.3–1.8 °C (1.1 °C on average; Table E.1). These values are small relative to the 10–20 °C temperature tolerance ranges reported from

laboratory studies of planktonic foraminifera (Bijma, Faber, et al. 1990) or the up-to-30 °C range in ocean MAT both horizontally and vertically.

4.3.2 Phylogenetic ecotype comparisons

The degree of niche lability can vary among species, and ecological and life history traits could explain some of this variation. For instance, foraminifera that actively prey upon other plankton might incur higher metabolic expenses than autotrophic species. Consequently, these heterotrophs might be more sensitive to ambient temperature changes and less flexible in adapting their thermal niche. Symbiont-bearing planktonic foraminifera experience faster speciation rates than heterotrophs (Ezard, Aze, et al. 2011), which could be explained if niches of autotrophs are more labile and if niche lability facilitates speciation (Kozak and Wiens 2010; Gómez-Rodríguez, Baselga, et al. 2015; Castro-Insua, Gómez-Rodríguez, et al. 2018). Alternatively, survivorship and speciation differences among ecotypes could be due in part to an advantage for symbiont-bearing foraminifera in warm intervals, when nutrient supply is lower (Hannisdal, Haaga, et al. 2017).

I used analysis of variance (ANOVA) to investigate differences in niche lability among ecotypes, whilst incorporating phylogenetic relationships to mitigate pseudoreplication due to shared evolutionary history (Methods section 4.5.3). Ecotype explained little variation in niche lability among species either in the sea-surface approach ($F(4, 22) = 3.31$, phy $p = 0.34$, Table E.2A) or depth-habitat approach ($F(4, 22) = 2.55$, phy $p = 0.44$, Table E.2B).

4.3.3 Across-lineage evolution models

Although within-lineage trait models fit to time series of occupied temperature values supported widespread niche stasis on scales of hundreds of thousands of years, these niches must evolve at some point, since planktonic foraminifera species have distinct temperature preferences (e.g. compare species' long-term means in Fig. E.1). Punctuated equilibrium offers one reconciliation of abiotic niche stability within lineages and niche differences across a tree, by attributing niche

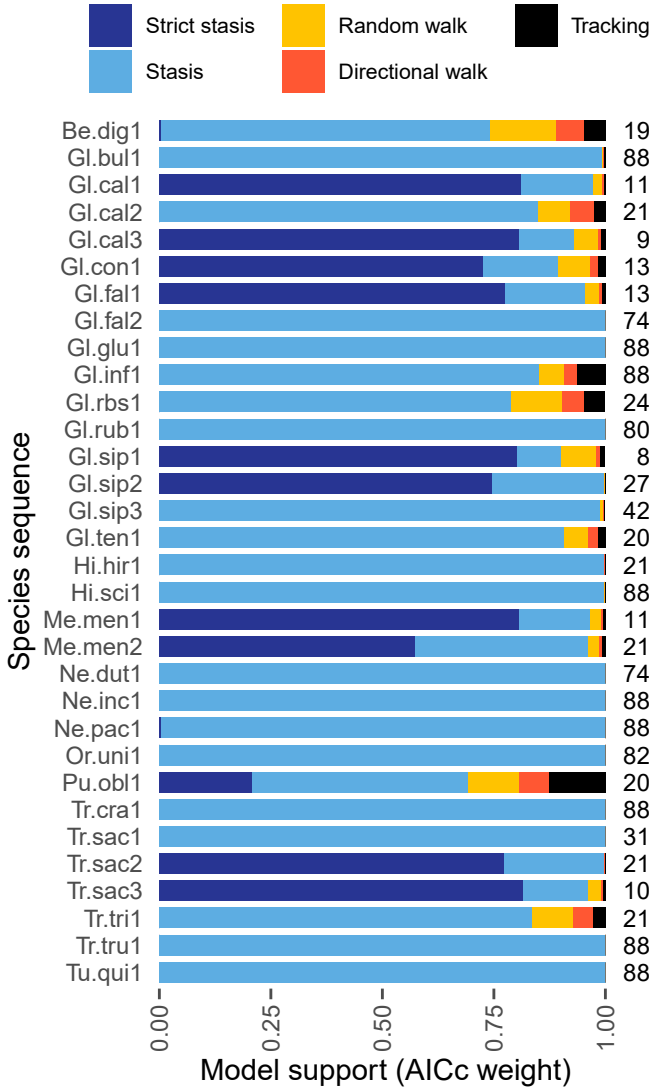


Figure 4.2: Relative support for five evolutionary models fit to each species' time series of mean occupied temperature at the sea surface (Fig. E.1). Blue indicates support for stasis (*sensu stricto* or *sensu lato*), whereas yellow and orange imply changes in central tendency of occupied temperatures away from the starting value. Black represents a model where species' occupied temperature correlates with global temperature. The preponderance of blue supports stasis as the dominant evolutionary mode of those tested. Parameter estimates from the most-supported model for each species are listed in Table E.1. The length of each series (number of 8-ka time bins) is listed at right of the figure. Binomial names are abbreviated at left as the first two and three letters of the genus and species epithets, respectively. Discontinuous time series for a single species are distinguished with numeric suffices. (Note that *rubescens* is abbreviated *rbs* to distinguish it from *ruber*.) Table C.1 lists full scientific names. AICc, Akaike Information Criterion corrected for sample size.

change to speciation events (Gould and Eldredge 1977). To test for punctuated or other macroevolutionary patterns directly, I explored niche variation across the phylogenetic tree of planktonic foraminifera (Fig. 4.1).

I fit seven competing phylogenetic models to occupied temperature from each study species, i.e. the 'tips' of the tree branches: (i) white noise, (ii) Brownian motion, (iii) *lambda*, (iv) *delta*, (v) *kappa*, (vi) Ornstein-Uhlenbeck (OU), and (vii) early burst (Methods section 4.5.4 for detailed model descriptions). Brownian motion is commonly taken as a null model, in which evolution occurs at a constant rate and following a random walk along each branch; trait covariance between species is directly proportional to their length of shared ancestry. White noise is an even simpler null model, where traits follow a normal frequency distribution without any phylogenetic signal. The *lambda* model considers the continuum between white noise and Brownian motion and estimates the position on this spectrum as the model's namesake parameter. The *delta* model parameterises an acceleration or decay in evolutionary rate through time. The *kappa* model is traditionally interpreted as support for punctuated equilibrium when its parameter is near zero. OU models are similar to Brownian motion in that evolutionary rates are constant and trait change advances as a random walk along each branch, but this walk tends to revert to its central tendency or ancestral mean with an attraction strength parameterised as *alpha*. Lastly, early-burst models are motived by assuming evolutionary rates decay as speciation fills niche space, similar to *delta* models with low *delta* values. I compared all models for each of 18 datasets, corresponding to tip values measured at every time step in which all species were sampled since the penultimate glacial maximum (156 ka).

White-noise models received the most weight for all but one time step (68 ka), at which *delta* was more supported (Fig. 4.3, Table E.3A). However, there was little power to discriminate white-noise from the next-most supported model, OU; if the data were truly generated by an OU process, only three datasets had more than 50% probability of correctly rejecting a white-noise model, allowing a 5% false positivity rate (Table E.3A). Results were similar for temperatures extracted

within habitat depths instead of at the sea surface (Fig. E.4, Table E.3B). When the error term in models was specified as the empirical standard error associated with temperature means, results became highly dependent on sample size (Fig. E.5). Models for trait data from the two most recent time bins continued to assign the largest weight to white noise, but models on older data had negligible power to distinguish alternative modes of evolution (Table E.4).

In addition to constructing datasets from synchronously-sampled species' mean temperatures at chronological time steps, I assembled 1,000 replicate datasets of mean temperatures sampled randomly along the last 156 ka of each species' history. Using sea-surface temperatures and default estimation of error, I fit the same seven evolutionary models to each asynchronous niche-trait dataset and investigated the spread of support each model received across replicates (Fig. 4.4). White-noise was the best-supported model in 94.4% of iterations, followed by OU in 2.3%, Brownian motion in 2.0%, and *delta* in 1.3%. Hence, the favour towards white-noise models was consistent whether evaluated with synchronous or asynchronous fossil data.

Inspection of coefficient estimates provides additional nuance beyond model selection. Although the *lambda* model ranked third by many weighted comparisons, the point estimate and lower bound for *lambda* were trivially small, implying the *lambda* model collapsed into another representation of white noise (Table E.5A). Small *lambda* values could result either from a true white-noise process, where phylogeny is unimportant to trait evolution, or from a small tree, which has insufficient power to estimate the value accurately (Boettiger, Coop, et al. 2012).

In contrast to *lambda*, *delta* estimates capped out at the maximum bound of estimation (Table E.5B). The full confidence interval for *delta* was greater than unity, estimating an acceleration of niche evolution through time. This estimated acceleration explains the negligible support for an early burst model (Figs. 4.3 and 4.4), which assumes decay in evolutionary rates. If thermal niches did evolve according to a *delta* model, one would expect the fastest rates along the youngest branches. However, direct estimation of niche change in within-lineage models demonstrated strong stasis in niches from the most recent 700 ka of evolution, in

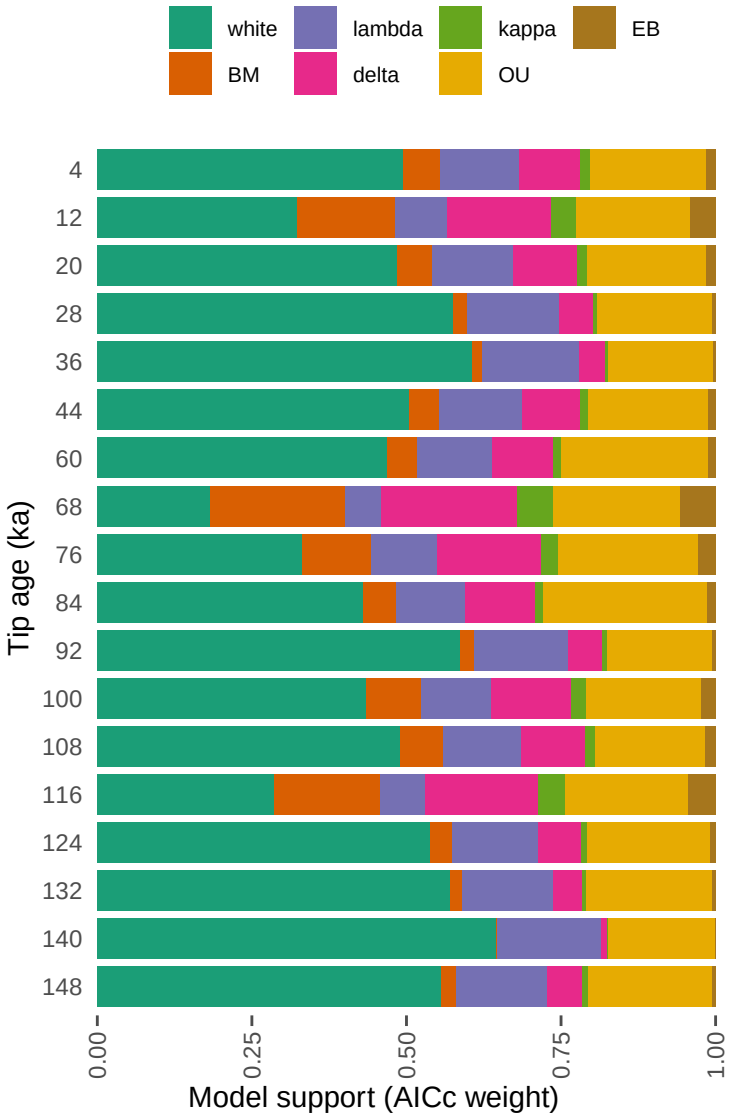


Figure 4.3: Relative support for seven phylogenetic models fit to mean occupied temperature for each species (i.e. the trait at tree tips). Bar heights are listed in Table E.3A. Tips were defined at successive time steps, numbered at left. The best-supported model for most time bins was white noise, in which phylogenetic signal is inferred absent altogether. Temperature was reconstructed at the sea surface; Fig. E.4 plots model support from temperature within habitat depths. BM, Brownian motion; OU, Ornstein-Uhlenbeck; EB, early burst; AICc, Akaike Information Criterion corrected for sample size.

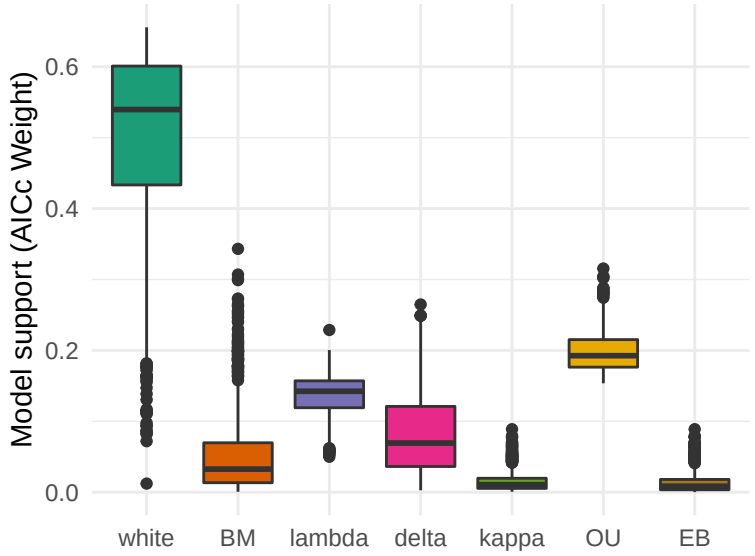


Figure 4.4: Boxplots of relative support for a given model across 1,000 datasets of tip values observed at random points along branches since the penultimate interglacial maximum (156 ka). Colours as in Fig. 4.3. BM, Brownian motion; OU, Ornstein-Uhlenbeck; EB, early burst; AICc, Akaike Information Criterion corrected for sample size.

contradiction to this expectation from *delta* models. Thus, evolutionary acceleration is far from a parsimonious explanation of within- and across-lineage observations.

Finally, the *alpha* parameter of selection strength in OU models was estimated in the vicinity expected under a Brownian motion process rather than that of white noise (Table E.5C). These values should be interpreted with caution, however, because *alpha* estimates can become inflated to absorb intraspecific variation or measurement error in data (Cooper, Thomas, et al. 2016). Overall, estimates for *lambda*, *delta*, and *alpha* parameters confirm the conclusion of bootstrapped likelihood ratio tests: thermal niche data from planktonic foraminifera have insufficient power to resolve the evolutionary process(es) acting on the clade.

4.4 Discussion

Quantifying patterns of abiotic niche evolution holds consequence for understanding macroevolutionary as well as biogeographic processes. Planktonic foraminifera exemplified conservatism in thermal niche occupancy within extant morphospecies

for at least 700 ka. This within-lineage stasis could be consistent with punctuated equilibrium, where change occurred predominantly at speciation events. Unfortunately, phylogenetic models lacked sufficient power to distinguish signal of punctuated equilibrium or other macroevolutionary processes. The variation between species was explained poorly by ecotype and instead could be modelled as coming from a single normal distribution, i.e. white-noise evolution.

Support for white noise could arise from at least three scenarios. In the first two cases, white-noise might reflect a biological reality. Thermal niches might lack phylogenetic signal, perhaps due to an absence of selection or heritability. Conversely, selection could be so strong that evolution has overwritten historical signatures rapidly relative to branch length (Cooper, Thomas, et al. 2016). However, absence of evidence is not evidence of absence (Altman and Bland 1995). A third possibility is abiotic niches carry phylogenetic signal, but variance or error overwhelmed the limited ability to detect the nature of this signal.

There were multiple sources of error in thermal niche data for foraminifera. First, there was unconstrained measurement error in palaeotemperature reconstruction. There was also large variance for a given species within a time bin, because a species could span sampling sites with different temperatures; the standard error of the mean temperature across these sites depended on the species' niche breadth and the number of sediment cores. A spread of observed values is unavoidable in any wide-ranging species because of the Cenozoic temperature gradient of planet Earth. It is a large simplifying assumption of phylogenetic models to reduce natural trait variance within a species to a single point estimate, the mean (Ives, Midford, et al. 2007). The models then lose statistical power from the small number of study species. Ultimately, sample size of taxa is limited by the diversity of the clade.

It would be possible to expand analysis of planktonic foraminifera by including rarer taxa or splitting 'pink' and 'white' genetic variants of *Globigerinoides ruber* (Aurahs, Treis, et al. 2011). However, an enlargement of the dataset by 1–10 species is unlikely to yield large benefit; in comparative analysis, each additional taxon provides only a fractional increase in effective degrees of freedom because of shared

history with existing study taxa. The size of the tree in this study was within the range of those used to fit OU models elsewhere; the mean across a literature review only five years ago was 44 tips, and the minimum a mere six (Cooper, Thomas, et al. 2016). While a small tree need not outright preclude investigation via phylogenetic models, it is imperative that authors of small datasets report statistical power and avoid over-interpretation of results.

Bayesian methods offer an alternative approach to maximum-likelihood phylogenetic modelling, with the advantage of allowing simultaneous consideration of within- and across-lineage observations. Bayesian implementations are available from multiple R packages (FitzJohn 2012; Pennell, Eastman, et al. 2014) and proprietary software (Pagel, Meade, et al. 2004), in which one can integrate fossil data into trait evolution models through informative prior distributions on internal tree branches and nodes (Slater, Harmon, et al. 2012; Uyeda and Harmon 2014). No matter the estimation approach, however, low statistical power from a small tree is likely to remain an intractable limitation to inference.

Even in absence of phylogenetic models, stability of empirical niche data within lineages has been taken as evidence for punctuated equilibrium (e.g. Hunt 2007). If one were to make an analogous claim for macroperforate foraminifera, a question arises about how speciation in this group generates niche change—or even how speciation occurs at all (Lazarus 1983). Original formulations of punctuated equilibrium assumed an allopatric or peripatric model of speciation (Gould 1989; Mayr 1989). However, it is difficult to argue for repeated isolation of populations in the marine realm, particularly for organisms with a planktonic life stage, given how much rarer and more porous geographic barriers are in oceans than on land. Although shallow-water passageways might present impassable barriers to polar species during glacioeustatic regressions (Darling, Kucera, et al. 2007), the bulk of planktonic foraminifera diversity is partitioned across ocean basins at large scale, even relative to other plankton (Sommeria-Klein, Watteaux, et al. 2021).

An alternative mode of reproductive isolation in planktonic foraminifera could be behavioural. Foraminifera change position in the water column during matura-

tion and breed synchronously with the full moon (Schiebel and Hemleben 2017). Therefore, a change in the timing or depth of spawning could instantly isolate a (sub)population from the rest of its kin (Isabel Fenton, pers. commun.). There are many uncertainties with this hypothesis, though, including large proportions of populations deviating from the canonical breeding cycle (Meilland, Siccha, Kaffenberger, et al. 2021) and recent documentation of asexual reproduction in two species (Davis, Livsey, et al. 2020; Takagi, Kurasawa, et al. 2020). The dearth of live observations on planktonic foraminifera in either the field or lab leaves much of the basic life history and reproductive biology of these enigmatic organisms still to be discovered.

4.5 Methods

4.5.1 Foraminifera data

The temperature niche data here were those described in Chapter 3 Methods section 3.5.3. Fossil occurrences stem from the Triton database (Fenton, Woodhouse, et al. 2021) and palaeotemperatures were reconstructed by Atmosphere–Ocean Global Circulation Models (Valdes, Armstrong, et al. 2017). As in Chapter 3, I excluded any species not observed to cross at least six consecutive time bin boundaries; this cut-off was necessary to ensure sufficient per-species sample size to fit within-lineage evolution models and to adequately estimate means of H distances for phylogenetic ANOVA.

For ecotype comparisons and across-lineage evolution models, I paired temperature niche data with a phylogenetic tree of the clade. *Globigerinita glutinata* and *Neogloboquadrina incompta* were missing from the phylogeny and so were omitted from these two analyses. A bifurcating morphospecies phylogeny of macroperforate planktonic foraminifera (Fig. 4.1) was available from Aze, Ezard, et al. (2011), which can be downloaded from the **timetree** R package, version 3.3.25 (Bapst 2012). I dropped from this tree any anagenetic ancestors, which have terminal edge lengths of zero, so all hypothesised speciations were associated with splitting events.

Similarly, species unsampled in the thermal niche data were pruned from the tree, so trait data and phylogenetic data contained identical lists of study taxa ($n = 22$).

4.5.2 Within-lineage evolution models

I fit five competing models of trait evolution (described below) to time series of each species' MAT occurrences to characterise niche change as static, white noise, wandering, directional, or related to global temperature. I used Akaike weights to compare the five models for each series and defined the 'best fit' as the mode of trait change receiving the most support (Figs. 4.2 and E.3). Parameters were estimated with a maximum likelihood approach (Hunt 2006; Hunt 2012) implemented in the **palaeoTS** package, version 0.5.2 (Hunt 2019). The functions in this package take as arguments the mean, variance, and sample size of a trait at each time step. These summary statistics must be calculated from the original, finite set of MAT values associated with species' occurrences. Optimum temperature arguably has more direct biological meaning than mean temperature among a focal species' occupied sites, and it is possible to estimate optima with kernel density estimation as in Chapter 3. However, inputting continuous density functions into trait models would require substantial rewrites of the **paleoTS** package; I chose to follow precedent instead and use mean values with existing code. In practice, the difference between these two metrics is minor; the Pearson correlation coefficient between mean and optimum MAT across all species-bin combinations was 0.85 (95% CI = [0.83, 0.86]) at the sea surface.

The **paleoTS** models allow several user-specified arguments. I chose not to pool the variance from time steps in a series, since some species' sample variances were unequal. The singular exception was the covariate-tracking model for *Globigerinoides conglobatus* under the depth habitat approach (marked with an asterisk in Fig. E.3), for which I pooled variances—the specific trait and covariate values for this series by chance resulted in an otherwise unsolvable problem for the default optimisation method, a limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints (L-BFGS-B). For all series, I specified a restricted

maximum likelihood approach ('ancestor–descendant') instead of a full likelihood approach ('joint'), which treats a series as a joint sample from a multivariate normal distribution. Although the joint approach is more sensitive to directional trends in noisy data (Hunt 2008), package documentation warns that ancestor–descendant parameterisation is 'generally safer' for covariate-tracking models because it avoids issues of induced autocorrelation (Hunt 2019).

The modelled evolutionary modes were: (i) strict stasis, (ii) stasis with white noise, (iii) unbiased random walk, (iv) generalised (i.e. directional) random walk, and (v) tracking of global temperature as a covariate. A strict stasis model assumes that the mean of a trait is constant through time, whereas a general stasis model assumes that the observed mean in each step is drawn from a distribution with a constant mean but positive variance. In other words, in a model of stasis *sensu latu*, observed trait means tend to fluctuate about a constant underlying mean. An unbiased random walk allows the mean in each time bin to move away from the mean in the previous bin by a step length drawn from a distribution with mean of zero and a positive variance; observed trait means migrate from the initial value in a random direction. A generalised random walk modifies the distribution for the step length such that the mean is constant and non-zero; observed trait means tend to increase or decrease depending on the sign of the step-length mean.

For the covariate-tracking model, I identified the series of global mean temperatures during time bins corresponding to the focal species' series. In the depth-habitat approach, I extracted global temperature values at the preferred depth of the focal species in the respective time bins. The model then estimated a parameter for the correlation between the trait series (species' occupied temperature) and the covariate series (global temperature). Hence, the covariate-tracking model was akin to the 'Time series correlation' analysis of Chapter 3, which found species' preferred temperature to be independent of mean global temperature. Based on this earlier result, the covariate-tracking model might be expected to receive poor support. However, by including covariate-tracking in the trait evolution models

analysis irrespective of earlier results, I could quantify the relative amount by which other models of niche change were more favoured.

It is possible to fit more complicated models than the five considered here. I refrained from fitting step-changes and other complex model types on grounds of statistical rigour and relevance to my research questions. My primary focus was mode and rate of niche change, which the five simple trait models already quantified. Incidentally, there is a known bias towards selecting a complicated model when time series are long (Hunt, Hopkins, et al. 2015), and the length of time series varied among species in the foraminifera dataset. If I had included complicated models in comparisons, sampling inconsistencies might have led simple models to be favoured more often for rare species and complicated models to be favoured more often for common ones.

4.5.3 Phylogenetic ecotype comparisons

The diversity of ecological strategies in foraminifera includes autotrophy (via association with algal symbionts), herbivory, omnivory, and carnivory (Schiebel and Hemleben 2017). These ecotypic traits were considered during phylogenetic analysis of variance in niche lability among species. Each species was classed into one of five ecotypes (Table C.1) following Aze, Ezard, et al. (2011): (i) tropical/subtropical, mixed layer, bearing symbionts, (ii) tropical/subtropical, mixed layer, without symbionts, (iii) open ocean, thermocline, (iv) open ocean, sub-thermocline, or (v) high latitude. The Triton foraminifera database also contained a category for upwelling/high productivity (Fenton, Woodhouse, et al. 2021), but no species in this group occurred in the subset of data used here, from the last 700,000 years.

To explore intraspecific variation in niche lability in relation to ecotype, I first calculated the mean Hellinger's H distance across all consecutive time bin pairs in which a focal species was observed. Dissimilarity was calculated across entire niche distributions as reconstructed by kernel density estimation (detailed in Chapter 3 and Appendix D). I merged the species-level, mean H dataset with the morphospecies phylogeny to perform phylogenetic ANOVA. This analysis partitioned

variation in H values as a function of ecotype, following the method of Garland Jr, Dickerman, et al. (1993) as implemented in the R package **geiger**, version 2.0.7 (Pennell, Eastman, et al. 2014). In this approach, the test statistic is calculated in the standard way for ANOVA, whilst the null distribution is generated by simulating data along the tree. I set the number of simulations at 10,000 to calculate Wilks test values. The evolutionary rate was estimated from the mean squared independent contrasts under Brownian motion.

4.5.4 Across-lineage evolution models

Besides investigating niche evolution within lineages, described above, I fit phylogenetic models of trait evolution to observations across the tree to assess macroevolutionary patterns. The seven models fit to each tip dataset of thermal niches were: (i) white noise, i.e. no phylogenetic signal; (ii) Brownian motion; (iii) *lambda*, (iv) *delta*, and (v) *kappa* transformations of branch lengths; (vi) Ornstein-Uhlenbeck (OU); and (vii) early burst. Details and biological interpretations of individual models are listed at the bottom of this section. I fit likelihood models of continuous character evolution with the **geiger** package, using default optimisation algorithms and parameter search bounds (Pennell, Eastman, et al. 2014).

Much of phylogenetic model development is tailored towards data from living species, i.e. data sampled at a single, synchronous time step. Fossil data, in rare cases where they are incorporated in phylogenies, usually contribute a point observation from one or few specimens. The structure of data analysed here—continuously-sampled niche traits along all branches for dozens of time steps—presents an unprecedented opportunity to fit phylogenetic models based on data from a variety of time steps. I varied the subset of model-training data according to two frameworks (Fig. 4.5).

First, I changed the age at which tree tips were defined, fitting models to traits sampled synchronously at progressive chronological ages (Fig. 4.5A,B). For fair comparison of results from different time intervals, the topology of the tree must remain constant: all branches must be sampled. Therefore, I restricted modelling

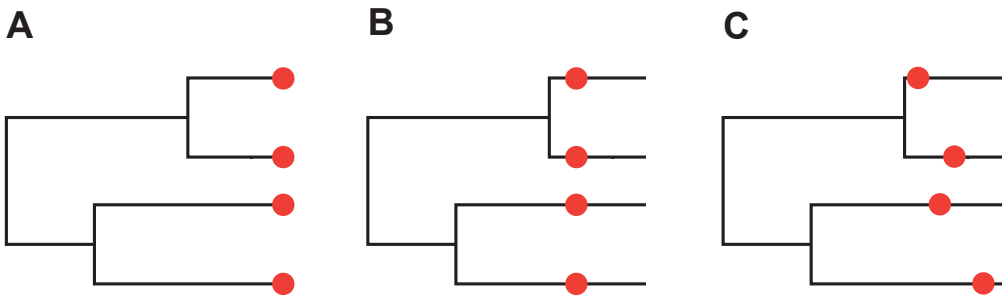


Figure 4.5: Sampling strategies for tree tips. Red points mark the age at which a tip is sampled on a simplified phylogeny of four taxa. **(A)** Tips are defined at the most recent time step, for instance from living species. **(B)** Traits are sampled synchronously at a past time step. **(C)** Traits are sampled randomly along each lineage.

to the last 156 ka (the last two glacial cycles; Table C.2), an interval of dense sampling for species time series (Figs. E.1 and E.2). Two bins in this timespan lacked complete observations of all study taxa (52 and 156 ka) and were therefore omitted. I fit all seven models to data from the full complement of species sampled in each of the 18 remaining time bins (Fig. 4.3, Table E.3). I did not adjust the branch lengths of the phylogeny to account for earlier sampling of tips, because even the oldest time bin considered represented only 0.2% of total tree height (Fig. 4.1).

Changing the time step used to define tree tips relaxes one of the constraints of how evolutionary models are applied to living species' data. As a further step, one could sample branches asynchronously (Fig. 4.5C). Arguably, if species from the same time step tend to share ecological dynamics or spatial/environmental sampling structure, then data from multiple time steps might be more informative (less autocorrelated) than those from a single step. Therefore, as another way to define the ages of tree tips, I randomly drew an occupied temperature value for each species from the 20 intervals spanning the last 156 ka. Since *Beella digitata* was too poorly sampled (<6 observations) to determine mean occupied temperature at 156 ka and *Globigerinoides conglobatus* at 52 ka, the sampling interval for these species was restricted to the 19 time bins where their niches were characterised. I iterated the random sampling procedure 1,000 times to generate a variety of data recombinations from the fossil record. I fit and compared the seven evolutionary

models on each replicate dataset, and then I summarised as boxplots the weights received for each model type across all iterations (Fig. 4.4).

As in within-lineage analysis, I used sample-size-corrected Akaike weights to compare models for a given set of observations and defined the best fit as the mode of trait change receiving the most support (Table E.3). The benefits of this model selection framework are its interpretability and ability to compare many models simultaneously. However, information criteria such as Akaike's have known drawbacks as well; for instance, small trees can bias results towards near-zero *lambda* estimates and favour OU over Brownian motion models (Boettiger, Coop, et al. 2012; Cooper, Thomas, et al. 2016). Therefore, I also bootstrapped likelihood ratios to test pairwise hypotheses discriminating between the models that received the most weight. Likelihood ratio statistics are often constructed based on asymptotic χ^2 distributions, but this is inappropriate for phylogenetic data (Boettiger, Coop, et al. 2012; Cooper, Thomas, et al. 2016). Instead, I calculated likelihood ratios from bootstrapped Monte Carlo simulations of trait evolution according to the two competing models being compared.

A bootstrap approach to likelihood ratios generates distributions of the test statistic (*delta*), which can be used to calculate p-values and power. The method has three steps (Boettiger, Coop, et al. 2012). First, the maximum likelihood parameter estimates from each of the two models under comparison are used to simulate trait values at the tips of the tree; this process is iterated to produce n new datasets. Second, for each simulated dataset, the approach estimates the parameters of each model anew, as well as the likelihood ratio statistic between the two model fits. The final step is comparison of the distribution of likelihood ratio statistics ultimately derived from traits simulated under the first model against those from the second model. The simpler model is usually considered a null hypothesis and the more complicated model an alternative hypothesis. In that framing, the proportion of ratio statistics from the null distribution that fall to the right of the ratio statistic estimated from the empirical data (Fig. 4.6A) represents the probability of observing a value at least as high as in the empirical data, if

those data came from the null model. This proportion corresponds to a p-value in Frequentist interpretation. Meanwhile, the proportion of ratio statistics from the alternative model that fall to the right of the 95% quantile of the null model distribution (Fig. 4.6B) represents the probability of correctly rejecting the null hypothesis if the data came from the alternative model, allowing a 5% false positive rate. This probability is the power to distinguish between the competing models. I generated bootstrap distributions with the **pmc** package, version 1.0.4 (Boettiger, Coop, et al. 2012). Since the most recent version of the package lacks functionality to calculate p-values, statistical power, and confidence intervals (discussed below), I estimated these values from density functions with Gauss-Kronrod adaptive numerical integration (Appendix D section 6.1).

A bootstrap approach also generates probability distributions for model variables, from which to derive confidence intervals. I reported confidence intervals as the inner 95% quantile from distributions for *lambda*, *delta*, and *alpha* (the selection coefficient from OU models)—the parameters of the three models with most weight. On account of the computationally intensive nature of bootstrapping, I estimated confidence intervals only for the datasets of tips sampled synchronously at chronological time bins, and not for the 1,000 replicates of tips sampled randomly along the last 156 ka of branches. A faster confidence interval estimation to bootstrapping is to use the Hessian matrix from the original model fit to empirical data; however, inverting the Hessian matrix is sometimes an impossible problem to solve and can generate widely inaccurate upper and lower bounds for some parameters (Ives and Garland Jr 2010; Boettiger, Coop, et al. 2012).

Model functions in **geiger** contain the option to specify standard error associated with trait data. Questions about the meaning of standard error in phylogenetic models trace back more than 30 years (Felsenstein 1985; Grafen 1989; Ives, Midford, et al. 2007). When sampling variation is high relative to comparative differences between means, perpetuating the sampling error in a model could wash out the power to detect a phylogenetic signal. Grafen (1989) called this a problem 'with no truly principled solution' (p. 147). I fit models both with and without specifying the

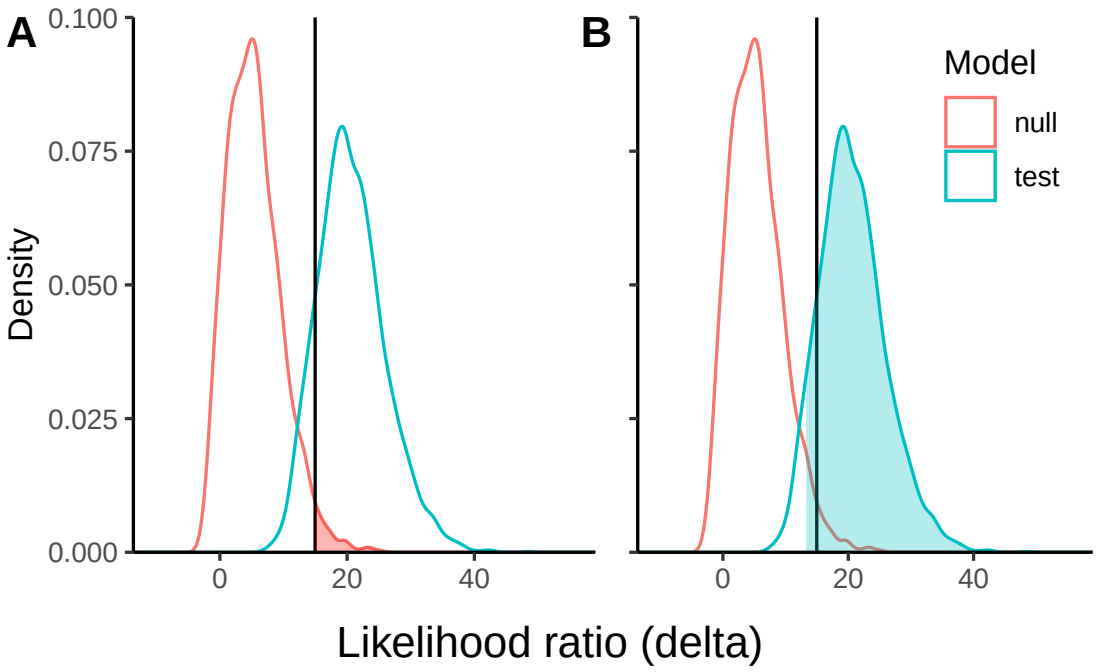


Figure 4.6: Distribution of the likelihood ratio statistic observed from 3,000 bootstrap simulations under a null model (Brownian motion, red line) or alternative model (Ornstein-Uhlenbeck with three optima, blue line). The likelihood ratio estimated from the empirical data is drawn as a vertical black line. The data are body sizes of *Anolis* from Butler and King (2004). **(A)** The shaded area under the orange density curve is the probability of generating a test statistic greater than observed if the data came from the null model. Interpreted as a p-value, this is the probability the null model cannot be rejected. **(B)** The statistical power of the test is the shaded area under the blue density curve. Allowing a 5% false positive rate, the left bound of the integral is then the 95% quantile of the null distribution. The p-value and power estimated with code from this study were 0.027 and 93.3%, respectively, whilst the values reported by Boettiger, Coop, et al. (2012) were 0.025 and 93.6%.

empirical standard error associated with per-species temperature means, calculated from the standard deviation and sample size of observations for a given species and time bin. Including the error in this way is comparable to the fitting of within-lineage evolution models above. However, model selection results from Akaike weights conflicted with those of bootstrap likelihood tests when models included empirical standard error. Furthermore, the direction of this conflict reversed in large *vs.* small sample sizes, associated with younger *vs.* older time bins (Table E.4): weights favoured white-noise models for 4 ka and 12 ka but OU for older time steps; bootstrap results inclined towards OU for the two youngest bins but had no power to reject white-noise in any older steps. Since default error estimation facilitated

agreement across datasets and selection approaches, I left error unspecified in the model formulation for results figured in the main text.

White noise

A white-noise model is non-phylogenetic, so it is uncommonly fit in comparative analyses. However, this simplicity also makes white noise a strong null model against which to compare phylogenetic models of various complexity. A white-noise model parameterises the distribution of traits among taxa as if the trait values were drawn from a normal distribution, rather than make any assumptions about covariance between taxa based on shared evolutionary history. If one were to specify a variance–covariance matrix for white noise as for the other models below, one would arrive at an identity matrix.

Brownian motion

Brownian motion is the most common model fit to continuous trait data in a phylogenetic context and follows the same statistical description as the unbiased random walk for within-lineage evolution above. Trait change at each time step is assumed to be independent and equally likely to be positive or negative (Felsenstein 1973). Over time, the trait mean will tend to wander from its initial value. The most intuitive explanation for Brownian motion is neutral genetic drift and random mutation—that is, a scenario where the trait is free from selection (Lande 1976). However, it is also possible for a trait to evolve under Brownian motion when subject to constant selection. Selection could be directional, but with a direction and magnitude that change randomly every generation, or selection could be stabilising, but with an optimum value that changes randomly every generation (Hansen and Martins 1996). Without estimates for parameters such as population size and selection strength to compare with a Brownian rate estimate, it is impossible to distinguish between these and other alternative biological explanations (Harmon 2019).

The statistical distribution of trait values at tree tips expected under Brownian motion is described by a phylogenetic variance–covariance matrix (Felsenstein

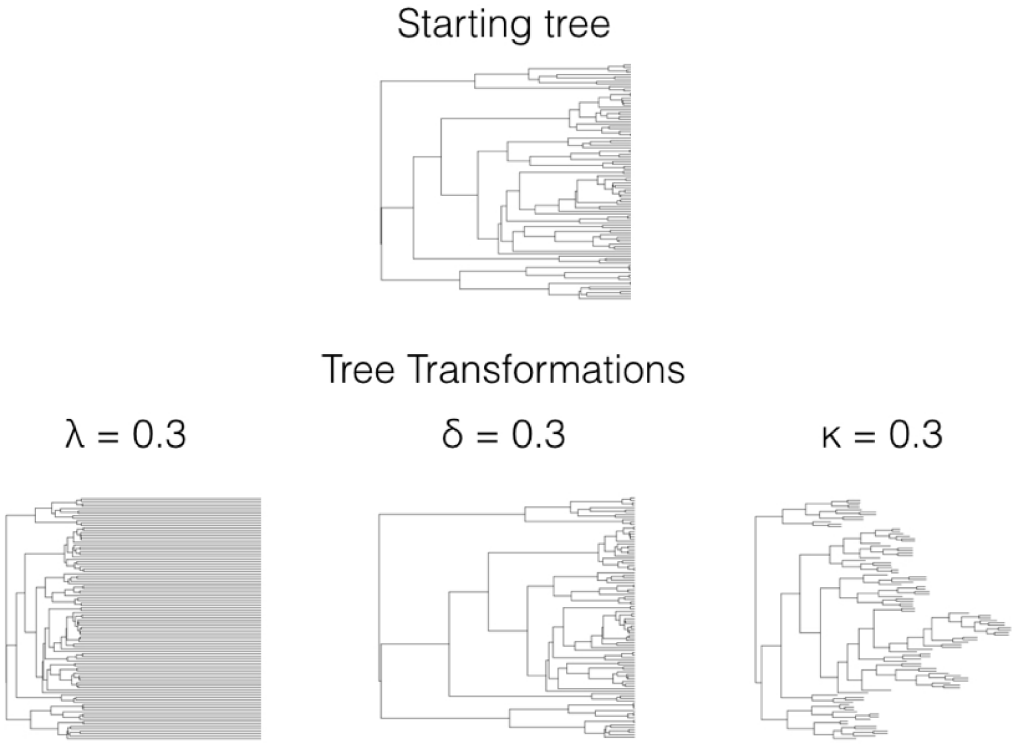


Figure 4.7: Example branch-length transformations on an ultrametric tree (top) for equal parameter values of *lambda* (left), *delta* (middle), and *kappa* (right). Reproduced from Figure 6.1 of Harmon (2019), CC-BY-4.0 licence.

1985). Each row and column corresponds to a lineage, and the diagonal elements are the age of each lineage tip from the last common ancestor. The off-diagonal elements represent the branch length shared between every pair of lineages. The phylogenetic matrix simplifies the simulation and model-fitting of Brownian motion and is a useful basis to construct other models of evolution. The *lambda*, *delta*, and *kappa* models of trait evolution result from transformations of the matrix by different constants (Fig. 4.7) (Pagel 1999).

lambda transformation

The *lambda* transformation of a phylogenetic variance–covariance matrix multiplies all off-diagonal elements by a constant (*lambda*) between zero and one inclusive. Effectively, *lambda* represents the degree to which traits are more independent than expected given shared evolutionary history ($\lambda < 1$) *vs.* changing at a constant rate

across all branches ($\lambda = 1$)—that is, *lambda* allows a scalar between end-point models of pure white noise and Brownian motion. The *lambda* transformation can be imagined graphically as an extension or contraction of the terminal branches of the tree (Fig. 4.7). Note that although *lambda* values near one can result from unconstrained evolution under Brownian motion, inference from model results is complicated by the abundance of biological processes capable of generating large or small *lambda* values (Revell, Harmon, et al. 2008).

delta transformation

Whereas the *lambda* transformation multiples off-diagonal elements of a phylogenetic variance–covariance matrix, by a constant factor, the *delta* transformation raises every element by a constant power. A *delta* value less than one models a slowdown of evolution across the tree, similar to an early-burst model, whilst a *delta* value greater than one represents an acceleration of evolutionary rates through time (Pagel 1999). Visualisation of the transformation involves rescaling branches disproportionately, with deep branches either extended ($\delta < 1$) or contracted ($\delta > 1$) relative to shallow branches (Fig. 4.7).

kappa transformation

The *kappa* model multiplies all branch lengths by a constant from zero to one inclusive. Because each element in the phylogenetic matrix is a sum of branch lengths, this numerical effect of this transformation is difficult to intuit. Biologically, however, a *kappa* estimate of zero corresponds to a tree where all branch lengths are equal (with value of one), which is interpretable as a scenario in which evolutionary change is fully and equally concentrated at speciation events—punctuated equilibrium (Pagel 1999; Harmon 2019). The *kappa* model assumes all speciation events are captured in the tree; the fossil record of many study systems is too sparse to meet this assumption, leading to the invention of alternative models that account for sampling gaps (e.g. Bokma 2008; Goldberg and Igić 2012). The Cenozoic fossil record of planktonic foraminifera is so prolific and has been examined so intensively that the morphospecies phylogeny for the clade does encompass extinct lineages.

Unfortunately, to measure thermal niches at deeper nodes or extinct branches in the tree would require age estimates and climate data that may be non-independent of foraminifera occurrences themselves; this problem is discussed in Chapter 5 in relation to ancestral state reconstructions.

Ornstein-Uhlenbeck

An OU process is a type of random walk that tends to revert to its mean value, unlike the unconstrained random walk of Brownian motion. One biological pressure that generates an OU process is stabilising selection on a trait towards an optimum value. Hence, OU models estimate an optimum trait value (*theta*) and a selection strength coefficient (*alpha*) in addition to the two parameters of Brownian motion, an initial trait mean (X_0) and trait variance across each time step (squared *sigma*). In contrast, *lambda*, *delta*, and *kappa* models each add only their namesake parameter to those of Brownian motion. The **geiger** algorithm used here models *theta* and the initial trait value as equal, a 'single stationary peak' model (Harmon, Losos, et al. 2010). Estimating these parameters separately (a 'single peak' process) can lead to poor model performance in some cases (Cooper, Thomas, et al. 2016). To report *alpha* on a standardised range, I rescaled the tree length (the distance from the common ancestor node to present-day tips) to be unity. The result is given on a $-log_{10}$ scale, where values below -4 correspond to negligible phylogenetic signal (white noise) and a value of zero corresponds to Brownian motion (Ives and Garland Jr 2010). Positive values could correspond to strong phylogenetic signal and conservatism, such as if an acting OU process were very weak, although values beyond 4 lose distinct interpretability.

Early burst

The model formulations described above developed out of widely-observed types of stochastic processes or experimental manipulations of the phylogenetic variance–covariance matrix. In contrast, the early-burst model stems directly from a biological hypothesis. An early-burst process represents adaptive radiation, premised on a limit to the number of species that can coexist in a given volume of niche space.

When a clade accesses a new area of niche space, either through environmental opportunity, novel innovation, or competitor reduction, species are expected to originate and diverge in trait space. As species accumulate and reach a hypothesised occupancy limit, however, trait evolution is expected to slow. Early-burst models parameterise the rate of trait evolution decay; in **geiger** the decay is exponential, but an accelerating-decelerating (ACDC) version of the model allows other rate functions (Blomberg, Garland Jr, et al. 2003). As reviewed in Chapter 1 section 1.4.4 and by Erwin (2015b), adaptive radiations are a fraught theoretical concept, and the early burst model has few empirical cases of support (Harmon, Losos, et al. 2010).

4.5.5 Data and code availability

Analyses were conducted in an R statistical computing environment, version 4.1.1 (R Core Team 2021). Data are archived at Zenodo (DOI: 10.5281/zenodo.4658884) or cited herein. Scripts for within-lineage models and ecotype comparisons are available at `https://github.com/GwenAntell/ForamNiches` and for across-lineage evolution models are available at `https://github.com/GwenAntell/ForamEvo`.

Oh, knowledge exploring is oh so lyrical
When you think thoughts that are empirical

Oh, let's name the species, the species, the species
Let's name the species that live in the sea

Come on, sing with me!

— Mr. Ray, *Finding Nemo* (2003)

5

General discussion and conclusions

Contents

5.1 Contribution to knowledge 109
5.2 Limitations and opportunities 112
5.2.1 Ancestral state reconstruction 112
5.2.2 Multiple sources of autocorrelation 114
5.3 Closing remarks on ecological and evolutionary scaling 116

This thesis examined macroecological and macroevolutionary patterns in the sea with regard to putative causal factors. I opened with a synthesis of major theories in evolutionary ecology for the structure of marine ecosystems past and present. In data chapters that followed, I quantified biotic and abiotic influences on species' distributions through time, testing three major hypotheses: (i) competition restricts geographic ranges on large scales, (ii) climate change pressures species to adapt environmental preferences, and (iii) niche evolution occurs primarily via speciation. In the present chapter I discuss major findings of my research and opine on remaining unknowns that merit further study.

5.1 Contribution to knowledge

Chapter 1 reviewed what is known about the history of marine ecosystems since the origin of metazoan communities. I focussed on four time intervals with pronounced ecological or taxonomic turnover: the Ediacaran–Cambrian Radiation, Great Ordovician Biodiversification Event, Devonian Nekton Revolution, and Mesozoic Marine Revolution. Two of the populous clades mentioned in this history, Brachiopoda and Bivalvia, received further attention in Chapter 2. Beyond summarising the annals of past events, I identified major hypotheses about putative causes of successive increases in functional and taxonomic diversity, primary production and total biomass, and body size and metabolic rates. Each hypothesis boasts a long bibliography of supporting and contesting works on its own, but I recognised a need to assimilate the variety of individual ideas—originating from a variety of disciplines—into a cohesive framework. In reading the literature, I developed the impression ecologists have too rarely engaged with global nutrient cycles and the magnitude of past ecosystem perturbations, while palaeobiologists have failed to assimilate the full wealth of recent knowledge derived from living ecosystems.

My goal in writing a review paper for a general biology journal was to introduce each life-science discipline to related ideas from other fields, by synthesising theory and empirical study from palaeobiology, geochemistry, food-web ecology, evolutionary biology, and oceanography. The result was a formulation of eleven 'stepping-stones' (Fig. 1.3, Table 1.1) by which changes in nutrient availability and primary producer efficiency might cascade upwards to affect consumer ecology and evolution at higher trophic tiers of marine ecosystems. Walking through the evidence for each stepping-stone, I identified gaps and untested assumptions in our collective knowledge base. Research programmes that could bridge these gaps include: primary geobiology data collection from the late Neoproterozoic through Cambrian; physiological experiments that manipulate food availability, temperature, and water oxygenation; and longitudinal observations at intermediate spatiotemporal scales. The latter topic is explored in the final section below. In addition, I emphasised

a cross-cutting need for studies to invoke explicit and mechanistic hypotheses so later researchers can test and sort through alternative explanations of data.

In the first data chapter, I tested a simple ecological hypothesis with profound evolutionary and biogeographic implications: species restrict each other's distributions in proportion to the amount they compete. I maintained Phanerozoic-wide scope and evaluated the nature and history of the relationship between geographic range size and species richness, a proxy for competitive pressure. Drawing on the dense sampling record of fossil brachiopods and bivalves, I found the number of potential competitors in a region did not explain how widespread species were. Furthermore, although ecological release predicts species to have expanded after mass extinctions, such as the end-Cretaceous, survivors of these events did not invade the space of lost competitors. My three novel analyses indicated that competitive exclusion has not dominated as a pattern in marine seafloor communities at the species level. This finding suggests the relatively greater importance of abiotic than biotic conditions in shaping biogeographic distributions.

Many recent palaeontology papers continue to couch data interpretations in terms of competitive exclusion or ecological release, even where those papers did not attempt direct measures or tests of competitive pressure (e.g. Button, Lloyd, et al. 2017; Foster, Heindel, et al. 2019). Individual taxa thought to have proliferated after mass extinction events, such as *Claraia* or *Lystrosaurus*, are easily cast as examples of geographic release, despite opposing evidence or alternative theory explaining such 'disaster taxa' as opportunists of perturbed environmental conditions (Hallam and Wignall 1997; Modesto 2020). The holistic examination I undertook of many taxa, time intervals, and geographic regions serves as a cautionary tale against cherry-picking observations from the fossil record. Although geographic shifts in individual taxa may well have occurred in response to competition and been observable on large scales, this phenomenon is replicated too rarely to be generalisable. The benefit of a quantitative approach to palaeobiology is the ability to weigh the bulk of evidence in lieu of qualitative story-telling.

A critical component of the analytical approach to testing competitive exclusion was spatial standardisation of sampling. Preliminary data exploration made clear that if I had omitted accounting for heterogeneous extent and area of sampling, the study would have come to different—and in this case erroneous—results. Unfortunately, although there is growing appreciation for the importance of spatial standardisation (Close, Benson, et al. 2017; Benson, Butler, et al. 2021), many macroecology studies have yet to control both area and dispersion of fossil sites. It is unwise to consider these unstandardised studies when drawing conclusions about the weight of support for hypotheses involving either biodiversity or geographic distributions. In the same vein, the first step in analysing planktonic foraminifera occurrences in Chapter 3 was development of sampling-standardisation methods. I wrote the R package **kerneval**, documented in Appendix D, to fill a need for a mathematically rigorous, accessible way to estimate niche distributions in the presence of observation bias.

Using sampling-corrected density estimates for thermal niche distributions, I tested the degree to which planktonic foraminifera kept pace with changing environmental conditions across the last seven glacial–interglacial cycles. The exceptional microfossil record allowed me to hone measurements to the scale of 8-ka time bins, a nearly unprecedented resolution for such a long timespan. In three statistical analyses, I found niches demonstrated negligible response to global temperature change, even when comparing glacial maxima and minima. The findings of prevalent niche stasis suggest warming during coming centuries could redistribute and reduce populations of foraminifera and other calcifying plankton, which, as explored in Chapter 1, are primary components of marine food webs and biogeochemical cycles. Even if these conclusions do not extend to other taxa, the specific result of thermal niche stability in foraminifera carries important consequences for paleoclimatology: 'transfer functions,' used to infer past climate conditions from foraminifera fossil isotopes, assume invariant thermal tolerances (Kucera, Weinelt, et al. 2005; Kucera 2007). My ecological study supports this assumption.

In Chapter 4 I turned from ecological to evolutionary questions about abiotic niches. I modelled thermal niche means as a trait in evolutionary models both within individual lineages and across a phylogenetic tree of macroperforate foraminifera. Although I had discovered in Chapter 3 that niches within lineages were uncorrelated to climate cycles, there remained a possibility that niches did change over the study interval, but in directional or wandering ways. Trait evolution models repudiated this possibility: niches exemplified stasis over the last 700 ka of extant morphospecies' history. The observation of stasis within lineages is consistent with a punctuated mode of evolution, in which niche change might accrue in large part at splitting events. Unfortunately, evolution models across a bifurcating phylogeny of the clade lacked statistical power to distinguish support for punctuated, non-phylogenetic, or other models of niche evolution.

The quantitative results presented in these chapters stand on the shoulders of thousands of workers who collected, described, or entered fossil information in the Paleobiology Database or who crewed ocean drilling program vessels or identified foraminifera from sediment core repositories. The publication associated with Chapters 3 and 4 is the first to have built on the most recent and extensive compilation of Cenozoic planktonic foraminifera records, the Triton database (Fenton, Woodhouse, et al. 2021), but it will not be the last. Quantitative palaeobiology is a burgeoning field, and collaborative, NSF-funded projects are underway to link data sources such as those used here. For instance, the in-progress Extending Ocean Drilling Pursuits (eODP) project aims to collate, standardise, and migrate plankton records such as those in Triton to the Paleobiology Database. The integration of digital repository platforms will facilitate further investigations into myriad aspects of biodiversity across space and time.

5.2 Limitations and opportunities

5.2.1 Ancestral state reconstruction

The combination of palaeoclimate reconstructions, fossil occurrences, and phylogenetic trees offers opportunities to continue unravelling the reciprocal interactions

between ecological and evolutionary processes and Earth and life systems. Taking macroperforate foraminifera as an example, one next step could be to compare temperature tolerances inferred back in time along an evolutionary tree against states observed directly in early Cenozoic ancestor species—something that is impossible to do from living species' traits or genetic data alone. Simulation studies have issued cautions about the precision and accuracy of ancestral state reconstructions, particularly those that rely solely on extant data (Litsios and Salamin 2012; Holland, Ketelaar-Jones, et al. 2020; Keating, Donoghue, et al. 2021); an empirical comparison of inferred *vs.* observed past states could be an affirming extension of that work.

One challenge for investigating ancestral foraminifera niches, besides the limited sample size of tree tips and nodes as discussed in Chapter 4, is the non-independence of deep-time foraminifera occurrences and climate estimates. Isotopes from foraminifera fossils are primary proxies for palaeotemperature estimates on which global circulation models are calibrated. In Chapters 3 and 4 I circumvented this dependency by restricting study to the last 700 ka, a recent enough timespan climate models were calibrated on ice core records and orbital positions. In deeper time, one must take care when calculating rates of climatic niche change because climate estimates may be built on an original assumption of thermal niche stasis in foraminifera. A further potential tautology is a dependence of time itself on foraminifera occurrences: age estimates in the early Cenozoic record may be calculated from microfossil biozones rather than geochemical dating, and original biozone delimitation may assume constant rates of evolution in the taxa defining zonation.

The non-independence of foraminifera occurrences and Earth system reconstructions are pernicious but surmountable. One could omit from analysis any species that serve as proxies, or separate out results by proxy status. Careful simulation of null expectations could help confirm the validity of results. Alternatively, one could turn to a different study clade altogether. The limiting data type is phylogenetic information; many taxa lack robust trees. The hyperdiverse gastropod group Conidae could be a candidate for comparative palaeobiology studies, as its Cenozoic

diversification is well studied (Puillandre, Bouchet, et al. 2014; Abalde, Tenorio, et al. 2017; Phuong, Alfaro, et al. 2019). This line of research could provide fundamental insight into processes of speciation and adaptation to Earth system change.

5.2.2 Multiple sources of autocorrelation

A recurrent complication in statistical ecology is the non-independence of observations. Species–environment data can contain spatial, temporal, and phylogenetic autocorrelation: sites closer together tend to be more similar in environmental conditions than sites that are far apart, events in quick succession tend to be more similar than events that occur after a long separation, and species or other evolvable entities that share a more recent common ancestor tend to be more similar in trait values than less-related taxa. The number of independent observations is the basis for calculating statistical degrees of freedom; if an observation is counted towards this value when it pseudo-replicates other observations, then one will overestimate degrees. This inaccuracy will perpetuate into every statistical metric that evaluates degrees of freedom. Autocorrelation can lead to overconfidence in a result, or worse, confidence in a wrong result. On the other hand, by accounting for autocorrelation a true signal could emerge that otherwise might remain hidden.

I addressed each type of autocorrelation separately in analyses throughout the data chapters here. When extracting environmental conditions associated with species occurrences (Chapter 3), I corrected for over- and under-sampling of different regions and analysed environmental occupancy in the form of density distributions instead of raw values. In linear mixed-effects models (Chapter 2), I added time bin identity as a random effect to control for coeval species experiencing similar conditions. When analysing time series (Chapters 2 and 3), I estimated a temporal autocorrelation coefficient and filtered out this dependency before correlating pairs of variables. When exploring niche trait values across an evolutionary tree (Chapter 4), I used comparative analyses to mitigate any phylogenetic signal. The reason I used different methods to account for autocorrelation in each context

is a lack of tools to control for multiple correlation sources simultaneously and in a symmetrical and rigorous way.

Of the possible methods to account for multiple autocorrelation structures, the simplest in theory is Generalised Least Squares (Dormann, McPherson, et al. 2007, for review). However, the optimisation of error structure matrices can be intractable in current implementations, which also have scant documentation. Generalised Estimating Equations also format correlation structures as a matrix but improve optimisation by subsetting the data as a preliminary step (Paradis and Claude 2002). Nevertheless, there are many more examples in the ecological literature of Generalised Least Squares than Estimating Equations. Another alternative is Eigenvector Maps and the many variations thereupon, e.g. Principal Coordinate Analysis of Neighbour Matrices, Asymmetric Eigenvector Maps, and Phylogenetic Eigenvector Regression/Mapping (Dray, Legendre, et al. 2006). All eigenvector approaches suffer from complexity in interpretation. Each response variance is partitioned into four components, representing different combinations of causal factors; moreover, although it may seem intuitive to match single eigenvectors to response data patterns, the correct approach is to identify sets of eigenvectors that are relevant for a given scale. With so many possible model implementations, macroecology faces difficulty training students and ensuring peer reviewers have sufficient knowledge to critique empirical studies.

Macroecology would benefit from development of a new, versatile method to handle the complex dependencies present in many biological and environmental datasets. Regression model innovation is a low priority for academic statistics at present, but a demonstrated need from the ecology community could initiate partnerships to tackle the topic in either a Bayesian framework (Paul Bürkner, pers. comm.) or a frequentist framework such as maximum-likelihood mixed-effects models (Matteo Tanadini, pers. comm.).

5.3 Closing remarks on ecological and evolutionary scaling

The transition from population ecology to macroecology is often discussed in terms of geographic scale (Brown 1995); analogously, the jump from microevolution to macroevolution is often discussed in terms of temporal scale (Serrelli and Gontier 2015). Nonetheless, units of space, time, and biological organisation all map onto each other to some degree. At small scale, populations within subsets of a geographic range respond to adaptive pressures and undergo genetic drift. At large scale, the union of populations across a range distribution constitutes a species, which is the replicating unit of macroevolution. The degree of continuity *vs.* step-change between these end-points is a longstanding issue of contention in ecology and evolutionary biology.

Some evolutionary biologists have argued for a gradation from microevolution to macroevolution (Simons 2002), while others theorise a disjunction, with separate selection acting on traits expressed in individuals and emergent at the species level (Vrba 1984; Futuyma 2015). Even these alternatives oversimplify evolutionary processes, which also act on lower levels of biological hierarchy, such as genes and chromosomes. Gene Regulatory Networks compartmentalise genetic expression and constrain evolvability in a way beyond either micro- or macro-evolutionary theory (Erwin and Davidson 2009). To a gross approximation, however, the crux of debate is the extent to which clade histories can be reduced to the mechanics of adaptation observable in populations in the lab or field.

I conducted all analyses at the level of species, as defined by a morphological species concept, and therefore results cannot directly address questions involving population dynamics. Instead, evidence at the species level can be interpreted in conjunction with knowledge from studies at finer spatiotemporal and biological scales. In Chapter 2, I searched for a signal of ecological release on large scales, a pattern expected from extrapolating the strong effect of competition observed in population studies. I found species' geographic ranges invariant to inferred competitive pressure, which suggests a disconnect between biotic influences at the levels of populations and

species. In interpreting this geographic stasis with respect to competitor diversity, I noted the possibility that conspecific populations experienced different competitive and environmental pressures and therefore were asynchronous in growth. This same idea is a keystone in the hypothesis for phenotypic stasis in punctuated equilibrium (Eldredge, Thompson, et al. 2005) and also could explain the abiotic niche stasis I documented in plankton species in Chapter 4. Another potential reason for niche stasis, reviewed in the discussion to Chapter 3, is asymmetric dispersal between central and distal populations in a geographic range. Either high or low gene flow to the periphery could limit adaptive potential across populations (Bridle, Polechová, et al. 2010; Polechová and Barton 2015). Population demographics studies across species distributions will be necessary to distinguish these possibilities.

Allopatric or peripatric speciation offers a rare moment of alignment between micro and macro scales. For the fleeting time after a group of individuals has become reproductively isolated but not yet widespread, this single population could be considered a species. Biotic or abiotic pressures that cause the population to adapt, including competition or climatic stress, could leave a lasting imprint as the lineage later spreads in numbers and distribution. Unfortunately, such ephemeral 'nascent species' are unlikely to leave a trace in the fossil record (Rabosky 2013). Neither fossil descriptions nor laboratory experiments will be able to test these scenarios alone. To unravel the full complexity of interactions between Earth and life systems across scales will require the united efforts of ecology, palaeobiology, environmental science, and statistics.

Appendices

A

Supporting charts for Chapter 2

A.1 Figures

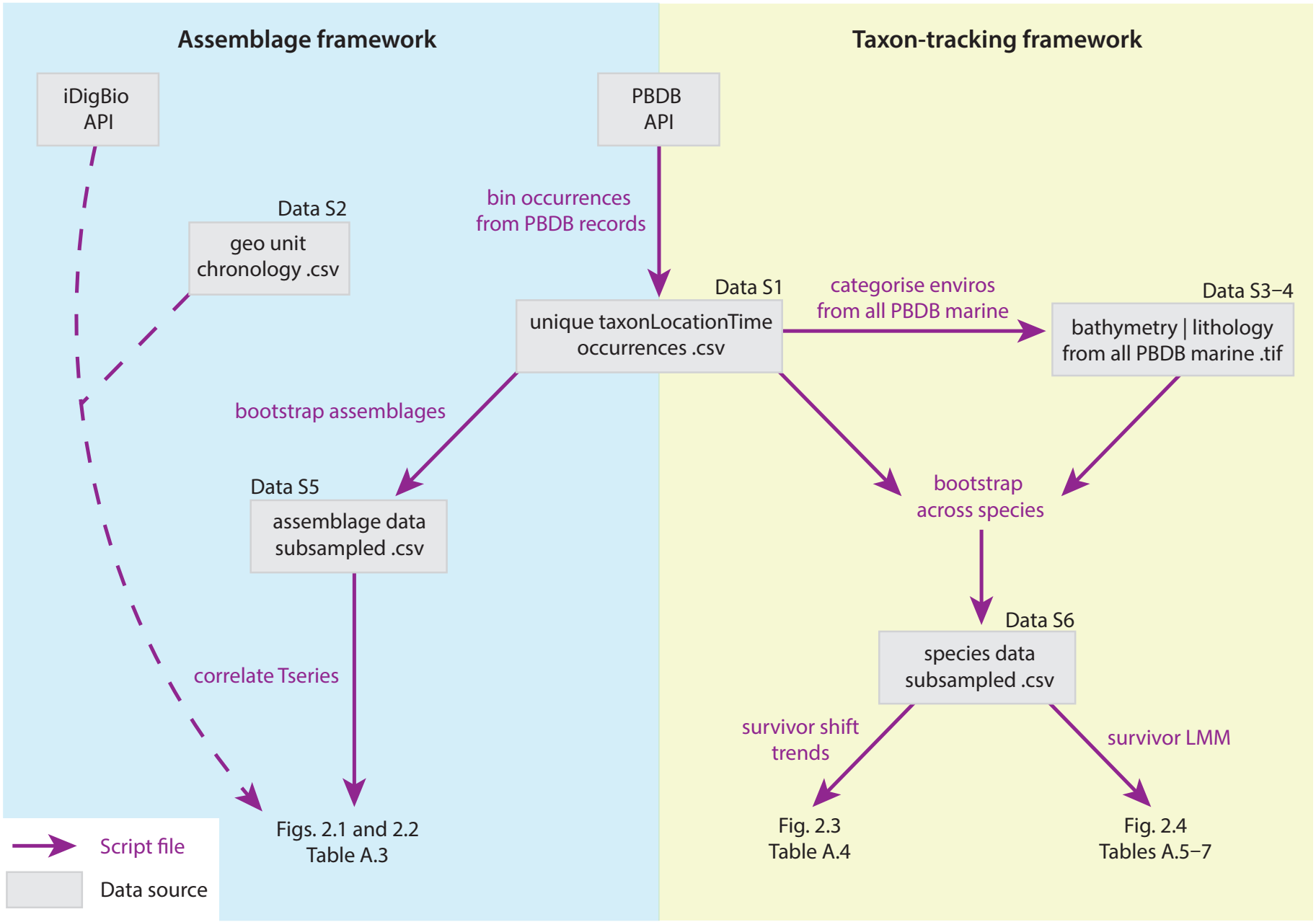

Assemblage framework
Taxon-tracking framework
iDigBio API
PBDB API
Data S2
geo unit chronology .csv
bin occurrences from PBDB records
Data S1
unique taxonLocationTime occurrences .csv
categorise enviros from all PBDB marine
Data S3–4
bathymetry | lithology from all PBDB marine .tif
bootstrap assemblages
Data S5
assemblage data subsampled .csv
bootstrap across species
Data S6
species data subsampled .csv
correlate Tseries
survivor shift trends
survivor LMM
Figs. 2.1 and 2.2 Table A.3
Fig. 2.3 Table A.4
Fig. 2.4 Tables A.5–7
Script file
Data source

Figure A.1: Workflow of analysis. Data sources are boxed in grey and are supplied as .csv and .tif files. Computational steps are indicated with purple arrows and are supplied as .R files. Dashed lines indicate the workflow for the northwest Atlantic dataset. (Note that spaces are substituted with underscores in file names.)

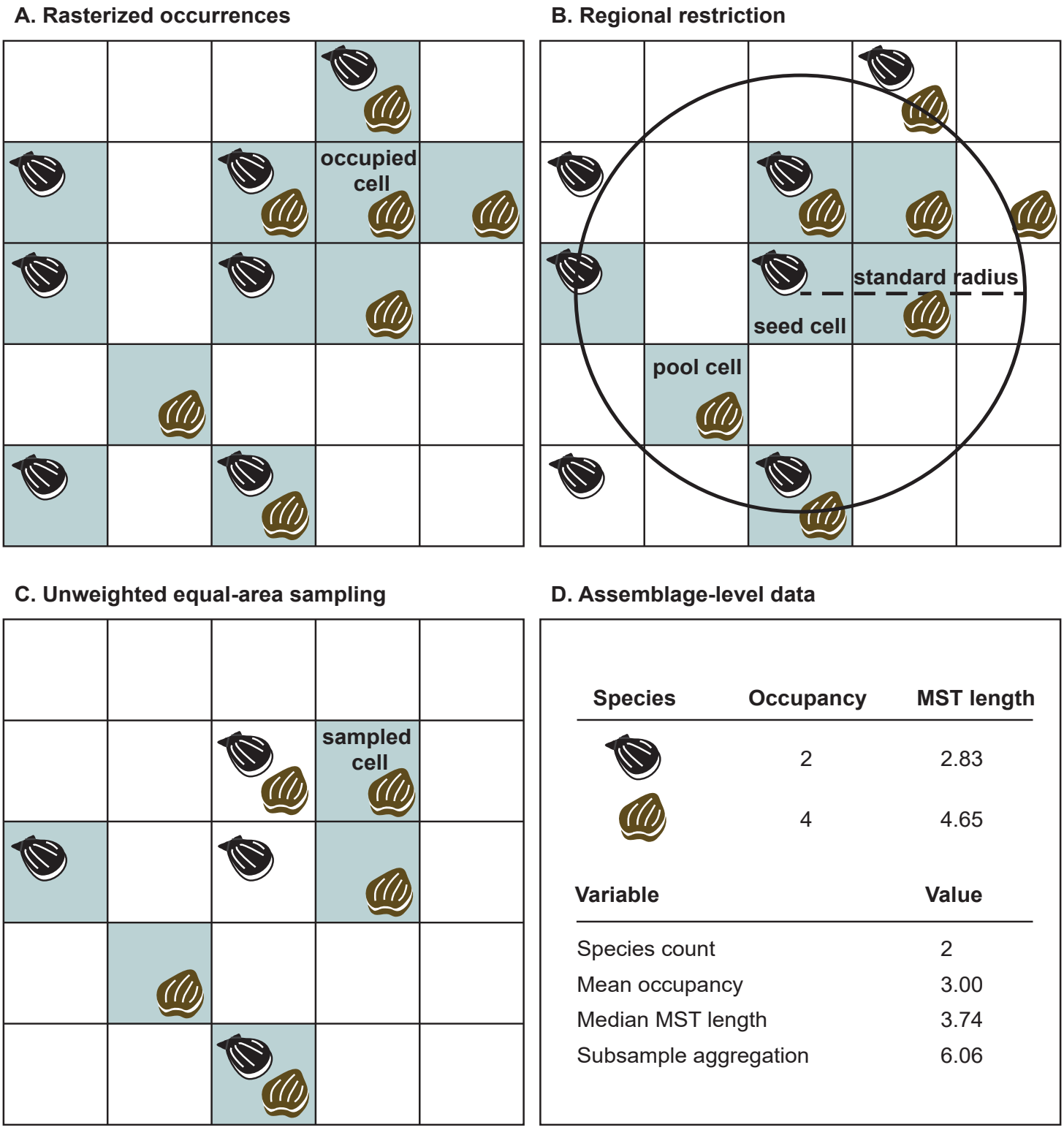
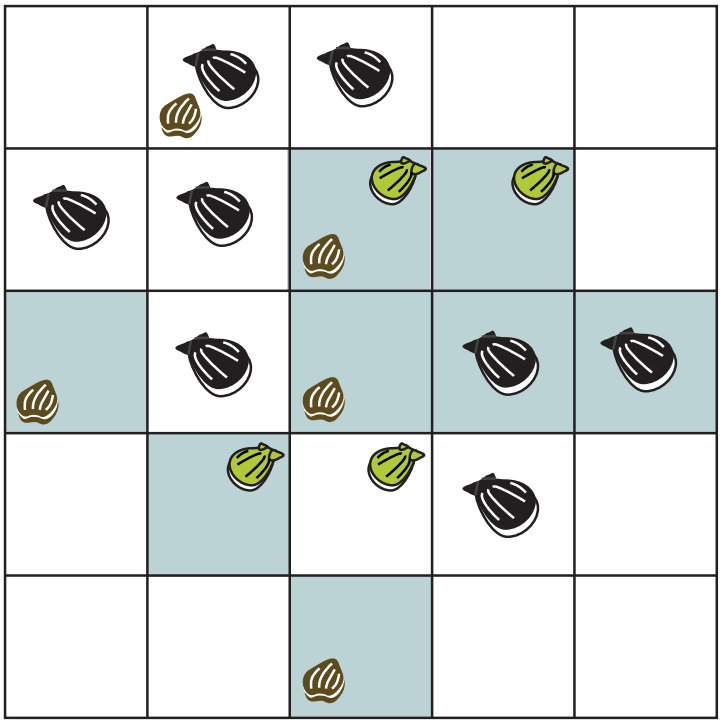
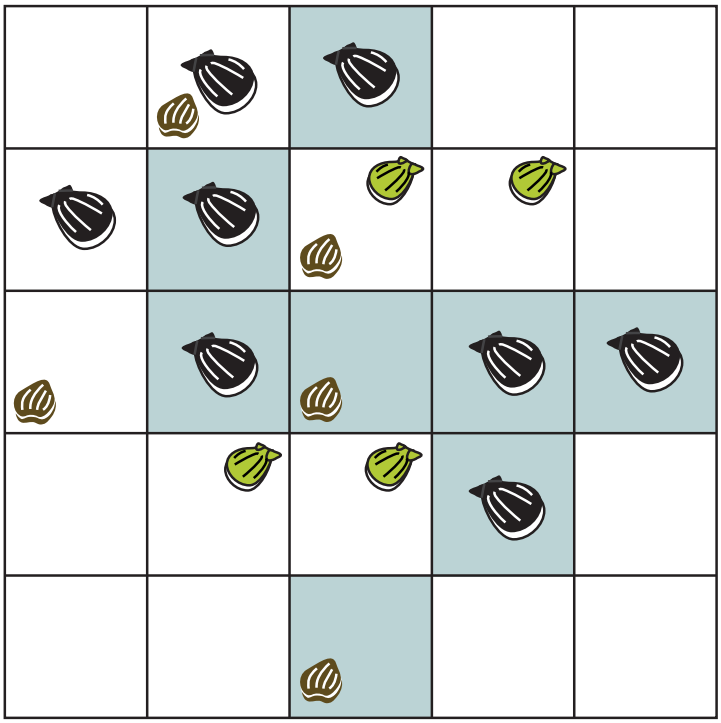


Figure A.2: Simplified diagram of assemblage subsampling. **(A)** Occurrence coordinates were rasterised. Each silhouette represents a species' presence in a grid cell. Every occupied cell (shaded) is a potential 'seed cell'. **(B)** A circular buffer of given radius is circumscribed around the centroid of a seed cell. Occupied cells with centroids lying inside the buffer (shaded) comprise the 'subsampling pool'. **(C)** A given number of cells (here, 5) are drawn from the pool to form a subsample (shaded). **(D)** For each species, range size is measured as occupancy and summed length of the minimum spanning tree that connects the centroids of occupied cells (MST). The lower table reports all assemblage-based variables used in subsequent analysis.

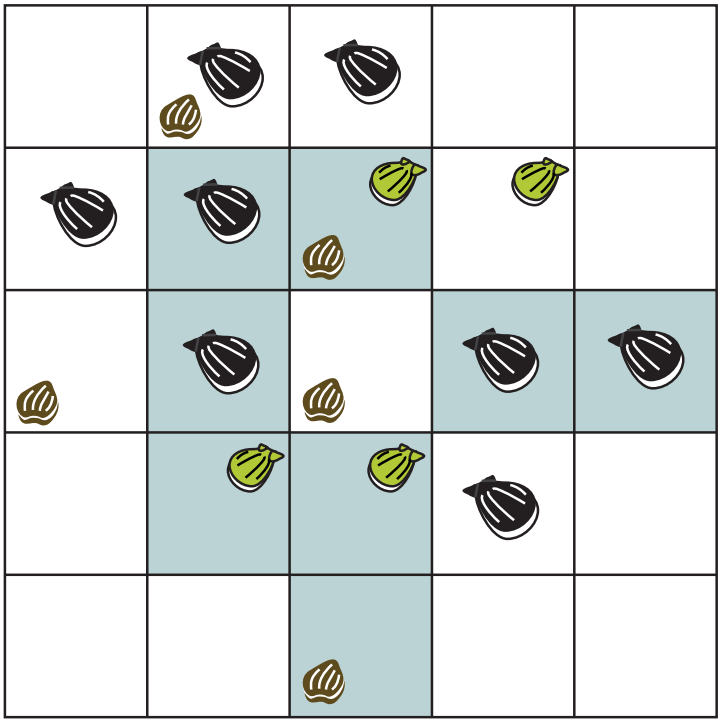
A. First iteration



B. Second iteration



C. Third iteration



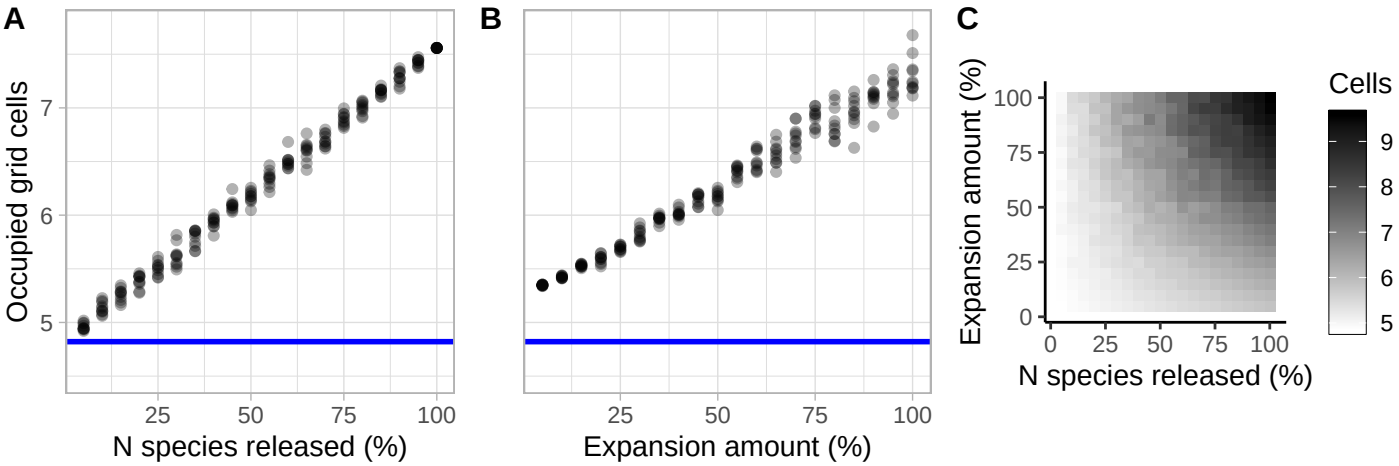
D. Focal species data

Iteration	Occupancy	MST length	Species count
1	2	1.00	3
2	6	6.41	2
3	4	4.00	3

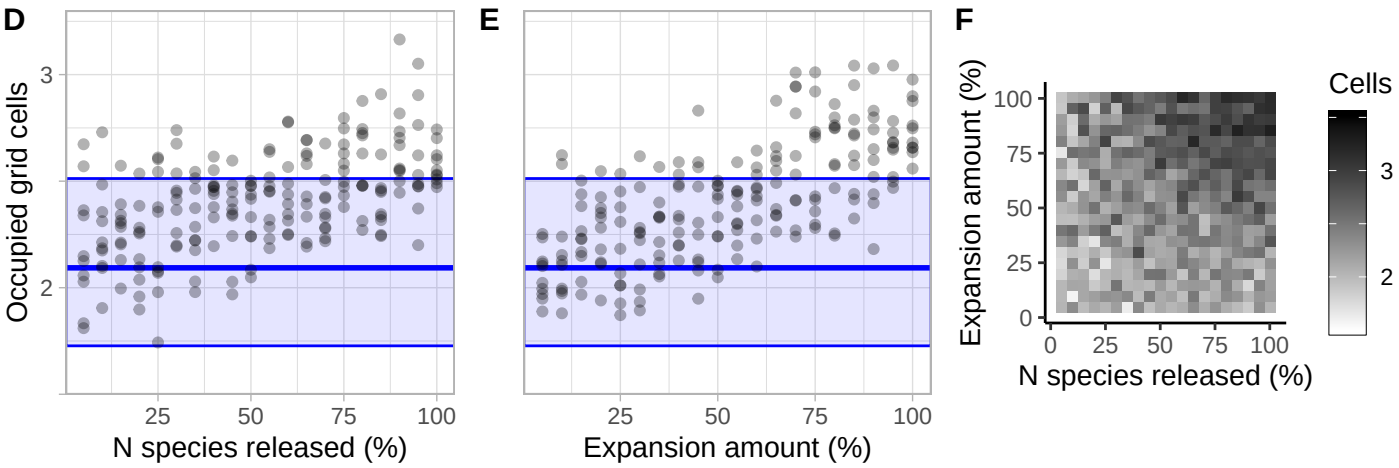
Variable	Median
Occupancy	4
MST length	4.00
Species count	3

Figure A.3: Simplified diagram of taxon-tracking subsampling. **(A–C)** Three different realisations of the subsampling procedure illustrated for identical occurrence grids of a focal species (black pectinid). Shading indicates cells selected in a given a subsample (for simplicity, each subsample in this diagram contained 8 cells). **(D)** Two metrics of range area were calculated for each iteration: occupancy and summed length of the minimum spanning tree that connects the centroids of occupied cells (MST). The range size of the species, as used in subsequent analysis, was defined as the median metric value among iterations.

Simulated geographic expansion



Observed geographic expansion after subsampling



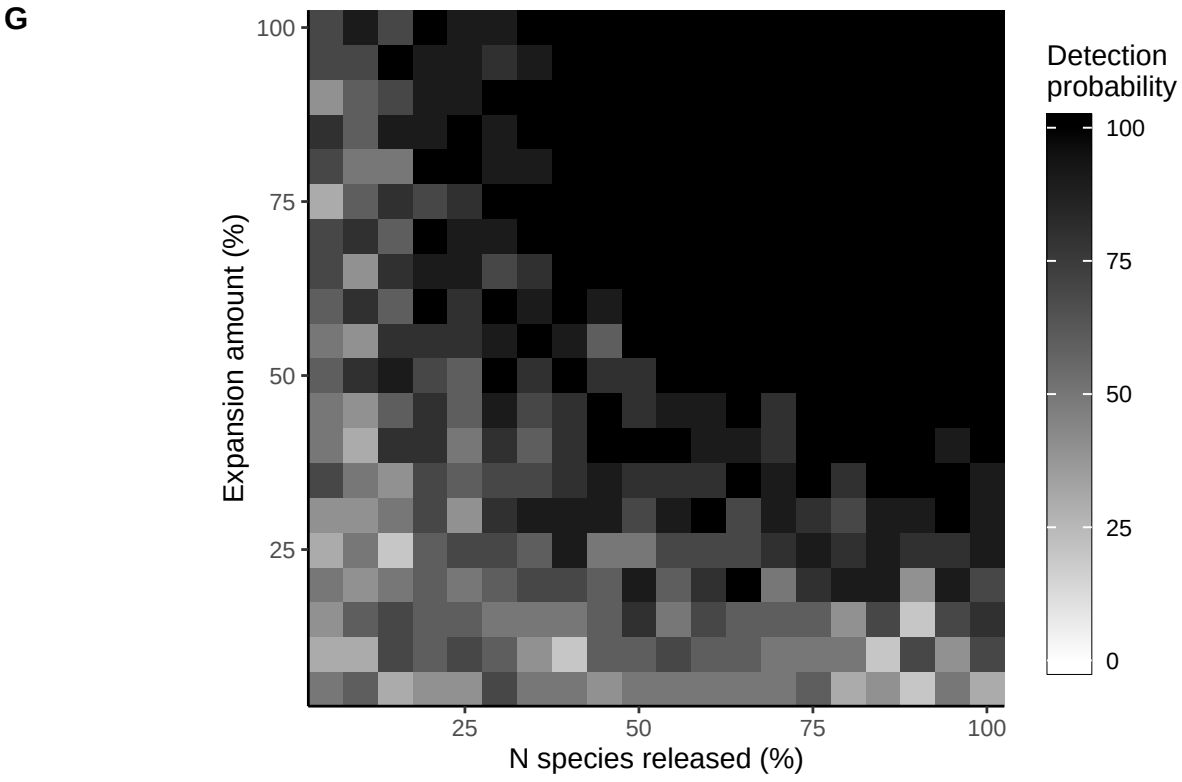


Figure A.4: (Overleaf) Power analysis from simulated geographic expansion. Occupancy increases were simulated without subsampling **(A–C)** and with subsmapling **(D–F)**. Simulated occupancy increased independently as a function of **(A, D)** the percentage of species that experienced release or **(B, E)** the magnitude of release as a percentage of initial range size, from 0–100% expansion. **(C, F)** Occupancy increased jointly with both parameters. For **A–F**, simulation was iterated 10 times for each set of parameters. In **A** and **D**, the amount of release was constant at 50% expansion. In **B** and **E**, the frequency of release was constant at 50% of species. The blue lines mark mean occupancy prior to range expansion: 4.8 grid cells without subsampling **(A, B)**, and 2.1 grid cells with subsampling **(D, E;** 95% interval among subsamples in light blue shading, 1.7–2.5 grid cells). **(G)** The probability of observing release in a subsample is plotted jointly as a function of the frequency and magnitude of range expansion. Detection is calculated as the percentage of iterations ($n = 10$) in which mean occupancy after simulated release exceeded the initial value (2.1 grid cells) for a given frequency and magnitude of release.

A.2 Tables

Table A.1: (A) The global Phanerozoic dataset was divided into 63 time bins. A species' occurrence record was included in a time bin if the name or age of both its maximum and minimum occurrence estimates fell within the onset and terminus ages (in Ma) for the bin. During binning, ages were rounded to the nearest 0.01 Ma for boundaries younger than 10 Ma, 0.1 Ma for boundaries 50–150 Ma, or 1 Ma for boundaries older than 150 Ma. The number of unique occurrences for each species is tallied for each time bin. **(B)** The northwest Atlantic dataset was divided into 9 time bins. Consult Appendix B.1.2 for explication of how species occurrences were assigned to time bins.

Name	Onset	Terminus	Duration	Occurrences
A. Global Phanerozoic				
Tremadocian, Floian	485.4	470.0	15.4	196
Dapingian, Darriwilian	470.0	458.4	11.6	347
Sandbian, Katian	458.4	445.2	13.2	1226
Hirnantian	445.2	443.8	1.4	394
Rhuddanian, Aeronian	443.8	438.5	5.3	411
Telychian	438.5	433.4	5.1	337
Sheinwoodian, Homerian	433.4	427.4	6.0	1072
Gorstian, Ludfordian	427.4	423.0	4.4	924
Pridoli	423.0	419.2	3.8	313
Lochkovian, Pragian	419.2	407.6	11.6	1141
Emsian, Eifelian	407.6	387.7	19.9	914
Givetian	387.7	382.7	5.0	1293
Frasnian	382.7	372.2	10.5	937
Famennian	372.2	358.9	13.3	371
Tournaisian	358.9	346.7	12.2	641
Visean	346.7	330.9	15.8	607
Serpukhovian, Bashkirian	330.9	315.2	15.7	344
Moscovian, Kasimovian	315.2	303.7	11.5	561
Gzhelian	303.7	298.9	4.8	613
Asselian, Sakmarian	298.9	290.1	8.8	1716
Artinskian, Kungurian	290.1	272.9	17.2	2937
Roadian	272.9	268.8	4.1	1596
Wordian	268.8	265.1	3.7	1621
Capitanian	265.1	259.1	6.0	1342
Wuchiapingian	259.1	254.1	5.0	2451
Changhsingian	254.1	251.9	2.2	2411
Induan, Olenekian	251.9	247.2	4.7	430
Anisian	247.2	242.0	5.2	782
Ladinian	242.0	237.0	5.0	284
Carnian	237.0	227.0	10.0	503
Norian	227.0	208.5	18.5	689
Rhaetian	208.5	201.3	7.2	481

... Table A.1 continued

Name	Onset	Terminus	Duration	Occurrences
Hettangian, Sinemurian	201.3	190.8	10.5	581
Pliensbachian	190.8	182.7	8.1	1296
Toarcian	182.7	174.1	8.6	866
Aalenian, Bajocian	174.1	168.3	5.8	588
Bathonian	168.3	166.1	2.2	659
Callovian	166.1	163.5	2.6	719
Oxfordian	163.5	157.3	6.2	437
Kimmeridgian	157.3	152.1	5.2	304
Tithonian	152.1	145.0	7.1	486
Berriasian	145.0	139.8	5.2	317
Valanginian, Hauterivian, Barremian	139.8	125.0	14.8	656
Aptian	125.0	113.0	12.0	509
Albian	113.0	100.5	12.5	522
Cenomanian	100.5	93.9	6.6	590
Turonian	93.9	89.8	4.1	187
Coniacian, Santonian	89.8	83.6	6.2	709
Campanian	83.6	72.1	11.5	422
Maastrichtian	72.1	66.0	6.1	3197
Danian	66.0	61.6	4.4	363
Selandian, Thanetian	61.6	56.0	5.6	272
Ypresian	56.0	47.8	8.2	661
Lutetian, Bartonian	47.8	37.8	10.0	1893
Priabonian	37.8	33.9	3.9	993
Rupelian	33.9	27.8	6.1	773
Chattian, Aquitanian	27.8	20.4	7.4	632
Burdigalian	20.4	16.0	4.5	817
Langhian, Serravallian	16.0	11.6	4.3	2056
Tortonian, Messinian	11.6	5.3	6.3	1760
Zanclean	5.3	3.6	1.7	934
Piacenzian	3.6	2.6	1.0	1333
Pleistocene	2.6	0.0	2.6	2204
B. Northwest Atlantic				
Burdigalian	20.44	15.97	4.47	1019
Middle Miocene	15.97	11.63	4.34	639
Tortonian	11.63	7.25	4.38	968
Messinian	7.25	5.33	1.91	923
Zanclean	5.33	3.60	1.73	1447
Piacenzian	3.60	2.58	1.02	2557
Gelasian	2.58	1.80	0.78	1423
Calabrian	1.80	0.78	1.02	1039
Middle–Late Pleistocene	0.78	0.01	0.77	2020

Table A.2: Values of parameters used in each sensitivity analysis of the assemblage subsampling method: number of equal-area cells in a subsample, radius of regional buffer, n iterations of subsampling in each time bin, probability of selecting a cell within the subsample region, type of rarefaction applied to species count in the subsample, whether species with a single spatiotemporal occurrence were included, and taxonomic rank of analysis. Boldface indicates values that differ from those of the baseline method (third row). The left-most column lists the table location of results for the given sensitivity tests, which are discussed in subsections of B.3.3.

Table	Cells	Radius (km)	Iterations	Cell selection	Rarefaction	Assemblage singletons	Taxonomic rank	Other notes
None	**10**	1500	500	Even	None	Yes	Species	**'Accepted names' only**
None	**10**	1500	500	Even	None	Yes	Species	**No space-time singletons**
A.3A	12	1500	500	Even	None	Yes	Species	Baseline method
None	**6**	1500	500	Even	None	Yes	Species	
A.3B	**6**	**750**	500	Even	None	Yes	Species	
None	12	1500	**1000**	Even	None	Yes	Species	
A.3C	12	1500	500	**Weighted**	None	Yes	Species	Main text results
A.3D	12	1500	500	Even	**Coverage = 0.4**	Yes	Species	
A.3E	12	1500	500	Even	**Coverage = 0.7**	Yes	Species	
A.3F	12	1500	500	Even	**Coverage = 0.7**	**No**	Species	
None	12	1500	500	Even	**Count = 50**	Yes	Species	
A.3G	12	1500	500	Even	None	Yes	**Genus**	
A.3H	**15**	**Atlantic**	500	Even	None	Yes	Species	Main text results

Table A.3: Kendall's *tau* correlation for every pair of predictor and response time series, for parameter combinations listed in Table A2. The metrics of range size were grid cell occupancy (occ) and summed length of minimum spanning tree (MST).

	Correlation with predictor time series			
	Species count		Sampling aggregation	
Response series	Mean	95%	Mean	95%
A. Baseline				
occ with singletons, min	-0.15	(-0.18, -0.10)	-0.15	(-0.18, -0.06)
occ with singletons, med	-0.05	(-0.23, 0.14)	-0.18	(-0.36, 0.02)
occ with singletons, max	0.06	(-0.07, 0.19)	-0.19	(-0.32, -0.02)
occ with singletons, mean	-0.09	(-0.23, 0.06)	-0.23	(-0.35, -0.08)
occ no singletons, min	-0.15	(-0.21, -0.04)	0.01	(-0.16, 0.17)
occ no singletons, med	-0.15	(-0.33, 0.03)	-0.16	(-0.32, 0.03)
occ no singletons, max	0.06	(-0.07, 0.19)	-0.19	(-0.32, -0.02)
occ no singletons, mean	-0.12	(-0.25, 0.02)	-0.18	(-0.33, -0.03)
MST, min	-0.21	(-0.39, -0.03)	0.13	(-0.05, 0.30)
MST, med	-0.05	(-0.18, 0.11)	0.06	(-0.08, 0.21)
MST, max	0.15	(0.01, 0.29)	0.23	(0.11, 0.37)
MST, mean	-0.04	(-0.17, 0.11)	0.13	(-0.01, 0.26)
species count	1.00	(1.00, 1.00)	0.04	(-0.09, 0.18)
B. 6 cells, 750 km radius				
occ no singletons, mean	-0.08	(-0.23, 0.10)	-0.18	(-0.33, -0.03)
MST, median	-0.05	(-0.20, 0.11)	0.17	(0.01, 0.33)
species count	1.00	(1.00, 1.00)	0.04	(-0.12, 0.19)
C. Weighted cell selection				
occ no singletons, mean	-0.11	(-0.25, 0.01)	-0.17	(-0.32, -0.03)
MST, median	-0.06	(-0.19, 0.08)	0.09	(-0.07, 0.24)
species count	1.00	(1.00, 1.00)	-0.01	(-0.14, 0.12)

... Table A.3 continued

Response series	Mean	95%	Mean	95%
D. SQS, coverage of 0.4				
occ no singletons, mean	-0.12	(-0.26, 0.02)	-0.19	(-0.34, -0.04)
MST, median	-0.06	(-0.20, 0.08)	0.07	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	0.04	(-0.09, 0.18)
E. SQS, coverage of 0.7				
occ no singletons, mean	-0.11	(-0.25, 0.01)	-0.19	(-0.34, -0.04)
MST, median	-0.05	(-0.20, 0.09)	0.06	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	0.02	(-0.12, 0.15)
F. SQS, coverage of 0.7, no assemblage singletons				
occ no singletons, mean	0.11	(-0.03, 0.25)	-0.18	(-0.33, -0.05)
MST, median	-0.08	(-0.23, 0.05)	0.07	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	-0.09	(-0.23, 0.05)
G. Genus-level				
occ no singletons, mean	-0.10	(-0.25, 0.03)	-0.18	(-0.31, -0.04)
MST, median	-0.05	(-0.19, 0.10)	0.04	(-0.11, 0.19)
species count	1.00	(1.00, 1.00)	0.03	(-0.12, 0.17)
H. Northwest Atlantic				
occ no singletons, mean	0.18	(-0.17, 0.53)	-0.51	(-0.67, -0.33)
MST, median	-0.13	(-0.59, 0.31)	0.06	(-0.33, 0.50)
species count	1.00	(1.00, 1.00)	-0.18	(-0.56, 0.21)

Table A.4: Species count dynamic and geographic range-size shift as measured in **(A)** occupancy or **(B)** minimum spanning tree length (square-root km), after excluding newly-originated and nearly-extinct species. To calculate geographic range size trends for an interval, multiple species had to occur in successive time bins that contained at least 10 cells of suitable habitat. For minimum spanning tree calculations, a species had to occur more than once in each bin. The shift is the average and standard deviation (SD) of change in number of grid cells among survivors. Each time interval was characterised by either a rise or fall in the number of species, depending on the median change in species count associated with survivors. Each interval is referenced by the oldest geologic stage included.

Early bin	Species count	Range shift	Shift SD	Survivor n
A. Occupied grid cells				
Rhuddanian	Rise	0.50	0.58	4
Telychian	Rise	1.00	1.46	18
Emsian	Rise	1.36	1.57	11
Moscovian	Rise	0.89	2.62	9
Gzhelian	Rise	-0.92	2.06	13
Asselian	Rise	-0.60	1.34	45
Artinskian	Fall	1.33	2.31	3
Wordian	Fall	0.31	0.95	13
Wuchiapingian	Rise	0.21	0.73	56
Changhsingian	Fall	-0.57	1.81	7
Anisian	Fall	-1.43	1.79	14
Hettangian	Fall	-0.40	2.12	10
Pliensbachian	Fall	-0.15	0.84	39
Toarcian	Fall	0.25	1.50	4
Aalenian	Rise	2.00	2.83	2
Bathonian	Fall	0.75	0.96	4
Callovian	Rise	-0.50	1.87	6
Coniacian	Rise	-1.00	0.89	11
Ypresian	Rise	-0.11	1.88	18
Langhian	Fall	-0.08	1.08	12
Piacenzian	Fall	0.70	1.85	43

. . . Table A.4 continued

Early bin	Species count	Range shift	Shift SD	Survivor n
	B. Minimum spanning tree length			
Rhuddanian	Rise	0.00	0.00	2
Telychian	Rise	7.19	8.18	8
Emsian	Rise	8.06	3.89	4
Moscovian	Rise	3.98	11.43	5
Gzhelian	Rise	-9.32	15.26	8
Asselian	Fall	-8.98	6.24	3
Wordian	Fall	1.64	4.03	4
Wuchiapingian	Rise	3.85	8.98	11
Changhsingian	Fall	-4.13	3.24	3
Anisian	Fall	-7.22	6.33	6
Pliensbachian	Rise	-6.81	7.64	5
Bathonian	Fall	3.95	7.85	3
Callovian	Rise	-3.90	13.39	4
Coniacian	Fall	-10.12	14.29	7
Ypresian	Rise	6.19	11.54	11
Langhian	Fall	9.13	20.72	2
Piacenzian	Fall	5.36	17.18	10

Table A.5: Relative performance of non-trivial linear mixed-effects models. The response variable was **(A)** the change in number of occupied grid cells, including singletons, or **(B)** minimum spanning tree length (square-root km). For **A**, models were built on 708 changes in range size recorded for survivors from 26 Phanerozoic boundaries. For **B**, models were built on 239 changes in range size recorded for survivors from the 23 intervals for which tree length could be calculated. Please read Appendix B.4.4 for model comparison details. Abbreviations: delta, change in species count; time, age of boundary (Ma); point, categorical variable with the values first, middle, and last to indicate newly originated, mid duration, or nearly extinct; clade, membership in Brachiopoda or Mollusca; bath, bathymetric preference for deep, shallow, or both water depths; lith, lithological preference for carbonate, siliciclastic, or both lithologies.

Rank	Predictors	n predictors	ΔAIC	Weight
	A. Occupied grid cells			
1	delta + point + time	7	0.00	23.64
2	point + delta	6	0.96	14.63
3	delta + point + clade + time	8	2.00	8.70
4	delta + point + lith + time	9	2.15	8.07
5	delta + point + lith	8	2.20	7.87
6	delta + point + clade	7	2.74	6.01
7	delta + point + bath + time	9	3.43	4.25
8	delta + point + clade + lith + time	10	4.08	3.07
9	delta + point + clade + lith	9	4.18	2.92
10	delta + point + bath	8	4.33	2.71
11	time + point	6	4.49	2.50
12	point + lith + time	8	5.13	1.82
13	delta + point + clade + bath + time	10	5.42	1.57
14	lith + point	7	5.50	1.51
15	delta + point + lith + bath + time	11	5.54	1.48
16	delta + point + lith + bath	10	5.56	1.47
17	delta + point + clade + bath	9	6.17	1.08
18	point	5	6.18	1.08
19	point + clade + time	7	6.47	0.93
20	point + clade + lith + time	9	7.04	0.70
21	delta + point + clade + lith + bath + time	12	7.44	0.57
22	point + clade + lith	8	7.49	0.56

... Table A.5 continued

Rank	Predictors	*n* predictors	ΔAIC	Weight
23	delta + point + clade + lith + bath	11	7.55	0.54
24	clade + point	6	7.73	0.50
25	point + bath + time	8	7.83	0.47
26	point + lith + bath + time	10	8.47	0.34
27	point + lith + bath	9	8.85	0.28
28	bath + point	7	9.52	0.20
29	point + clade + bath + time	9	9.83	0.17
	B. Minimum spanning tree length			
1	point + delta	6	0.00	13.03
2	delta + point + bath	8	0.50	10.15
3	delta + point + time	7	1.10	7.52
4	delta + point + clade + time	8	1.21	7.12
5	delta + point + clade	7	1.37	6.57
6	delta + point + lith + bath	10	1.62	5.80
7	delta + point + clade + bath	9	1.75	5.43
8	delta + point + clade + lith + bath + time	12	1.78	5.35
9	delta + point + lith + bath + time	11	1.86	5.14
10	delta + point + bath + time	9	1.94	4.94
11	delta + point + clade + bath + time	10	2.12	4.52
12	delta + point + clade + lith + time	10	2.16	4.43
13	delta + point + lith + time	9	2.27	4.19
14	delta + point + lith	8	2.38	3.97
15	delta + point + clade + lith + bath	11	3.12	2.74
16	delta + point + clade + lith	9	3.96	1.80
17	point + lith + bath	9	5.61	0.79
18	bath + point	7	5.82	0.71
19	point + clade + bath	8	6.05	0.63
20	point + clade + lith + bath + time	11	6.14	0.61
21	point + clade + lith + bath	10	6.60	0.48
22	point + lith + bath + time	10	6.66	0.47

... Table A.5 continued

Rank	Predictors	n predictors	ΔAIC	Weight
23	point	5	6.92	0.41
24	point + clade + bath + time	9	7.08	0.38
25	clade + point	6	7.27	0.34
26	point + bath + time	8	7.75	0.27
27	point + clade + time	7	7.88	0.25
28	lith + point	7	8.37	0.20
29	point + clade + lith + time	9	8.60	0.18
30	time + point	6	8.71	0.17
31	point + lith + time	8	9.22	0.13
32	point + clade + lith	8	9.41	0.12
33	delta	4	9.90	0.09
34	delta + lith + bath	8	9.95	0.09

Table A.6: The log-likelihood estimates and 95% confidence intervals were extracted for each predictor in the 5 models with largest weights (Table A.5), with change in geographic range size as response, measured either as **(A)** number of occupied grid cells (including singletons) or **(B)** minimum spanning tree length (square-root km). Abbreviations and sample sizes as in Table A.5.

Rank	Predictor	Estimate	95%
	A. Occupied grid cells		
1	(Intercept)	-0.489	(-1.278, 0.276)
	delta	0.002	(0.001, 0.004)
	point First	0.732	(0.362, 1.110)
	point Last	-0.355	(-0.624, -0.082)
	time	0.002	(-0.000, 0.005)
2	(Intercept)	0.099	(-0.275, 0.492)
	point First	0.757	(0.385, 1.135)
	point Last	-0.331	(-0.600, -0.061)
	delta	0.002	(0.001, 0.004)
3	(Intercept)	-0.498	(-1.363, 0.351)
	delta	0.002	(0.001, 0.004)
	point First	0.731	(0.359, 1.111)
	point Last	-0.355	(-0.624, -0.082)
	clade Mollusca	0.007	(-0.332, 0.343)
	time	0.003	(-0.000, 0.006)
4	(Intercept)	-0.361	(-1.208, 0.470)
	delta	0.002	(0.000, 0.004)
	point First	0.742	(0.370, 1.121)
	point Last	-0.349	(-0.618, -0.077)
	lith Clastic	-0.017	(-0.399, 0.345)
	lith Silic.	-0.338	(-0.793, 0.142)
	time	0.002	(-0.001, 0.005)
5	(Intercept)	0.174	(-0.227, 0.588)
	delta	0.002	(0.000, 0.004)
	point First	0.765	(0.394, 1.144)
	point Last	-0.327	(-0.596, -0.058)
	lith Clastic	-0.046	(-0.424, 0.320)
	lith Silic.	-0.394	(-0.847, 0.069)

... Table A.6 continued

Rank	Predictor	Estimate	95%
	B. Minimum spanning tree length		
1	(Intercept)	0.156	(-3.034, 3.441)
	point First	3.904	(-0.546, 8.440)
	point Last	-5.336	(-8.808, -1.816)
	delta	0.038	(0.013, 0.062)
2	(Intercept)	-0.828	(-4.114, 2.522)
	delta	0.034	(0.009, 0.058)
	point First	3.259	(-1.188, 7.838)
	point Last	-5.690	(-9.166, -2.193)
	bath Deep	4.007	(-3.224, 11.229)
	bath Shallow	3.247	(-0.636, 7.205)
3	(Intercept)	2.725	(-3.535, 9.124)
	delta	0.039	(0.015, 0.064)
	point First	4.040	(-0.365, 8.660)
	point Last	-5.375	(-8.831, -1.859)
	time	-0.011	(-0.034, 0.012)
4	(Intercept)	6.079	(-1.760, 13.928)
	delta	0.037	(0.013, 0.062)
	point First	4.123	(-0.262, 8.721)
	point Last	-5.190	(-8.646, -1.689)
	clade Mollusca	-3.120	(-7.595, 1.341)
	time	-0.019	(-0.044, 0.007)
5	(Intercept)	0.979	(-2.784, 4.817)
	delta	0.036	(0.011, 0.061)
	point First	3.856	(-0.581, 8.387)
	point Last	-5.238	(-8.713, -1.715)
	clade Mollusca	-1.686	(-5.713, 2.391)

Table A.7: Log-likelihood estimates and confidence intervals for global linear mixed-effects model. Scaled estimates are plotted in Fig. 2.4. The 95% confidence intervals were derived from profiling. Estimates are relative to mid-existence (*vs.* early or late) brachiopod species (*vs.* bivalve) with no preference for habitat (*vs.* preference for substrate or water depth). Substrate preference (subst.) is either carbonate (CaCO3, e.g. reef) or siliciclastic (clastic, e.g. sand). Separate models were built for geographic distributions measured in occupancy (Occ, $n = 708$) or square-root minimum spanning tree length (MST, $n = 239$).

Response	Predictor	Estimate	2.5 %	97.5 %
Occ	Phase: Early	0.733	0.359	1.117
Occ	Phase: Late	-0.354	-0.623	-0.081
Occ	N spp. change	0.002	0.000	0.004
Occ	Time	0.002	-0.001	0.006
Occ	Clade: Bivalve	0.065	-0.298	0.413
Occ	Subst: CaCO3	-0.020	-0.420	0.353
Occ	Subst: Clastic	-0.358	-0.819	0.133
Occ	Depth: Deep	0.239	-0.368	0.852
Occ	Depth: Shallow	0.032	-0.273	0.344
Occ	(Intercept)	-0.448	-1.385	0.480
MST	Phase: Early	3.317	-1.065	8.121
MST	Phase: Late	-5.145	-8.621	-1.705
MST	N spp. change	0.032	0.007	0.057
MST	Time	-0.027	-0.056	0.002
MST	Clade: Bivalve	-3.297	-8.124	1.248
MST	Subst: CaCO3	-2.138	-7.729	2.871
MST	Subst: Clastic	-7.405	-13.653	0.147
MST	Depth: Deep	3.891	-3.446	11.223
MST	Depth: Shallow	3.992	-0.118	8.015
MST	(Intercept)	8.785	-0.633	18.047

A.3 Data availability

All data and code necessary to produce the final results are archived at `https://doi.org/10.5281/zenodo.3491853` and maintained at `https://github.com/GwenAntell/EcoRelease`. All computations were processed in an R statistical environment (R Core Team 2018).

Data S1. Spreadsheet of unique spatiotemporal occurrences of brachiopod and bivalve species across the Phanerozoic.

Data S2. Lithochronology chart for northwest Atlantic geologic units.

Data S3. Raster of water depth conditions for each Phanerozoic time bin.

Data S4. Raster of substrate conditions for each Phanerozoic time bin.

Data S5. Phanerozoic occurrence data subsampled in the assemblage framework.

Data S6. Phanerozoic occurrence data subsampled in the taxon-tracking framework.

B

Supporting methods for Chapter 2

Contents

B.1 Fossil occurrences compilation 141
B.1.1 Global Phanerozoic dataset 141
B.1.2 Northwest Atlantic Neogene–Quaternary dataset 144
B.2 Standardisation of spatial coverage 146
B.2.1 Further notes on proportional occupancy 150
B.3 Assemblage framework 153
B.3.1 Assemblage subsampling routine 154
B.3.2 Time series analysis 155
B.3.3 Sensitivity to subsampling parameters 156
B.4 Taxon-tracking framework 163
B.4.1 Environmental preferences 164
B.4.2 Taxon-tracking subsampling routine 166
B.4.3 Biotic crises *vs.* background intervals 167
B.4.4 Mixed-effects models 168
B.5 Power analysis 169
B.5.1 Justification 169
B.5.2 Simulation of ecological release 170
B.5.3 Power to detect release 171

There are three prongs to the study. The first approach, based on assemblages, tested the degree to which the number of species varies inversely with geographic distribution areas. The other two approaches were embedded in a taxon-tracking framework: first I tested whether species that survive intervals of net species loss

expanded in distribution within areas of suitable habitat, and then I modelled change in geographic distribution as a function of change in the number of species present in a region, environmental specificity, and clade membership.

B.1 Fossil occurrences compilation

Each unit of the biological hierarchy functions at a unique spatiotemporal resolution. For instance, individual study animals may live for years to decades, populations may persist for decades to centuries, and species may persist for hundreds of thousands to millions of years. Biotic competition causes change across all levels: direct competition for resources (e.g. larval settlement on an exclusive patch of seafloor) occurs between individual organisms, but consequential differentiation in survival and reproductive success compound at the population level, and ensuing changes in range size occur at the species level. Since these different processes operate at different spatiotemporal scales, I compiled fossil occurrences at two resolutions and extents: global occurrences of bivalves and brachiopods across the Phanerozoic (section B.1.1) and northwest Atlantic occurrences of bivalves from the Neogene through Quaternary (section B.1.2). With this combination of data, analyses were able to search for signals of ecological release and competitive over a spectrum of scales.

B.1.1 Global Phanerozoic dataset

Global bivalve and brachiopod occurrences across the Phanerozoic were downloaded from the Paleobiology Database (PBDB; `paleobiodb.org`) in July 2018, according to the following criteria: 'taxa to include' as 'Brachiopoda, Bivalvia'; 'taxonomic resolution' as 'genus'; 'identification' as 'latest'; 'modifiers' as excluding the specific modifiers 'aff.,' 'cf.,' 'sensu lato,' 'informal,' 'ex gr.,' '?' or empty from genus and species names; 'interval' as Cambrian–Holocene, with 'time rule' as 'overlap'; and 'environment' as excluding 'terrestrial.' I chose not to restrict records to 'accepted names only' in the download step, because this filter would have removed a large number of records that contained valid coordinates despite lacking minor fields

associated with taxonomy (e.g. the year of species description). Furthermore, when this filter was applied in a sensitivity test, results were equivalent (section B.3.3). I excluded records that (i) lacked palaeocoordinates, (ii) contained metadata that indicated a geographic resolution greater than local area, (iii) were identified as 'sp,' 'spp.,' or a single letter, (iv) were flagged as 'nomen dubium,' 'nomen vanum,' 'nomen nudum,' or 'indet.,' (v) were identified at a rank above species level, or (vi) lacked taxonomic rank information above the genus level.

The frequency of synonyms could be high in the PBDB, because taxonomists have revised nomenclature after many of the names were entered. Therefore, I took precautions to validate taxonomic identifications. In particular, vetting included (i) cross-referencing entries against published lists (Kiessling and Aberhan 2007a; Williams 1997-2002) and the World Register of Marine Species (`marinespecies.org`), (ii) hand-checking the original publication record for all species with durations longer than 50 Ma, and (iii) searching for synonyms based on spelling errors (e.g. *Saccostrea cuculata* instead of *S. cucullata*) and suffix variants (e.g. *-us* instead of *-um*).

I treated data at the level of species for the main text, because I was interested in the potential effects of inter- and intra-specific interactions (Hendricks, Saupe, et al. 2014). Limitations of the fossil record include the potential to overlook cryptic species, because taxonomic definitions for fossils depend on morphology alone. This issue, however, likely affects both brachiopods and bivalves, from all time periods, and therefore is unlikely to result in directional bias. Nevertheless, I also repeated analyses at the level of genera (Table A.3G). To minimise the contributions of waste-bin taxa, only genera with species-level identifications were used; that is, I aggregated the species-level records to their encompassing genera.

I binned data at the finest temporal resolution (stages or clustered stages; Table A.1) that retained sufficient occurrences in every time bin to calculate the variables of interest. Boundary dates followed the International Chronostratigraphic Chart from the International Commission on Stratigraphy (Cohen, Harper, et al. 2017) (`stratigraphy.org`). Spatial sampling in the Cambrian was too sparse to achieve

sub-epoch resolution, so the Phanerozoic dataset was truncated to Ordovician–Pleistocene records. Time bins were < 20 Ma in duration (Table A.1); the median was 5.1 Ma (interquartile range 3.6–7.7 Ma). A specimen record was included in a time bin if the name or age of both its maximum and minimum occurrence estimate fell within the start and end dates for the bin. Dates were inclusive of rounding error to the nearest 0.01 Ma for boundaries younger than 10 Ma, 0.1 Ma for boundaries 50–150 Ma, and 1 Ma for boundaries older than 150 Ma. Where age estimates did not constrain records to single time bins, I matched the 'early interval' and 'late interval' names to time bin names where possible.

The PBDB internally derives palaeo-latitude and -longitude under plate tectonic model described by Wright, Zahirovic, et al. (2013). I used these coordinates to rasterise occurrences in each of 55,800 global grid cells. To ensure that all cells contained the same area, an initial 1° grid was transformed to Eckert VI projection, using the **raster** (Hijmans 2019), **rgdal** (Bivand, Keitt, et al. 2018), and **sp** (Bivand, Pebesma, et al. 2013) packages in R. The centroid coordinates of each of these equal-area cells were used for all subsequent calculations involving the spatial location of records (e.g. minimum spanning tree length). The transformation to ensure equal-area among cells meant that the shape of the cells was distorted (i.e. the sides differed in length), but average width of a cell was 1×10^2 km. This spatial grain of analysis was chosen to smooth differences between densely and sparsely sampled localities while still capturing interspecific differences in regional occupancy (Close, Benson, et al. 2017). Finally, I eliminated occurrence records for any species observed in exactly one grid cell from a single time period (please read section B.3.3 for sensitivity to space-time singletons); this step reduced the influence of Lagerstätten and the incidence of invalid species identifications.

The rasterised occurrence records are available as Dataset S1. The dataset contains 1.7×10^4 species (8.1×10^3 brachiopods and 8.9×10^3 bivalves) and 5.8×10^4 unique species-location-time points (2.7×10^4 for brachiopods and 3.1×10^4 for bivalves).

B.1.2 Northwest Atlantic Neogene–Quaternary dataset

The northwest Atlantic contains a dense record of fossil molluscs. This system is also of interest to test for ecological release because the Gulf Coast is the only region where taxa 'bloomed' after the end-Cretaceous mass extinction (Jablonski 1998). I downloaded occurrences from the Integrated Digitized Biocollections database in October 2018, according to the following criteria: 'geopoint' entered; 'specificepithet' entered; 'basisofrecord' as 'FossilSpecimen'; and 'class' as 'Bivalvia'. After download, I excluded records where the species epithet was entered as 'sp.,' the taxon rank was entered as 'doubtful,' or where the radius of uncertainty exceeded 50 km. I cross-referenced the species list against the World Register of Marine Species database to identify freshwater species, which were removed. As in the PBDB dataset, I synonymised species names with spelling errors or suffix variants. I also checked that no set of coordinates fell exactly at the origin (i.e. 0° latitude and longitude), which could indicate a default entry for an unknown location. I removed records with an era assigned to the Mesozoic, Paleozoic, or Proterozoic. The iDigBio records were combined with the subset of global PBDB records from Miocene–Late Pleistocene molluscs and subset to the northwest Atlantic.

The extent of the study area (6–42° North latitude, 50–100° West longitude; Fig. 2.2A) corresponds to the Gulf of Mexico and the western part of the Central North Atlantic biozone defined by Sutton, Clark, et al. (2017). The northern limit of this area matches the meso-pelagic ecoregion boundary quantified elsewhere (Backus, Craddock, et al. 1977). The southern limit is set by the Isthmus of Panama and coastline of South America.

PBDB and iDigBio contain almost entirely different sets of specimens, because the PBDB draws primarily on published records whereas iDigBio draws primarily on unpublished museum specimens. There are rare cases where specimens could have been entered into both databases (e.g. some digitized holotypes). This duplication in individual occurrence records would not affect analysis, however, because only unique space-time-taxa combinations were retained during rasterisation.

Time bins for the northwest Atlantic dataset (Table A.1B) were shorter than those used in the global dataset. All bins were < 5 Ma in duration, with a median of 1.8 Ma (interquartile range 1.0–4.4 Ma). Sampling in the Aquitanian was so sparse in comparison to the other intervals that the dataset was truncated to Burdigalian–Late Pleistocene records only. Many iDigBio records lacked precise age information, either as a numeric estimate or geologic stage name. I therefore researched the chronostratigraphy of sedimentary units in the northwest Atlantic to determine the minimum and maximum age estimates for 485 formations and members, expanding the correlation chart of Saupe, Hendricks, et al. (2014). The compilation (Dataset S2) includes referenced opinions of age estimates from isotopic dating, sequence stratigraphy matches with sea level curves, and stratigraphic correlations and relationships. Age estimates for each occurrence record that lacked sufficient temporal information were inferred from the geologic unit with which it was associated. I searched for geologic context in the 'group,' 'formation,' 'member,' 'verbatimLocality,' and 'lithostratigraphicTerms' fields of iDigBio records; and the 'stratgroup,' 'formation,' 'member,' 'collection_name,' 'collection_aka,' and 'localsection' fields of PBDB records. I hand-checked any records that included mention of multiple geologic units to determine a unique assignment.

The age range of deposition in many geologic units spanned multiple time bins. Therefore, I assigned a record to a time bin if more than two-thirds of its estimated range fell within the bin. Records that spanned more than 3 bins were excluded altogether. This two-thirds overlap rule assigned 70% of geologic units to time bins; a stricter binning rule would have failed to maintain a sufficient amount for analysis. Some records could not be assigned to a time bin using the methods above, but had been identified with a stage or North American Land Mammal Age (e.g. Hemingfordian). These temporal units were converted to numeric age ranges based on the literature (Bell, Lundelius Jr, et al. 2004; Tedford, Albright III, et al. 2004; Cohen, Harper, et al. 2017). It is probable that the two-thirds overlap rule assigned some occurrences to incorrect time bins. I account for this error later, in time series analysis, which controlled for similarity between time bins that

occurred close together regardless of the source of similarity (e.g. founder effects, environmental stasis, or methodological artefacts).

The iDigBio platform provides coordinates at their modern-day positions. Therefore, I reconstructed the palaeocoordinates for all records (iDigBio and PBDB) in the northwest Atlantic dataset. Modern coordinates were partitioned among plates and rotated according to PaleoAtlas models from EarthByte (Scotese 2016b) (version 2, accessed October 2018) within the GPlates computing environment (Müller, Cannon, et al. 2018) (version 2.1.0, released August 2018). The coordinates of all records within a time bin were reconstructed at the median age of the bin. I rasterised occurrences into equal-area grid cells at a resolution that was twice as fine as that of the global dataset: I used a 0.5° grid, resulting in equal-area cells of 5x10 km width on average, and removed species that occurred in a single cell in only one time bin.

The rasterised occurrence records for the northwest Atlantic are available as Dataset S3. The dataset contains $1.9\text{x}10^3$ species and $6.1\text{x}10^4$ unique species-location-time points.

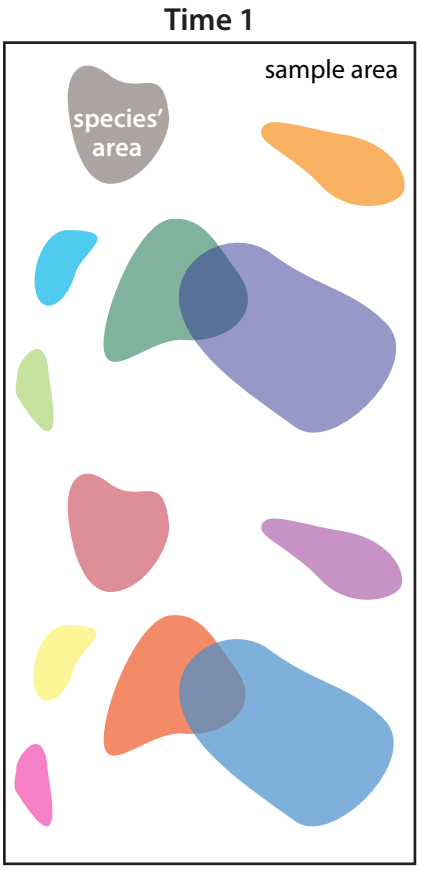
B.2 Standardisation of spatial coverage

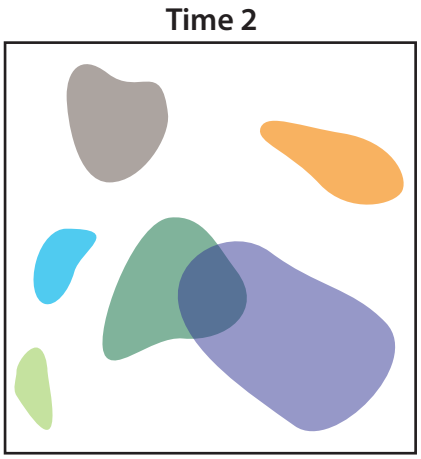
Statistical comparisons of spatiotemporal occurrences assume that sampling universes are equivalent across space and through time. Fossil data violate this assumption consistently, because preservation potential, field sampling intensity, and amount of preserved rock area structure the record of species' distributions and diversity (Sheehan 1977; Signor, Lipps, et al. 1982; Behrensmeyer and Kidwell 1985; Kiessling 2005; Kowalewski and Bambach 2008; Patzkowsky and Holland 2012; Vilhena and Smith 2013). Furthermore, biogeographic responses vary by geographic region (Jablonski 1998; Button, Lloyd, et al. 2017), and diversity patterns vary with the spatial scale of analysis, from local sites (α diversity) to a whole study region or biosphere (γ diversity). Turnover, the factor by which γ exceeds mean α, is termed β diversity (Tuomisto 2010). β diversity increases as either the total area of sampling or the dispersion of sampling locations increases (Connor and

McCoy 1979). This biological link between diversity and spatial coverage, referred to as the species–area effect, pervades all spatial grains throughout the Phanerozoic (Sepkoski Jr 1976; Barnosky, Carrasco, et al. 2005). Because the research question concerns the relationship between range size and diversity through time, spatial coverage standardisation in terms of both total sampling area and dispersion of sampling sites was exigent to analysis.

A simple way to account for unequal sample area is to divide a species' distribution area by the total sample area (e.g. Foote, Crampton, et al. 2007; Harnik, Simpson, et al. 2012; Finnegan, Anderson, et al. 2015). This standardised measure, sometimes called proportional occupancy, could account adequately for unequal sampling area if species' ranges were compared contemporaneously, or if species composition were homogeneous among sites. In practice, palaeontologists often track geographic ranges through time, and species turn over among sites. A proportional measure of occupancy may not be desirable in these cases (Kiessling and Kocsis 2016). In the context of the present hypothesis of ecological release, converting range area to a proportion of total sampling area could induce a negative correlation between time series of range size and species count as an artefact of sampling (Fig. B.1). A large sampling area will tend to contain more species than a small area. Meanwhile, if the frequency distribution of species' geographic range sizes is identical between the two regions, but these sizes are converted to proportions of the total sampling area, then the large study area will appear to contain species with small geographic ranges. Furthermore, in section B.2.1 I prove the statement that a difference in proportional occupancy of widespread species will be greater than that of restricted species when a study area increases beyond the extent of a species' geographic range.

Even with constant sample area, variation in spatial dispersion of sampling locations could cause a negative correlation between range area and sampling amount (Close, Benson, et al. 2017). When localities are rarefied, each site could reflect a distinct ecological community, resulting in estimates of small distribution areas and large counts of species. In contrast, when localities are aggregated, species' distributions might be sampled more completely, but many more species would





Variable	Time 1	Time 2
Sample area	large	medium
Species count	12	6
Mean species' area	large	large
Mean proportional species' area	tiny	medium

Figure B.1: In this conceptual figure of the species–area effect, each coloured shape represents a species' geographic distribution in map view. The sizes of outlining boxes indicate sampling area, which is twice as large at observation Time 1 than Time 2. The mean species' range area is unchanged, but proportional range area decreases as sampling area increases. A correlation between species count and proportional range area from these two time steps would be negative due to the difference in sampling area alone.

be shared among sample sites. In this latter scenario, one would observe large distribution areas and small counts of species.

To account for the heterogeneity of preserved fossil occurrences, I devised a subsampling method that controls spatial extent and dispersion jointly. Each iteration picks a random occurrence point as a centre around which to circumscribe a buffer region with a predetermined radius; the procedure then draws a given number of equal-area grid cells within the circular buffer (Fig. A.2). These cells form a subsample, within which range size and species count are calculated. I performed sensitivity analyses on every parameter of the routine, including the radius and

number of cells of each subsample (section B.3.3). For the analysis results presented in Chapter 2, I set a 1500 km radius to confine each subsample, which agrees well with previous quantitative thresholds (Close, Benson, et al. 2017) and the common approach of constraining the dispersion of a sample to one continent or ocean basin (e.g. Jablonski 1998; Saupe, Hendricks, et al. 2014; Patzkowsky and Holland 2016). Consequently, the circle that constrained a random sample covered $7x10^6$ km^2, an area approximately equivalent to the size of Australia. Within this region, each subsample iteration drew a given number of cells (itemised in sections B.3.1 and B.4.2); the quota was within the range that others have set when sampling grid cells (Darroch and Wagner 2015). Although any individual subsample from the Phanerozoic dataset may have captured only some of the fossil species present in the global record at the time, all available data contributed to analysis because the regional subsample procedure was replicated 500 times in each time bin.

The details of the subsampling procedure, and the calculation of variables from subsampled data, are explained below (sections B.3.1 and B.4.2). I constructed buffers with the **raster** package (Hijmans 2019) and manipulated vector geometries with the **sp** package (Bivand, Pebesma, et al. 2013). For range size calculations, great circle and spanning tree distances were calculated with the **vegan** package in R (Oksanen, Blanchet, et al. 2019). I did not rarefy PBDB 'collection' or 'reference' groups as done in earlier studies (for review, consult Close, Benson, et al. 2017). This was because the spatial subsampling procedure already rarefied the number of grid cells in each sample, and it has been found that there is a 1:1 correlation between reference counts and grid cell counts in global PBDB datasets of marine invertebrates (Alroy, Aberhan, et al. 2008). Indeed, the spatial standardisation procedure rarefies species count as a consequence of rarefying sampling area. Additional rarefaction on species count would compress the distribution of observed values further, perhaps to the point of obscuring any true signal, i.e. causing type II error.

B.2.1 Further notes on proportional occupancy

Section B.2 reviews proportional occupancy as a common approach to address variation in spatial sampling through time. Problems arise, however, when the metric is used to compare range sizes across time intervals, rather than within an interval. I explore these problems mathematically in this section. To begin, I specify two criteria that a range size metric must meet in order to be suitable for analysis that is simultaneously spatial and temporal: (i) the metric should be equal among all species that do not change in distribution, regardless of the size of the species' range, and (ii) the effect of sampling area should be minimal.

When total sampling area increases, the difference in proportional occupancy of widespread species will be greater than that of restricted species, a violation of the first criterion for a fair metric. Calculating changes in values as a relative difference (rather than absolute difference) avoids this situation. However, calculation of a relative change in proportions exaggerates the influence of sampling area, which violates the second criterion. The proof follows.

Statement. Let species X and Y have distributions with area x and y, respectively, such that $x < y$ (Fig. B.2). Let both species occur exclusively within the preserved rock area in each of two time periods. Let the outcrop areas be c_1 and c_2, such that $c_1 < c_2$.

Then:

1. The difference in species' proportional occupancy of areas c_1 and c_2 will be smaller for species X than for species Y. Let us call this difference ΔA_i for a species' area i.

2. The relative change in species' proportional occupancy of areas c_1 and c_2 will be equal for species X and Y. Let us call this difference δA_i.

3. The magnitude of the change in proportional occupancy of areas c_1 and c_2 for either species will be larger when this change is calculated as a proportion than as a difference.

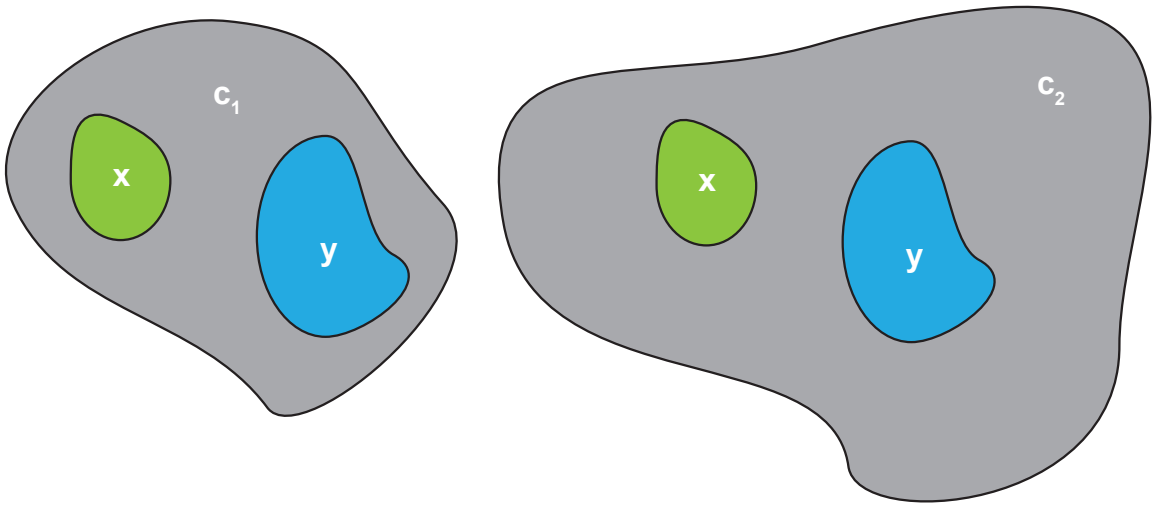


Figure B.2: The geographic distributions of species X and Y are shaded in green and blue, respectively. Two possible distributions of sampling are shaded in grey.

Proof.

Proof. I prove each of the 3 parts of the above statement.

1. For any species with area i, the proportional occupancy of area c is i/c. Using the definition of ΔA_i above, the difference in proportional occupancy of species X and Y in areas c_1 and c_2 can be written

$$\Delta A_i = \frac{i}{c_1} - \frac{i}{c_2} = \frac{ic_2 - ic_1}{c_1 c_2} = \frac{i(c_2 - c_1)}{c_1 c_2} \quad \forall i \in [x, y].$$

   Let k be a constant defined as $k = (c_2 - c_1)/(c_1 c_2)$. We can write $\Delta A_x = xk$ and $\Delta A_y = ay$. It is given that $0 < c_1 < c_2$, so $k > 0$. It is also given that $0 < x < y$. Thus, $\Delta A_x < \Delta A_y$.

2. The relative change in proportional occupancy of species X and Y in areas c_1 and c_2 can be written

$$\delta A_i = \frac{i}{c_1} \div \frac{i}{c_2} = \frac{ic_2}{ic_1} = \frac{c_2}{c_1} \quad \forall i \in [x, y].$$

   This is equivalent to stating that $\delta A_x = \delta A_y$.

3. Using the expressions for ΔA_i and δA_i in parts 1 and 2 above,

$$\Delta A_i = \frac{i(c_2 - c_1)}{c_1 c_2} \text{ and } \delta A_i = \frac{c_2}{c_1}$$

It is given that $0 < i < c_1 < c_2 \quad \forall i \in [x, y]$. Therefore,

$$c_1 c_2(\Delta A_i) = i(c_2 - c_1) = ic_2 - ic_1 < ic_2 < c_2 c_2 = c_1 c_2(\delta A_i)$$

It follows that $\delta A_i > \Delta A_i$ for any species with distribution area i smaller than c_1.

□

Any study that aims to quantify spatial patterns through time fairly must standardise the area of sampling. I prove in this section that transforming geographic distributions to a proportion of sampling area is not a fair way to mitigate the heterogeneity of fossil occurrences. Therefore, I developed a randomised subsampling routine that standardises spatial coverage without invoking proportions. The procedure draws subsamples within the confines of a single region; the size of species' distributions in a subsample does not depend on the size of the sampling universe outside this region (Fig. A.2). As a result of the subsampling approach, a species' distribution that does not change from one time to another will be captured as stable, satisfying the first criterion above. Furthermore, the procedure ensures that subsamples are equal in area, which is equivalent to forcing the denominator of a proportion to be a constant through time. This satisfies the second criterion. The subsampling routine is a type of count-based rarefaction, and therefore its fairness comes with the trade-off of compression (i.e. lessened effect size). Nevertheless, I demonstrate that analysis remains powerful enough to detect the direction of a signal accurately (section B.5).

B.3 Assemblage framework

The first analysis approach was built to relate the number of species in a regional marine assemblage to the geographic distributions of the assemblages' constituent species, in each of a series of consecutive time intervals. To begin, I implemented a randomised subsampling procedure to select a given number of equal-area grid cells within a constrained region, across a time series and in replication. The steps of the routine follow in the next section (B.3.1). Next, I calculated time series correlations between every pair of study variables. The whole analysis was conducted on both the global and the northwest Atlantic datasets.

The spatial standardisation procedure can be amended in many ways, for instance by selecting grid cells according to either unweighted or proximity-weighted probabilities. In the unweighted cell selection method, occupied cells were selected at random within each radius-buffered area. In the proximity-weighted method, cells were selected with scaled probability inversely proportional to their squared distance from the center of the buffer. The proximity-weighted sampling method is more computationally intensive since the weights must be calculated for every occupied cell in all iterations of the subsampling procedure. However, this method clusters subsampled cells more compactly (i.e. reduces spatial dispersion) while retaining a random component to subsampling. In cases where sampling is dense and more even across space and time, the proximity-weighted method might reduce the grain of study and hence lead to more precise estimates of assemblage-level characters, such as mean species' occupancy.

Due to the stochastic nature of the subsampling procedure, it is possible that an assemblage could consist entirely of species represented by a single occurrence. In such a case it would not be possible to calculate any minimum spanning tree (MST) lengths. Therefore, I excluded any replicate time series that included this type of assemblage. There were only two such series in the realised subsampling runs for the analysis presented in Chapter 2.

B.3.1 Assemblage subsampling routine

1. **Regional restriction** (Fig. A.2A,B). Select one grid cell populated with a fossil (i.e. an 'occupied cell'). Draw a circular buffer with a radius of 1500 km from the centroid of this 'seed cell'. Define the subsampling pool as the contemporaneous occupied cells (including seed cell) that have centroids within the buffered area. If the sample size of the pool meets or exceeds 12 cells, save the position of the seed cell. Repeat this step for all occupied cells in each time bin.

2. a. **Unweighted equal-area sampling** (Fig. A.2C). Randomly select one seed cell saved from step 1. Draw randomly 12 cells without replacement from the subsampling pool associated with the seed cell. Repeat this step once in every time bin.

   b. **Proximity-weighted equal-area sampling.** Randomly select one seed cell saved from step 1. Compute the great-circle distance matrix between the centroid of each pool cell and the centroid of the seed cell. Define the probability of sampling a pool cell as proportional to the inverse of its squared distance from the seed cell. Include the seed cell in the subsample in addition to 11 pool cells drawn with weighted probabilities (without replacement).

3. **Assemblage-level data** (Fig. A.2D). For each subsample, calculate the square-root of the length of the MST connecting all subsampled cells ('sampling aggregation'). For each species that occurs more than once, count the number of occupied cells and calculate the square-root of the length of the MST connecting occupied cell centroids. Save the sampling aggregation, species count, mean occupancy among species, and median MST length among species.

4. **Iterate.** Repeat steps 2–3 for a total of 500 replicates. Each time series of subsamples represents one subsampling iteration.

Note that if rarefaction methods are used to calculate species count, additional components exist for steps 2 and 3. In step 2 (a or b), check that the subsample

contains the minimum number of records for classical rarefaction or the coverage level required for shareholder quorum subsampling. If the subsample is not large enough, take a new subsample. In step 3, use the distribution of presences (occupancy counts) to subsample the number of species to a desired coverage level or count.

B.3.2 Time series analysis

Time series analysis aimed to establish the relationship between species' distributions and species count, and to evaluate the success of spatial sampling standardisation. For each subsampled dataset, I calculated the correlations between predictor time series (species count or sampling aggregation) and response time series (occupancy or minimum spanning tree length). I used Kendall's *tau* to estimate correlation strength because this non-parametric metric can be modified to account for ties, which occurred in time series of median occupancy.

Taxon counts were similar in adjacent time bins, and so I corrected the time series analysis for this autocorrelation. I inspected autocorrelation and partial autocorrelation plots for a representative iteration of each variable's time series and tested the assumption of stationarity with an augmented Dickey-Fuller test from the **fUnitRoots** package (Wuertz, Setz, et al. 2017) in R. Based on these explorations, I chose an autoregressive (AR) error structure of order 1 to model the predictor time series (either species count or sampling aggregation) in each iteration, using the 'arima' function from the base **stats** package in R (R Core Team 2018). The first-order coefficient estimated for species count was 0.33 and for sampling aggregation was 0.38. I calculated *tau* based on the predictor time series' AR(1) model residuals instead of the original series to prevent induced correlation (Probst, Stelzenmüller, et al. 2012). The mean and 95% confidence interval of Kendall's *tau* coefficient are reported for each pair-wise correlation. The bounds of the confidence interval are the 2.5% and 97.5% quantiles of values obtained across bootstrap iterations. I used the 'prewhiten' function in the **TSA** package (Chan and Ripley 2018) to check for correlations at positive time lags of 1–5 steps.

B.3.3 Sensitivity to subsampling parameters

The spatial standardisation procedure (sections B.3.1 and B.4.2) contains a variety of user-defined parameters (e.g. number of iterations). I varied each such parameter in isolation, as summarised in Table A.2, to test the sensitivity of the analysis to the subsampling method. For instance, I experimented with stricter taxonomic vetting, different grid cell quotas, classical or coverage-based species richness rarefaction, and aggregation by genus names. I analysed time series from each variation of the subsampled dataset.

As a baseline against which to compare results from sensitivity analyses, I report correlations from subsamples generated under the following parameters: 1500 km radial constraint, 12 grid cells, 500 iterations, and unweighted cell selection (Table A.3A). The area of 12 grid cells totals 1.1×10^5 km^2, approximately the land area of Honduras. The baseline considered many metrics to summarise range size among species: minimum, median, maximum, and average of occupancies or MST lengths; for occupancy, values are reported both when single-occurring species ('singletons') were included or excluded. The parameter combination from which correlations were derived for the results presented in Chapter 2 was identical to the baseline, except with proximity-weighted instead of unweighted cell selection (Table A.3C).

Taxonomic data quality

The first sensitivity test concerned the size and quality of the original PBDB dataset. In particular, I downloaded records in the same way specified above (section B.1.1), except that records were filtered to those with 'accepted names' at the level of species. This filter could remove some synonyms and misspellings, but it also severely restricted the number of records retrieved. In many of these excluded records, an enterer may have correctly specified the species' name, locality, stratigraphic data, and reference, but not performed the additional verification steps for the PBDB to recognise the taxonomy as 'accepted' (e.g. specifying the year and author of the species' description). These valid records would be retained in a download that does not distinguish between accepted and unaccepted taxonomic

status. In light of the high potential for synonyms and misspellings when including 'unaccepted' names, I still implemented manual vetting procedures to validate taxonomic names (section B.1.1).

The accepted-only data were too sparse in some time bins to reach a spatial coverage of 12 grid cells, so I subsampled only 10 cells at a time. The radius of constraint was 1500 km. The subsampling procedure was replicated 500 times. The same parameters were applied when subsampling the full dataset.

The magnitude and direction of correlations from the restricted dataset were equivalent to those generated from the larger dataset: species count correlated only very weakly with every range metric. Sampling aggregation correlated negatively with occupancy but positively with MST, although weakly in both cases. Because the larger dataset provided equivalent results to the more restricted version, I used the larger dataset for all further sensitivity analyses and Chapter 2 results. The additional records allowed for finer temporal resolution; the resultant binning is given in Table A.1A.

Singletons in space-time

I tested how the correlation between species count and range size might change if I included *vs.* excluded species that occurred in exactly one grid cell of one time bin ('space-time singletons'). Species names that were misspelled or otherwise wrong would likely occur in the dataset as space-time singletons, and thus removing these species should improve the quality of the dataset. Space-time singletons accounted for $1.3\text{x}10^4$ of $3.0\text{x}10^4$ taxa.

Including singleton taxa to test sensitivity to data quality did not change the correlation between species count and geographic distribution. Similarly, species count and range size series correlated to the same degree with spatial sample coverage whether or not space-time singletons were included.

I removed space-time singletons for all further sensitivity analyses and Chapter 2 results. There are three motivations for this decision. Firstly, I prioritised data quality above data abundance. Secondly, space-time singletons would be eliminated

a priori during taxon-tracking analyses, which considered only species that occurred in consecutive time bins. Finally, space-time singletons could exacerbate type I error in sensitivity analyses that performed coverage-based rarefaction, as discussed in section B.3.3.

Number of cells

The next parameter I tested was the size of a subsample (i.e. number of cells selected) (Table A.2). Increasing subsample size might reduce variability among subsampling iterations. After removing a filter on taxonomic status as 'accepted', sufficient occurrence records were available to increase the number of subsampled cells from 10 to 12. I tested the influence of this increase in cells whilst keeping all other subsampling parameters (e.g. radius) constant. I also experimentally reduced the number of subsampled cells by half, from 12 to 6.

The correlation between species count and range size was weaker (less negative, i.e. closer to zero) for subsamples of 12 cells compared to those with 10 cells. Spatial area was better controlled for with the larger sample size, as reflected in smaller correlations between sampling aggregation and either species count or geographic distribution time series. Hence, the slightly negative correlation between species count and range size in subsamples of only 10 cells could be attributed primarily to bias in spatial coverage.

Results from subsamples of 6 cells support the interpretation that the strength of spatial bias decreases as more cells are selected. In particular, with only 6 cells, the correlation between spatial aggregation and either species count or any of the distribution metric time series was as great as, or larger than, the same correlation based on subsamples with 10 or 12 cells.

The width of the confidence intervals for correlation coefficients did not change in a uniform way or to a large degree when either more or fewer cells were selected. This might indicate that the variability in subsamples depends on which geographic

province they capture, rather than on which cells are chosen from within the radius-constrained area. Alternatively, a change in the number of cells from 6 to 12 might not be large enough to induce a measurable effect on variance.

I included 12 cells in subsamples for all further sensitivity analyses and Chapter 2 results to minimize the influence of spatial bias.

Radius of regional constraint

While testing sensitivity to the area (number of cells) subsampled within each region, I also varied the region area (radius of constraint). Specifically, I reduced the radius of the buffer (Fig. A.2) from 1500 km to 750 km, while selecting 6 grid cells within the region. The magnitude of correlation between sampling aggregation and range size increased for MST length, while the magnitude of other correlations remained constant (Table A.3B). I set the radius at 1500 km for all further sensitivity analyses and Chapter 2 results to increase the number of cells that could be selected and minimise the influence of spatial bias.

Number of subsampling iterations

I noted above that the variability obtained among subsamples did not seem to depend on the number of cells included in a subsample. However, substantive variation among iterations could exist due to ecosystem differences among provinces, of which there have been up to 12 at times in the Phanerozoic (Kocsis, Reddin, et al. 2018). It is difficult to estimate *a priori* how many iterations are necessary to sample this regional variation thoroughly, especially because there are different numbers of possible starting points for subsampling (i.e. number of available 'seed cells') among time bins. Therefore, I experimented with an increase in the number of iterations, from 500 to 1000 replicates (Table A.2). Doubling the number of iterations, however, did not change the mean or confidence interval for correlations. Therefore, it is likely that 500 iterations was sufficient to sample most or all combinations of regions and site (cell) configurations. For this reason, I continued to replicate subsampling over 500 iterations for further sensitivity analyses and Chapter 2 results.

Proximity-weighted cell selection

The spatial coverage standardisation routine for assemblage data (section B.3.1) contained the option to select grid cells with scaled probability inversely proportional to the square of their distance from the regional center. In contrast, occupied cells were selected at random within each radius-buffered area for the unweighted cell selection method used in the above sensitivity analyses. The weighted method was designed to pick cells in a tighter cluster than the unweighted method (i.e. reduce spatial dispersion) whilst retaining a random component to subsampling. In cases where sampling was dense and homogeneous across space and time, the proximity-weighted method might reduce the grain of study and hence lead to more precise estimates of assemblage-level traits, such as mean species' occupancy.

I calculated time series correlations from weighted subsamples, for 500 iterations of 12 cell subsamples (Table A.3C). Weighting the selection probability reduced the correlation between species count metrics and sample aggregation, as predicted. However, weighting did not reduce the correlation between range size metrics and sample aggregation. The correlation between range size and species count time series was equivalent whether cell selection was weighted or unweighted.

Species count rarefaction

The number of species observed in a region depends strongly on fossil preservation and collection effort. The spatial aspects of sampling (area and extent) are discussed above (section B.2). In this section, I consider the possibility that even after standardising spatial coverage, heterogeneous sampling effort (e.g. intensity of fossil collecting) could affect the number of species recovered. To standardise this variability, I implemented two common interpolation-based methods of (non-spatial) richness rarefaction: classical rarefaction and coverage-based rarefaction. Classical rarefaction standardises sample size. Coverage-based rarefaction, also called shareholder quorum subsampling (SQS) or coverage-based richness interpolation, standardises the estimated amount of the underlying frequency distribution represented in a sample (Alroy 2010; Chao and Jost 2012). Thus, sample size

does not influence SQS estimates. Species count curves produced by SQS tend to be more stable than those from classical rarefaction or extrapolation methods (Close, Evers, et al. 2018).

Coverage-based rarefaction. In the first rarefaction approach, I calculated SQS estimates of species richness within subsamples, using the **iNEXT** package (Hsieh, Ma, et al. 2016). The implementation of **iNEXT** finds the same species count estimate as John Alroy's exact Perl scripts from version 4.0 onwards (Close, Evers, et al. 2018). Since my spatial subsampling procedure rarefied the number of grid cells in each sample, and since grid cell counts correlate directly with reference counts in global PBDB datasets of marine invertebrates (Alroy, Aberhan, et al. 2008), I did not rarefy by PBDB 'collections' or 'references'. Neither did I transform species count estimates, because correlations considered only relative rank of counts, rather than absolute magnitude.

Close, Evers, et al. (2018) suggested that rarefied richness is sensitive to coverage level and thus recommend using the highest possible quorum value. After increasing the number of cells in a subsample from 10 to 12 in the global dataset (section B.3.3), it was possible to increase coverage from 40% to 70% of the incidence frequency distribution in every time bin. To test the influence of coverage level, I spatially subsampled the global data while applying either 40% or 70% coverage standardisation (Table A.2). As in the non-rarefied data, the correlation between range size and species count (as estimated by coverage-based rarefaction) was negative (Table A.3D,E). The magnitude, however, was not greater than that estimated for the correlation between sampling aggregation and geographic range size. In other words, the confounding effect of spatial coverage in samples could explain the slightly negative relationship between the variables of interest, and hence no support was found for ecological release when species count was standardised by frequency distribution coverage. Correlations were similar regardless of whether 40% or 70% coverage was used, although this similarity likely occurred because of the previous removal of space-time singletons, which could influence coverage estimates.

Singleton taxa in assemblages. Frequency distribution coverage (derived mathematically from an expression named Good's u) is based in part on the relative frequency of taxa that occur only once (Good 1953; Alroy 2010). Singletons result in a smaller estimate of Good's u, which would lead to more sampling of the coverage distribution to achieve the desired quorum level, and ultimately a higher estimate of richness. Hence, singletons could upwardly bias coverage-based richness estimates except in the unlikely case that the number of singletons were evenly distributed through time and space. Further confounding analysis, assemblage singletons cannot influence mean occupancy or median MST, because definitions of both of those range size metrics here excluded singletons. To evaluate the effect of this asymmetry, I modified the subsampling protocol to remove single-occurring species from each assemblage before performing SQS. This added protocol reduced the correlation between sample aggregation and the SQS estimate of species count to the smallest value found in any of the sensitivity analyses. Indeed, time series correlations recovered a weakly negative relationship between species count and sampling aggregation (Table A.3F), whereas otherwise this relationship was positive (Table A.3D,E). All else equal, this inverse pattern would bias the correlation between species count and mean occupancy towards type II error, because the relationship between occupancy and sampling aggregation is also consistently negative.

Classical rarefaction. The classical form of rarefaction standardises the number of sampled species units (i.e. presence counts in rasterised datasets). Here, I selected 50 presences (species-cell combinations) in a spatial subsample and estimated the expected number of species in that selection (there were 12 cells in each subsample, within a 1500 km radius, and selected with unweighted probability). The time series was truncated to the Middle Ordovician onward because fossils in the Lower Ordovician were too sparse to obtain 50 occurrences in any given subsample. The results from classically-rarefied assemblages were similar to those derived from analyses where SQS was used to rarefy at 70% coverage including singleton taxa in assemblages (Table A.3E).

Genus-level data

Although I was interested in processes that occur at the level of species, I repeated the analysis at the level of genera as well (Table A.2). The latter approach increased the number of occurrences from which to reconstruct each geographic distribution and provides comparability with previous palaeontological studies (e.g. Foote, Crampton, et al. 2007; Harnik, Simpson, et al. 2012; Finnegan, Anderson, et al. 2015). I did not elevate subgenera to the rank of genera. Changing the taxonomic rank of analysis affects both taxon count and geographic range size: taxon count in any time period will decrease when assessed at higher levels, whereas geographic range sizes will increase. The empirical evidence from data aggregated by genus name suggests that the results presented in Chapter 2 are insensitive to taxonomic level. That is, the *tau* correlations were equivalent between analyses run on species (Table A.3A) and those run on genera (Table A.3G).

Northwest Atlantic dataset

The Neogene–Quaternary dataset was already restricted to a single ocean basin in extent, so I did not use a radial constraint during subsampling. I did, however, still remove space-time singletons and standardise spatial area with the cell selection step of subsampling (Fig. A.2C). The high density of occurrences allowed inclusion of 15 cells in the subsamples from every time bin, compared to 12 cells in the Phanerozoic dataset (Table A.2).

The time series analysis from the finer-resolution, northwest Atlantic data produced wide confidence intervals (Table A.3H). Correlation between species count and any of the distribution metrics was negligible, as in the main results from the Phanerozoic dataset.

B.4 Taxon-tracking framework

I applied a taxon-tracking approach to build a dataset of temporal changes in range size and species count from the perspective of species that cross stage boundaries (i.e. survivors). I used this dataset (i) to test whether survivors contracted in

distribution over intervals when species counts grew compared to intervals when species count fell (section B.4.3), and (ii) to estimate the effect on range dynamics from environmental affinity, clade membership, evolutionary dynamics, and changes in species count (section B.4.4).

B.4.1 Environmental preferences

The position and extent of different habitat types (e.g. shallow water carbonate settings) have changed through time. Many species have specific habitat preferences, in which case individuals may be constrained in the areas where they can occur. Since environmental preferences are defined at the level of species, it was not straightforward to account for habitat within an assemblage framework. Therefore, I accounted for suitable habitat only in the taxon-tracking framework. Environmental data were unavailable and unstandardised for most iDigBio specimen records, so I could determine environmental affinities for taxa in the global (PBDB) dataset but not in the northwest Atlantic dataset.

I considered environmental preferences along two axes: substrate (carbonate, siliciclastic, or generalist for both) and bathymetry (deep water, shallow water, or generalist for both). I did not code preferences for latitude or sediment grain size. Carbonate substrates already reflect the extent of tropical biomes to a large degree, and perhaps more accurately than would latitudinal coordinates. Similarly, variation in substrate type and bathymetry also capture some of the differences between fine- and coarse-grained habitats. Determining grain size affinity would improve the quality of niche characterisations only moderately, while at the same time reducing sample size substantially—grain size was unrecorded in many PBDB records.

I coded binary environmental conditions based on associated data fields in the PBDB records (prior to rasterisation), according to the definitions provided by Nürnberg and Aberhan (2013). In particular, 'carbonate' included any instance of 'carbonate,' 'limestone,' 'reef rocks,' 'bafflestone,' 'bindstone,' 'dolomite,' 'framestone,' 'grainstone,' 'lime mudstone,' 'packstone,' 'rudstone,' 'floatstone,' or 'wackestone' in the PBDB primary lithology field. In contrast, 'siliciclastic' included

any instance of 'shale,' 'siliciclastic,' 'volcaniclastic,' 'claystone,' 'conglomerate,' 'mudstone,' 'phyllite,' 'quartzite,' 'sandstone,' 'siltstone,' 'slate,' or 'schist' in primary lithology. For water depth, 'deep' was assigned for any environment field in the set of 'basinal (carbonate),' 'basinal (siliceous),' 'basinal (siliciclastic),' 'deep-water indet.,' 'deep subtidal indet.,' 'deep subtidal ramp,' 'deep subtidal shelf,' 'offshore,' 'offshore indet.,' 'offshore shelf,' 'slope,' 'submarine fan,' 'offshore ramp,' 'basin reef,' and 'slope/ramp reef'. Otherwise, 'shallow' was assigned for 'coastal indet.,' 'delta front,' 'delta plain,' 'deltaic indet.,' 'estuary/bay,' 'foreshore,' 'interdistributary bay,' 'lagoonal,' 'lagoonal/restricted shallow subtidal,' 'marginal marine indet.,' 'open shallow subtidal,' 'fluvial-deltaic indet.,' 'paralic indet.,' 'peritidal,' 'prodelta,' 'sand shoal,' 'shallow subtidal indet.,' 'shoreface,' 'transition zone/lower shoreface,' 'intrashelf/intraplatform reef,' 'reef, buildup or bioherm,' 'perireef or subreef,' or 'platform/shelf-margin reef' in the environment field.

The environment in a grid cell was classified as a single condition if $\geq 80\%$ of the palaeoenvironmental records in that cell were indicative of that condition; otherwise, the cell was inferred to contain both conditions (e.g. mixed carbonate and siliciclastic substrate). Where both conditions occurred in the reference raster cell, however, I made no inference from that cell about the occupying species' habitat preferences.

Some occupied grid cells lacked sufficient information to assign an environmental condition. I referenced these cells against an environmental raster derived from an independent PBDB download of all brachiopod and bivalve occurrences (not restricted to species-level identification, but excluding fossils from terrestrial environments). If the grid cell in question was characterised by a single environmental condition in the reference raster, I assigned that condition to the species' occurrence records from the corresponding cell in the vetted dataset. Cells where environmental conditions still could not be inferred along one or both axes were not considered: occurrences that fell in these cells were excluded.

After grading the environmental state for every occurrence record, I inferred the environmental preferences of each species. For each axis, a species with $\geq 80\%$ of occurrence records classed as only one category was considered a specialist for

that environmental condition. A species with $> 20\%$ of occurrence records in both categories was treated as a generalist along that axis.

B.4.2 Taxon-tracking subsampling routine

The below procedure applies to a focal species in a time bin. I iterated the procedure across species and time bins to generate a taxon-tracking dataset.

1. **Suitable habitat.** Cull the raster of occupied cells in the time bin such that only cells with habitat suitable for the focal species are included. (If the species is a generalist along both axes, no action is necessary.)

2. **Regional restriction.** This step is the same as in the assemblage-based procedure (Fig. A.2A,B). Select one occupied cell. Draw a circular buffer with a radius of 1500 km from the centroid of this 'seed cell'. Define the subsampling pool as the contemporaneous occupied cells (including seed cell) that have centroids within the buffer area. If the sample size of the pool meets or exceeds 10 cells, save the position of the seed cell. Repeat this step for all occupied cells in the time bin.

3. **Unweighted equal-area sampling** (Fig. A.3A–C). Randomly select one seed cell saved from step 1, and then randomly draw 9 additional cells without replacement from the pool cells associated with the seed cell. Repeat this step 500 times in every time bin.

4. **Species-level data** (Fig. A.3D). In each time bin, find the subsamples where the focal species occurs. In each of those subsamples, calculate:

    - the number of cells occupied by the focal species (once by including and once by excluding single occurrences)
    - the square-root of the length of the MST connecting centroids of cells occupied by the focal species
    - the number of species in the subsample (including the focal species)

Save the median value of each of these three metrics across subsamples associated with the focal species.

B.4.3 Biotic crises *vs.* background intervals

To test the prediction that species expand their ranges when released from competition after major drops in richness, I compared 'range-size shift' between two sets of time intervals: intervals when species count increased and intervals when species count decreased. I defined range-size shift for a temporal interval as the mean change in range size among survivors of the corresponding stage boundary. To classify each boundary as associated with either an increase or decrease in species count, I calculated whether the median change in species count experienced by survivors of a given boundary was positive or negative (Table A.4).

Unbalanced numbers of newly-originated and nearly-extinct species between extinction intervals and background intervals (Darroch and Wagner 2015; Button, Lloyd, et al. 2017) could bias analysis, because taxa at these times often experience predictable changes in geographic range size unrelated to potential effects of ecological release (Foote, Crampton, et al. 2007; Liow and Stenseth 2007). The typical trajectory of geographic distribution over the duration of a lineage is arched like a rainbow (mathematically, a convolution of two top hat functions). That is to say, lineages tend to expand after appearing in the fossil record and contract prior to disappearing from the record. Moreover, species in the middle of their existence, at peak occupancy, are hypothesised to experience the strongest competitive pressure (Žliobaitė, Fortelius, et al. 2017). I therefore calculated range-size shift (i) on all species, and (ii) only for species in the middle interval of their temporal durations. In the latter case, if a species was sampled across three consecutive boundaries (i.e. occurred in four consecutive time bins), the change in range size over the middle interval (but not either end interval) would contribute to the calculation of range-size shifts. Survivors of exactly one interval behaved equivalently to those that were in the middle of their duration, so these temporal doubletons were included in all range-size shift measurements.

Phanerozoic intervals were included only when successive bins contained ≥ 10 raster cells of the same habitat type inhabited by multiple surviving species. Species from the Pleistocene were not assessed because no modern occurrence data were included. More species crossed certain stage boundaries than others (Table A.4), which resulted in unequal variances among intervals. Therefore, I weighted each range-size shift by the number of survivors that contributed to that shift and compared the difference in the two groups with a two-sided Mann-Whitney (Wilcoxon rank-sum) test.

B.4.4 Mixed-effects models

I explored the nature of the relationship (i.e. estimated degrees of freedom) between change in species count and change in range size with a generalised additive mixed-effects model (GAMM). The additive component of the model fit a penalised thin plate regression spline on species count, implemented with the **mgcv** (Wood 2003) and **gamm4** (Wood and Scheipl 2017) packages in R. I inspected this spline to determine whether the smoothing term was justified, relative to a linear model. In addition, the identity of the time interval associated with each survivor was included as a random effect in the GAMM. I included this random effect because species from the same time bin were expected to pseudoreplicate each other due to shared environmental conditions and spatial sampling structure. Moreover, this random effect obviated the need to correct for multiple testing, because a single coefficient was calculated for each fixed effect over the entire time series, instead of for every time bin as in Finnegan, Anderson, et al. (2015), for example.

The estimates for the spline in the GAMM (degrees of freedom = 0.23) gave no justification to parameterise species count with a higher-order polynomial than a linear regression. Therefore, I performed model selection on linear mixed-effects models (LMM) rather than a more complicated regression structure. As in the GAMM, time interval was included as a random effect in the LMM. I considered the influence of the following six fixed effects: change in species count (integer), clade (categorical: Brachiopoda or Bivalvia), point in lineage duration (categorical: Early,

Middle, or Late), time bin age (continuous and positive), bathymetric preference (categorical: carbonate, siliciclastic, or both), and lithological preference (categorical: shallow, deep, or both). I then built a global linear mixed-effects model (all predictor variables included; Fig. 2.4 and Table A.7) and compared all 64 possible models, including the model with no fixed effects added (that is, a horizontal line). I implemented mixed-effects models with functions in the **lme4** package (Venables and Ripley 2002; Bates, Mächler, et al. 2015). Following default settings, I fit models based on restricted maximum likelihood for estimation purposes and based on maximum likelihood for testing purposes.

I estimated Akaike's Information Criterion (AIC) instead of corrected AIC, given the large number of species in the dataset, to determine the subset of models that received non-trivial weight ($\Delta\text{AIC} < 10$); for derivation and discussion of weights, see Burnham and Anderson (2010). Models are listed in order of relative performance in Table A.5. For fixed effects of these supported models, I recorded point estimates and the 95% confidence interval bounds from the likelihood profile (Table A.6). Log-likelihood estimates can be interpreted on the original scale of the variable. I chose to report log-likelihood instead of marginal likelihood estimates because (i) interpretation of coefficients is more straightforward and (ii) statistical theory to calculate marginal likelihood of parameters over the random effect(s) in mixed-effects models is poorly developed.

B.5 Power analysis

B.5.1 Justification

There are many reasons why a species might have had a larger geographic range than observed in the dataset for analysis: (i) fossils did not preserve in certain areas, (ii) fossiliferous sediment was eroded or overlain in certain areas, (iii) palaeontologists did not collect fossils in certain areas, or (iv) an iteration of the spatial subsampling procedure did not happen to select all occurrences of collected fossils for the species. The first three biases exist for almost all quantitative palaeobiology studies and have received intensive study already (e.g. Behrensmeyer and Kidwell 1985; Kiessling

2005; Patzkowsky and Holland 2012; Darroch, Laflamme, et al. 2018). The fourth source of underestimation in range size is specific to my novel analytical approach.

The spatial standardisation procedure of this study reduces the size of measured geographic distributions by truncating the area of observation. This compression could reduce the effect size of an underlying biological signal of ecological release. In addition, the subsampling procedure is probabilistic, and so measured range sizes for a given species in a given time interval vary from one iteration to the next. In light of these limitations, the possibility of type II error in analyses merits scrutiny.

B.5.2 Simulation of ecological release

I simulated ecological release at a range of magnitudes and frequencies to calculate the probability of observing release in a subsample of the simulated data. I modelled range expansion based on the Middle–Late Pleistocene record of bivalves in the northwest Atlantic. This was the interval with the densest occurrences (70 occupied grid cells), and so one would expect the compression from the subsampling procedure would be particularly severe. To control other biases on fossil range size, such as outcrop area, I used the same Pleistocene area before and after simulated release as the setting in which to calculate occupancy.

There were two experimental parameters in the simulation framework: the frequency with which species experience ecological release, and the percentage expansion of geographic range size that occurred for a species. I varied the frequency of release from 0–100% of species and the magnitude of release from 0–100% increase in occupancy for any given species. For each set of parameter values, I performed 10 simulations (4000 simulations in total). Considering a simulation where 50% of species expanded by 25%, for instance, I selected half of the observed species ($n = 423$) at random. For each of these 'released species,' I calculated an integer number of cells equivalent to 25% of current occupancy, and I generated that number of new occurrences for the species. These occurrences were placed in grid cells that were already occupied by other species, so as not to change the pool of cells available for spatial subsampling.

In the first set of simulation experiments, no subsampling procedure was applied to occurrences. In a second set of experiments, species' occupancy was subsampled before and after ecological release. From the latter simulations, I calculated the probability of observing release for a given magnitude and frequency of expansion. Subsampling consisted of selecting 15 occupied grid cells from which to calculate per-species range size (as in section B.3.3).

B.5.3 Power to detect release

When no subsampling procedure was used, species expanded as a linear function of the frequency and magnitude of ecological release, as intended (Fig. A.4A–C). When grid cells were rarefied prior to calculating occupancy, the same trend emerged, although mean occupancy was lower and varied more among iterations (Fig. A.4D–F). A signal of release appeared in every simulation where at least two-thirds of species expanded in occupancy by at least 50% from their initial value (Fig. A.4G). The same result held when at least half of species expanded by at least two-thirds of their initial occupancy. The probability of detecting release decreased at lower frequencies or magnitudes of expansion. Nevertheless, even when only half of species expanded by 25%, the probability of detecting this expansion reached 60% or more.

The simulations provide an estimate of statistical power for detecting release in species that cross a single time bin. The three prongs of my empirical analysis, however, quantified patterns across many species and time bins. Therefore, even if the subsampling procedure produced a false negative of release for certain species that crossed a stage boundary, a positive signal would still emerge in the final analysis. Furthermore, when characterising the relationship between time series of geographic range size and species count (section B.3.2), I employed a non-parametric correlation coefficient (Kendall's *tau*). Correlation strength then depended only on the direction of range shifts relative to species counts; the magnitude of range expansion during any intervals of ecological release would not affect correlation strength. In other words, although the spatial subsampling procedure might reduce the magnitude of observed geographic expansion in individual species relative to the

true change, type II error in analyses is unlikely. The high probability of detecting a signal of release in simulations gives confidence in my empirical results.

C

Supporting charts for Chapter 3

C.1 Figures

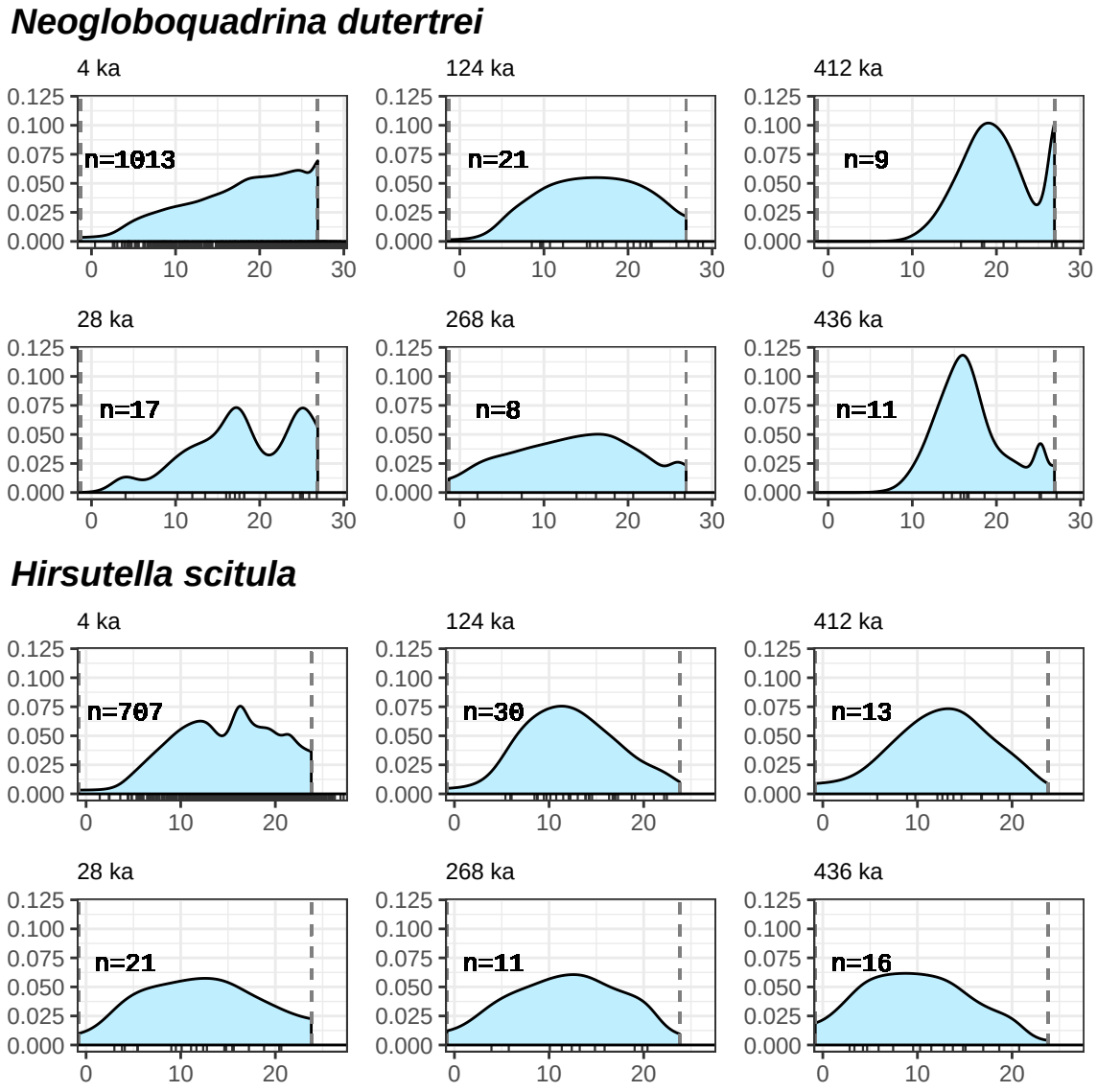


Figure C.1: Kernel density estimates for two species in three interglacial peaks (top row for each species) and three glacial maxima (bottom row for each species). Temperature was extracted from the depth habitat of each species: surface for *Neogloboquadrina dutertrei* and subsurface for *Hirsutella scitula*. The original observations used for kernel density estimation are plotted as a rug at bottom of each plot. To compare niches from different time intervals, all densities of a given habitat depth were cut (and rescaled) to the same upper and lower bounds, marked by gray dashed lines. The interval of comparison was −1.3–26.9 °C for surface species and 0.8–23.8 °C for subsurface species; the limits for all habitat depths are listed in the caption to Fig. C.2. Note the much larger sample size at 4 ka than any other age. The density estimate is correspondingly rougher for some species (e.g. *H. scitula*), which could artificially inflate the Hellinger distance between niches in the most recent time bin and older interglacial intervals. Chapter 3 section 3.5.8, 'Niche lability among extreme climate intervals,' discusses this; Figs. 3.3C and C.6B repeat analysis excluding the youngest interglacial (4 ka) and oldest glacial (668 ka).

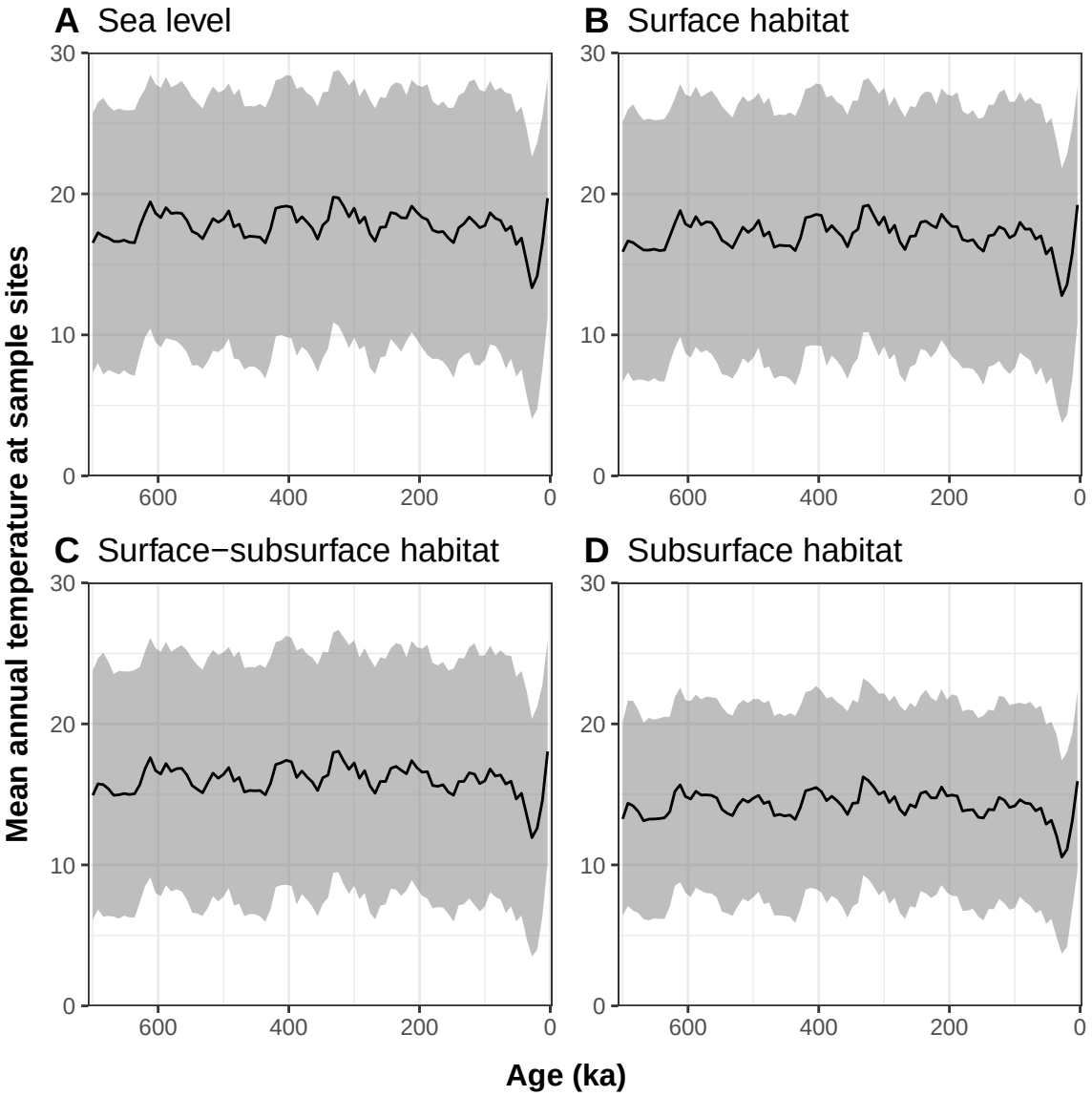
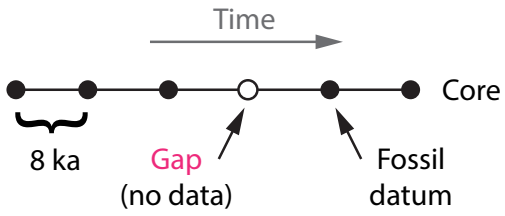
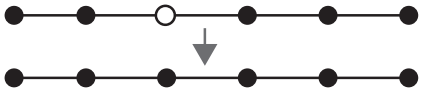


Figure C.2: Mean temperature across sampling sites is plotted as a line within an envelope of one standard deviation to indicate the spread of observed global temperatures. The threshold for maximum temperature of niche comparisons, in order of depth layers, was 27.1, 26.9, 25.9, and 23.8 °C. The threshold for minimum temperature was $-1.4, -1.3, -0.6$, and 0.8 °C, respectively. See Fig. C.3 and Chapter 3 section 3.5.7, 'Niche dissimilarity measure,' for discussion of threshold calculations.

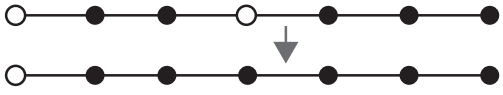
A. Each sediment core records nearly-continuous deposition over its length. Sometimes there are gaps in the fossil record reported from the core. Many gaps are due to failure to report a complete absence of fossils (or the presence of only one species) from sediment examined at that point.



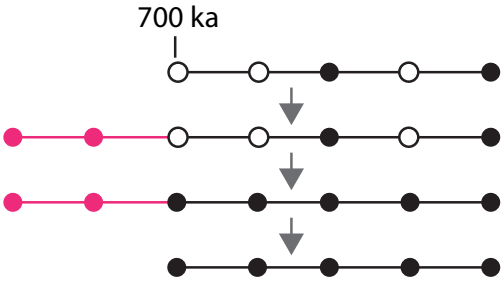
B. One can infer that sediment may have been examined (but not reported) by filling in the points that 'range through' an interval bracketed by observations.



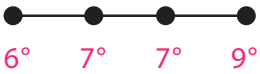
C. A range-through approach can only interpolate in the middle of a sequence, not extrapolate past the end-points. This can mean the ends of the series look less sampled than the middle.

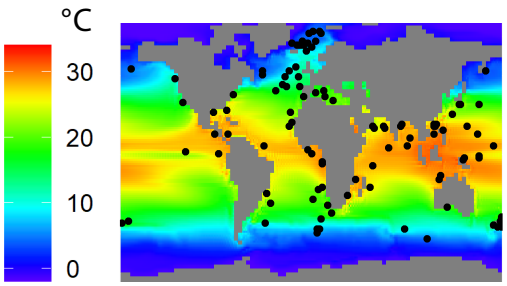


D. The young end of the sediment cores (near 4 ka) is very well sampled, but the old end (near 700 ka) is not. Therefore, for the range-through interpolation step, I added fossil and climate data from 800–700 ka, beyond the end of the study interval, so the amount of inference was spread equally across time intervals.



E. I combined the ages and coordinates of each core sample with Atmosphere-Ocean General Circulation Models to convert observations to estimated environmental conditions (mean annual temperature, MAT).





F. I compared core records from across the world to estimate the available and observable temperature range for species in each time bin. The segment of the temperature axis shared by all time bins is the standard interval across which species' niches can be fairly compared.

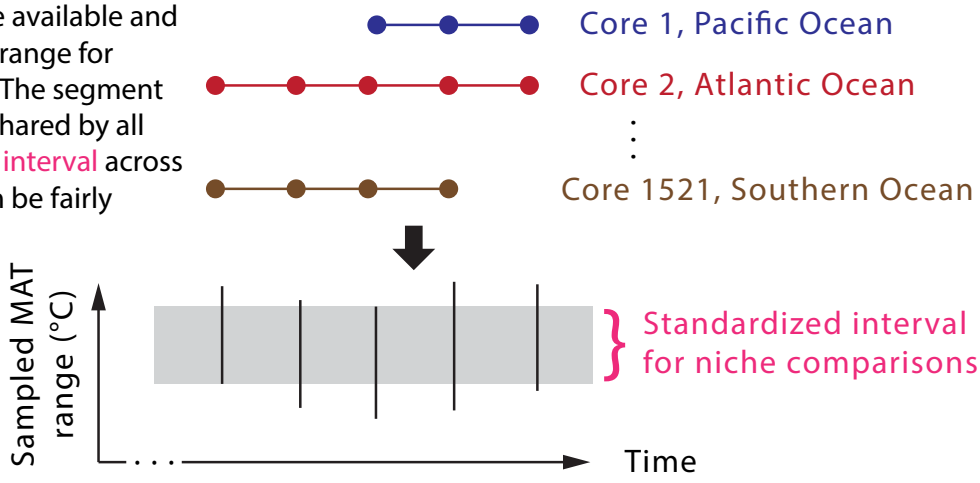


Figure C.3: This diagram explains the steps to standardise the temperature interval over which thermal niches were compared. The method uses range-through inference on all available foraminifera occurrence data (presence and absence records) to determine the breadth of environmental conditions (i.e. range in mean annual temperature) that was available and observable for species in each time bin. The comparison interval is the range of environmental conditions shared across every time bin (grey band in **F**). The bounds of the comparison interval were calculated separately for each depth habitat and are listed in the caption of Fig. C.2. Range-through interpolation adds only 'pseudo-absence' points to the estimation of sampling distributions, not presence points to species distributions. Most of the pseudo-absence and recorded absence sampling locations were from polar regions during interglacial intervals. Hence, by including all absence points in the definition of sampling distributions, the lower limit of niche estimation was extended to colder temperature values. One reason for gaps in reported data is samples that contain no species (as in present-day Antarctica) or only the polar species *Neogloboquadrina pachyderma*. A single species' abundance is useless for many palaeoclimate applications, so records for samples with single species might not have been entered even where sediment was present and observed.

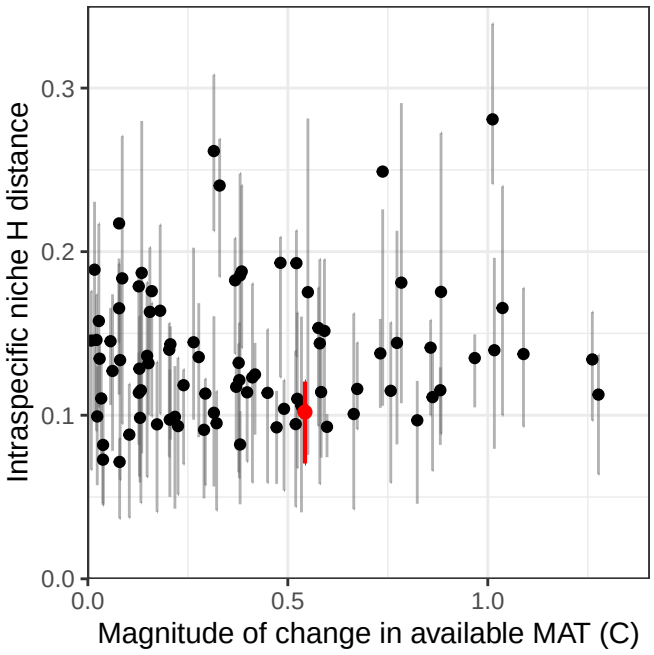


Figure C.4: Hellinger's H distance (mean and interquartile range among species) between niches in each pair of consecutive time bins is plotted against absolute differences in available temperature (surface Mean Annual Temperature averaged among global marine sites). Niches were constructed from temperature extracted within depth habitats (see Fig. 3.2 for sea-surface data). Correcting for temporal autocorrelation in the H time series, Pearson's r was 0.18 (95% CI = $[-0.03, 0.38]$). Without pre-whitening, $r = 0.098$ (95% CI = $[-0.12, 0.30]$). The most recent interval (12–4 ka ago) is plotted in red.

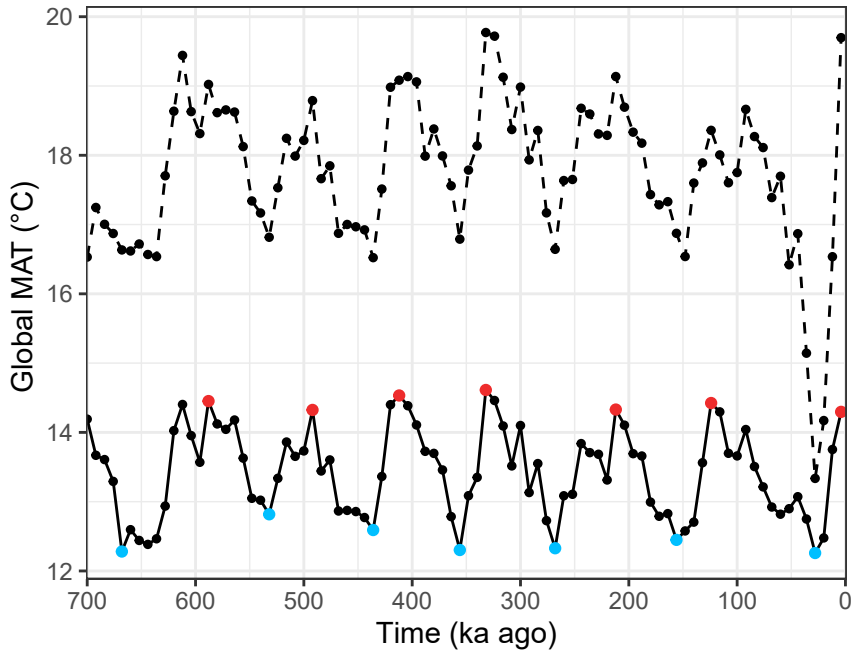


Figure C.5: Time series of mean annual temperature (MAT) averaged over either a global marine grid of 10°-resolution (solid line) or all sampled sediment core sites (dashed line). Blue points identify glacial maxima, and red points highlight interglacial warm peaks (ages listed in Table C.2).

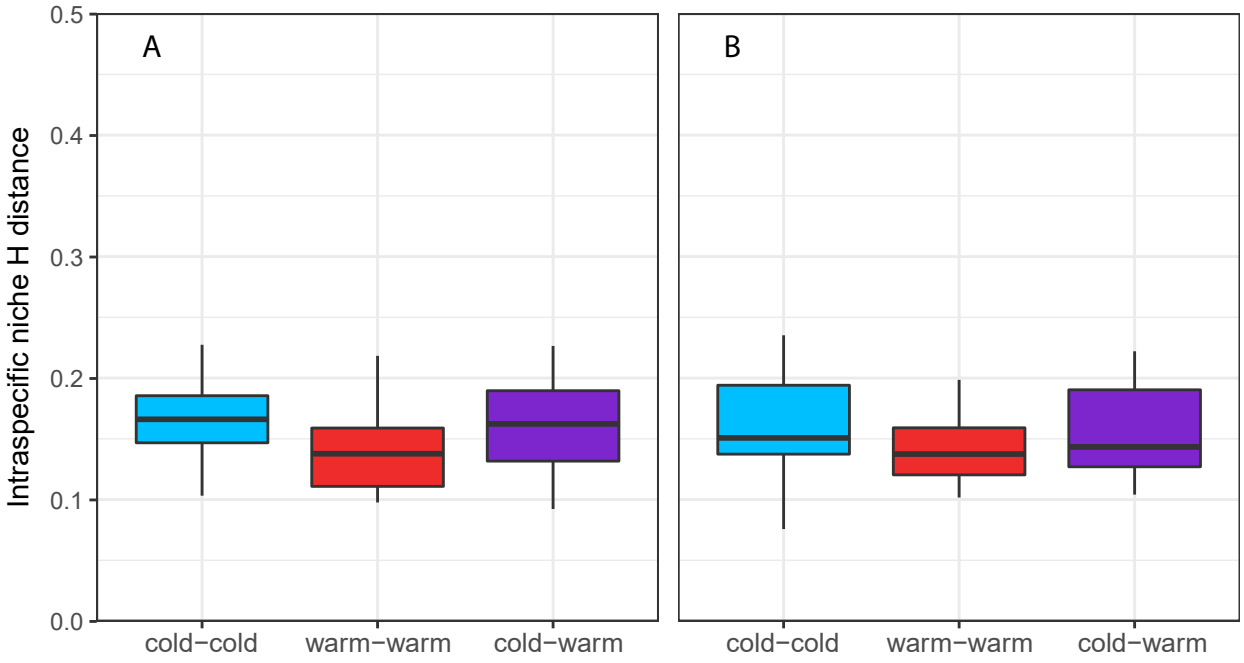


Figure C.6: Hellinger's H distances were computed for (i) each different pair of glacial periods ('cold–cold'), (ii) each different pair of glacial periods ('warm–warm'), or (iii) each different pair of one glacial and one interglacial period ('cold–warm'). Temperature estimates were extracted in species' habitat depths (see Fig. 3.3 for sea-surface data). Boxplots depict the variation in mean species' H among all pairs of time bins in each category. In **A**, calculations were based on all 7 glacial and interglacial intervals (Table C.2). In **B**, the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were omitted. The exceptionally dense and sparse sampling in these intervals, respectively, could influence H as described in Chapter 3 section 3.5.8, 'Niche lability among extreme climate intervals'.

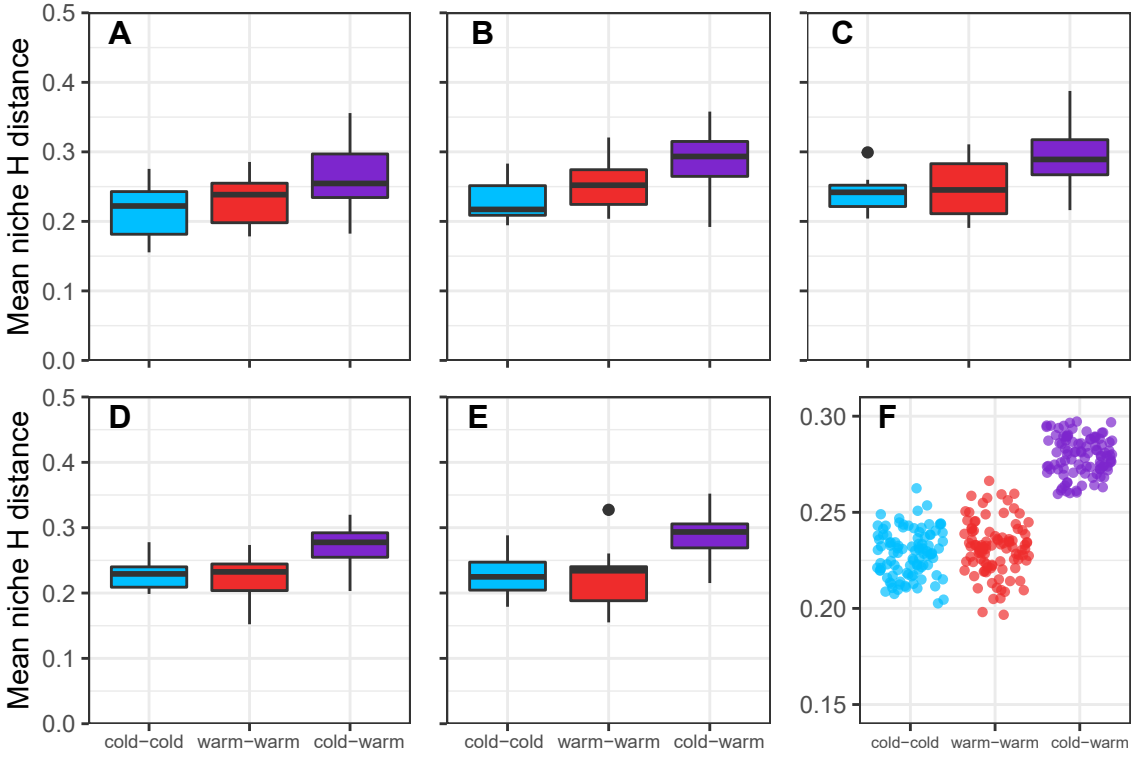


Figure C.7: (A–E) The first 5 (of 100) replicates of mean Hellinger's H distance comparisons simulated under a hypothesis of strict geographic stasis (i.e. complete niche adaptation to local conditions during environmental change). Boxplots are as in Figs. 3.3 and C.6: H was averaged among all species sampled in a given pair of time bins, and then averaged among all time bin pairs in each of three climate-comparison categories. In **F**, the mean for each climate-comparison category from every simulation replicate is plotted as a point, e.g. the middle line in each boxplot from **A**–**E** appears again as a point in **F**. Note the expanded y-axis in **F**. In each of 100 simulations, the mean H distance was larger on average between a glacial and interglacial interval ('cold–warm') than between two glacial intervals ('cold–cold') or two interglacial intervals ('warm–warm'). The simulation steps are detailed in Chapter 3 section 3.5.8, 'Test for type II error'. Temperature values were extracted from the sea surface, and the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were excluded from comparison, as in Fig. 3.3C.

C.2 Tables

Species	Ecotype	Ecotype ref.	Depth	Depth ref.	H	Optimum
Beella digitata	Subthermocline	Coxall et al. 2007	Subsurface	Schiebel and Hemleben 2017	0.10	22.6
Globigerina bulloides	Mix layer heterotroph	Vergnaud-Grazzini 1976, Durazzi 1981	Surface-subsurf	Rebotim et al. 2017	0.09	13.0
Globigerina falconensis	Mix layer symbiotic	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.13	16.8
Globigerinella calida	Thermocline	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.16	20.2
Globigerinella siphonifera	Thermocline	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.14	19.8
Globigerinita glutinata	-		Surface-subsurf	Rebotim et al. 2017	0.09	13.5
Globigerinoides conglobatus	Mix layer symbiotic	Keller 1985	Surface	Watkins et al. 1996	0.19	24.2

Globigerinoides ruber	Mix layer symbiotic	Keller 1985, Pearson et al. 2001, Pearson and Shackleton 1995	Surface	Rebotim et al. 2017	0.15	20.5
Globoconella inflata	Thermocline	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.10	13.2
Globoturborotalita rubescens	Mix layer heterotroph	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.12	22.3
Globoturborotalita tenella	Mix layer heterotroph	Aze et al. 2011	Surface	Rebotim et al. 2017	0.12	23.7
Hirsutella hirsuta	Subthermocline	Aze et al. 2011	Subsurface	Rebotim et al. 2017	0.16	16.5
Hirsutella scitula	Subthermocline	Shackleton and Vincent 1978, D. R. M. Stewart unpublished data	Subsurface	Rebotim et al. 2017	0.10	14.7

Menardella menardii	Thermocline	Shackleton and Vincent 1978, Pearson and Shackleton 1995, D. R. M. Stewart unpublished data	Surface	Watkins et al. 1996	0.17	23.3
Neogloboquadrina dutertrei	Thermocline	Kahn 1979, Shackleton and Vincent 1978	Surface	Rebotim et al. 2017	0.16	19.5
Neogloboquadrina incompta	High latitude	Darling et al 2006	Surface-subsurf	Rebotim et al. 2017	0.10	9.0
Neogloboquadrina pachyderma	High latitude	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.08	7.7
Orbulina universa	Mix layer symbiotic	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.13	16.4
Pulleniatina obliquiloculata	Thermocline	Shackleton and Vincent 1978, Pearson and Shackleton 1995	Surface	Rebotim et al. 2017	0.22	25.3

Trilobatus sacculifer	Mix layer symbiotic	Keller 1985, D. R. M. Stewart unpublished data	Surface	Rebotim et al. 2017	0.18	22.4
Trilobatus trilobus	Mix layer symbiotic	Keller 1985, Pearson et al. 2001, Pearson and Shackleton 1995	Surface	Loncaric et al. 2006	0.17	20.1
Truncorotalia crassaformis	Subthermocline	D. R. M. Stewart unpublished data	Surface-subsurf	Ezard et al. 2015, Rebotim et al. 2017, Schiebel 2017, Meilland et al. 2019	0.15	18.5
Truncorotalia truncatulinoides	Subthermocline	Vergnaud-Grazzini 1976, Shackleton and Vincent 1978	Surface-subsurf	Rebotim et al. 2017	0.14	15.7
Turborotalita quinqueloba	High latitude	Pearson and Wade 2009	Subsurface	Rebotim et al. 2017	0.11	10.9

Table C.1: Attributes are listed and referenced for each species in the dataset. H is the mean Hellinger's distance between pairs of niches in consecutive time bins; $H = 0$ if identical. Optimum is the mean annual temperature (°C) associated with the largest probability of occurrence, averaged over the time bins in which the species was sampled.

Table C.2: Median age of time bins (ka) associated with climatic extremes. In Figs. 3.3A and C.5, glacial maxima are marked in blue and interglacial peaks are marked in red. Interglacial onset estimates are from Lisiecki and Raymo (2005).

Glacial maximum	Interglacial peak	Interglacial onset
28	4	14
156	124	130
268	212	243
356	332	337
436	412	424
532	492	533
668	588	621

D

Kernel density estimation methods for Chapter 3



D.1 Overview of KDE and kerneval

D.1.1 Why use kernels?

Kernel density estimation (KDE) is a foundational method in nonparametric statistics. Practical applications abound, from ecology to engineering. There are also theoretical applications of kernels to other statistical methods, such as spline smoothing and machine learning. KDE has achieved its prominence because the basic method is a straightforward, elegant, and effective way to generate a smooth probability distribution function (PDF) from a finite vector of numeric observations. Many R users will be familiar already with kernel-smoothed data visualisations, such as `ggplot2::geom_density()`.

D.1.2 How does it work?

One can find a brief summary of KDE in nearly any basic statistics textbook (e.g. Wasserman (2006)). Silverman (1986) gives a more thorough introduction to the topic and writes for accessibility towards readers at a range of experience levels. Many explanations of KDE start with histograms, which are a simple variation on KDE wherein the observations are binned. Graphically, a probability function can be visualised by stacking a rectangle in the bin containing each observation. The width of the rectangle along the x-axis is the 'bandwidth' in KDE terminology. The height is scaled such that the probability sums to unity across a given interval. One need not bin the axis, however; one could instead place a rectangle directly at the x-value of each observation. The result is a step-wise function (the summed height of all rectangles at any given x-value), which is more jagged than a regular histogram and is called a naive estimator. One step further, one could replace rectangles with Gaussian curves (or triangles, or another distribution for the kernel). These options are available in the `density()` function from the base **stats** package. There are more than 100 other R packages that also involve KDE. For a comparison of the functionality and performance of 15 major KDE packages, a starting place is Deng and Wickham's (2011) review.

D.1.3 When should one use kerneval?

Most density estimation methods assume that the probability of observing an event (independent of the event occurring) is constant across the interval of estimation. However, many empirical datasets violate this assumption. For instance, in some cases the observation rate increases linearly as a function of x, so-called length bias or size bias.

kerneval implements methods to account for observation bias (also called selection bias) when estimating probability densities. The package can accommodate length-biased data or data with more complicated bias

functions defined for any x in the study interval. The estimation steps will not eliminate bias, but they will reduce it, and the error of the estimate will decrease with sample size—as is the case for most statistical methods for biased data.

D.2 Mathematical framework

The point of KDE is to estimate a probability density function, $f(x)$. The estimate is notated $\hat{f}(x)$. Without selection bias (a case equivalent to a uniform bias function), observations are drawn directly from distribution F. However, when there exists selection bias, observations are drawn from G, which can be described as the scaled product of $f(x)$ and a weight function, $w(x)$:

$$g(x) = \frac{w(x)f(x)}{\mu_w} \tag{1}$$

$$\text{where } u_w = \int w(u)f(u)du < \infty.$$

Note that this requires $w(u) > 0$ for all u. In the case of length-biased data, $w(x) = x$.

Many statisticans attribute the first treatment of the length-bias problem to Cox (1969), who investigated an example in textile fibre manufacturing. Cristobal and Alcala (2001) trace the ideas to earlier works by Fisher and Rao, however. Vardi (1982) proved that Cox's approach was the nonparametric maximum likelihood estimator. Jones (1991) found a kernel density estimator for this NPMLE, which has been widely adopted and extended. Borrajo and others (2017) developed a bootstrapping approach to bandwidth selection for Jones' weighted KDE. Barmi and Simonoff (2000) evaluated an alternative approach: transforming the biased data, conducting classical KDE, and then back-transforming and rescaling the density. **kerneval** contains tools for both the weighting and transformation approaches to calculate $\hat{f}(x)$ based on a random sample of n observations drawn from G.

D.3 Transformed density estimation

Barmi and Simonoff (2000) proposed an elegant strategy to estimate density functions from selection-biased samples. The premise is that one may transform the empirical sample to an unbiased space, perform any type of established density estimation, and then back-transform to the original scale to interpret the estimate. This approach has been described as 'adjust-first, smooth-later' because there is no need to modify the density

estimation step. For instance, one could perform KDE with a Gaussian kernel and rule-of-thumb bandwidth, so long as KDE is done in the transformed space. The alternative approach involves weighting the kernels (described below) to adjust the estimation concurrent with smoothing.

Continuing with the mathematical notation of Eq. 1, let $X_1, X_2, ..., X_n$ be a random sample of n observations from G. To transform these empirical data, calculate the cumulative distribution function of the bias function, $W(x) = \int_0^x w(u)du$. Define $Y_i = W(X_i)$ for $i = 1, 2, ..., n$. To estimate the true probability distribution function f,

1. Perform density estimation on the sample $Y_1, Y_2, ..., Y_n$
2. Back-transform from the y to x axis, using $W^{-1}(y)$
3. Scale the density by a constant ($\hat{\mu}$) such that the function integrates to unity

The `transdens()` function evaluates these steps, calling the `inverse()` function from **GoFKernel** (Pavia 2015) to estimate the inverse cumulative distribution function, W^{-1}.

Depending on the shape of the true PDF and the bias function, analysts would be wise to inspect kernel density plots on the transformed scale, just as one might plot estimates (on the original scale) when selecting a bandwidth. In particular, one should consider whether the transformed distribution has a long tail or otherwise is difficult to estimate. If the density estimation problem seems more straightforward on the original than transformed scale, one could weight the KDE instead of transforming the data. Example 1 below illustrates these trade-offs. As another alternative, the original authors of the transformation method also suggest a local quadratic likelihood density estimator for distributions where the tails may be difficult to estimate. Kernels and local likelihood estimators can each vastly outperform the other in different contexts (Hall and Tao 2002). **kerneval** implements only KDE methods, but includes the option for boundary reflection (`reflect = TRUE`) to mitigate edge effects at the tails.

D.4 Weighted density estimation

The weighting operation of Jones (1991) is the KDE method for selection-biased data that has received the most attention in the statistics literature. The method is a slight modification of classical KDE and does not add onerous calculation, so it is an attractive approach. However, the choice of bandwidth is complicated and potentially time-consuming (e.g. bootstrap selection, described below).

To account for selection bias, the kernel on each datum is weighted by the inverse of the observation probability at that point. An informal, geometric way to visualise this is to imagine a one-dimensional axis with several

Gaussian kernels along it, one at each datum. In regions of the axis with higher intensity of observation, pinch off some of the tops of the kernels. Then, add this area to the tops of the kernels in regions where observation is less probable. In a formal mathematical framework, weighting can be expressed by adding a scaling factor to the estimated density function. The classical definition of the KDE is:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i}^{n} K\left(\frac{x - x_i}{h}\right)$$

where K is a kernel function and h is bandwidth.

The weighted estimate is:

$$\hat{f}(x) = \frac{1}{n} \hat{u} \sum_{i}^{n} \frac{1}{w(x_i)} K\left(\frac{x - x_i}{h}\right)$$

$$\text{where } \hat{u} = \left[\frac{1}{n} \sum_{i} \frac{1}{w(x_i)}\right]^{-1}.$$

Weighted kernel estimation should not be confused with adaptive kernel estimation, which is most relevant when a dataset is large and the probability function is rougher in some areas than others. (The interested reader would do well to consult Wasserman (2006) chapter 9.10, 'Do adaptive methods work?') Both approaches modify the individual kernels that contribute to an estimate. However, in adaptive KDE the badwidth of each kernel is adjusted, whereas in weighted KDE the bandwidth is constant while the height (total probability density) of each kernel is adjusted.

D.4.1 Bootstrap bandwidth selection

The internal `selectbw()` function in **kerneval** selects a bandwidth following the bootstrap method of Borrajo and others (2017). The theory behind bandwidth selection can be summarised as a series of three, potentially iterative steps. Luckily, in practice, there is a computational shortcut to find the bandwidth h^* that minimises bootstrapped mean integrated squared error (BMISE).

1. Pick an initial ('pilot') bandwidth λ and kernel L. Smooth the empirical distribution F_n with these parameters to get $\hat{f}_\lambda$. Bose and Dutta (2013) recommended the use of a Gaussian kernel with $\lambda = \frac{1}{8} n^{-1/9}$.
2. Resample n observation from $\hat{f}_\lambda$, for B replicates. This is equivalent to resampling with replacement from F_n, then adding ϵ to each resampled set of observations, X^*. The ϵ are i.i.d. from $\frac{K(\cdot/\lambda)}{\lambda}$. (This step is not run in practice.)
3. Calculate the MISE of the bootstrap samples (section D.6.3). $BMISE(h, \lambda) = \frac{1}{B} \sum_{i=1}^{B} \int [f^*_{h,i}(x) - \hat{f}_\lambda(x)]^2 dx = V^*(h) + B^{*2}(h)$.

The variance and bias terms of BMISE can be expressed in terms of K_h and K_λ convolution, which do not depend on X^*. Therefore, it is possible to find the h^* that minimises BMISE without actually drawing bootstrap samples. One must still pick a pilot bandwidth, however.

D.5 Examples

This vignette gives examples that examine the effect of sample size and compare the performance of weighting vs. transformation approaches to KDE on selection-biased data. The data in the first example are length-biased, while the data in the second example are generated according to a custom (fourth-order polynomial) bias function.

D.5.1 Length bias and inverse length bias

The first example replicates the simulation study of Barmi and Simonoff (2000). We will draw biased samples from known distributions and then estimate PDFs with transformed vs. weighted kernels. The simulations make use of the convenient result that when sampling is length biased, $w(x) = x$, samples drawn from a true distribution of χ^2_k (chi-square distribution with k degrees of freedom) will come from an observed distribution (G) of χ^2_{k+2}. If there is inverse length bias, $w(x) = 1/x$, then G is χ^2_{k-2}.

There are eight simulation scenarios in total. The first four scenarios contain length-bias, and the true PDF F is set as a chi-square distribution with either 2 or 12 degrees of freedom. (Thus, G is a chi-square distribution with 4 or 14 degrees of freedom, respectively.) For each PDF, sample size is set at either 50 or 200 observations.

```
nVals <- c(50, 200)
kValsLin <- c(2, 12)
wLin <- function(x){ x }
degLin <- function(k){ k + 2 }
```

The remaining four scenarios contain inverse-length-bias, and F is a chi-square distribution with either 3 or 16 degrees of freedom. (Thus, G is a chi-square distribution with 1 or 14 degrees of freedom, respectively.) As before, $n = 50$ or 200.

```
kValsInv <- c(3, 16)
wInv <- function(x){ 1/x }
degInv <- function(k){ k - 2 }
```

The original study replicated each scenario with 500 samples. Depending on processing speed, and without parallelising calculations, the full simulation and estimation could run for an hour or more. For a quick exploration of the example, one could run fewer iterations.

```
nreps <- 25 # 500
```

To compare the performance of weighted vs. transformed KDE in each scenario, the standard approach is to calculate the mean integrated squared error, MISE. First, draw a sample of n observations from a chi-square distribution with given degrees of freedom. Then, calculate $\hat{f}$ with each KDE method. Finally, integrate the squared difference between the estimated and true PDF over the study integral. The ultimate goal is to calculate the *expectation* of this integrated error, however, so it is helpful to wrap all of these steps into a function that can be iterated. The mean ISE will converge on the MISE with more replicates.

```
miser1 <- function(w, n, k, deg){
  df <- deg(k)
  x <- rchisq(n = n, df = df)

  # weighted KDE
  wtEst <- wdens(x, w)
  wtFun <- approxfun(wtEst$x, wtEst$y)
  wtSE <- function(x) (wtFun(x) - dchisq(x, df=k))^2
  wtISE <- integral(wtSE, min(wtEst$x), max(wtEst$x))

  # transformed KDE
  transEst <- transdens(x, w)
  transFun <- approxfun(transEst$x, transEst$y)
  transSE <- function(x) (transFun(x) - dchisq(x, df=k))^2
  transISE <- integral(transSE, min(transEst$x), max(transEst$x))

  c(wtISE, transISE)
}
```

There are many ways one could iterate through each sample and simulated distribution. Here are nested for-loops to produce the results in the same format as the MISE values from Table 1 of Barmi and Simonoff (2000).

```
# tbl1a <- tbl1b <- data.frame()
# for (n in nVals){
#  for (k in kValsLin){
#    mises <- replicate(n = nreps,
#                       miser1(w = wLin, n = n, k = k, deg = degLin)
#                       )
#    miseAvg <- rowMeans(mises)
#    iter <- data.frame(k, n, t(miseAvg))
#    tbl1a <- rbind(tbl1a, iter)
#  }
#  for (k in kValsInv){
#    mises <- replicate(n = nreps,
#                       miser1(w = wInv, n = n, k = k, deg = degInv)
#                       )
#    miseAvg <- rowMeans(mises)
#    iter <- data.frame(k, n, t(miseAvg))
#    tbl1b <- rbind(tbl1b, iter)
#  }
# }
```

Average ISE values from 500 replicate estimates of MISE are listed in Table D.1 and agree with the original trends reported by Barmi and Simonoff, reproduced in Table D.2. KDE is more accurate with more symmetric distributions (chi-square distributions with 12 or 16 degrees of freedom instead of 2 or 3). In addition, the transformation approach performs better for inverse-length-biased data than for length-biased data, because the transformed axis in the former case (a log-x scale) is a 'generally favourable' estimation problem.

Table D 1: Replication of published MISE results, using wdens() and transdens() functions in kerneval

	k	n	Weighted	Transformed
w(x)=x				
	2	50	0.0182	0.0507
	2	200	0.0154	0.0521

k	n	Weighted	Transformed
12	50	0.0022	0.0029
12	200	0.0009	0.0012
w(x)=1/x			
3	50	0.0562	0.0130
3	200	0.0601	0.0048
16	50	0.1535	0.0022
16	200	0.0371	0.0007

Table D 2: Barmi and Simonoff (2000) Table 1, using weighted KDE and transformed local likelihood estimators

k	n	Weighted	Transformed
w(x)=x			
2	50	0.0312	0.0309
2	200	0.0155	0.0317
12	50	0.002033	0.00434
12	200	0.000891	0.00184
w(x)=1/x			
3	50	0.0155	0.00568
3	200	0.00744	0.00216
16	50	0.00179	0.00105
16	200	0.000691	0.000293

Although relative patterns in Tables D.1 and D.2 match, exact values of respective entries differ. At the second and third significant digits, some variation seems to occur due to stochasticity in the replicates. Perhaps 1000 replicates would be more suitable to approximate the MISE. Additional discrepancies for weighted kernel MISE might be due to differences in bandwidth selection method. The authors chose the bandwidth 'in each replication to minimise ISE,' but this leaves some ambiguity. In the code above, bandwidth was selected with a bootstrap estimate and rule-of-thumb pilot. For the transformed estimates, the authors used a local quadratic likelihood density estimator, while here transformed kernels are used instead.

D.5.2 Polynomial bias function

Length bias is the type of selection bias that appears most frequently in the statistical literature. However, there are many situations where observation probability could behave in a more complicated way, perhaps non-monotonically. The **kerneval** user may work with any bias function that is positive and continuously defined over the study interval.

The present example is based loosely on an empirical dataset of marine species' ecological niches. The x-axis represents a resource or environmental gradient (e.g. temperature), and the PDF is the probability of a species' occurrence for a given resource amount/environmental state. Species tend to achieve peak abundance at an optimum condition, and to become rarer in habitats farther from the optimum. The true PDF F in this example is a chi-square distribution with 5 degrees of freedom, representing a species with a broad tolerance.

Unfortunately for ecologists, it is impossible to measure F directly. Instead, one might have the spatial coordinates where a focal species has been observed, along with the environmental conditions at those locations. KDE on these data will provide an estimate $\hat{f}$ of the species' niche along the given habitat axes. Unless the project managers took great care with the original sampling design, however, the data will be biased by the relative availability of habitat, the way field researchers took observations, or other spatio-temporal effects. (For instance, GPS trackers tend to lose satellite connection when an animal wanders into dense vegetation, leading to a larger observation probability in open terrain (Frair et al. (2010)).) The bias function in this example is a 4th-order polynomial, defined over the interval [0,10]. This function reflects (1) the larger surface area of temperate zones relative to the equatorial or polar zones, (2) asymmetries between the Northern and Southern hemisphere, and (3) the most common routes of research vessels.

```
w <- function(x){
  x <- 0.3*(x - 5.5)
  -0.8 * x^4 + 0.7 * x^3 + 1.5 * x^2 - 2 * x + 2
}
g <- function(x){
  g1 <- function(x) w(x) * dchisq(x, df=5)
  u <- integral(g1, 0, 10)
  w(x) * dchisq(x, df=5) / u
}
```

As in Example 1 above, the MISE is a useful way to compare the performance of different KDE methods. The weighted and transformed approaches both reduce the estimation error compared to the base **stats**

method, which does not consider selection bias. This result is unsurprising; it is more interesting to examine the magnitude of the improvement, and how the performance depends on sample size.

```
miser2 <- function(g, n){
  x <- random.function(n = n, g, lower=0, upper=10)

  # weighted KDE
  wtEst <- wdens(x, w)
  wtFun <- approxfun(wtEst$x, wtEst$y)
  wtSE <- function(x) (wtFun(x) - dchisq(x, df=5))^2
  wtISE <- integral(wtSE, min(wtEst$x), max(wtEst$x))

  # transformed KDE
  transEst <- transdens(x, w)
  transFun <- approxfun(transEst$x, transEst$y)
  transSE <- function(x) (transFun(x) - dchisq(x, df=5))^2
  transISE <- integral(transSE, min(transEst$x), max(transEst$x))

  # KDE without accounting for selection bias
  baseEst <- density(x, cut=0)
  baseFun <- approxfun(baseEst$x, baseEst$y)
  baseSE <- function(x) (baseFun(x) - dchisq(x, df=5))^2
  baseISE <- integral(baseSE, min(baseEst$x), max(baseEst$x))

  c(wtISE, transISE, baseISE)
}

# iseSmolN <- replicate(10, miser2(g = g, n = 50))
# round( rowMeans(iseSmolN), 4)

# weighted, transformed, biased
#> [1] 0.0082 0.0059 0.0169
```

If one repeats this exercise with $n = 500$, both the weighted and transformed estimates improve several times

over, but the uncorrected estimate does not.

```
# iseBigN <- replicate(10, miser2(g = g, n = 500))
# round( rowMeans(iseBigN), 4)


# weighted, transformed, biased
#> [1] 0.0029 0.0018 0.0127
```

Although MISE is a useful metric, as a single number it cannot indicate the manner in which the estimate errs along the study interval. A graphical approach provides this complementary information, as demonstrated in Figs. D.1 and D.2. The weighted (in orange) and transformed (in red) kernel estimates approach the true density as sample size increases from 50 to 500 observations. The weighted estimate undersmoothes at both sample sizes, however. The uncorrected estimate (in blue) follows the selection-biased PDF (in grey) regardless of sample size.

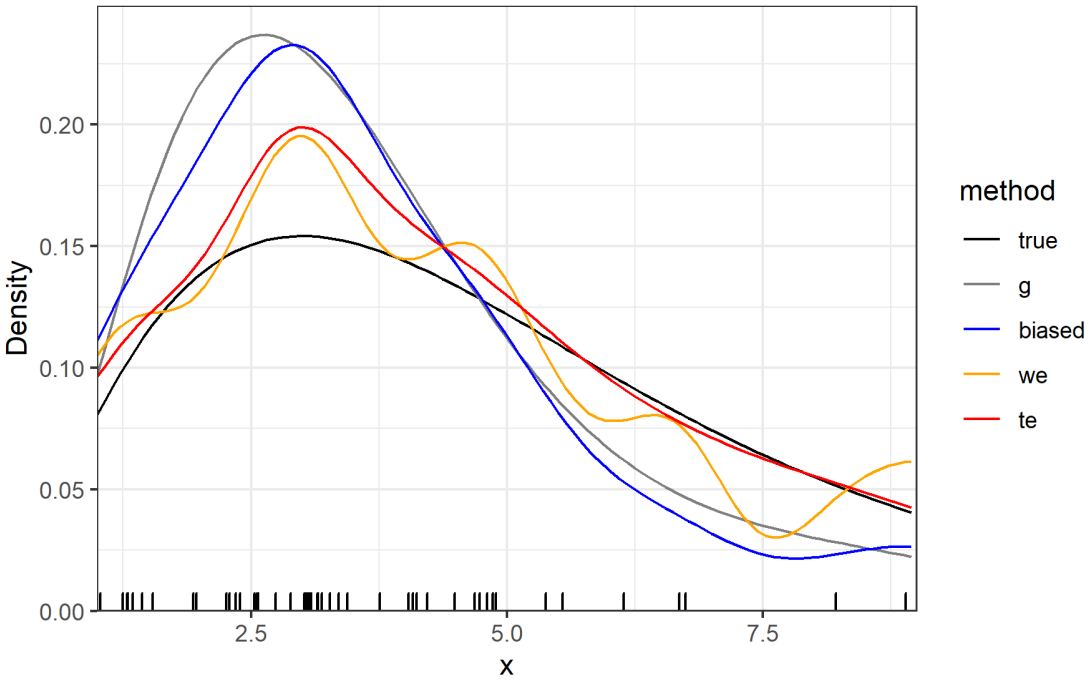


Figure D 1: The true PDF is a chi-square distribution with 5 degrees of freedom is and is plotted in black. The distribution from which observations are drawn is in grey and reflects the influence of selection bias. A random sample of 50 observations from the grey PDF is plotted as a rug at bottom. The kernel density estimates based on the sample are drawn in colour: the blue curve is a standard KDE, the orange curve is a weighted KDE, and the red curve is a transformed KDE.

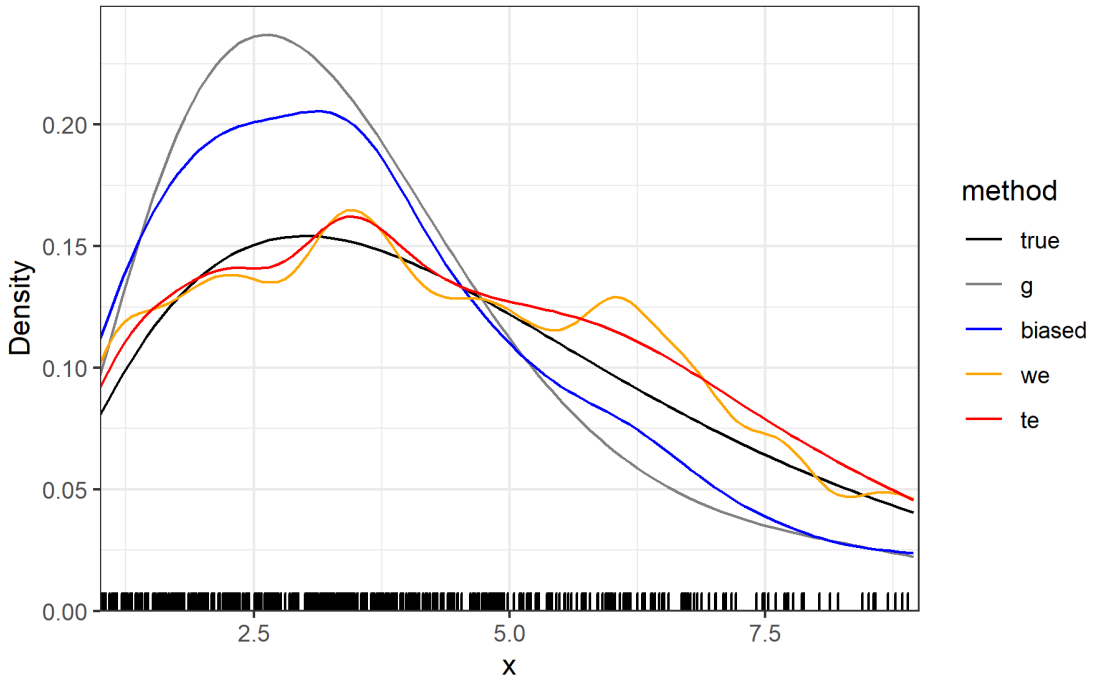


Figure D 2: The PDF colours are as in Fig. D.1. The sample size in this case is 500 observations instead of 50.

D.6 Technical details

D.6.1 Approximating functions and integration

In R, probability density estimates are evaluated at discrete points. However, to work with the probability density (e.g. calculate the probability over an interval or the overlap between distributions), one usually needs a description of the probability density as a continuous function; point estimates on their own have little use besides plotting. One could use a linear (`stats::approxfun()`), spline (`stats::splinefun()`), or polynomial function for approximation. Due to the mostly concave-down, unimodal shape of many niches, linear interpolation tends to under-estimate the probability when integrated compared to spline interpolation. However, the difference is small, and splines can lead to negative probability values in niche tails. **kerneval** interpolates linearly between adjacent point estimates. The default discretisation ($N = 2^9$ for `base::density()`) gives good behaviour in the author's experience, but one could specify a different resolution. There is a trade-off between scale and computation time.

Analagously, one could use any of a variety of functions to test that the PDF integral sums to unity. Common approximations of the integral include rectangular, trapezoidal, Gauss-Kronrod, or Simpson adaptive numerical integration. In the author's experience with simulated and empirical data, all integration options are equivalent

to at least 3 significant figures, so the default `stats::integrate()` implementation is sufficient for ecological analysis.

Comparison with other R packages

Both the **GoFKernel** and **ecospat** packages integrate PDFs to rescale the density estimate. The `density.reflected()` function in **GoFKernel** (Pavia 2015) approximates the PDF integral with rectangles, as illustrated in Fig. D.3. For comparison, a trapezoial approximation treats the point estimates of the PDF as the top vertices of adjacent trapezoids (Fig. D.4). Where the curve is concave (in the middle), trapezoids under-estimate the integral. Where the curve is convex (in the tails), trapezoids over-estimate the integral. In general, over the entire distribution, trapezoids approximate functions well.

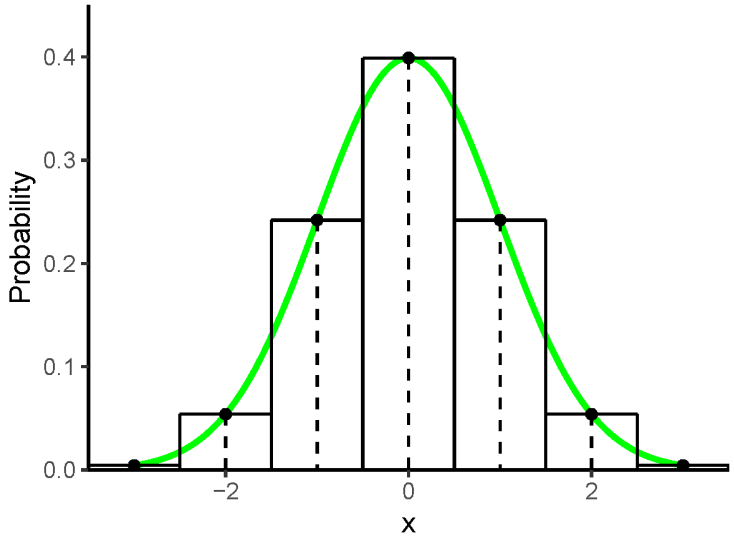


Figure D 3: The green curve traces a normal distribution with $\mu = 0$ and $\sigma = 1$, over the interval $[-3, 3]$. The PDF is estimated at 7 equidistant points (dashed lines) for a rectangular approximation of the integral over the interval. Each rectangle is centred on a point estimate. In the **GoFKernel** implementation, the estimate includes both endpoints of the KDE, which in this example extends the integration to the interval $[-3.5, 3.5]$.

One can quickly calculate the true integral of the normal function on the interval $[-3, 3]$ (the green curve in Figs. D.3 and D.4):

```
pnorm(3) - pnorm(-3)
#> [1] 0.9973002
```

The rectangular estimate (Fig. D.3) is 0.9997294 and the trapezoidal estimate (Fig. D.4) is 0.9952975. Increasing the discretisation from 7 points to 100, the estimates improve to 0.9975607 and 0.9972921, respectively. Recall that the default discretisation in `base::density()` is $N = 512$.

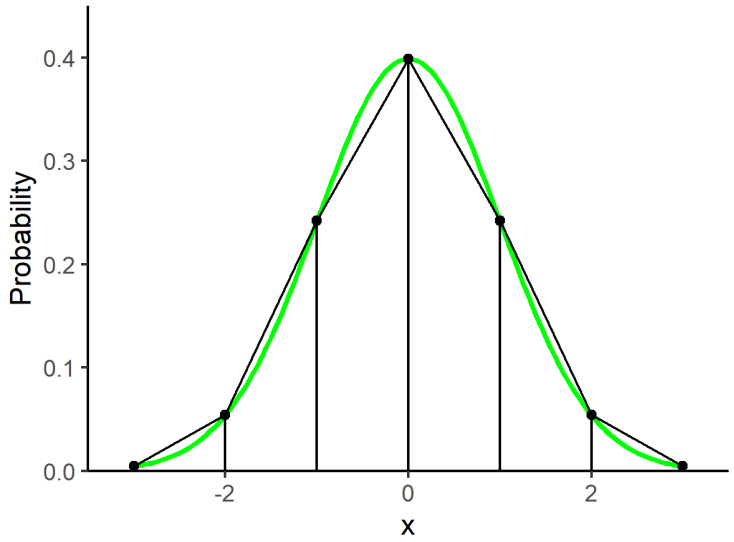


Figure D 4: The green curve is as in Fig. D.3. The PDF is estimated at 7 equidistant points (vertical lines) for a trapezoidal approximation of the integral over the interval $[-3, 3]$.

The **ecospat** package does not estimate integrals with polygons, but rather treats the PDF as a probability mass function (Di Cola et al. 2017; Broennimann, Di Cola, and Guisan 2020). (To review, a PDF describes probability along a continuous axis, whereas a PMF describes probability at discrete intervals. For a PDF, the probability of an event occurring in a given interval is the integral of the PDF over that interval, and the probability at any single point is zero. For a PMF, probability is defined only at individual points, and the sum of the function across all categories/discrete points is unity.) The `ecospat.grid.clim.dyn()` function estimates the density of a species' niche (`sp.dens$y`) at 100 discrete points, and re-scales the estimate at those points by the sum of all discrete density estimates. The result is scaled again by the number of observations for the species (`nrow(sp)`) and then by the maximum estimate.

```
# z <- sp.dens$y * nrow(sp)/sum(sp.dens$y)
# z.uncor <- z/max(z)
```

D.6.2 Boundary reflection

For some density estimation problems it may be desirable to reflect data across the estimation bounds so that the density does not drop towards zero as it approaches those bounds. In particular, if the range of sampled values is much narrower than the suspected true breadth of the distribution, then boundary reflection could improve KDE accuracy greatly. In Fig. D.5, observations are taken from only the middle region of a broad distribution; boundary reflection (blue) leads to an estimate much closer to the true shape of the

PDF (black) than an unreflected kernel estimate (red). Note that the mode/peak of the KDE is similar regardless of reflection - the main differences are in the tails of the estimated distributions. **kerneval** allows boundary reflection with the argument `reflect=TRUE` and specified bounds `a` and `b`, which the `wdens()` and `transdens()` functions will pass on internally to `GoFKernel::density.reflected()`.

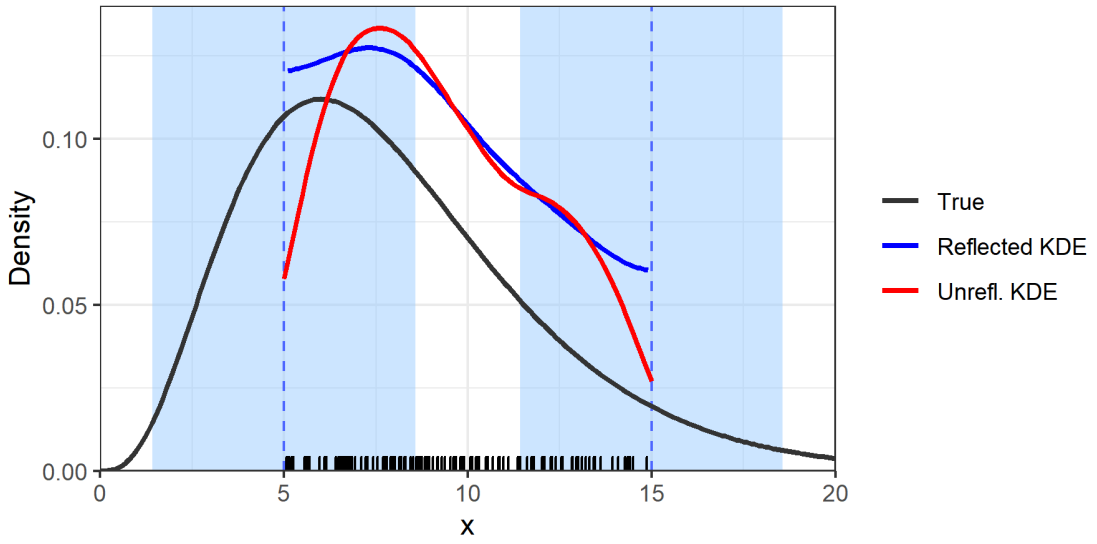


Figure D 5: The true PDF is a chi-square distribution with 8 degrees of freedom (black line), from which a random sample of 200 observations were drawn. Imagine a sampling bias that allows observations only on the interval $[5, 15]$; in this case, 145 observations remain (plotted as a rug at bottom) from which to estimate the density. The red curve plots the basic KDE `density(X, from = 5, to = 15)`. The blue curve is a KDE based on reflected data: all ovservations within the shaded blue region (4 times as wide as the bandwidth) were reflected across the boundaries (dashed lines).

D.6.3 Mean integrated squared error

The most common way to define the fitting error of an estimated density function is mean squared error, MSE. MSE is defined at a single point, while the error accumulated over the entire study interval is the mean integrated squared error, MISE. Most bandwidth selection methods attempt to estimate and then minimise the MISE as a function of the bandwidth, h. MISE can be written as the sum of systematic error (bias) and variance, and so the chosen bandwidth is a trade-off between these two terms. A large h will reduce variance but increase bias, and a small h will increase variance but reduce bias. This is a more formal description of the intuitive idea that an overly-narrow bandwidth will undersmooth the data, and an overly-broad bandwidth will oversmooth. The following section is based on notes from Silverman (1986) and summarises the mathematical explanation of the variance-bandwidth trade-off and the criteria for an optimal bandwidth.

The sum of the squared bias and variance at point x is:

$$MSE_x(\hat{f}) = E\{\hat{f}(x) - f(x)\}^2$$

$$= \{E\hat{f}(x) - f(x)\}^2 + var\hat{f}(x)$$

For a global estimate, the mean integrated squared error is defined as

$$MISE(\hat{f}) = E\int\{\hat{f}(x) - f(x)\}^2 dx$$

$$= \int MSE_x(\hat{f})dx$$

$$= \int\{E\hat{f}(x) - f(x)\}^2 dx + \int var\hat{f}(x)dx$$

Exploring the bias term individually,

$$bias_h(x) = E\hat{f}(x) - f(x) = \int \frac{1}{h}K\big(\frac{x-y}{h}\big)f(y)dy - f(x)$$

With some assumptions about the kernel function K,

$$bias_h(x) = \frac{1}{2}h^2 f''(x)k_2 + \text{ higher-order terms in } h$$

where k_2 is a constant, for instance the variance of a Gaussian distribution K.

Therefore,

$$\int bias_h(x)^2 dx \approx \frac{1}{4}h^4 k_2 \int f''(x)^2 dx$$

After some mathematical acrobatics, one can also derive an approximation of the variance term:

$$var\hat{f}(x) \approx \frac{1}{nh}f(x)\int K(t)^2 dt$$

$$\int var\hat{f}(x)dx \approx \frac{1}{nh}\int K(t)^2 dt$$

Combining the above expressions for MISE, bias, and variance, we have:

$$MISE \approx \frac{1}{4}h^4 k_2 \int f''(x)^2 dx + \frac{1}{nh}\int K(t)^2 dt$$

Note that the variance term depends on sample size n, while the bias term depends on the second derivative of the PDF. The bandwidth h affects both the variance and the bias. The value of h that minimises MISE is:

$$h_{opt} = k_2^{-2/5} \left\{ \int K(t)^2 dt \right\}^{1/5} \left\{ \int f''(x)^2 dx \right\}^{-1/5} n^{-1/5}$$

From inspection of this last result, we can tell that, unfortunately, h_{opt} depends on the (unknown) true density distribution f. This is why it is impossible to calculate the optimal bandwidth exactly given only a finite random sample from f. Inspection also reveals $h_{opt} \to 0$ as $n \to \infty$, but slowly. Lastly, smaller values of h are better for rougher (more 'wiggly') f. Roughness is indicated by the expression $\int f''^2$.

E

Supporting charts for Chapter 4

E.1 Figures

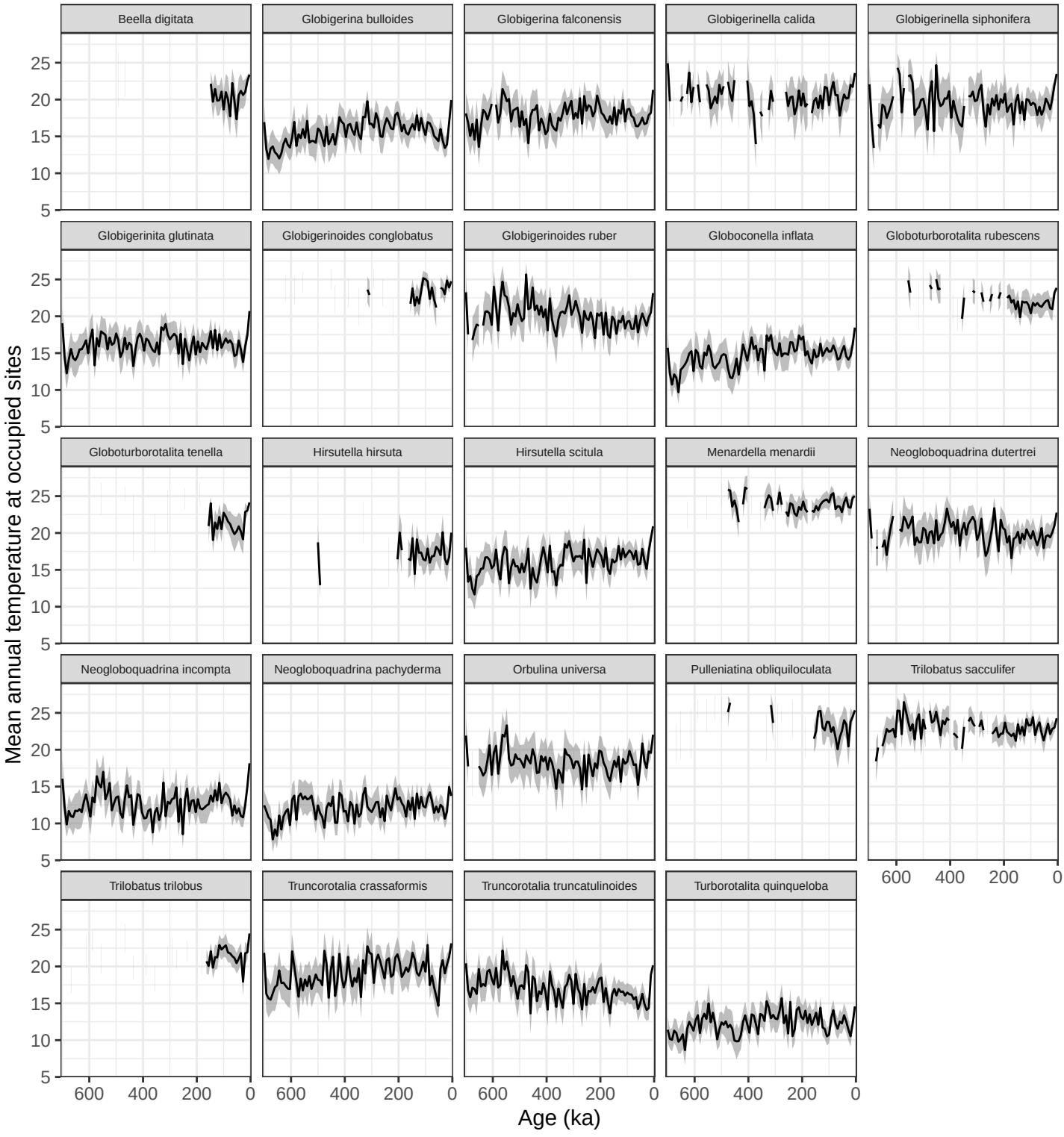


Figure E.1: Time series of mean annual temperature at the sea surface of all occupied sites for a focal species. The grey band covers one standard error, which trait evolution models use to estimate parameter values and confidence intervals.

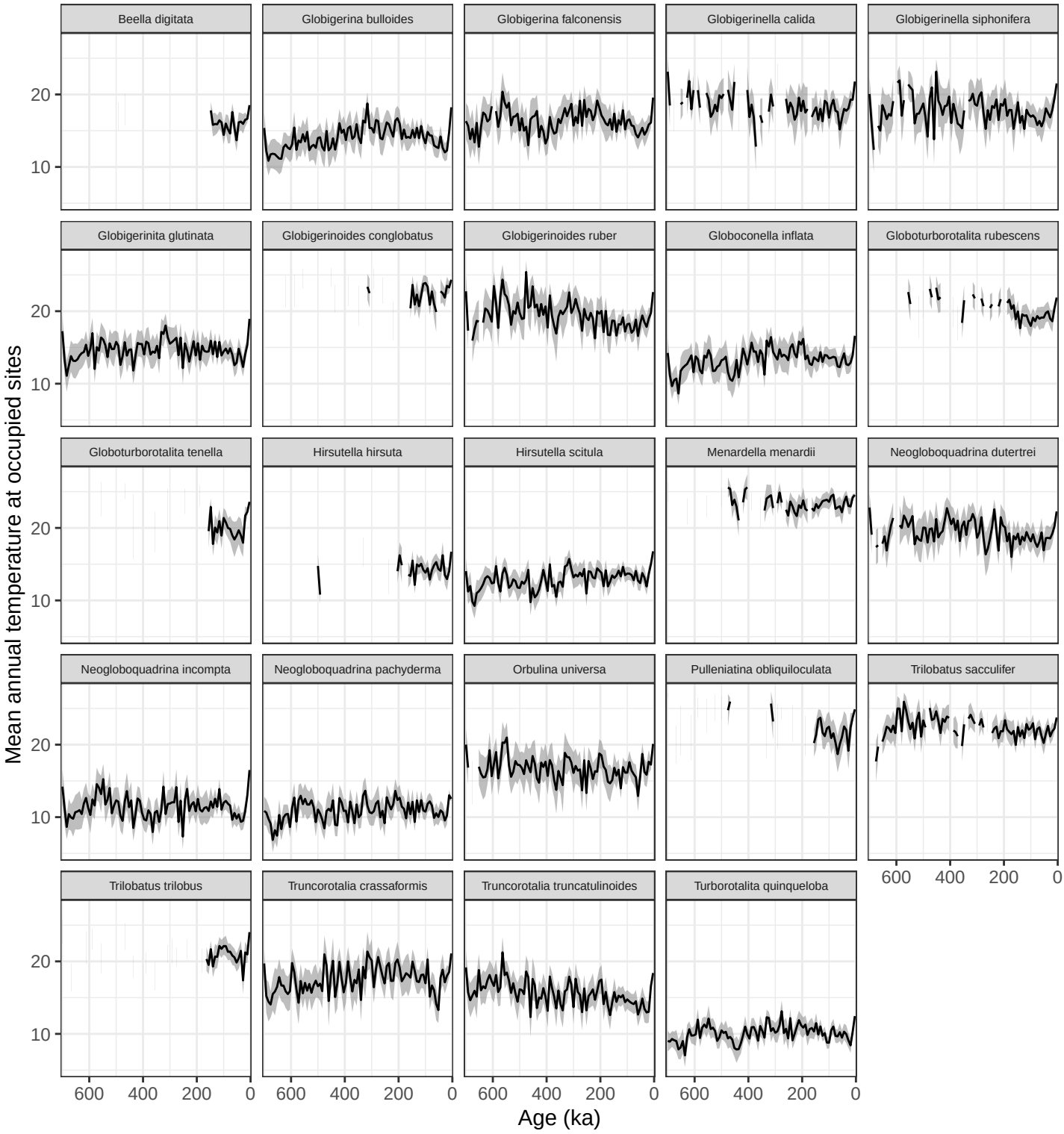


Figure E.2: Time series of mean annual temperature at all occupied sites for a focal species, within habitat depths. The grey band covers one standard error, which trait evolution models use to estimate parameter values and confidence intervals.

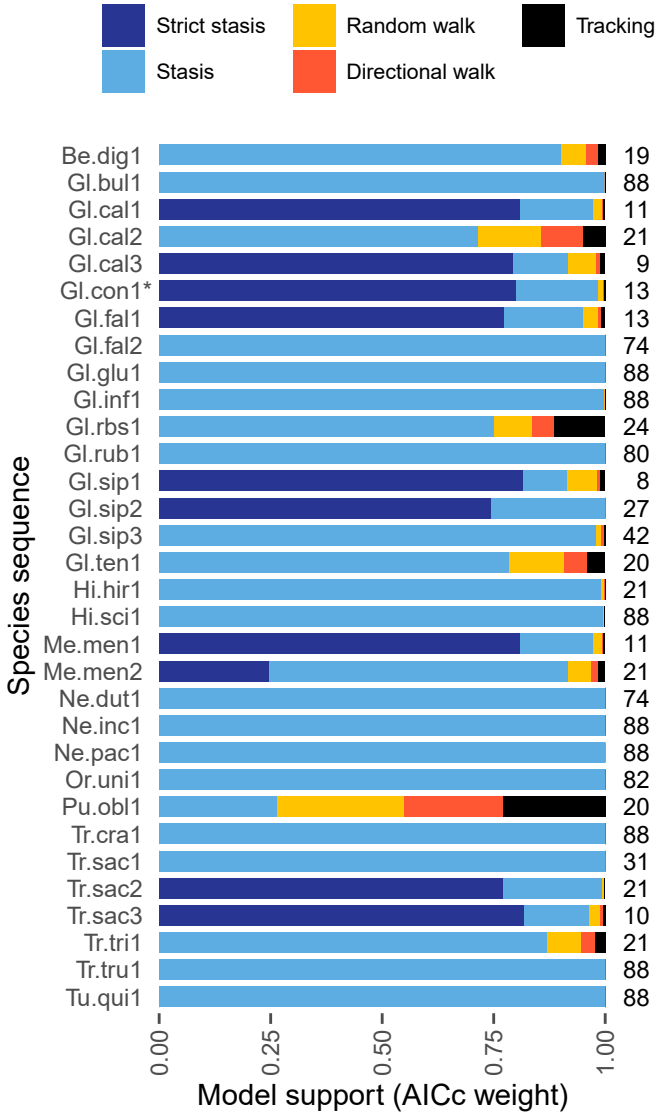


Figure E.3: Relative support for five evolutionary models fit to time series of species' occupied temperature values under the depth-habitat approach. Fig. 4.2 presents results under the sea-surface approach. The length of each sequence (number of 8-ka time bins) is listed at right. Abbreviated species names are listed at left; discontinuous time series for a single species are distinguished with numeric suffices. (Note that *rubescens* is abbreviated rbs to distinguish it from *ruber*.) Table C.1 lists full scientific names. An asterisk marks *Globigerinoides conglobatus* because the covariate-tracking model for this species was modified by pooling variances, for successful optimisation. AICc, Akaike Information Criterion corrected for sample size.

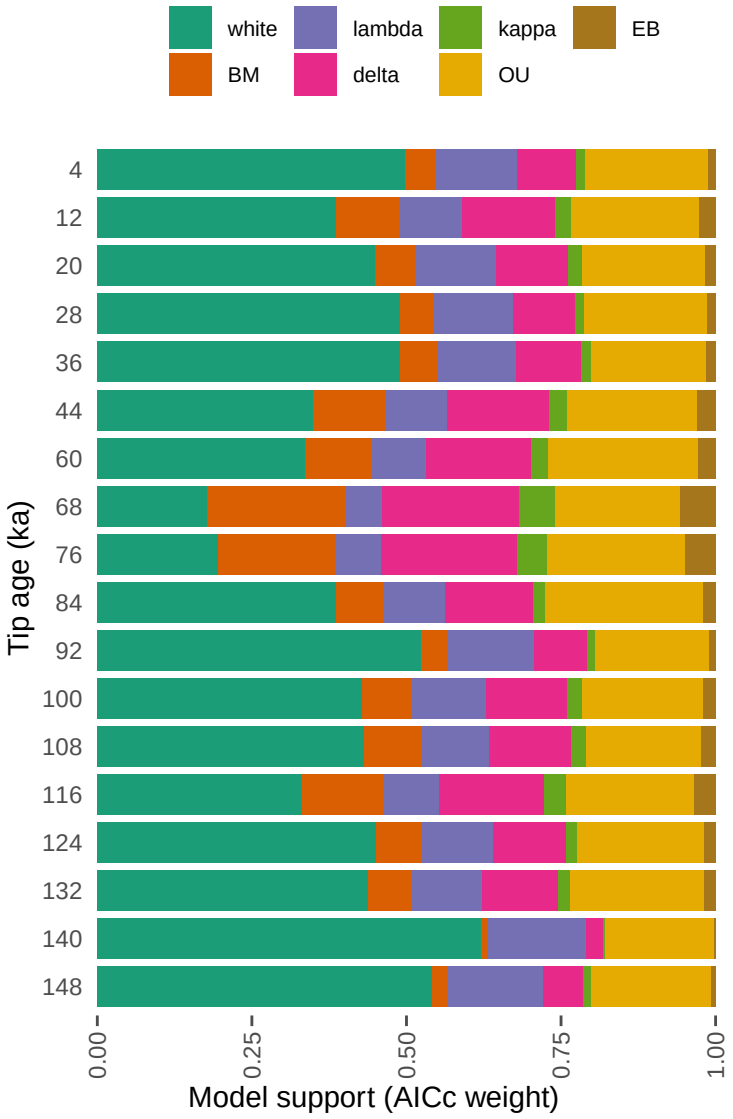


Figure E.4: Relative support for seven phylogenetic models fit to mean occupied temperature for each species (i.e. the trait at tree tips). Tips were defined at successive time steps, numbered at left. Bar heights are listed in Table E.3B. Temperature was reconstructed within habitat depths; Fig. 4.3 presents results under the sea-surface approach. BM, Brownian motion; OU, Ornstein-Uhlenbeck; EB, early burst; AICc, Akaike Information Criterion corrected for sample size.

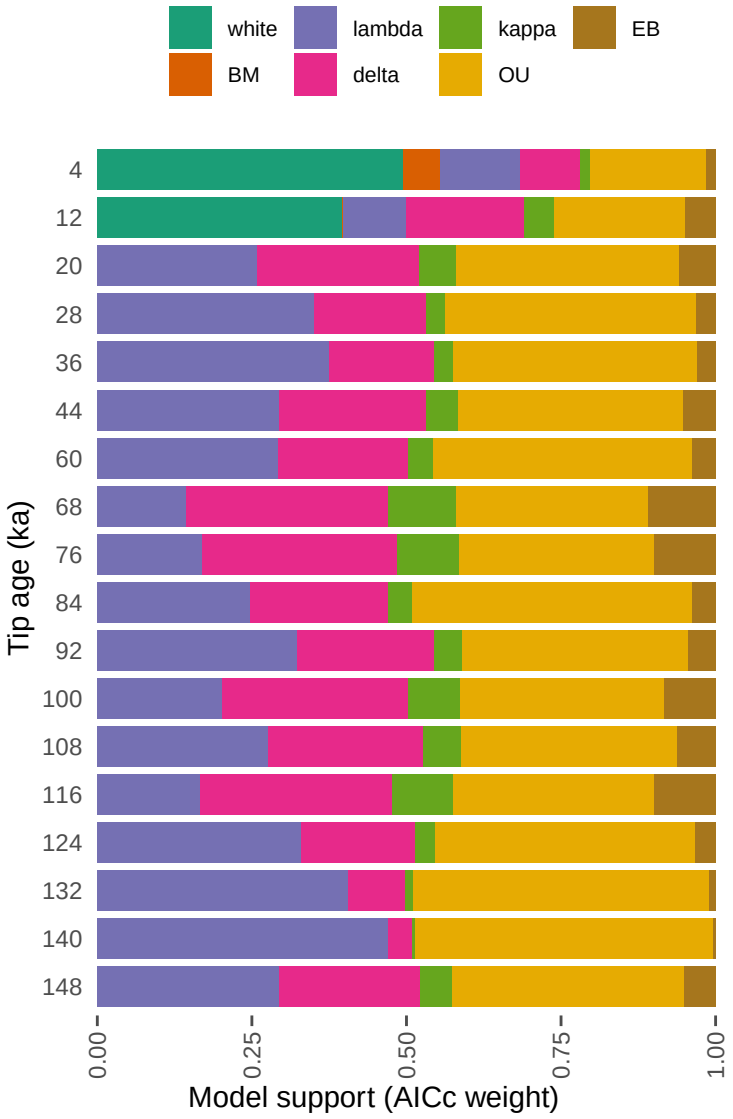


Figure E.5: Relative support for seven phylogenetic models fit to mean occupied temperature for each species (i.e. the trait at tree tips), with standard error specified for each species from empirical sample size and variance. Tips were defined at successive time steps, numbered at left. Temperature was reconstructed at the sea surface. Bar heights are listed in Table E.4. BM, Brownian motion; OU, Ornstein-Uhlenbeck; EB, early burst; AICc, Akaike Information Criterion corrected for sample size.

E.2 Tables

Table E.1: Parameter estimates are listed for trait evolution models that received the most support (Fig. 4.2). The 'best-fit' models for species' mean occupied sea surface temperature were all either strict stasis or broad-sense stasis. The parameters for these models are (i) the estimated static mean and (ii) the estimated variance about that mean, in °C. For strict stasis models (listed first), the estimated variance is zero.

	Est. Mean	Est. Variance
Globigerina falconensis1	16.89	0.00
Globigerinella calida1	20.67	0.00
Globigerinella calida2	19.84	0.00
Globigerinella siphonifera1	18.74	0.00
Globigerinella siphonifera2	20.12	0.00
Globigerinoides conglobatus1	23.79	0.00
Menardella menardii11	23.31	0.00
Menardella menardii12	24.77	0.00
Trilobatus sacculifer2	23.60	0.00
Trilobatus sacculifer3	23.92	0.00
Beella digitata1	21.20	1.33
Globigerina bulloides1	15.82	1.31
Globigerina falconensis2	18.15	0.92
Globigerinella calida3	21.10	1.28
Globigerinella siphonifera3	19.90	1.59
Globigerinita glutinata1	16.38	1.19
Globigerinoides ruber1	20.39	1.09
Globoconella inflata1	14.85	1.18
Globoturborotalita rubescens1	22.08	0.65
Globoturborotalita tenella1	22.03	1.28
Hirsutella hirsuta1	17.59	1.27
Hirsutella scitula1	16.43	1.52
Neogloboquadrina dutertrei1	20.39	0.70
Neogloboquadrina incompta1	12.77	1.81
Neogloboquadrina pachyderma1	12.20	0.55
Orbulina universa1	18.61	1.11
Pulleniatina obliquiloculata1	24.23	0.51
Trilobatus sacculifer1	23.02	0.32
Trilobatus trilobus1	21.82	1.20
Truncorotalia crassaformis1	19.34	1.55
Truncorotalia truncatulinoides1	17.10	1.02
Turborotalita quinqueloba1	12.38	0.56

Table E.2: Phylogenetic ANOVA compared variation in mean species' H distance among ecotypes. Coefficient values are relative to the reference group of open-ocean thermocline species. Negative estimates imply smaller Hellinger's H values, i.e. greater niche conservatism. Mixed layer species are from tropical/subtropical regions and are separated into those with symbionts and those without them. Ecotypes are plotted on the phylogenetic tree in Fig. 4.1.

Ecotype	Estimate	Std. Err.
A. Sea surface		
Intercept	0.160	0.012
Thermocline	0	-
Mix layer symbiotic	-0.001	0.017
Mix layer heterotroph	-0.051	0.021
Subthermocline	-0.029	0.018
High latitude	-0.066	0.025
B. Depth habitat		
Intercept	0.155	0.014
Thermocline	0	-
Mix layer symbiotic	0.019	0.020
Mix layer heterotroph	-0.048	0.025
Subthermocline	-0.008	0.021
High latitude	-0.047	0.029

Table E.3: Estimates of relative support for across-lineage evolution models, displayed graphically in Figs. 4.3 and E.4. Support was calculated as sample-size-corrected Akaike weight. The seven models were fit to trait data (species' occupied temperatures) separately for each time bin. The model receiving the most weight for a given dataset is listed in column 9 ('best'). Additionally, a Monte Carlo bootstrapping approach ($n = 500$) calculated pairwise likelihood ratio tests between Ornstein-Uhlenbeck (OU) and white-noise models. The probability of generating a test statistic greater than observed if the data truly came from a white-noise process is the stated p-value; values should not be compared to the common cut-off of 0.05 on account of multiple testing. The statistical power of the test (i.e. the probability of correctly rejecting a white-noise model if the data were generated by an OU process) is given in the final column, with an allowed false positivity rate of 5%. Standard error in models was left unspecified to be calculated by default methods. Temperature values were reconstructed at **(A)** the sea surface or **(B)** within depth habitats. BM, Brownian motion; EB, early burst.

Bin	White	BM	lambda	delta	kappa	OU	EB	Best	p-val	Power
A. Sea surface										
4	0.495	0.060	0.128	0.098	0.016	0.188	0.016	white	0.178	0.335
12	0.323	0.159	0.084	0.168	0.041	0.184	0.041	white	0.092	0.467
20	0.484	0.058	0.130	0.103	0.018	0.192	0.015	white	0.167	0.325
28	0.575	0.023	0.149	0.054	0.007	0.185	0.006	white	0.265	0.135
36	0.607	0.016	0.157	0.041	0.006	0.169	0.004	white	0.395	0.063
44	0.505	0.049	0.133	0.095	0.013	0.193	0.013	white	0.194	0.290
60	0.469	0.049	0.121	0.099	0.014	0.236	0.013	white	0.140	0.223
68	0.183	0.219	0.057	0.221	0.057	0.207	0.057	delta	0.030	0.609
76	0.331	0.112	0.106	0.168	0.029	0.225	0.029	white	0.065	0.511
84	0.429	0.055	0.111	0.112	0.014	0.264	0.014	white	0.071	0.323
92	0.586	0.024	0.152	0.055	0.008	0.170	0.006	white	0.326	0.106
100	0.435	0.088	0.114	0.130	0.023	0.187	0.023	white	0.142	0.492
108	0.491	0.068	0.127	0.103	0.018	0.176	0.018	white	0.221	0.316
116	0.287	0.170	0.074	0.182	0.044	0.199	0.044	white	0.065	0.519
124	0.538	0.036	0.139	0.069	0.009	0.199	0.009	white	0.162	0.242
132	0.571	0.019	0.148	0.047	0.005	0.205	0.005	white	0.215	0.185
140	0.645	0.003	0.167	0.010	0.001	0.173	0.001	white	0.527	0.090
148	0.556	0.023	0.147	0.058	0.009	0.200	0.006	white	0.187	0.143

... Table E.3 continued

Bin	White	BM	lambda	delta	kappa	OU	EB	Best	*p*-val	Power
					B. Depth habitat					
4	0.498	0.050	0.132	0.094	0.016	0.197	0.013	white	0.181	0.271
12	0.386	0.104	0.100	0.150	0.027	0.206	0.027	white	0.105	0.431
20	0.450	0.065	0.130	0.116	0.023	0.199	0.017	white	0.159	0.384
28	0.490	0.055	0.127	0.101	0.015	0.198	0.014	white	0.161	0.371
36	0.491	0.058	0.127	0.107	0.016	0.185	0.015	white	0.187	0.406
44	0.350	0.115	0.100	0.165	0.030	0.210	0.030	white	0.072	0.539
60	0.337	0.107	0.087	0.170	0.028	0.243	0.028	white	0.060	0.472
68	0.178	0.224	0.058	0.222	0.058	0.202	0.058	BM	0.026	0.639
76	0.195	0.192	0.073	0.219	0.050	0.222	0.050	OU	0.023	0.629
84	0.387	0.076	0.100	0.141	0.020	0.256	0.020	white	0.077	0.358
92	0.524	0.041	0.142	0.086	0.013	0.183	0.011	white	0.223	0.321
100	0.428	0.081	0.121	0.131	0.024	0.195	0.021	white	0.139	0.445
108	0.431	0.092	0.112	0.132	0.024	0.186	0.024	white	0.162	0.407
116	0.332	0.133	0.089	0.171	0.034	0.208	0.034	white	0.066	0.515
124	0.452	0.072	0.117	0.118	0.019	0.204	0.019	white	0.117	0.326
132	0.438	0.072	0.113	0.124	0.019	0.217	0.019	white	0.091	0.449
140	0.621	0.009	0.161	0.027	0.003	0.176	0.002	white	0.408	0.113
148	0.542	0.026	0.154	0.065	0.011	0.195	0.007	white	0.209	0.189

Table E.4: Estimates of relative support for across-lineage evolution models in which error was specified as the empirical standard error for occupied temperature means. Relative support values are displayed graphically in Fig. E.5. Columns as in Table E.3.

Bin	White	BM	lambda	delta	kappa	OU	EB	Best	*p*-val	Power
4	0.494	0.061	0.128	0.098	0.016	0.187	0.016	white	0.186	0.212
12	0.397	0.000	0.103	0.190	0.050	0.211	0.050	white	0.102	0.470
20	0.000	0.000	0.258	0.263	0.060	0.360	0.059	OU	0.971	0.000
28	0.000	0.000	0.352	0.180	0.031	0.406	0.031	OU	0.998	0.000
36	0.000	0.000	0.375	0.170	0.031	0.395	0.029	OU	0.952	0.000
44	0.000	0.000	0.294	0.237	0.052	0.364	0.052	OU	0.982	0.000
60	0.000	0.000	0.293	0.209	0.042	0.417	0.038	OU	0.999	0.000
68	0.000	0.000	0.143	0.328	0.109	0.312	0.109	delta	0.999	0.000
76	0.000	0.000	0.170	0.316	0.099	0.316	0.099	OU	0.987	0.000
84	0.000	0.000	0.247	0.225	0.037	0.454	0.037	OU	0.856	0.000
92	0.000	0.000	0.323	0.222	0.045	0.364	0.045	OU	0.954	0.000
100	0.000	0.000	0.201	0.302	0.084	0.329	0.084	OU	0.988	0.000
108	0.000	0.000	0.276	0.251	0.062	0.349	0.062	OU	0.582	0.000
116	0.000	0.000	0.167	0.310	0.099	0.325	0.099	OU	0.971	0.000
124	0.000	0.000	0.331	0.183	0.033	0.419	0.033	OU	1.000	0.000
132	0.000	0.000	0.407	0.091	0.013	0.478	0.011	OU	0.947	0.000
140	0.000	0.000	0.471	0.038	0.006	0.482	0.003	OU	0.993	0.000
148	0.000	0.000	0.294	0.228	0.051	0.375	0.051	OU	0.713	0.000

Table E.5: Parameters for *lambda*, *delta*, and Ornstein-Uhlenbeck models fit to species' occupied sea-surface temperatures at each time step (bin). Fig. 4.3 plots the corresponding weights of the models. Each parameter point estimate is from maximum likelihood, and lower and upper bounds are the 95% confidence interval from Monte Carlo simulation bootstrapping. **(A)** *lambda* is also known as 'phylogenetic signal' and is parameterised in its namesake model. **(B)** *delta* parameterises an increase (for $\delta > 1$) or decrease (for $\delta < 1$) in evolutionary rates through time. The upper bound allowed during model fitting was $\delta = 3$. **(C)** *alpha* is given on a $-log_{10}$ scale. For the purpose of reporting *alpha* estimates, the phylogenetic tree was scaled to height of unity.

Bin	Estimate	2.5% CI	97.5%
	A. lambda		
4	1.30e-184	1.41e-217	0.466
12	1.14e-217	1.58e-217	0.466
20	6.59e-02	1.56e-217	0.573
28	1.84e-199	1.50e-217	0.491
36	5.20e-16	1.63e-217	0.500
44	3.89e-02	1.48e-217	0.481
60	1.33e-217	1.57e-217	0.490
68	1.00e+00	1.79e-217	1.000
76	1.88e-01	1.69e-217	0.739
84	2.56e-187	1.44e-217	0.429
92	1.07e-03	1.41e-217	0.454
100	4.17e-02	1.54e-217	0.651
108	2.63e-103	1.54e-217	0.362
116	6.18e-02	1.46e-217	0.804
124	1.56e-217	1.35e-217	0.337
132	9.66e-17	1.38e-217	0.573
140	6.45e-176	1.42e-217	0.423
148	3.95e-02	1.36e-217	0.584
	B. delta		
4	3.000	1.322	3.000
12	3.000	1.349	3.000

... Table E.5 continued

Bin	Estimate	2.5% CI	97.5%
20	3.000	1.463	3.000
28	3.000	1.153	3.000
36	3.000	1.143	3.000
44	3.000	1.105	3.000
60	3.000	1.198	3.000
68	3.000	1.396	3.000
76	3.000	1.272	3.000
84	3.000	1.072	3.000
92	3.000	1.226	3.000
100	3.000	1.197	3.000
108	3.000	1.570	3.000
116	3.000	1.092	3.000
124	3.000	1.080	3.000
132	3.000	1.086	3.000
140	3.000	1.141	3.000
148	3.000	1.366	3.000
	C. $-log$(alpha)		
4	1.098	1.560	-0.434
12	1.297	1.723	-0.434
20	1.161	1.606	-0.434
28	0.911	1.395	-0.434
36	0.714	1.354	-0.434
44	1.119	1.562	-0.434
60	1.073	1.474	-0.434
68	1.422	1.946	-0.434
76	1.295	1.772	-0.434
84	1.109	1.566	-0.434
92	0.844	1.419	-0.434

... Table E.5 continued

Bin	Estimate	2.5% CI	97.5%
100	1.240	1.880	-0.434
108	1.118	1.485	-0.434
116	1.327	1.846	-0.434
124	0.947	1.487	-0.434
132	0.899	1.466	-0.434
140	0.593	1.294	-0.434
148	0.940	1.414	-0.434

References

Abalde, Samuel, Manuel J Tenorio, Carlos M L Afonso, Juan E Uribe, Ana M Echeverry, and Rafael Zardoya (2017). "Phylogenetic relationships of cone snails endemic to Cabo Verde based on mitochondrial genomes". In: *BMC Evolutionary Biology* 17, 231.

Abrams, Peter A (1995). "Monotonic or unimodal diversity-productivity gradients: What does competition theory predict?" In: *Ecology* 76.7, pp. 2019–2027.

Aguilée, Robin, Fanny Gascuel, Amaury Lambert, and Regis Ferriere (2018). "Clade diversification dynamics and the biotic and abiotic controls of speciation and extinction rates". In: *Nature Communications* 9, 3013.

Algeo, Thomas J and Stephen E Scheckler (1998). "Terrestrial-marine teleconnections in the Devonian: Links between the evolution of land plants, weathering processes, and marine anoxic events". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 353.1365, pp. 113–130.

Allmon, Warren D and Ronald E Martin (2014). "Seafood through time revisited: The Phanerozoic increase in marine trophic resources and its macroevolutionary consequences". In: *Paleobiology* 40.2, pp. 256–287.

Alroy, John (2010). "The shifting balance of diversity among major marine animal groups". In: *Science* 329.5996, pp. 1191–1194.

Alroy, John, Martin Aberhan, David J Bottjer, Michael Foote, Franz T Fürsich, Peter J Harries, et al. (2008). "Phanerozoic trends in the global diversity of marine invertebrates". In: *Science* 321.5885, pp. 97–100.

Altman, Douglas G and J Martin Bland (1995). "Absence of evidence is not evidence of absence". In: *British Medical Journal* 311.7003, p. 485.

Antell, Gwen S, Isabel S Fenton, Paul J Valdes, and Erin E Saupe (2021). "Thermal niches of planktonic foraminifera are static throughout glacial-–interglacial climate change". In: *Proceedings of the National Academy of Sciences* 118.18.

Antell, Gwen S, Wolfgang Kiessling, Martin Aberhan, and Erin E Saupe (2020). "Marine biodiversity and geographic distributions are independent on large scales". In: *Current Biology* 30.1, 115–121.e5.

Antell, Gwen S and Erin E Saupe (2021). "Bottom-up controls, ecological revolutions and diversification in the oceans through time". In: *Current Biology* 31.19, R1237–R1251.

Aurahs, Ralf, Yvonne Treis, Kate Darling, and Michal Kucera (2011). "A revised taxonomic and phylogenetic concept for the planktonic foraminifer species Globigerinoides ruber based on molecular and morphometric evidence". In: *Marine Micropaleontology* 79.1-2, pp. 1–14.

Aze, Tracy, Thomas HG Ezard, Andy Purvis, Helen K Coxall, Duncan RM Stewart, Bridget S Wade, et al. (2011). "A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data". In: *Biological Reviews* 86.4, pp. 900–927.

Backus, RH, JE Craddock, RL Haedrich, and BH Robison (1977). "Atlantic mesopelagic zoogeography". In: *Fishes of the Western North Atlantic*. Vol. 7. 1. Sears Foundation for Marine Research.

Bambach, Richard K (1983). "Ecospace utilization and guilds in marine communities through the Phanerozoic". In: *Biotic Interactions in Recent and Fossil Benthic Communities*. Springer, pp. 719–746.

— (1993). "Seafood through time: Changes in biomass, energetics, and productivity in the marine ecosystem". In: *Paleobiology*, pp. 372–397.

— (1999). "Energetics in the global marine fauna: A connection between terrestrial diversification and change in the marine biosphere". In: *Geobios* 32.2, pp. 131–144.

Bambach, Richard K, Andrew M Bush, and Douglas H Erwin (2007). "Autecology and the filling of ecospace: Key metazoan radiations". In: *Palaeontology* 50.1, pp. 1–22.

Banks, William E, Francesco d'Errico, A Townsend Peterson, Masa Kageyama, and Guillaume Colombeau (2008). "Reconstructing ecological niches and geographic distributions of caribou (Rangifer tarandus) and red deer (Cervus elaphus) during the Last Glacial Maximum". In: *Quaternary Science Reviews* 27.27-28, pp. 2568–2575.

Banks, William E, Francesco d'Errico, A Townsend Peterson, Masa Kageyama, Adriana Sima, and Maria-Fernanda Sánchez-Goñi (2008). "Neanderthal extinction by competitive exclusion". In: *PLoS One* 3.12, e3972.

Bapst, David W (2012). "paleotree: An R package for paleontological and phylogenetic analyses of evolution". In: *Methods in Ecology and Evolution* 3.5, pp. 803–807.

Barmi, Hammou El and Jeffrey S Simonoff (2000). "Transformation-based density estimation for weighted distributions". In: *Journal of Nonparametric Statistics* 12.6, pp. 861–878.

Barnes, Christopher R (2004). "Was there an Ordovician superplume event?" In: *The Great Ordovician Biodiversification Event*. Ed. by B D Webby, F Paris, Mary L Droser, and I G Percival. New York, NY: Columbia University Press. Chap. 8, pp. 77–80.

Barnosky, Anthony D, Marc A Carrasco, and Edward B Davis (2005). "The impact of the species–area relationship on estimates of paleodiversity". In: *PLoS Biology* 3.8, e266.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). "Fitting linear mixed-effects models using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48.

Baumiller, Tomasz K, Mariusz A Salamon, Przemysław Gorzelak, Rich Mooi, Charles G Messing, and Forest J Gahn (2010). "Post-Paleozoic crinoid radiation in response to benthic predation preceded the Mesozoic marine revolution". In: *Proceedings of the National Academy of Sciences* 107.13, pp. 5893–5896.

Behrenfeld, Michael J (2014). "Climate-mediated dance of the plankton". In: *Nature Climate Change* 4.10, pp. 880–887.

Behrensmeyer, Anna K and Susan M Kidwell (1985). "Taphonomy's contributions to paleobiology". In: *Paleobiology* 11.1, pp. 105–119.

Bell, Christopher J, Ernest L Lundelius Jr, AD Barnosky, RW Graham, EH Lindsay, DR Ruez Jr, et al. (2004). "The Blancan, Irvingtonian, and Rancholabrean mammal ages". In: *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology*. New York, NY: Columbia University Press, pp. 232–314.

Benitez-Nelson, Claudia R (2000). "The biogeochemical cycling of phosphorus in marine systems". In: *Earth-Science Reviews* 51.1-4, pp. 109–135.

Bennett, Elena M, Stephen R Carpenter, and Nina F Caraco (2001). "Human impact on erodable phosphorus and eutrophication: A global perspective". In: *BioScience* 51.3, pp. 227–234.

Benson, Roger BJ, Richard J Butler, Roger A Close, Erin E Saupe, and Daniel L Rabosky (2021). "Biodiversity in space and time—insights from the fossil record". In: *Current Biology* 31.19, R1225–R1236.

Benton, Michael J (1996). "Testing the roles of competition and expansion in tetrapod evolution". In: *Proceedings of the Royal Society B: Biological Sciences* 263.1370, pp. 641–646.

— (2009). "The Red Queen and the Court Jester: Species diversity and the role of biotic and abiotic factors through time". In: *Science* 323.5915, pp. 728–732.

Berner, Robert A (1997). "The rise of plants and their effect on weathering and atmospheric CO2". In: *Science* 276.5312, pp. 544–546.

Bijma, Jelle, Walter W Faber, and Christoph Hemleben (1990). "Temperature and salinity limits for growth and survival of some planktonic foraminifers in laboratory cultures". In: *The Journal of Foraminiferal Research* 20.2, pp. 95–116.

Bivand, Roger, Tim Keitt, and Barry Rowlingson (2018). *rgdal: Bindings for the 'geospatial' data abstraction library*. URL: `https://CRAN.R-project.org/package=rgdal`.

Bivand, Roger S, Edzer Pebesma, and Virgilio Gómez-Rubio (2013). *Applied Spatial Data Analysis with R*. 2nd ed. New York, NY: Springer Science and Business Media.

Blomberg, Simon P, Theodore Garland Jr, and Anthony R Ives (2003). "Testing for phylogenetic signal in comparative data: Behavioral traits are more labile". In: *Evolution* 57.4, pp. 717–745.

Blount, Zachary D, Christina Z Borland, and Richard E Lenski (2008). "Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli". In: *Proceedings of the National Academy of Sciences* 105.23, pp. 7899–7906.

Boersma, Maarten and James J Elser (2006). "Too much of a good thing: On stoichiometrically balanced diets and maximal growth". In: *Ecology* 87.5, pp. 1325–1330.

Boettiger, Carl, Graham Coop, and Peter Ralph (2012). "Is your phylogeny informative? Measuring the power of comparative methods". In: *Evolution* 66.7, pp. 2240–2251.

Bokma, Folmer (2008). "Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny". In: *Evolution* 62.11, pp. 2718–2726.

Borrajo, Maria Isabel, Wenceslao González-Manteiga, and María Dolores Martínez-Miranda (2017). "Bandwidth selection for kernel density estimation with length-biased data". In: *Journal of Nonparametric Statistics* 29.3, pp. 636–668.

Bose, Arup and Santanu Dutta (2013). "Density estimation using bootstrap bandwidth selector". In: *Statistics & Probability Letters* 83.1, pp. 245–256.

Botting, Joseph P and Lucy A Muir (2008). "Unravelling causal components of the Ordovician Radiation: The Builth Inlier (central Wales) as a case study". In: *Lethaia* 41.2, pp. 111–125.

Boyce, C Kevin and Jung-Eun Lee (2011). "Could land plant evolution have fed the marine revolution?" In: *Paleontological Research* 15.2, pp. 100–105.

Boyd, Philip W, Timothy Jickells, CS Law, S Blain, EA Boyle, KO Buesseler, et al. (2007). "Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions". In: *Science* 315.5812, pp. 612–617.

Bridle, Jon R, Jitka Polechová, Masakado Kawata, and Roger K Butlin (2010). "Why is adaptation prevented at ecological margins? New insights from individual-based simulations". In: *Ecology Letters* 13.4, pp. 485–494.

Bridle, Jon R and Timothy H Vines (2007). "Limits to evolution at range margins: When and why does adaptation fail?" In: *Trends in Ecology & Evolution* 22.3, pp. 140–147.

Briggs, Robert D, David Pollard, and Lev Tarasov (2014). "A data-constrained large ensemble analysis of Antarctic evolution since the Eemian". In: *Quaternary Science Reviews* 103, pp. 91–115.

Bringué, Manuel, Vera Pospelova, and David B Field (2014). "High resolution sedimentary record of dinoflagellate cysts reflects decadal variability and 20th century warming in the Santa Barbara Basin". In: *Quaternary Science Reviews* 105, pp. 86–101.

Brocks, Jochen J, Amber JM Jarrett, Eva Sirantoine, Christian Hallmann, Yosuke Hoshino, and Tharika Liyanage (2017). "The rise of algae in Cryogenian oceans and the emergence of animals". In: *Nature* 548.7669, pp. 578–581.

Broennimann, Olivier, Matthew C Fitzpatrick, Peter B Pearman, Blaise Petitpierre, Loïc Pellissier, Nigel G Yoccoz, et al. (2012). "Measuring ecological niche overlap from occurrence and spatial environmental data". In: *Global Ecology and Biogeography* 21.4, pp. 481–497.

Brombacher, Anieke, Paul A Wilson, Ian Bailey, and Thomas HG Ezard (2018). "Temperature is a poor proxy for synergistic climate forcing of plankton evolution". In: *Proceedings of the Royal Society B: Biological Sciences* 285, 20180665 (1883).

Brown, James H (1995). *Macroecology*. Chicago, IL: The University of Chicago Press, 269 pp.

Burnham, Kenneth P and David R Anderson (2010). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd. New York, NY: Springer Science and Business Media, 488 pp.

Burrows, Michael T, David S Schoeman, Lauren B Buckley, Pippa Moore, Elvira S Poloczanska, Keith M Brander, et al. (2011). "The pace of shifting climate in marine and terrestrial ecosystems". In: *Science* 334.6056, pp. 652–655.

Bush, Andrew M and Richard K Bambach (2011). "Paleoecologic megatrends in marine metazoa". In: *Annual Review of Earth and Planetary Sciences* 39, pp. 241–269.

Bush, Andrew M, Gene Hunt, and Richard K Bambach (2016). "Sex and the shifting biodiversity dynamics of marine animals in deep time". In: *Proceedings of the National Academy of Sciences* 113.49, pp. 14073–14078.

Butler, Marguerite A and Aaron A King (2004). "Phylogenetic comparative analysis: A modeling approach for adaptive evolution". In: *The American Naturalist* 164.6, pp. 683–695.

Butterfield, Nicholas (2018). "Oxygen, animals and aquatic bioturbation: An updated account". In: *Geobiology* 16.1, pp. 3–16.

Butterfield, NJ (2009). "Oxygen, animals and oceanic ventilation: An alternative view". In: *Geobiology* 7.1, pp. 1–7.

Button, David J, Graeme T Lloyd, Martín D Ezcurra, and Richard J Butler (2017). "Mass extinctions drove increased global faunal cosmopolitanism on the supercontinent Pangaea". In: *Nature Communications* 8, 733.

Cahill, Abigail E, Matthew E Aiello-Lammens, M Caitlin Fisher-Reid, Xia Hua, Caitlin J Karanewsky, Hae Yeong Ryu, et al. (2013). "How does climate change cause

extinction?" In: *Proceedings of the Royal Society B: Biological Sciences* 280.1750, 20121890.

Campbell, Ian H and Richard J Squire (2010). "The mountains that triggered the Late Neoproterozoic increase in oxygen: The Second Great Oxidation Event". In: *Geochimica et Cosmochimica Acta* 74.15, pp. 4187–4206.

Canfield, Don E, Simon W Poulton, and Guy M Narbonne (2007). "Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life". In: *Science* 315.5808, pp. 92–95.

Carlson, Sandra J (2016). "The evolution of Brachiopoda". In: *Annual Review of Earth and Planetary Sciences* 44, pp. 409–438.

Carpenter, Stephen R and James F Kitchell (1988). "Consumer control of lake productivity". In: *BioScience* 38.11, pp. 764–769.

Castro-Insua, Adrián, Carola Gómez-Rodríguez, John J Wiens, and Andrés Baselga (2018). "Climatic niche divergence drives patterns of diversification and richness among mammal families". In: *Scientific Reports* 8.1, pp. 1–12.

Chan, Kung-Sik and Brian Ripley (2018). *TSA: Time series analysis*. URL: `https://CRAN.R-project.org/package=TSA`.

Chao, Anne and Lou Jost (2012). "Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size". In: *Ecology* 93.12, pp. 2533–2547.

Clapham, Matthew E and David J Bottjer (2007). "Permian marine paleoecology and its implications for large-scale decoupling of brachiopod and bivalve abundance and diversity during the Lopingian (Late Permian)". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 249.3-4, pp. 283–301.

CLIMAP Project (1981). *Seasonal reconstructions of the Earth's surface at the last glacial maximum*. Geological Society of America.

Close, Roger A, Roger B.J. Benson, Paul Upchurch, and Richard J Butler (2017). "Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification". In: *Nature Communications* 8, 15381.

Close, Roger A, Serjoscha W Evers, John Alroy, and Richard J Butler (2018). "How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves". In: *Methods in Ecology and Evolution* 9.6, pp. 1386–1400.

Cohen, K.M., D.A.T. Harper, and P.L. Gibbard (2017). *International Chronostratigraphic Chart 2017/02*. Chart or Table. URL: `www.stratigraphy.org`.

Connor, Edward F and Earl D McCoy (1979). "The statistics and biology of the species-area relationship". In: *American Naturalist* 113.6, pp. 791–833.

Cooney, Christopher R, Nathalie Seddon, and Joseph A Tobias (2016). "Widespread correlations between climatic niche evolution and species diversification in birds". In: *Journal of Animal Ecology* 85.4, pp. 869–878.

Cooper, Natalie, Gavin H Thomas, Chris Venditti, Andrew Meade, and Rob P Freckleton (2016). "A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies". In: *Biological Journal of the Linnean Society* 118.1, pp. 64–77.

Cox, David (1969). "Some sampling problems in technology". In: *New Developments in Survey Sampling*. Ed. by NL Johnson and H Smith. New York: Wiley, pp. 506–527.

Cox, Peter M (2001). *Description of the "TRIFFID" dynamic global vegetation model*. Report. Hadley Centre.

Cristóbal, José A and José T Alcalá (2001). "An overview of nonparametric contributions to the problem of functional estimation from biased data". In: *Sociedad de Estadística e Investigación Operativa Test* 10.2, pp. 309–332.

Cuthill, Jennifer F Hoyal, Nicholas Guttenberg, and Graham E Budd (2020). "Impacts of speciation and extinction measured by an evolutionary decay clock". In: *Nature* 588.7839, pp. 636–641.

D'Antonio, Michael P, Daniel E Ibarra, and C Kevin Boyce (2020). "Land plant evolution decreased, rather than increased, weathering rates". In: *Geology* 48.1, pp. 29–33.

Dahl, Tais W and Susanne KM Arens (2020). "The impacts of land plant evolution on Earth's climate and oxygenation state—An interdisciplinary review". In: *Chemical Geology* 547, 119665.

Dahl, Tais W, Emma U Hammarlund, Ariel D Anbar, David PG Bond, Benjamin C Gill, Gwyneth W Gordon, et al. (2010). "Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish". In: *Proceedings of the National Academy of Sciences* 107.42, pp. 17911–17915.

Darling, Kate F, Michal Kucera, and Christopher M Wade (2007). "Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist". In: *Proceedings of the National Academy of Sciences* 104.12, pp. 5002–5007.

Darroch, Simon A.F. and Peter J Wagner (2015). "Response of beta diversity to pulses of Ordovician–Silurian mass extinction". In: *Ecology* 96.2, pp. 532–549.

Darroch, Simon AF, Marc Laflamme, and Peter J Wagner (2018). "High ecological complexity in benthic Ediacaran communities". In: *Nature Ecology & Evolution* 2.10, pp. 1541–1547.

Davies-Barnard, Taraka, Andy Ridgwell, Joy Singarayer, and Paul Valdes (2017). "Quantifying the influence of the terrestrial biosphere on glacial–interglacial climate dynamics". In: *Climate of the Past* 13.10, pp. 1381–1401.

Davis, Catherine V, Caitlin M Livsey, Hannah M Palmer, Pincelli M Hull, Ellen Thomas, Tessa M Hill, et al. (2020). "Extensive morphological variability in asexually produced planktic foraminifera". In: *Science Advances* 6.28, eabb8930.

De Baar, Hein JW, Philip W Boyd, Kenneth H Coale, Michael R Landry, Atsushi Tsuda, Philipp Assmy, et al. (2005). "Synthesis of iron fertilization experiments: From the Iron Age in the Age of Enlightenment". In: *Journal of Geophysical Research: Oceans* 110.C9.

De Boer, B, Lucas J Lourens, and Roderik SW Van De Wal (2014). "Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the Plio-Pleistocene". In: *Nature Communications* 5, 2999.

De Boer, B, RSW Van de Wal, LJ Lourens, R Bintanja, and TJ Reerink (2013). "A continuous simulation of global ice volume over the past 1 million years with 3-D ice-sheet models". In: *Climate Dynamics* 41.5-6, pp. 1365–1384.

DeAngelis, Donald L (1995). "Relationships between the energetics of species and large-scale species richness". In: *Linking Species & Ecosystems*. Springer, pp. 263–272.

Deng, Henry and Hadley Wickham (2011). "Density estimation in R". Unpublished Work. URL: `http://www2.cs.uh.edu/~ceick/7362/T2-4.pdf`.

Derry, Louis A (2015). "Causes and consequences of mid-Proterozoic anoxia". In: *Geophysical Research Letters* 42.20, pp. 8538–8546.

Di Cola, Valeria, Olivier Broennimann, Blaise Petitpierre, Frank T Breiner, Manuela d'Amen, Christophe Randin, et al. (2017). "ecospat: An R package to

support spatial analyses and modeling of species niches and distributions". In: *Ecography* 40.6, pp. 774–787.

Dickman, Elizabeth M, Jennifer M Newell, María J González, and Michael J Vanni (2008). "Light, nutrients, and food-chain length constrain planktonic energy transfer efficiency across multiple trophic levels". In: *Proceedings of the National Academy of Sciences* 105.47, pp. 18408–18412.

Donelson, Jennifer M, Jennifer M Sunday, Will F Figueira, Juan Diego Gaitán-Espitia, Alistair J Hobday, Craig R Johnson, et al. (2019). "Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1768, 20180186.

Dormann, Carsten F, Jana M McPherson, Miguel B Araújo, Roger Bivand, Janine Bolliger, Gudrun Carl, et al. (2007). "Methods to account for spatial autocorrelation in the analysis of species distributional data: A review". In: *Ecography* 30.5, pp. 609–628.

Dray, Stéphane, Pierre Legendre, and Pedro R Peres-Neto (2006). "Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)". In: *Ecological Modelling* 196.3-4, pp. 483–493.

Dresow, Max (2019). "Macroevolution evolving: Punctuated equilibria and the roots of Stephen Jay Gould's second macroevolutionary synthesis". In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 75, pp. 15–23.

Droser, Mary L, Lidya G Tarhan, and James G Gehling (2017). "The rise of animals in a changing environment: Global ecological innovation in the late Ediacaran". In: *Annual Review of Earth and Planetary Sciences* 45, pp. 593–617.

Dunhill, Alexander M and Matthew A Wills (2015). "Geographic range did not confer resilience to extinction in terrestrial vertebrates at the end-Triassic crisis". In: *Nature Communications* 6, 7980.

Dunn, Frances S, Alexander G Liu, Dmitriy V Grazhdankin, Philip Vixseboxse, Joseph Flannery-Sutherland, Emily Green, et al. (2021). "The developmental biology of Charnia and the eumetazoan affinity of the Ediacaran rangeomorphs". In: *Science Advances* 7.30, eabe0291.

Dunne, Jennifer A, Richard J Williams, Neo D Martinez, Rachel A Wood, and Douglas H Erwin (2008). "Compilation and network analyses of Cambrian food webs". In: *PLoS Biol* 6.4, e102.

Dutkiewicz, Stephanie, J Jeffrey Morris, Michael J Follows, Jeffery Scott, Orly Levitan, Sonya T Dyhrman, et al. (2015). "Impact of ocean acidification on the structure of future phytoplankton communities". In: *Nature Climate Change* 5.11, pp. 1002–1006.

Dyer, Lee A and Deborah Letourneau (2003). "Top-down and bottom-up diversity cascades in detrital vs. living food webs". In: *Ecology Letters* 6.1, pp. 60–68.

Eldredge, Niles and Stephen Jay Gould (1972). "Punctuated equilibria: An alternative to phyletic gradualism". In: *Models in Paleobiology*. Ed. by T J M Schopf. San Francisco: Cooper & Co., pp. 82–115.

Eldredge, Niles, John N Thompson, Paul M Brakefield, Sergey Gavrilets, David Jablonski, Jeremy BC Jackson, et al. (2005). "The dynamics of evolutionary stasis". In: *Paleobiology* 31.S2, pp. 133–145.

Elser, James J, John H Schampel, Marcia Kyle, James Watts, Evan W Carson, Thomas E Dowling, et al. (2005). "Response of grazing snails to phosphorus

enrichment of modern stromatolitic microbial communities". In: *Freshwater Biology* 50.11, pp. 1826–1835.

Elser, James J, James Watts, John H Schampel, and Jack Farmer (2006). "Early Cambrian food webs on a trophic knife-edge? A hypothesis and preliminary data from a modern stromatolite-based ecosystem". In: *Ecology Letters* 9.3, pp. 295–303.

Erwin, Douglas H (2015a). "A public goods approach to major evolutionary innovations". In: *Geobiology* 13.4, pp. 308–315.

— (2015b). "Novelty and innovation in the history of life". In: *Current Biology* 25.19, R930–R940.

— (2015c). "Was the Ediacaran–Cambrian radiation a unique evolutionary event?" In: *Paleobiology* 41.1, pp. 1–15.

Erwin, Douglas H and Eric H Davidson (2009). "The evolution of hierarchical gene regulatory networks". In: *Nature Reviews Genetics* 10.2, pp. 141–148.

Erwin, Douglas H, Marc Laflamme, Sarah M Tweedt, Erik A Sperling, Davide Pisani, and Kevin J Peterson (2011). "The Cambrian conundrum: Early divergence and later ecological success in the early history of animals". In: *Science* 334.6059, pp. 1091–1097.

Erwin, Douglas H and James W Valentine (2013). "The Cambrian Explosion". In: Genwodd Village, Colorado: Roberts and Company, 416 p.

Erwin, Douglas H, James W Valentine, and J John Sepkoski Jr (1987). "A comparative study of diversification events: The early Paleozoic versus the Mesozoic". In: *Evolution* 41.6, pp. 1177–1186.

Estes, James A, John Terborgh, Justin S Brashares, Mary E Power, Joel Berger, William J Bond, et al. (2011). "Trophic downgrading of planet Earth". In: *Science* 333.6040, pp. 301–306.

Ezard, Thomas HG, Tracy Aze, Paul N Pearson, and Andy Purvis (2011). "Interplay between changing climate and species' ecology drives macroevolutionary dynamics". In: *Science* 332.6027, pp. 349–351.

Felsenstein, Joseph (1973). "Maximum-likelihood estimation of evolutionary trees from continuous characters". In: *American Journal of Human Genetics* 25.5, p. 471.

— (1985). "Phylogenies and the comparative method". In: *The American Naturalist* 125.1, pp. 1–15.

Fenton, Isabel S, Paul N Pearson, Tom Dunkley Jones, and Andy Purvis (2016). "Environmental predictors of diversity in recent planktonic foraminifera as recorded in marine sediments". In: *PloS One* 11.11, e0165522.

Fenton, Isabel S, Adam D Woodhouse, Tracy Aze, David Lazarus, Johan Renaudie, Alexander M Dunhill, et al. (2021). "Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences". In: *Scientific Data* 8.160, pp. 1–9.

Field, Christopher B, Michael J Behrenfeld, James T Randerson, and Paul Falkowski (1998). "Primary production of the biosphere: Integrating terrestrial and oceanic components". In: *Science* 281.5374, pp. 237–240.

Filippelli, Gabriel M (2010). "Phosphorus and the gust of fresh air". In: *Nature* 467.7319, pp. 1052–1053.

Finkel, Zoe Vanessa (2007). "Does phytoplankton cell size matter? The evolution of modern marine food webs". In: *Evolution of Primary Producers in the Sea*. Ed. by Paul G Falkowski and Andrew H Knoll. Elsevier. Chap. 15, pp. 333–350.

Finnegan, Seth, Sean C Anderson, Paul G Harnik, Carl Simpson, Derek P Tittensor, Jarrett E Byrnes, et al. (2015). "Paleontological baselines for evaluating extinction risk in the modern oceans". In: *Science* 348.6234, pp. 567–570.

Finnegan, Seth and Mary L Droser (2008). "Body size, energetics, and the Ordovician restructuring of marine ecosystems". In: *Paleobiology* 34.3, pp. 342–359.

Finnegan, Seth, Craig M McClain, Matthew A Kosnik, and Jonathan L Payne (2011). "Escargots through time: An energetic comparison of marine gastropod assemblages before and after the Mesozoic Marine Revolution". In: *Paleobiology* 37.2, pp. 252–269.

FitzJohn, Richard G (2012). "diversitree: Comparative phylogenetic analyses of diversification in R". In: *Methods in Ecology and Evolution* 3.6, pp. 1084–1092.

Folk, Ryan A, Rebecca L Stubbs, Mark E Mort, Nico Cellinese, Julie M Allen, Pamela S Soltis, et al. (2019). "Rates of niche and phenotype evolution lag behind diversification in a temperate radiation". In: *Proceedings of the National Academy of Sciences* 116.22, pp. 10874–10882.

Foote, Michael, James S Crampton, Alan G Beu, Bruce A Marshall, Roger A Cooper, Phillip A Maxwell, et al. (2007). "Rise and fall of species occupancy in Cenozoic fossil mollusks". In: *Science* 318.5853, pp. 1131–1134.

Foster, William J, Katrin Heindel, Sylvain Richoz, Jana Gliwa, Daniel J Lehrmann, Aymon Baud, et al. (2019). "Suppressed competitive exclusion enabled the proliferation of Permian/Triassic boundary microbialites". In: *The Depositional Record* 00, pp. 1–13.

Frair, Jacqueline L, John Fieberg, Mark Hebblewhite, Francesca Cagnacci, Nicholas J DeCesare, and Luca Pedrotti (2010). "Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1550, pp. 2187–2200.

Fretwell, Stephen D (1987). "Food chain dynamics: The central theory of ecology?" In: *Oikos*, pp. 291–301.

Futuyma, Douglas J (2015). "Can modern evolutionary theory explain macroevolution?" In: *Macroevolution: Explanation, Interpretation and Evidence*. Ed. by Emanuele Serrelli and Nathalie Gontier. Cham, Switzerland: Springer, pp. 29–85.

Gaines, Steven D, Sarah E Lester, Ginny Eckert, Brian P Kinlan, Rafe Sagarin, and Brian Gaylord (2009). "Dispersal and geographic ranges in the sea". In: *Marine Macroecology*. Chicago, IL: The University of Chicago Press, pp. 227–249.

Garland Jr, Theodore, Allan W Dickerman, Christine M Janis, and Jason A Jones (1993). "Phylogenetic analysis of covariance by computer simulation". In: *Systematic Biology* 42.3, pp. 265–292.

Gaston, Kevin J (2003). *The structure and dynamics of geographic ranges*. Oxford University Press, 278 pp.

Gaston, Kevin J, Tim M Blackburn, and John H Lawton (1997). "Interspecific abundance-range size relationships: An appraisal of mechanisms". In: *Journal of Animal Ecology*, pp. 579–601.

Gaston, Kevin J and John H Lawton (1990). "Effects of scale and habitat on the relationship between regional distribution and local abundance". In: *Oikos*, pp. 329–335.

Gibling, Martin R and Neil S Davies (2012). "Palaeozoic landscapes shaped by plant evolution". In: *Nature Geoscience* 5.2, pp. 99–105.

Giraud, Fabienne (2009). "Calcareous nannofossil productivity and carbonate production across the Middle-Late Jurassic transition in the French Subalpine Basin". In: *Geobios* 42.6, pp. 699–714.

Godsoe, William and Bradley S Case (2015). "Accounting for shifts in the frequency of suitable environments when testing for niche overlap". In: *Methods in Ecology and Evolution* 6.1, pp. 59–66.

Goldberg, Emma E and Boris Igić (2012). "Tempo and mode in plant breeding system evolution". In: *Evolution* 66.12, pp. 3701–3709.

Gómez-Rodríguez, Carola, Andrés Baselga, and John J Wiens (2015). "Is diversification rate related to climatic niche width?" In: *Global Ecology and Biogeography* 24.4, pp. 383–395.

Good, Irving J (1953). "The population frequencies of species and the estimation of population parameters". In: *Biometrika* 40.3-4, pp. 237–264.

Götzenberger, Lars, Francesco de Bello, Kari Anne Bråthen, John Davison, Anne Dubuis, Antoine Guisan, et al. (2012). "Ecological assembly rules in plant communities—approaches, patterns and prospects". In: *Biological Reviews* 87.1, pp. 111–127.

Gould, Stephen Jay (1980). "The promise of paleobiology as a nomothetic, evolutionary discipline". In: *Paleobiology* 6.1, pp. 96–118.

— (1989). "Punctuated equilibrium in fact and theory". In: *Journal of social and biological structures* 12.2-3, pp. 117–136.

Gould, Stephen Jay and C Bradford Calloway (1980). "Clams and brachiopods—ships that pass in the night". In: *Paleobiology* 6.4, pp. 383–396.

Gould, Stephen Jay and Niles Eldredge (1977). "Punctuated equilibria: The tempo and mode of evolution reconsidered". In: *Paleobiology*, pp. 115–151.

Grafen, Alan (1989). "The phylogenetic regression". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 326.1233, pp. 119–157.

Gyldenfeldt, Anna-B v, Jörn Carstens, and Jens Meincke (2000). "Estimation of the catchment area of a sediment trap by means of current meters and foraminiferal tests". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 47.9-11, pp. 1701–1717.

Hairston, Nelson G, Frederick E Smith, and Lawrence B Slobodkin (1960). "Community structure, population control, and competition". In: *American Naturalist*, pp. 421–425.

Hall, Peter and Terence Tao (2002). "Relative efficiencies of kernel and local likelihood density estimators". In: *Journal of the Royal Statistical Society B: Statistical Methodology* 64.3, pp. 537–547.

Hallam, Anthony and Paul B Wignall (1997). *Mass Extinctions and Their Aftermath*. Oxford, UK: Oxford University Press, 328 pp.

Hannisdal, Bjarte, Kristian Agasøster Haaga, Trond Reitan, David Diego, and Lee Hsiang Liow (2017). "Common species link global ecosystems to climate change: Dynamical evidence in the planktonic fossil record". In: *Proceedings of the Royal Society B: Biological Sciences* 284.1858, 20170722.

Hansen, Thomas F and Emilia P Martins (1996). "Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data". In: *Evolution* 50.4, pp. 1404–1417.

Hansen, Thor A (1980). "Influence of larval dispersal and geographic distribution on species longevity in neogastropods". In: *Paleobiology*, pp. 193–207.

Harmon, Luke J (2019). *Phylogenetic Comparative Methods*. v1.4. CC-BY-4.0. URL: https://ecoevorxiv.org/e3xnr/download.

Harmon, Luke J, Jonathan B Losos, T Jonathan Davies, Rosemary G Gillespie, John L Gittleman, W Bryan Jennings, et al. (2010). "Early bursts of body size and shape evolution are rare in comparative data". In: *Evolution* 64.8, pp. 2385–2396.

Harnik, Paul G, Carl Simpson, and Jonathan L Payne (2012). "Long-term differences in extinction risk among the seven forms of rarity". In: *Proceedings of the Royal Society B: Biological Sciences* 279.1749, pp. 4969–4976.

Harper, David AT, Ren-Bin Zhan, and Jisuo Jin (2015). "The Great Ordovician Biodiversification Event: reviewing two decades of research on diversity's big bang illustrated by mainly brachiopod data". In: *Palaeoworld* 24.1-2, pp. 75–85.

Hartmann, Jens, Nils Moosdorf, Ronny Lauerwald, Matthias Hinderer, and A Joshua West (2014). "Global chemical weathering and associated P-release—The role of lithology, temperature and soil properties". In: *Chemical Geology* 363, pp. 145–163.

Hayes, John M, Harald Strauss, and Alan J Kaufman (1999). "The abundance of 13C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma". In: *Chemical Geology* 161.1-3, pp. 103–125.

Heim, Noel A, Matthew L Knope, Ellen K Schaal, Steve C Wang, and Jonathan L Payne (2015). "Cope's rule in the evolution of marine animals". In: *Science* 347.6224, pp. 867–870.

Hendricks, Jonathan R, Erin E Saupe, Corinne E Myers, Elizabeth J Hermsen, and Warren D Allmon (2014). "The generification of the fossil record". In: *Paleobiology* 40.4, pp. 511–528.

Herrera-Alsina, Leonel, Alex L Pigot, Hanno Hildenbrandt, and Rampal S Etienne (2018). "The influence of ecological and geographic limits on the evolution of species distributions and diversity". In: *Evolution* 72.10, pp. 1978–1991.

Hijmans, Robert J (2019). *raster: Geographic data analysis and modeling*. URL: https://CRAN.R-project.org/package=raster.

Hoffman, Paul F, Dorian S Abbot, Yosef Ashkenazy, Douglas I Benn, Jochen J Brocks, Phoebe A Cohen, et al. (2017). "Snowball Earth climate dynamics and Cryogenian geology-geobiology". In: *Science Advances* 3.11, e1600983.

Holland, Barbara R, Saan Ketelaar-Jones, Aiden R O'Mara, Michael D Woodhams, and Gregory J Jordan (2020). "Accuracy of ancestral state reconstruction for non-neutral traits". In: *Scientific Reports* 10, 7644.

Holland, Steven M and Andrew Zaffos (2011). "Niche conservatism along an onshore-offshore gradient". In: *Paleobiology* 37.2, pp. 270–286.

Horne, Jon S, Edward O Garton, and Kimberly A Sager-Fradkin (2007). "Correcting home-range models for observation bias". In: *The Journal of Wildlife Management* 71.3, pp. 996–1001.

Hsieh, TC, KH Ma, and Anne Chao (2016). "iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers)". In: *Methods in Ecology and Evolution* 7.12, pp. 1451–1456.

Hunt, G (2019). *paleoTS: Analyze paleontological time-series*. URL: https://CRAN.R-project.org/package=paleoTS.

Hunt, Gene (2006). "Fitting and comparing models of phyletic evolution: Random walks and beyond". In: *Paleobiology* 32.4, pp. 578–601.

Hunt, Gene (2007). "The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages". In: *Proceedings of the National Academy of Sciences* 104.47, pp. 18404–18408.

— (2008). "Evolutionary patterns within fossil lineages: Model-based assessment of modes, rates, punctuations and process". In: *From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century*. Ed. by Richard K Bambach and PH Kelley. The Paleontological Society Papers. New Havent, CT: The Paleontological Society, pp. 117–131.

— (2012). "Measuring rates of phenotypic evolution and the inseparability of tempo and modemeasuring rates of phenotypic evolution". In: *Paleobiology* 38.3, pp. 351–373.

Hunt, Gene, Melanie J Hopkins, and Scott Lidgard (2015). "Simple versus complex models of trait evolution and stasis as a response to environmental change". In: *Proceedings of the National Academy of Sciences* 112.16, pp. 4885–4890.

Hunter, Mark D (2001). "Multiple approaches to estimating the relative importance of top-down and bottom-up forces on insect populations: Experiments, life tables, and time-series analysis". In: *Basic and Applied Ecology* 2.4, pp. 295–309.

Hunter, Mark D and Peter W Price (1992). "Playing chutes and ladders: Heterogeneity and the relative roles of bottom-up and top-down forces in natural communities". In: *Ecology*, pp. 724–732.

Hutchins, David A and Feixue Fu (2017). "Microorganisms and ocean global change". In: *Nature Microbiology* 2.6, p. 17058.

Irwin, Andrew J, Zoe V Finkel, Frank E Müller-Karger, and Luis Troccoli Ghinaglia (2015). "Phytoplankton adapt to changing ocean environments". In: *Proceedings of the National Academy of Sciences* 112.18, pp. 5762–5766.

Ives, Anthony R and Theodore Garland Jr (2010). "Phylogenetic logistic regression for binary dependent variables". In: *Systematic Biology* 59.1, pp. 9–26.

Ives, Anthony R, Peter E Midford, and Theodore Garland Jr (2007). "Within-species variation and measurement error in phylogenetic comparative methods". In: *Systematic Biology* 56.2, pp. 252–270.

Jablonski, David (1986). "Larval ecology and macroevolution in marine invertebrates". In: *Bulletin of Marine Science* 39.2, pp. 565–587.

— (1998). "Geographic variation in the molluscan recovery from the end-Cretaceous extinction". In: *Science* 279.5355, pp. 1327–1330.

Jablonski, David and Kaustuv Roy (2003). "Geographical range and speciation in fossil and living molluscs". In: *Proceedings of the Royal Society B: Biological Sciences* 270.1513, pp. 401–406.

Jackson, Jeremy BC (2008). "Ecological extinction and evolution in the brave new ocean". In: *Proceedings of the National Academy of Sciences* 105, pp. 11458–11465.

Jackson, Jeremy BC and Alan H Cheetham (1999). "Tempo and mode of speciation in the sea". In: *Trends in Ecology & Evolution* 14.2, pp. 72–77.

Jones, Michael C (1991). "Kernel density estimation for length biased data". In: *Biometrika* 78.3, pp. 511–519.

Jonkers, Lukas, Helmut Hillebrand, and Michal Kucera (2019). "Global change drives modern plankton communities away from the pre-industrial state". In: *Nature* 570.7761, pp. 372–375.

Keating, Joseph N, Philip CJ Donoghue, and Michael J Benton (2021). *In search of ancestors: Benchmarking parsimony and likelihood methods of ancestral state*

estimation. Conference Paper. URL: https://www.palass.org/sites/default/files/media/annual_meetings/2021/abstracts%20booklet%202021.mod_.pdf.

Kemp, WM and WR Boynton (2004). "Productivity, trophic structure, and energy flow in the steady-state ecosystems of Silver Springs, Florida". In: *Ecological Modelling* 178.1-2, pp. 43–49.

Kiessling, Wolfgang (2005). "Habitat effects and sampling bias on Phanerozoic reef distribution". In: *Facies* 51.1-4, pp. 24–32.

Kiessling, Wolfgang and Martin Aberhan (2007a). "Environmental determinants of marine benthic biodiversity dynamics through Triassic–Jurassic time". In: *Paleobiology* 33.3, pp. 414–434.

— (2007b). "Geographical distribution and extinction risk: Lessons from Triassic–Jurassic marine benthic organisms". In: *Journal of Biogeography* 34.9, pp. 1473–1489.

Kiessling, Wolfgang and Adam T Kocsis (2016). "Adding fossil occupancy trajectories to the assessment of modern extinction risk". In: *Biology Letters* 12.10, 20150813.

Klompmaker, Adiël A and Seth Finnegan (2018). "Extreme rarity of competitive exclusion in modern and fossil marine benthic ecosystems". In: *Geology* 46.8, pp. 723–726.

Klompmaker, Adiël A, Michał Kowalewski, John Warren Huntley, and Seth Finnegan (2017). "Increase in predator-prey size ratios throughout the Phanerozoic history of marine ecosystems". In: *Science* 356.6343, pp. 1178–1180.

Klug, Christian, Bjoern Kroeger, Wolfgang Kiessling, Gary L Mullins, Thomas Servais, Jiří Frýda, et al. (2010). "The Devonian nekton revolution". In: *Lethaia* 43.4, pp. 465–477.

Knoll, Andrew H and Michael J Follows (2016). "A bottom-up perspective on ecosystem change in Mesozoic oceans". In: *Proceedings of the Royal Society B: Biological Sciences* 283.1841, 20161755.

Kocsis, Adam T, Carl J Reddin, and Wolfgang Kiessling (2018). "The biogeographical imprint of mass extinctions". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1878, 20180232.

Kowalewski, Michał and Richard K Bambach (2008). "The limits of paleontological resolution". In: *High-resolution Approaches in Stratigraphic Paleontology*. Springer, pp. 1–48.

Kozak, Kenneth H and John J Wiens (2010). "Accelerated rates of climatic-niche evolution underlie rapid species diversification". In: *Ecology Letters* 13.11, pp. 1378–1389.

— (2016). "What explains patterns of species richness? The relative importance of climatic-niche evolution, morphological evolution, and ecological limits in salamanders". In: *Ecology and Evolution* 6.16, pp. 5940–5949.

Kucera, M (2007). "Planktonic foraminifera as tracers of past oceanic environments, Chapter 6". In: ed. by Claude Hillaire–Marcel and Anne De Vernal. Vol. 1. Developments in Marine Geology. Elsevier, pp. 213–262.

Kucera, Michal, Mara Weinelt, Thorsten Kiefer, Uwe Pflaumann, Angela Hayes, Martin Weinelt, et al. (2005). "Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: Multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans". In: *Quaternary Science Reviews* 24.7-9, pp. 951–998.

Laakso, Thomas A and Daniel P Schrag (2019). "A small marine biosphere in the Proterozoic". In: *Geobiology* 17.2, pp. 161–171.

Laakso, Thomas A, Erik A Sperling, David T Johnston, and Andrew H Knoll (2020). "Ediacaran reorganization of the marine phosphorus cycle". In: *Proceedings of the National Academy of Sciences* 117.22, pp. 11961–11967.

Laland, Kevin N, F John Odling-Smee, and Marcus W Feldman (1999). "Evolutionary consequences of niche construction and their implications for ecology". In: *Proceedings of the National Academy of Sciences* 96.18, pp. 10242–10247.

Laland, Kevin N and Kim Sterelny (2006). "Perspective: Seven reasons (not) to neglect niche construction". In: *Evolution* 60.9, pp. 1751–1762.

Lande, Russell (1976). "Natural selection and random genetic drift in phenotypic evolution". In: *Evolution*, pp. 314–334.

Laube, Irina, Catherine H Graham, and Katrin Böhning-Gaese (2013). "Intra-generic species richness and dispersal ability interact to determine geographic ranges of birds". In: *Global Ecology and Biogeography* 22.2, pp. 223–232.

Laundré, John W, Lucina Hernández, and William J Ripple (2010). "The landscape of fear: Ecological implications of being afraid". In: *The Open Ecology Journal* 3.1.

Lavergne, Sébastien, Margaret EK Evans, Ian J Burfield, Frederic Jiguet, and Wilfried Thuiller (2013). "Are species' responses to global change predicted by past niche evolution?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1610, 20120091.

Lazarus, David (1983). "Speciation in pelagic Protista and its study in the planktonic microfossil record: A review". In: *Paleobiology* 9.4, pp. 327–340.

Lefebvre, Vincent, Thomas Servais, Louis François, and Olivier Averbuch (2010). "Did a Katian large igneous province trigger the Late Ordovician glaciation?: A hypothesis tested with a carbon cycle model". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 296.3-4, pp. 310–319.

Lenoir, Jonathan, Romain Bertrand, Lise Comte, Luana Bourgeaud, Tarek Hattab, Jérôme Murienne, et al. (2020). "Species better track climate warming in the oceans than on land". In: *Nature Ecology & Evolution*, pp. 1–16.

Lenski, Richard E, Michael J Wiser, Noah Ribeck, Zachary D Blount, Joshua R Nahum, J Jeffrey Morris, et al. (2015). "Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1821, 20152292.

Lenton, Timothy M, Richard A Boyle, Simon W Poulton, Graham A Shields-Zhou, and Nicholas J Butterfield (2014). "Co-evolution of eukaryotes and ocean oxygenation in the Neoproterozoic era". In: *Nature Geoscience* 7.4, pp. 257–265.

Lenton, Timothy M, Michael Crouch, Martin Johnson, Nuno Pires, and Liam Dolan (2012). "First plants cooled the Ordovician". In: *Nature Geoscience* 5.2, pp. 86–89.

Liow, Lee Hsiang, Trond Reitan, and Paul G Harnik (2015). "Ecological interactions on macroevolutionary time scales: Clams and brachiopods are more than ships that pass in the night". In: *Ecology Letters* 18.10, pp. 1030–1039.

Liow, Lee Hsiang and Nils Chr Stenseth (2007). "The rise and fall of species: Implications for macroevolutionary and macroecological studies". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1626, pp. 2745–2752.

Lisiecki, Lorraine E and Maureen E Raymo (2005). "A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records". In: *Paleoceanography* 20.1.

Litsios, Glenn and Nicolas Salamin (2012). "Effects of phylogenetic signal on ancestral state reconstruction". In: *Systematic Biology* 61.3, pp. 533–538.

Lv, Xue, Lin Xia, Deyan Ge, Yongjie Wu, and Qisen Yang (2016). "Climatic niche conservatism and ecological opportunity in the explosive radiation of arvicoline rodents (Arvicolinae, Cricetidae)". In: *Evolution* 70.5, pp. 1094–1104.

Malzahn, Arne Michael, Nicole Aberle, Catriona Clemmesen, and Maarten Boersma (2007). "Nutrient limitation of primary producers affects planktivorous fish condition". In: *Limnology and Oceanography* 52.5, pp. 2062–2071.

Martin, Ronald and Antonietta Quigg (2012). "Evolving phytoplankton stoichiometry fueled diversification of the marine biosphere". In: *Geosciences* 2.2, pp. 130–146.

Martin, Ronald E (1996). "Secular increase in nutrient levels through the Phanerozoic: Implications for productivity, biomass, and diversity of the marine biosphere". In: *Palaios* 11.3, pp. 209–219.

— (1999). "Taphonomy and temporal resolution of foraminiferal assemblages". In: *Modern Foraminifera*. Springer, pp. 281–298.

Martin, Ronald E and Thomas Servais (2020). "Did the evolution of the phytoplankton fuel the diversification of the marine biosphere?" In: *Lethaia* 53.1, pp. 5–31.

Martínez-Meyer, E and A Townsend Peterson (2006). "Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition". In: *Journal of Biogeography* 33.10, pp. 1779–1789.

Martínez-Meyer, Enrique, A Townsend Peterson, and William W Hargrove (2004). "Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity". In: *Global Ecology and Biogeography* 13.4, pp. 305–314.

Mathes, Gregor H., Jeroen van Dijk, Wolfgang Kiessling, and Manuel J. Steinbauer (2021). "Extinction risk controlled by interaction of long-term and short-term climate change". In: *Nature Ecology & Evolution* 5, pp. 304–310.

Matthews, Jack J, Alexander G Liu, Chuan Yang, Duncan McIlroy, Bruce Levell, and Daniel J Condon (2021). "A chronostratigraphic framework for the rise of the Ediacaran macrobiota: New constraints from Mistaken Point Ecological Reserve, Newfoundland". In: *Geological Society of America Bulletin* 133.3-4, pp. 612–624.

Matthews, Kara J, Kayla T Maloney, Sabin Zahirovic, Simon E Williams, Maria Seton, and R Dietmar Mueller (2016). "Global plate boundary evolution and kinematics since the late Paleozoic". In: *Global and Planetary Change* 146, pp. 226–250.

Maurer, Brian A and M Philip Nott (2001). "Geographic range fragmentation and the evolution of biological diversity". In: *Biodiversity Dynamics*. Ed. by Michael L McKinney and James A Drake. Columbia University Press, pp. 31–50.

Mayr, Ernst (1989). "Speciational evolution or punctuated equilibria". In: *Journal of Social and Biological Structures* 12.2-3, pp. 137–158.

McGhee Jr, George R, Peter M Sheehan, David J Bottjer, and Mary L Droser (2004). "Ecological ranking of Phanerozoic biodiversity crises: Ecological and taxonomic severities are decoupled". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 211.3-4, pp. 289–297.

— (2012). "Ecological ranking of Phanerozoic biodiversity crises: The Serpukhovian (early Carboniferous) crisis had a greater ecological impact than the end-Ordovician". In: *Geology* 40.2, pp. 147–150.

McGill, Brian J, Jonathan M Chase, Joaquín Hortal, Isaac Overcast, Andrew J Rominger, James Rosindell, et al. (2019). "Unifying macroecology and

macroevolution to answer fundamental questions about biodiversity". In: *Global Ecology and Biogeography* 28.12, pp. 1925–1936.

Meilland, Julie, Michael Siccha, Maike Kaffenberger, Jelle Bijma, and Michal Kucera (2021). "Population dynamics and reproduction strategies of planktonic foraminifera in the open ocean". In: *Biogeosciences* 18.20, pp. 5789–5809.

Meilland, Julie, Michael Siccha, Manuel FG Weinkauf, Lukas Jonkers, Raphael Morard, Ulrike Baranowski, et al. (2019). "Highly replicated sampling reveals no diurnal vertical migration but stable species-specific vertical habitats in planktonic foraminifera". In: *Journal of Plankton Research* 41.2, pp. 127–141.

Miller, Arnold I and Shuguang Mao (1995). "Association of orogenic activity with the Ordovician radiation of marine life". In: *Geology* 23.4, pp. 305–308.

Miller, Arnold I and J John Sepkoski Jr (1988). "Modeling bivalve diversification: The effect of interaction on a macroevolutionary system". In: *Paleobiology* 14.4, pp. 364–369.

Mills, Daniel B and Donald E Canfield (2014). "Oxygen and animal evolution: Did a rise of atmospheric oxygen "trigger" the origin of animals?" In: *BioEssays* 36.12, pp. 1145–1155.

Mitchell, Emily G and Nicholas J Butterfield (2018). "Spatial analyses of Ediacaran communities at Mistaken Point". In: *Paleobiology* 44.1, pp. 40–57.

Modesto, Sean P (2020). "The disaster taxon Lystrosaurus: A paleontological myth". In: *Frontiers in Earth Science* 8, 617.

Mondal, Subhronil and Peter J Harries (2016). "Phanerozoic trends in ecospace utilization: The bivalve perspective". In: *Earth-Science Reviews* 152, pp. 106–118.

Moore, CM, MM Mills, KR Arrigo, I Berman-Frank, L Bopp, PW Boyd, et al. (2013). "Processes and patterns of oceanic nutrient limitation". In: *Nature Geoscience* 6.9, pp. 701–710.

Morard, Raphaël, Kate F Darling, Frédéric Mahé, Stéphane Audic, Yurika Ujiié, Agnes KM Weiner, et al. (2015). "PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution". In: *Molecular Ecology Resources* 15.6, pp. 1472–1485.

Morton, Brian (1990). "Corals and their bivalve borers—the evolution of a symbiosis". In: *Proceedings of a Memorial Symposium in Honour of Sir Charles Maurice Younge.* Ed. by Brian Morton. Hong Kong: Hong Kong University Press.

Moss, David K, Linda C Ivany, Emily J Judd, Patrick W Cummings, Claire E Bearden, Woo-Jun Kim, et al. (2016). "Lifespan, growth rate, and body size across latitude in marine Bivalvia, with implications for Phanerozoic evolution". In: *Proceedings of the Royal Society B: Biological Sciences* 283.1836, 20161364.

Müller, R Dietmar, John Cannon, Xiaodong Qin, Robin J Watson, Michael Gurnis, Simon Williams, et al. (2018). "GPlates: Building a virtual Earth through deep time". In: *Geochemistry, Geophysics, Geosystems* 19.7, pp. 2243–2261.

Munday, Philip L, Robert R Warner, Keyne Monro, John M Pandolfi, and Dustin J Marshall (2013). "Predicting evolutionary responses to climate change in the sea". In: *Ecology Letters* 16.12, pp. 1488–1500.

Myers, Corinne E and Bruce S Lieberman (2011). "Sharks that pass in the night: Using Geographical Information Systems to investigate competition in the Cretaceous Western Interior Seaway". In: *Proceedings of the Royal Society B: Biological Sciences* 278.1706, pp. 681–689.

Myers, Ransom A, Julia K Baum, Travis D Shepherd, Sean P Powers, and Charles H Peterson (2007). "Cascading effects of the loss of apex predatory sharks from a coastal ocean". In: *Science* 315.5820, pp. 1846–1850.

Novack-Gottshall, Philip M (2008). "Ecosystem-wide body-size trends in Cambrian—Devonian marine invertebrate lineages". In: *Paleobiology* 34.2, pp. 210–228.

Novack-Gottshall, Philip M and Michael A Lanier (2008). "Scale-dependence of Cope's rule in body size evolution of Paleozoic brachiopods". In: *Proceedings of the National Academy of Sciences* 105.14, pp. 5430–5434.

Nürnberg, Sabine and Martin Aberhan (2013). "Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves". In: *Paleobiology* 39.3, pp. 360–372.

Ochoa-Hueso, Raúl, Manuel Delgado-Baquerizo, Anita C Risch, Maarten Schrama, Elly Morriën, S Henrik Barmentlo, et al. (2021). "Ecosystem coupling: A unifying framework to understand the functioning and recovery of ecosystems". In: *One Earth* 4.7, pp. 951–966.

Oksanen, Jari, F Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, et al. (2019). *vegan: Community ecology package*. URL: `https://CRAN.R-project.org/package=vegan`.

Pagel, Mark (1999). "Inferring the historical patterns of biological evolution". In: *Nature* 401.6756, pp. 877–884.

Pagel, Mark, Andrew Meade, and Daniel Barker (2004). "Bayesian estimation of ancestral character states on phylogenies". In: *Systematic Biology* 53.5, pp. 673–684.

Paradis, Emmanuel and Julien Claude (2002). "Analysis of comparative data using generalized estimating equations". In: *Journal of Theoretical Biology* 218.2, pp. 175–185.

Patzkowsky, Mark E and Steven M Holland (2012). *Stratigraphic Paleobiology: Understanding the Distribution of Fossil Taxa in Time and Space*. Chicago, IL: University of Chicago Press, 256 pp.

— (2016). "Biotic invasion, niche stability, and the assembly of regional biotas in deep time: Comparison between faunal provinces". In: *Paleobiology* 42.3, pp. 359–379.

Pavia, Jose M. (2015). "Testing goodness-of-fit with the kernel density estimator: GoFKernel". In: *Journal of Statistical Software, Code Snippets* 66.1, pp. 1–27.

Payne, Jonathan L and Seth Finnegan (2006). "Controls on marine animal biomass through geological time". In: *Geobiology* 4.1, pp. 1–10.

Payne, Jonathan L, Noel A Heim, Matthew L Knope, and Craig R McClain (2014). "Metabolic dominance of bivalves predates brachiopod diversity decline by more than 150 million years". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1783, 20133122.

Pearman, Peter B, Christophe F Randin, Olivier Broennimann, Pascal Vittoz, Willem O van der Knaap, Robin Engler, et al. (2008). "Prediction of plant species distributions across six millennia". In: *Ecology Letters* 11.4, pp. 357–369.

Peltier, W Richard, DF Argus, and R Drummond (2015). "Space geodesy constrains ice age terminal deglaciation: The global ICE-6G C (VM5a) model". In: *Journal of Geophysical Research: Solid Earth* 120.1, pp. 450–487.

Pennell, Matthew W, Jonathan M Eastman, Graham J Slater, Joseph W Brown, Josef C Uyeda, Richard G FitzJohn, et al. (2014). "geiger v2.0: An expanded suite of

methods for fitting macroevolutionary models to phylogenetic trees". In: *Bioinformatics* 30.15, pp. 2216–2218.

Petersen, John E, W Michael Kemp, Rick Bartleson, Walter R Boynton, Chung-Chi Chen, Jeffrey C Cornwell, et al. (2003). "Multiscale experiments in coastal ecology: Improving realism and advancing theory". In: *BioScience* 53.12, pp. 1181–1197.

Peterson, A Townsend (2011). "Ecological niche conservatism: A time-structured review of evidence". In: *Journal of Biogeography* 38.5, pp. 817–827.

Peterson, A Townsend, Jorge Soberón, Richard G Pearson, Robert P Anderson, Enrique Martínez-Meyer, Miguel Nakamura, et al. (2011). *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton University Press, 314 pp.

Philippe, Nadège, Ludovic Pelosi, Richard E Lenski, and Dominique Schneider (2009). "Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with Escherichia coli". In: *Journal of Bacteriology* 191.3, pp. 909–921.

Phuong, Mark A, Michael E Alfaro, Gusti N Mahardika, Ristiyanti M Marwoto, Romanus Edy Prabowo, Thomas von Rintelen, et al. (2019). "Lack of signal for the impact of conotoxin gene diversity on speciation rates in cone snails". In: *Systematic Biology* 68.5, pp. 781–796.

Pigot, Alex L and Joseph A Tobias (2013). "Species interactions constrain geographic range expansion over evolutionary time". In: *Ecology Letters* 16.3, pp. 330–338.

Planavsky, Noah J, Olivier J Rouxel, Andrey Bekker, Stefan V Lalonde, Kurt O Konhauser, Christopher T Reinhard, et al. (2010). "The evolution of the marine phosphate reservoir". In: *Nature* 467.7319, pp. 1088–1090.

Plath, Klaus and Maarten Boersma (2001). "Mineral limitation of zooplankton: Stoichiometric constraints and optimal foraging". In: *Ecology* 82.5, pp. 1260–1269.

Pohl, Alexandre, David AT Harper, Yannick Donnadieu, Guillaume Le Hir, Elise Nardin, and Thomas Servais (2018). "Possible patterns of marine primary productivity during the Great Ordovician Biodiversification Event". In: *Lethaia* 51.2, pp. 187–197.

Polechová, Jitka and Nicholas H Barton (2015). "Limits to adaptation along environmental gradients". In: *Proceedings of the National Academy of Sciences* 112.20, pp. 6401–6406.

Polis, Gary A (1999). "Why are parts of the world green? Multiple factors control productivity and the distribution of biomass". In: *Oikos* 86 (1), pp. 3–15.

Polis, Gary A, Anna LW Sears, Gary R Huxel, Donald R Strong, and John Maron (2000). "When is a trophic cascade a trophic cascade?" In: *Trends in Ecology & Evolution* 15.11, pp. 473–475.

Porada, Philipp, TM Lenton, Alexandre Pohl, Bettina Weber, Luke Mander, Yannick Donnadieu, et al. (2016). "High potential for weathering and climate effects of non-vascular vegetation in the Late Ordovician". In: *Nature Communications* 7.1, pp. 1–13.

Probst, Wolfgang Nikolaus, Vanessa Stelzenmüller, and Heino Ove Fock (2012). "Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: An exemplary analysis of North Sea fish population indicators". In: *ICES Journal of Marine Science* 69.4, pp. 670–681.

Pu, Judy P, Samuel A Bowring, Jahandar Ramezani, Paul Myrow, Timothy D Raub, Ed Landing, et al. (2016). "Dodging snowballs: Geochronology of the Gaskiers glaciation and the first appearance of the Ediacaran biota". In: *Geology* 44.11, pp. 955–958.

Puillandre, N, P Bouchet, Thomas F Duda Jr, S Kauferstein, AJ Kohn, BM Olivera, et al. (2014). "Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea)". In: *Molecular Phylogenetics and Evolution* 78, pp. 290–303.

Pyron, R Alexander, Gabriel C Costa, Michael A Patten, and Frank T Burbrink (2015). "Phylogenetic niche conservatism and the evolutionary basis of ecological speciation". In: *Biological Reviews* 90.4, pp. 1248–1262.

Qiu, ZhongFeng, AM Doglioli, and François Carlotti (2014). "Using a Lagrangian model to estimate source regions of particles in sediment traps". In: *Science China Earth Sciences* 57.10, pp. 2447–2456.

Quigg, Antonietta, Zoe V Finkel, Andrew J Irwin, Yair Rosenthal, Tung-Yuan Ho, John R Reinfelder, et al. (2003). "The evolutionary inheritance of elemental stoichiometry in marine phytoplankton". In: *Nature* 425.6955, pp. 291–294.

Quintero, Ignacio and John J Wiens (2013). "Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species". In: *Ecology Letters* 16.8, pp. 1095–1103.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Computer Program. Vienna, Austria. URL: `https://www.R-project.org/`.

— (2020). *R: A Language and Environment for Statistical Computing*. Computer Program. Vienna, Austria. URL: `https://www.R-project.org/`.

— (2021). *R: A Language and Environment for Statistical Computing*. Computer Program. Vienna, Austria. URL: `https://www.R-project.org/`.

Rabosky, Daniel L (2013). "Diversity-dependence, ecological speciation, and the role of competition in macroevolution". In: *Annual Review of Ecology, Evolution, and Systematics* 44, pp. 481–502.

Racki, Grzegorz (2005). "Toward understanding Late Devonian global events: Few answers, many questions". In: *Understanding Late Devonian and Permian-Triassic Biotic and Climatic Events: Towards an Integrated Approach*. Ed. by D J Over, J R Morrow, and Paul B Wignall. Vol. 20. Elsevier. Chap. 2, pp. 5–36.

Ranjard, Louis, David Welch, Marie Paturel, and Stéphane Guindon (2014). "Modelling competition and dispersal in a statistical phylogeographic framework". In: *Systematic Biology* 63.5, pp. 743–752.

Rasmussen, Christian MØ, Clemens V Ullmann, Kristian G Jakobsen, Anders Lindskog, Jesper Hansen, Thomas Hansen, et al. (2016). "Onset of main Phanerozoic marine radiation sparked by emerging Mid Ordovician icehouse". In: *Scientific Reports* 6.1, pp. 1–9.

Raymond, Anne and Cheryl Metz (2004). "Ice and its consequences: Glaciation in the Late Ordovician, Late Devonian, Pennsylvanian-Permian, and Cenozoic compared". In: *The Journal of Geology* 112.6, pp. 655–670.

Record, Sydne, Matthew C Fitzpatrick, Andrew O Finley, Sam Veloz, and Aaron M Ellison (2013). "Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change". In: *Global Ecology and Biogeography* 22.6, pp. 760–771.

Reddin, Carl J, Paulina S Nätscher, Ádám T Kocsis, Hans-Otto Pörtner, and Wolfgang Kiessling (2020). "Marine clade sensitivities to climate change conform across timescales". In: *Nature Climate Change* 10.3, pp. 249–253.

Reinhard, Christopher T, Noah J Planavsky, Ben A Ward, Gordon D Love, Guillaume Le Hir, and Andy Ridgwell (2020). "The impact of marine nutrient abundance on early eukaryotic ecosystems". In: *Geobiology* 18.2, pp. 139–151.

Revell, Liam J, Luke J Harmon, and David C Collar (2008). "Phylogenetic signal, evolutionary process, and rate". In: *Systematic Biology* 57.4, pp. 591–601.

Reznick, David N, Ronald D Bassar, Corey A Handelsman, Cameron K Ghalambor, Jeff Arendt, Tim Coulson, et al. (2019). "Eco-evolutionary feedbacks predict the time course of rapid life-history evolution". In: *American Naturalist* 194.5, pp. 671–692.

Reznick, David N and Robert E Ricklefs (2009). "Darwin's bridge between microevolution and macroevolution". In: *Nature* 457.7231, pp. 837–842.

Richter, Daniel J, Romain Watteaux, Thomas Vannier, Jade Leconte, Paul Frémont, Gabriel Reygondeau, et al. (2019). "Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems". In: *bioRxiv*, 867739. URL: https://doi.org/10.1101/867739.

Ricklefs, Robert E (2004). "A comprehensive framework for global patterns in biodiversity". In: *Ecology Letters* 7.1, pp. 1–15.

Ridgwell, Andy and Richard E Zeebe (2005). "The role of the global carbonate cycle in the regulation and evolution of the Earth system". In: *Earth and Planetary Science Letters* 234.3-4, pp. 299–315.

Rillo, Marina C, Mauro TC Sugawara, Brenno Cabella, Lukas Jonkers, Ulrike K Baranowski, Michal Kučera, et al. (2019). "On the mismatch in the strength of competition among fossil and modern species of planktonic Foraminifera". In: *Global Ecology and Biogeography*.

Rohwer, Forest and Rebecca Vega Thurber (2009). "Viruses manipulate the marine environment". In: *Nature* 459.7244, pp. 207–212.

Rojas, Alexis, Joaquin Calatayud, Michał Kowalewski, Magnus Neuman, and Martin Rosvall (2021). "A multiscale view of the Phanerozoic fossil record reveals the three major biotic transitions". In: *Communications Biology* 4.1, pp. 1–8.

Roopnarine, Peter D and Ashley A Dineen (2018). "Coral reefs in crisis: The reliability of deep-time food web reconstructions as analogs for the present". In: *Marine Conservation Paleobiology*. Ed. by C Tyler and C Schneider. Topics in Geobiology. Springer, pp. 105–141.

Rosenzweig, Michael L (1995). *Species Diversity in Space and Time*. Cambridge University Press, 436 pp.

Roux, Simon, Jennifer R Brum, Bas E Dutilh, Shinichi Sunagawa, Melissa B Duhaime, Alexander Loy, et al. (2016). "Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses". In: *Nature* 537.7622, pp. 689–693.

Rutherford, Scott, Steven D'Hondt, and Warren Prell (1999). "Environmental controls on the geographic distribution of zooplankton diversity". In: *Nature* 400.6746, pp. 749–753.

Ruttenberg, Kathleen C and Robert A Berner (1993). "Authigenic apatite formation and burial in sediments from non-upwelling, continental margin environments". In: *Geochimica et Cosmochimica Acta* 57.5, pp. 991–1007.

Salariato, Diego L and Fernando O Zuloaga (2017). "Climatic niche evolution in the Andean genus Menonvillea (Cremolobeae: Brassicaceae)". In: *Organisms Diversity & Evolution* 17.1, pp. 11–28.

Sañudo-Wilhelmy, Sergio A, Adam B Kustka, Christopher J Gobler, David A Hutchins, Min Yang, Kamazima Lwiza, et al. (2001). "Phosphorus limitation of nitrogen fixation by Trichodesmium in the central Atlantic Ocean". In: *Nature* 411.6833, pp. 66–69.

Saupe, Erin E, Jonathan R Hendricks, Roger W Portell, Harry J Dowsett, A Haywood, S J Hunter, et al. (2014). "Macroevolutionary consequences of profound climate

change on niche evolution in marine molluscs over the past three million years". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1795, 20141995.

Saupe, Erin E and Corinne E Myers (2021). "Macroevolution". In: *Evolutionary Developmental Biology: A Reference Guide*. Ed. by L. Nuño de la Rosa and G. B. Müller. Springer Nature, pp. 149–167.

Saupe, Erin E, Huijie Qiao, Jonathan R Hendricks, Roger W Portell, Stephen J Hunter, Jorge Soberón, et al. (2015). "Niche breadth and geographic range size as determinants of species survival on geological time scales". In: *Global Ecology and Biogeography* 24.10, pp. 1159–1169.

Schiebel, Ralf and Christoph Hemleben (2017). *Planktic Foraminifers in the Modern Ocean*. Springer, 358 pp.

Schiffbauer, James D, John Warren Huntley, Gretchen R O'Neil, Simon AF Darroch, Marc Laflamme, and Yaoping Cai (2016). "The latest Ediacaran wormworld fauna: Setting the ecological stage for the Cambrian explosion". In: *GSA Today* 26.11, pp. 4–11.

Schindler, Daniel E, Jonathan B Armstrong, and Thomas E Reed (2015). "The portfolio concept in ecology and evolution". In: *Frontiers in Ecology and the Environment* 13.5, pp. 257–263.

Schmitz, Oswald J (2005). "Scaling from plot experiments to landscapes: Studying grasshoppers to inform forest ecosystem management". In: *Oecologia* 145.2, pp. 224–233.

Schoener, Thomas W (1983). "Field experiments on interspecific competition". In: *American Naturalist* 122.2, pp. 240–285.

Scotese, Christopher R (2009). "Late Proterozoic plate tectonics and palaeogeography: A tale of two supercontinents, Rodinia and Pannotia". In: *Geological Society of London, Special Publications* 326.1, pp. 67–83.

— (2016a). *PALEOMAP PaleoAtlas for GPlates and the PaleoData plotter program*. URL: `http://www.earthbyte.org/webdav/ftp/earthbyte/Scotese_PaleoAtlas_v3.zip`.

— (2016b). *Tutorial: PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program*. URL: `https://www.earthbyte.org/paleomap-paleoatlas-for-gplates/`.

Scott-Phillips, Thomas C, Kevin N Laland, David M Shuker, Thomas E Dickins, and Stuart A West (2014). "The niche construction perspective: A critical appraisal". In: *Evolution* 68.5, pp. 1231–1243.

Seeholzer, Glenn F, Santiago Claramunt, and Robb T Brumfield (2017). "Niche evolution and diversification in a Neotropical radiation of birds (Aves: Furnariidae)". In: *Evolution* 71.3, pp. 702–715.

Sepkoski Jr, J John (1976). "Species diversity in the Phanerozoic: Species-area effects". In: *Paleobiology* 2.4, pp. 298–303.

Serrelli, Emanuele and Nathalie Gontier (2015). *Macroevolution: Explanation, Interpretation and Evidence*. Cham, Switzerland: Springer, 403 pp.

Servais, Thomas and David AT Harper (2018). "The great Ordovician biodiversification event (GOBE): Definition, concept and duration". In: *Lethaia* 51.2, pp. 151–164.

Servais, Thomas, Oliver Lehnert, Jun Li, Gary L Mullins, Axel Munnecke, Alexander Nuetzel, et al. (2008). "The Ordovician Biodiversification: Revolution in the oceanic trophic chain". In: *Lethaia* 41.2, pp. 99–109.

Servais, Thomas, Vincent Perrier, Taniel Danelian, Christian Klug, Ronald Martin, Axel Munnecke, et al. (2016). "The onset of the 'Ordovician Plankton Revolution' in the late Cambrian". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 458, pp. 12–28.

Sexton, Jason P, Patrick J McIntyre, Amy L Angert, and Kevin J Rice (2009). "Evolution and ecology of species range limits". In: *Annual Review of Ecology, Evolution, and Systematics* 40, pp. 415–436.

Sharples, Jonathan, Jack J Middelburg, Katja Fennel, and Timothy D Jickells (2017). "What proportion of riverine nutrients reaches the open ocean?" In: *Global Biogeochemical Cycles* 31.1, pp. 39–58.

Sheather, Simon J and Michael C Jones (1991). "A reliable data-based bandwidth selection method for kernel density estimation". In: *Journal of the Royal Statistical Society B: Statistical Methodology* 53.3, pp. 683–690.

Sheehan, Peter M (1977). "A reflection of labor by systematists?" In: *Paleobiology* 3.3, pp. 325–328.

Siccha, Michael and Michal Kucera (2017). "ForCenS, a curated database of planktonic foraminifera census counts in marine surface sediment samples". In: *Scientific Data* 4, 170109.

Sidor, Christian A, Daril A Vilhena, Kenneth D Angielczyk, Adam K Huttenlocker, Sterling J Nesbitt, Brandon R Peecook, et al. (2013). "Provincialization of terrestrial faunas following the end-Permian mass extinction". In: *Proceedings of the National Academy of Sciences* 110.20, pp. 8129–8133.

Siegel, DA and WG Deuser (1997). "Trajectories of sinking particles in the Sargasso Sea: Modeling of statistical funnels above deep-ocean sediment traps". In: *Deep Sea Research Part I: Oceanographic Research Papers* 44.9-10, pp. 1519–1541.

Signor, Philip W and Carlton E Brett (1984). "The mid-Paleozoic precursor to the Mesozoic marine revolution". In: *Paleobiology* 10.2, pp. 229–245.

Signor, Philip W, Jere H Lipps, LT Silver, and PH Schultz (1982). "Sampling bias, gradual extinction patterns, and catastrophes in the fossil record". In: *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*. Ed. by Leon T Silver and Peter H Schultz. Vol. 190. Geological Society of America, pp. 291–296.

Silverman, Bernard W (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Bristol: Chapman and Hall Ltd., 175 pp.

Simons, Andrew M (2002). "The continuity of microevolution and macroevolution". In: *Journal of Evolutionary Biology* 15.5, pp. 688–701.

Sinclair, ARE, Simon Mduma, and Justin S Brashares (2003). "Patterns of predation in a diverse predator–prey system". In: *Nature* 425.6955, pp. 288–290.

Slater, Graham J, Luke J Harmon, and Michael E Alfaro (2012). "Integrating fossils with molecular phylogenies improves inference of trait evolution". In: *Evolution* 66.12, pp. 3931–3944.

Smith, Felisa A, Jonathan L Payne, Noel A Heim, Meghan A Balk, Seth Finnegan, Michał Kowalewski, et al. (2016). "Body size evolution across the Geozoic". In: *Annual Review of Earth and Planetary Sciences* 44, pp. 523–553.

Smith, J Travis and Jeremy B.C. Jackson (2009). "Ecology of extreme faunal turnover of tropical American scallops". In: *Paleobiology* 35.1, pp. 77–93.

Sommeria-Klein, Guilhem, Romain Watteaux, Federico M Ibarbalz, Juan José Pierella Karlusich, Daniele Iudicone, Chris Bowler, et al. (2021). "Global drivers of eukaryotic plankton biogeography in the sunlit ocean". In: *Science* 374.6567, pp. 594–599.

Spengler III, Robert N (2021). "Niche construction theory in archaeology: A critical review". In: *Journal of Archaeological Method and Theory* 28, pp. 925–955.

Sperling, Erik A, Christina A Frieder, Akkur V Raman, Peter R Girguis, Lisa A Levin, and Andrew H Knoll (2013). "Oxygen, ecology, and the Cambrian radiation of animals". In: *Proceedings of the National Academy of Sciences* 110.33, pp. 13446–13451.

Sperling, Erik A, Michael J Melchin, Tiffani Fraser, Richard G Stockey, Una C Farrell, Liam Bhajan, et al. (2021). "A long-term record of early to mid-Paleozoic marine redox change". In: *Science Advances* 7.28, eabf4382.

Sperling, Erik A and Richard G Stockey (2018). "The temporal and environmental context of early animal evolution: Considering all the ingredients of an "explosion"". In: *Integrative and Comparative Biology* 58.4, pp. 605–622.

Sprain, Courtney J, Paul R Renne, Loÿc Vanderkluysen, Kanchan Pande, Stephen Self, and Tushar Mittal (2019). "The eruptive tempo of Deccan volcanism in relation to the Cretaceous-Paleogene boundary". In: *Science* 363.6429, pp. 866–870.

Squire, Richard J, Ian H Campbell, Charlotte M Allen, and Christopher JL Wilson (2006). "Did the Transgondwanan Supermountain trigger the explosive radiation of animals on Earth?" In: *Earth and Planetary Science Letters* 250.1-2, pp. 116–133.

Srivastava, Diane S and John H Lawton (1998). "Why more productive sites have more species: An experimental test of theory using tree-hole communities". In: *American Naturalist* 152.4, pp. 510–529.

Stanley, Steven M (2008). "Predation defeats competition on the seafloor". In: *Paleobiology* 34.01, pp. 1–21.

Sterner, Robert W and James J Elser (2003). *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton, NJ: Princeton University Press, 464 pp.

Stigall, Alycia L (2014). "When and how do species achieve niche stability over long time scales?" In: *Ecography* 37.11, pp. 1123–1132.

Strotz, Luke C, Marianna Simoes, Matthew G Girard, Laura Breitkreuz, Julien Kimmig, and Bruce S Lieberman (2018). "Getting somewhere with the Red Queen: Chasing a biologically modern definition of the hypothesis". In: *Biology letters* 14.5, p. 20170734.

Sunday, Jennifer M, Amanda E Bates, and Nicholas K Dulvy (2012). "Thermal tolerance and the global redistribution of animals". In: *Nature Climate Change* 2.9, 686.

Sutton, Tracey T, Malcolm R Clark, Daniel C Dunn, Patrick N Halpin, Alex D Rogers, John Guinotte, et al. (2017). "A global biogeographic classification of the mesopelagic zone". In: *Deep Sea Research Part I: Oceanographic Research Papers* 126, pp. 85–102.

Takagi, Haruka, Atsushi Kurasawa, and Katsunori Kimoto (2020). "Observation of asexual reproduction with symbiont transmission in planktonic foraminifera". In: *Journal of Plankton Research* 42 (4), pp. 403–410.

Tappan, Helen (1968). "Primary production, isotopes, extinctions and the atmosphere". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 4.3, pp. 187–210.

— (1986). "Phytoplankton: Below the salt at the global table". In: *Journal of Paleontology*, pp. 545–554.

Tarasov, Lev, Arthur S Dyke, Radford M Neal, and W Richard Peltier (2012). "A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling". In: *Earth and Planetary Science Letters* 315, pp. 30–40.

Tarasov, Lev, Aanna Hughes, Richard Gyllencreutz, Oystein Strand Lohne, Jan Mangerud, and John-Inge Svendsen (n.d.). "The global GLAC-1c deglaciation chronology, melwater pulse 1-a, and a question of missing ice". In: *IGS Symposium: Contribution of Glaciers and Ice Sheets to Sea-Level Change*, pp. 26–30.

Tarasov, Lev and W Richard Peltier (2002). "Greenland glacial history and local geodynamic consequences". In: *Geophysical Journal International* 150.1, pp. 198–229.

Tarhan, Lidya G (2018). "The early Paleozoic development of bioturbation—evolutionary and geobiological consequences". In: *Earth-Science Reviews* 178, pp. 177–207.

Tedford, Richard H, L Barry Albright III, Anthony D Barnosky, Ismael Ferrusquia-Villafranca, Robert M Hunt Jr, John E Storer, et al. (2004). "Mammalian biochronology of the Arikareean through Hemphillian interval (late Oligocene through early Pliocene epochs)". In: *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology*. New York, NY: Columbia University Press, pp. 169–231.

Thingstad, T Frede, MD Krom, RFC Mantoura, GA Fonnes Flaten, S Groom, B Herut, et al. (2005). "Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean". In: *Science* 309.5737, pp. 1068–1071.

Title, Pascal O and Kevin J Burns (2015). "Rates of climatic niche evolution are correlated with species richness in a large and ecologically diverse radiation of songbirds". In: *Ecology Letters* 18.5, pp. 433–440.

Tuomisto, Hanna (2010). "A diversity of beta diversities: Straightening up a concept gone awry. Part 1: Defining beta diversity as a function of alpha and gamma diversity". In: *Ecography* 33.1, pp. 2–22.

Tyrrell, Toby (1999). "The relative influences of nitrogen and phosphorus on oceanic primary production". In: *Nature* 400.6744, pp. 525–531.

Tziperman, Eli, Itay Halevy, David T Johnston, Andrew H Knoll, and Daniel P Schrag (2011). "Biologically induced initiation of Neoproterozoic snowball-Earth events". In: *Proceedings of the National academy of Sciences* 108.37, pp. 15091–15096.

Uyeda, Josef C and Luke J Harmon (2014). "A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data". In: *Systematic Biology* 63.6, pp. 902–918.

Valdes, Paul J, Edaward Armstrong, Marcus PS Badger, Catherine D Bradshaw, Fran Bragg, Taraka Davies-Barnard, et al. (2017). "The BRIDGE HadCM3 family of climate models: HadCM@Bristol v1.0". In: *Geoscientific Model Development* 10, pp. 3715–3743.

Van Mooy, BAS, A Krupke, ST Dyhrman, HF Fredricks, KR Frischkorn, JE Ossolinski, et al. (2015). "Major role of planktonic phosphate reduction in the marine phosphorus redox cycle". In: *Science* 348.6236, pp. 783–785.

Vardi, Yehuda (1982). "Nonparametric estimation in the presence of length bias". In: *The Annals of Statistics* 10.2, pp. 616–620.

Velasco, Julian A, Enrique Martinez-Meyer, and Oscar Flores-Villela (2018). "Climatic niche dynamics and its role in the insular endemism of Anolis lizards". In: *Evolutionary Biology* 45.3, pp. 345–357.

Venables, WN and BD Ripley (2002). "Random and mixed effects". In: *Modern Applied Statistics with S*. Springer, pp. 271–300.

Vermeij, Geerat J (1977). "The Mesozoic marine revolution: Evidence from snails, predators and grazers". In: *Paleobiology*, pp. 245–258.

— (1995). "Economics, volcanoes, and Phanerozoic revolutions". In: *Paleobiology* 21 (2), pp. 125–152.

Vermeij, GJ (1987). *Evolution and Escalation*. Princeton, NJ: Princeton University Press, 527 pp.

Vilhena, Daril A and Andrew B Smith (2013). "Spatial bias in the marine fossil record". In: *PLoS One* 8.10, e74470.

Vitousek, Peter M, John D Aber, Robert W Howarth, Gene E Likens, Pamela A Matson, David W Schindler, et al. (1997). "Human alteration of the global nitrogen cycle: Sources and consequences". In: *Ecological applications* 7.3, pp. 737–750.

Vrba, Elisabeth S (1984). "What is species selection?" In: *Systematic Zoology* 33.3, pp. 318–328.

Wagner, Andreas (2017). "The white-knight hypothesis, or does the environment limit innovations?" In: *Trends in Ecology & Evolution* 32.2, pp. 131–140.

Wagner, Peter J, Matthew A Kosnik, and Scott Lidgard (2006). "Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems". In: *Science* 314.5803, pp. 1289–1292.

Waltari, Eric and Robert P Guralnick (2009). "Ecological niche modelling of montane mammals in the Great Basin, North America: Examining past and present connectivity of species across basins and ranges". In: *Journal of Biogeography* 36.1, pp. 148–161.

Waniek, Joanna, Wolfgang Koeve, and Ralf D Prien (2000). "Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic". In: *Journal of Marine Research* 58.6, pp. 983–1006.

Ward, Ben A, BB Cael, Sinead Collins, and C Robert Young (2021). "Selective constraints on global plankton dispersal". In: *Proceedings of the National Academy of Sciences* 118.10.

Ward, Ben A and Michael J Follows (2016). "Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux". In: *Proceedings of the National Academy of Sciences* 113.11, pp. 2958–2963.

Warke, Matthew R, Tommaso Di Rocco, Aubrey L Zerkle, Aivo Lepland, Anthony R Prave, Adam P Martin, et al. (2020). "The great oxidation event preceded a paleoproterozoic "snowball Earth"". In: *Proceedings of the National Academy of Sciences* 117.24, pp. 13314–13320.

Wasserman, Larry (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics. New York, NY: Springer, 270 pp.

Waterson, AM, KM Edgar, Daniela N Schmidt, and Paul J Valdes (2017). "Quantifying the stability of planktic foraminiferal physical niches between the Holocene and Last Glacial Maximum". In: *Paleoceanography* 32.1, pp. 74–89.

Westrop, Stephen R and Jonathan M Adrain (1998). "Trilobite alpha diversity and the reorganization of Ordovician benthic marine communities". In: *Paleobiology* 24.1, pp. 1–16.

Whalen, Christopher D and Derek EG Briggs (2018). "The Palaeozoic colonization of the water column and the rise of global nekton". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1883, 20180883.

Wiens, John J (2011). "The niche, biogeography and species interactions". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1576, pp. 2336–2350.

Wiens, John J and Catherine H Graham (2005). "Niche conservatism: Integrating evolution, ecology, and conservation biology". In: *Annual Review of Ecology, Evolution, and Systematics* 36, pp. 519–539.

Williams, A (1997-2002). *Part H: Brachiopoda*. Ed. by RL Kaesler. Vol. 1-4. Treatise on Invertebrate Paleontology. Geological Society of America.

Willis, Katherine J and Robert J Whittaker (2002). "Species diversity—scale matters". In: *Science* 295.5558, pp. 1245–1248.

Wisz, Mary Susanne, Julien Pottier, W Daniel Kissling, Loïc Pellissier, Jonathan Lenoir, Christian F Damgaard, et al. (2013). "The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling". In: *Biological Reviews* 88.1, pp. 15–30.

Wood, Simon and Fabian Scheipl (2017). *gamm4: Generalized additive mixed models using 'mgcv' and 'lme4'*. URL: `https://CRAN.R-project.org/package=gamm4`.

Wood, Simon N (2003). "Thin plate regression splines". In: *Journal of the Royal Statistical Society B: Statistical Methodology* 65.1, pp. 95–114.

Worm, Boris, Heike K Lotze, Helmut Hillebrand, and Ulrich Sommer (2002). "Consumer versus resource control of species diversity and ecosystem functioning". In: *Nature* 417.6891, pp. 848–851.

Worton, Brian J (1989). "Kernel methods for estimating the utilization distribution in home-range studies". In: *Ecology* 70.1, pp. 164–168.

Wright, N, S Zahirovic, RD Müller, and M Seton (2013). "Towards community-driven paleogeographic reconstructions: Integrating open-access paleogeographic and paleobiology data with plate tectonics". In: *Biogeosciences* 10.3, pp. 1529–1541.

Wuertz, Diethelm, Tobias Setz, and Yohan Chalabi (2017). *fUnitRoots: Rmetrics—modelling trends and unit roots*. URL: `https://CRAN.R-project.org/package=fUnitRoots`.

Xiao, Shuhai (2014). "Oxygen and early animal evolution". In: *Treatise on Geochemistry*. Ed. by Heinrich D Holland and Karl K Turekian. Vol. 6. Oxford, UK: Elsevier, pp. 231–250.

Zhang, Zixiang, Michael Augustin, and Jonathan L Payne (2015). "Phanerozoic trends in brachiopod body size from synoptic data". In: *Paleobiology* 41.3, pp. 491–501.

Zhong, Zhiwei, Xiaofei Li, Dean Pearson, Deli Wang, Dirk Sanders, Yu Zhu, et al. (2017). "Ecosystem engineering strengthens bottom-up and weakens top-down effects via trait-mediated indirect interactions". In: *Proceedings of the Royal Society B: Biological Sciences* 284.1863, 20170894.

Žliobaitė, Indrė, Mikael Fortelius, and Nils C Stenseth (2017). "Reconciling taxon senescence with the Red Queen's hypothesis". In: *Nature* 552.7683, pp. 92–95.


@palaeoins