PALEOECOLOGY OF THE LATE CRETACEOUS WESTERN INTERIOR SEAWAY AND GULF COASTAL PLAIN

Ceara Kathleen Quinn Purcell

University of New Mexico

Follow this and additional works at: https://digitalrepository.unm.edu/eps_etds

Part of the Geology Commons, Other Ecology and Evolutionary Biology Commons, and the Paleontology Commons

Recommended Citation

Purcell, Ceara Kathleen Quinn. "PALEOECOLOGY OF THE LATE CRETACEOUS WESTERN INTERIOR SEAWAY AND GULF COASTAL PLAIN." (2024). https://digitalrepository.unm.edu/eps_etds/402

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Earth and Planetary Sciences ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Peter Fawcett, Chairperson

Corinna Myers

Louis Scuderi

Jason Moore

Tyler Mackey
PALEOECOLOGY OF THE LATE CRETACEOUS WESTERN INTERIOR SEAWAY AND GULF COASTAL PLAIN

by

CEARA KATHLEEN QUINN PURCELL

B.S., Geological Sciences, Ohio University, Athens, 2021

DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Earth and Planetary Sciences

The University of New Mexico
Albuquerque, New Mexico

May 2024
I dedicate this dissertation to my husband, Austin Pollard. He knows what he did.
ACKNOWLEDGEMENTS

I would like to acknowledge the numerous awesome people who helped me get this far. Firstly, I’d like to thank the academic influences I’ve have throughout the years. Specifically, my advisor, Dr. Cori Myers. She suffered through my unreasonable program pace and the neuroses that came with it with more than just good grace, so thank you. Thank you as well to Dr. Louis Scuderi and Dr. Carli Pietsch for helping with publications and for always being such good sports about reading my extremely long supplementary materials. I would also like to thank my other committee members, Dr. Jason Moore and Dr. Tyler Mackey, the entire UNM Earth and Planetary Sciences Department, and the amazing Paula Pascetti and Cecilia Arias for always being so helpful and kind. To all the researchers, collectors, and explorers who made the dataset I used possible, thank you from the bottom of my heart. Not only could I not have compiled so much alone, I would never have dreamed of trying.

For helping me get as far as a Ph.D., I’d like to thank my Master’s advisor Dr. Alycia Stigall for her unrelenting enthusiasm and unfailing interest in my success. To all the professors who showed me the wonders of deserts, porphyry deposits, and caves, Todd, Russel, Hannah, and Greg, thank you so much. You guys are so cool, I hope I can be like you one day.

Additionally, without the financial support of the National Science Foundation through the Museum Research Traineeship and the UNM Department of Earth and Planetary Sciences this research would not be possible, and for that I am very grateful. Thank you especially to the hard work and dedication of Dr. Tom Turner, who loves all things science
and always made students feel welcome, and Loretta Esquibel, who constantly made the MRT a better and brighter group with her ingenuity and kindness. Thank you both.

Secondly, I would like to acknowledge the support and friendship of Rhiannon Nolan, Jone Naujokaityte, Alex Apgar, Joe Wislocki, Cole Cooper, Bri Gordon, Ian Forsythe, Sam Mignogna, Travis Heier, Stephanie Hunt, Alex Knoer, and Naveen Balaji. They helped me in so many ways, science and otherwise, but first and foremost they helped me have fun and stay sane.

To my family, Sheila, Kevin, Maura, Sean, Brenna, and Tara, thank you for always believing in me despite myself. I can finally say, in my professional opinion as a Doctor of Philosophy (which encompasses the study of truth): you guys are the inarguable best.

Finally, I’d like to thank my husband, Austin Pollard. You’re the one I always want to talk to first, and last. If there’s a better person than you, I’ve never met them, and I don’t believe there’s a better person out there for me. Clichés be damned — I couldn’t have done any of this without you.
PALEOECOLOGY OF THE LATE CRETACEOUS WESTERN INTERIOR SEAWAY AND GULF COASTAL PLAIN

by

CEARA KATHLEEN QUINN PURCELL

B.S., Geological Sciences, Ohio University, Athens, 2021

Ph.D. Earth and Planetary Sciences, University of New Mexico, 2024

ABSTRACT

This dissertation presents research into the paleobiogeography and paleoecology of the Western Interior Seaway (WIS) and Gulf Coastal Plain (GCP) during the Late Cretaceous. The dataset utilized here includes an extensive record of marine invertebrates from over 200 years of fossil collecting and sedimentary data collated from over 45 different literature sources, representing approximately 17 million years of deposition. The high-resolution of this dataset, and its extensive geographic distribution and temporal duration, make it ideal for exploring various ecological questions. Using these data, I have attempted to reconstruct paleobiogeographic provinces as published in prior works, explore functional diversity patterns, and test some of the fundamental assumptions related to niche modeling in the fossil record. This work represents one of the most extensive and dynamic analyses of the WIS, focusing on elucidating primary assumptions regarding broad scale ecological relationships and the methods used to explore them.
# TABLE OF CONTENTS

ABSTRACT .............................................................................................................................. vi

LIST OF TABLES ..................................................................................................................... xi

LIST OF FIGURES ................................................................................................................ xii

PREFACE ................................................................................................................................... xiv

CHAPTER 1: INTRODUCTION ................................................................................................. 1

Geologic Setting: The Western Interior Seaway and Gulf Coastal Plain .......................... 4

Research Chapters and Databases ..................................................................................... 7

CHAPTER 2: FAUNAL PROVINCIALITY IN THE LATE CRETAEOUS WESTERN
INTERIOR SEAWAY USING NETWORK MODELING ......................................................... 12

Abstract ............................................................................................................................... 12

Introduction ......................................................................................................................... 12

Methods ............................................................................................................................... 15

Results and Discussion ....................................................................................................... 16

WIS vs. GCP Provinces ....................................................................................................... 16

Decreasing Faunal Similarity and Sea Level Fall ................................................................ 19

Latitudinal Patterns ............................................................................................................. 22

Conclusions ......................................................................................................................... 23

Acknowledgements ............................................................................................................. 24

CHAPTER 3: A NETWORK ANALYSIS APPROACH TO UNDERSTANDING TRENDS IN
FUNCTIONAL DIVERSITY IN THE LATE CRETAEOUS OF NORTH AMERICA ............... 25

Introduction ......................................................................................................................... 26

Methods ............................................................................................................................... 30

Results .................................................................................................................................. 35

Relationship Between Functional and Generic Diversity .................................................... 35

Substage-level Diversity: ..................................................................................................... 36

WIS vs GCP Influence ....................................................................................................... 40

Class Influences .................................................................................................................. 42

Functional versus Generic Diversity Across Distance and Spatial Scales: ....................... 42

Paleolatitudinal Bins (unit size 3). ..................................................................................... 42

360km and 60km Units (unit size 4). .................................................................................. 45

Network models of Functional Assemblages: .................................................................... 47

Discussion ............................................................................................................................. 53

Functional Diversity in Classes: ........................................................................................ 53
Distance Comparisons: ..................................................................................................................................... 170
Average Dissimilarity by Network-Identified Components: ................................................................. 175
Minimum Spanning Trees: ..................................................................................................................... 177
Extended Biostratigraphic Chart (Extension of Figure 1): ................................................................. 184
SI References: ...................................................................................................................................... 185
Appendix A-2. Taxonomic References List (See Table S13) ................................................................. 187
Appendix A-3. R Code used for Analysis ........................................................................................... 223
Code for Comparing Generic Richness and Occurrence Number (Sampling Bias) ......................... 223
Code to configure Occurrence Database into format for EDENetworks Analysis ................................ 230
Code for Conducting Network Randomization Comparisons ........................................................... 252
Code for Manually Comparing Potential Network Subcomponents .................................................. 267
Code for Calculating Average Network Link Weights across Distance Bins ........................................ 284
APPENDIX B. Supplementary Materials for “A network analysis approach to understanding trends in functional diversity in the Late Cretaceous of North America” by Purcell and Myers (in press) ............................................................................................................................................. 318
Appendix B-1. Detailed Methods and Results ..................................................................................... 318
Relationship between FR and GR in Database: ....................................................................................... 318
Bootstrap results for variation in FR values: ........................................................................................... 321
SQS to determine lower sampling limits ............................................................................................... 322
Note on Shannon Diversity Index: .......................................................................................................... 323
Functional Diversity Patterns: .................................................................................................................. 324
Summary of Regional Functional vs Taxonomic Diversity change in Substages: ............................... 327
Basin-level Substage Trends: .................................................................................................................... 329
Class-level Influences: ............................................................................................................................. 339
Summary of Functional vs Taxonomic Diversity change in Latitudinal Bins: ................................. 352
Metrics in General: ................................................................................................................................. 352
Proportional Change in Metrics Across Latitude: .................................................................................. 359
Nestedness verses Turnover in Functional Entities over Paleolatitude .............................................. 364
Summary of Functional vs Taxonomic Diversity change in 360km² Nodes across Substages: ........ 366
Summary of Functional vs Taxonomic Diversity change in 60km² Nodes across Substages: ............ 367
Summary of Functional vs Taxonomic Diversity change in 360km² Nodes across Distance: ........... 374
Summary of Functional vs Taxonomic Diversity change in 60km² Nodes across Distance: .......... 379
Summary of 360km² Functional Entity Networks: ............................................................................... 384
Summary of 60km² Functional Entity Networks: ............................................................................... 389
SUMMARY OF 60Km2 CLASS-LEVEL FUNCTIONAL ENTITY NETWORKS: .................................................. 394

REFERENCES .................................................................................................................................. 401

APPENDIX B-2. FUNCTIONAL ECOLOGY REFERENCES FOR TABLE S36 ............................................ 402

APPENDIX B-3. R CODE .................................................................................................................. 417

APPENDIX C. SUPPLEMENTARY MATERIALS FOR CHAPTER 4 .......................................................... 542

APPENDIX C-1. DETAILED METHODS .......................................................................................... 542

- Sedimentary data aggregation ............................................................................................................... 542
- Sedimentary variable interpolation ......................................................................................................... 570
- Distance from seeps interpolation methods: ............................................................................................. 572
- Distance from strandlines interpolation: .................................................................................................. 574
- Masking by shoreline/outrcoup polygns: ................................................................................................. 574
- Environmental variables correlation: ....................................................................................................... 575
- Comparisons with substage and stage: .................................................................................................... 575
- Stratigraphic database: ............................................................................................................................ 578
- Results: ................................................................................................................................................ 579
- Ingroup comparison: ............................................................................................................................... 626
- Comparison taxa comparisons: ................................................................................................................. 629
- Species to genus (all occurrences) comparison: ....................................................................................... 634
- References: ........................................................................................................................................... 637

APPENDIX C-3. DETAILED REFERENCES FOR ENVIRONMENTAL RASTER LAYERS .................................. 637

- Sedimentological literature references: .................................................................................................... 637
- Seep localities literature sources: ............................................................................................................. 641
- Outcrop polygon references: .................................................................................................................... 643

APPENDIX C-4. R CODE .................................................................................................................. 644
LIST OF TABLES

Table 1. Taxonomic database summary........................................................................................................8
Table 2. The number of sedimentary data locality points at each level of biostratigraphic confidence in the twelve biozone intervals. See Chapter 4 for details...........................................................................11
Table 3. Average/median link weights within substages and between different substages. 95% confidence intervals (CI) of the mean indicated......................................................................................................................20
Table 4. Table of the FE present in the database and their code values. ...........................................................32
Table 5. List of taxonomic classes and the FE they represent. See Table S36 for details on genera. Asterisks denote FEs represented by a single class................................................................................................................38
Table 6. Table of regional generic richness, functional richness ($\alpha_f$), and functional evenness (SME) values in each substage. Values for the overall database are given first, followed by the values observed in the WIS and then GCP. Camp. = Campanian, Maastr. = Maastrichtian........................................41
Table 7. Table of paleolatitude bin generic richness, functional richness ($\alpha_f$), and functional evenness (SME) and mean, median, and standard deviation values for each substage. NA value indicate no value measured for that unit in that substage. LC = Early Campanian, MC = Middle Campanian, UM = Late Campanian, LM = Early Maastrichtian, UM = Late Maastrichtian.................................................................44
Table 8. Summary statistics for proportional change in generic richness, functional richness ($\alpha_f$), and functional evenness (SME) for paleolatitude bins across adjacent substages. Camp. = Campanian, Maastr. = Maastrichtian. Values with less than the SQS sample estimate have been removed from these summary statistics.................................................................44
Table 9. Summary of the number of species occurrences in each biozone interval. Note that some biozone intervals were not used for analysis, due to a lack of adequate species occurrence information........................................................................................................................................74
Table 10. Results (p-value, alpha=0.5) of greater niche similarity tests for temporal resolution comparisons. Bolded values indicate statistical significance at the p-value < 0.05 level..................................................................................80
Table 11. Summary of niche overlap and stability results. Sbst. D = duration of the substage/stage; b. D = duration of biozone interval; D Ratio = ratio of the biozone duration; N = number of comparisons.................................................................................................................................................82
Table 12. Results (p-value, alpha=0.5) of greater niche similarity tests for phylogenetic comparisons. Bold text and asterisks indicate significant results........................................................................................................85
Table 13. Results (p-value, alpha=0.5) of greater niche similarity tests for phylogenetic comparisons. Bold text and asterisks indicate significant results.................................................................................................88
LIST OF FIGURES

Figure 1. Campanian to Maastrichtian biostratigraphic chart showing polarity chorn intervals (Malinverno et al. 2020), biozones (Ogg et al. 2016; Lynds and Slattery 2017), sea level fluctuations (Haq 2014), global stable carbon and oxygen isotope trends (Gale et al. 2020), and stable oxygen and carbon isotope data collated for the Western Interior region (Dwyer, 2019). Ages marked with an asterisk have been added or updated based on Malinverno et al. (2020) from Slattery et al. (2013). WI isotope values modified from unpublished data by Dwyer (2019) which span the R9 regressive interval of the Bearpaw Cyclothem (D. cheyennenses – B. baculus ammonite zones).................. 3

Figure 2. Distribution of taxonomic (brown diamonds), stratigraphic (green circles), and hydrothermal seep (yellow triangles) from the WIS across the study interval (Campanian – Maastrichtian). See Chapters 1 and 4 for details................................................................. 9

Figure 3. A) Previously defined subprovinces and northernmost extent of tropical/subtropical faunas during transgressions, modified from Kauffman (1984). B) Occurrence map of data from this study. C) Global and regional sea levels curves with major transgressive-regressive events. D) Complete database 360 km network with Kauffman’s subprovinces indicated. A modified version of this figure with additional biostratigraphic and isotopic data is available as Appendix A, Figure S30. .................. 13

Figure 4. Substage and complete database networks at percolation points........................................ 17

Figure 5. Plots of average link weights within geographic distance bins (WGS84 ellipsoid). Bars indicate 95% confidence intervals............................................................. 21

Figure 6. Campanian to Maastrichtian data locations................................................................. 28

Figure 7. Line plots of substage values for A) generic richness (red, solid line) and functional richness (blue, dotted line) and B) generic richness (red, solid line) and SME (blue, dotted line). The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric. Proportional change in values is denoted as percentages. Camp. = Campanian, Maastr. = Maastrichtian .............................................................. 37

Figure 8. Plot of FE relative generic richness through time. Circle size indicates relative generic richness within each substage, color indicates a proportional increase (red) greater than 0.05, decrease (blue) greater than 0.05, change less than 0.05 (orange) from the previous substage. Grey indicates a value cannot be compared with a previous substage. Green rectangles denote FEs only found in the GCP and red squares denote FEs missing from the Late Maastrichtian WIS. Gastropod silhouettes denote gastropod-type FEs; larger green silhouettes indicate herbivorous FEs absent in the Late Maastrichtian 360km WIS network component. FEs that are present in every 60km unit just below percolation in each substage are denoted with an asterisk (GCP component) and a cross (WIS component). See Table 4 for FE naming key. .......................................................... 39

Figure 9. Beeswarm plot of proportional change in A) generic richness, B) functional richness (af), and C) functional evenness (SME). Values reflect change in metric between adjacent paleolatitude for each substage. Camp. = Campanian, Maastr. = Maastrichtian.............................................................. 43

Figure 10. Box and whisker plots of proportional change between substages for generic richness, functional richness (af), and functional evenness (SME) at the 360km and 60km unit scales. Triangles indicate mean values. Camp. = Campanian, Maastr. = Maastrichtian.......................................................... 46

Figure 11. Network models across paleolatitude bins. Colored points represent the paleolatitude (cool = higher paleolatitude, warm = lower paleolatitude). Size indicates betweenness centrality. Line color and thickness indicate degree of similarity (thinner/thicker lines = greater similarity). TH =Threshold. ........................................................... 48
Figure 12. Campanian 60km unit networks of FE assemblages plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Unit size represents Betweenness Centrality. Unit colors indicate paleolatitude as in Fig. 8 (warmer = low paleolatitudes and cooler = higher paleolatitudes). TH = Threshold.

Figure 13. Early Maastrichtian 60km unit networks plotted by FE assemblages (left) and based on paleo-coordinates (right). Unit colors of upper networks indicate paleolatitude (warmer = lower paleolatitudes and cooler = higher paleolatitudes). Lower networks are colored by groups of highly similar units or components. TH = Threshold.

Figure 14. Late Maastrichtian 60km unit networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Unit colors of upper networks indicate paleolatitude (warm = southern; cool = northern). Lower networks are colored by component. TH = Threshold.

Figure 15. Biostratigraphic chart showing collated biozone intervals; shading delineates each interval used in analyses (modified from Purcell et al. 2023).

Figure 16. Example interpolated maps within the twelve biozone intervals, representing percent mud. Cool colors indicate low percentage and warmer colors indicate high percentage of mud. Shorelines are denoted with black. Note that limitations in sedimentary data collected resulted in unmapped portions of the basin in some intervals.

Figure 17. Correlation plot of duration ratio to mean niche overlap (D value).

Figure 18. Example overlap plots of Inoceramus species against each other and against Ctena imbricatula during the B. reduncus-B. scotti biozone interval. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the second listed species has expanded beyond the niche of the first listed species), and green represents regions of niche unfilling (i.e., where the second listed species niche has moved away from space occupied by the first listed species). Red lines represent the extent of environmental space realized within the interval and the black arrow indicates the shift of the niche centroid between the two species. See Figures S5 and S6 for all species overlap plots.

Figure 19. Convex hull comparisons of the Inoceramus genus and Inoceramus species niches in PCS for three biozone intervals. Polygons represent occupied environmental spaces for each taxa within PC environmental space. Dark grey points represent environmental data observed in the interval. Colors indicate Inoceramus species (green = I. barabini, light blue = I. convexus, pink = I. sagensis, dark blue = I. oblongus, orange = I. sublaevis, yellow = I. saskatchewanensis); grey = Inoceramus genus estimation.

Figure 20. Examples of niche dynamics through time at the species and genus-level, which can be interpreted as representing the niche of an ecological entity. The genus-level niche is in black, species niches are colored.
This preface acknowledges the funding sources and outlines the roles that co-authors, and collaborators performed in each chapter pursuant to the requirements of the University of New Mexico and the Department of Earth and Planetary Sciences. Ceara K. Q. Purcell conducted greater than or equal to 51% of the work presented in each chapter of this dissertation and is therefore the primary author. This research was funded by NSF #2021744, #1924807, and #1601878.

Chapter 1 was published in the journal *Geology* with co-authors Dr. Corinne Myers and Dr. Louis Scuderi (Purcell et al., 2023). Dr. Myers and Dr. Scuderi contributed advice about methods and interpretations as well as providing textural edits for the manuscript. Ceara Purcell designed the project, conducted analyses, created tables and figures, wrote the manuscript, and interpreted the bulk of the results.

Chapter 2 was submitted (March 2024) to the journal *Paleobiology* and co-authored by Dr. Myers, Dr. Carlie Pietsch, and Mr. Naveen Balaji (Purcell et al., in review). Dr. Myers and Dr. Pietsch contributed advice about methods and interpretations as well as providing textural edits for the manuscript. Mr. Balaji assisted with a portion of R coding. Ceara Purcell designed the project, conducted analyses, created tables and figures, wrote the manuscript, and interpreted the bulk of the results.

Chapter 3 is unpublished work assessing fundamental biases of paleo-niche analysis that will be submitted for publication in the journal *Paleobiology* following the dissertation defense.
CHAPTER 1: INTRODUCTION

This analysis builds off the extensive research previously conducted on the Campanian and Maastrichtian paleontological record of the Western Interior Seaway (WIS) and adjacent Gulf Coastal Plain (GCP). It is intended to quantitatively test existing paleobiogeographic and ecological assumptions and build a foundational framework for future investigations into this interval and region. Specifically, my research utilizes relatively novel techniques applied to the fossil record that incorporate quantitative and qualitative methods to investigate fundamental biotic patterns and environmental associations that existed in the WIS and GCP during the upper Late Cretaceous, just prior to the end-Cretaceous mass extinction event (Figure 1).

The fundamental aspects of biotic associations present throughout Earth’s history are not only informative in understanding the ancient world, but they may also provide us with the tools and base-line assumptions necessary to predict and mitigate detrimental taxonomic responses in the world today to projected environmental change. Sediments and fossils from the WIS and GCP represent an extremely well-documented, geographically extensive, and diverse assemblage that is well suited to investigating these issues. Modern, quantitative, biostatistical techniques have in recent history applied to fossil data to expand on previous findings (i.e., Dudie and Stigall, 2010; Keil, 2017; Dean et al. 2019). However, it is not enough to be able to apply these techniques; we also need to establish the basic principles, biases, and viable methods to be sure these techniques are providing accurate results and can be interpreted in useful ways. Prior to their application, however, it is important to establish the basic principles underlying the phenomena recorded by paleontological and
# Table of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α&lt;sub&gt;F&lt;/sub&gt;</td>
<td>Functional richness</td>
</tr>
<tr>
<td>BC</td>
<td>Betweenness centrality</td>
</tr>
<tr>
<td>BC&lt;sub&gt;ave&lt;/sub&gt;</td>
<td>Average betweenness centrality</td>
</tr>
<tr>
<td>CAM</td>
<td>Campanian</td>
</tr>
<tr>
<td>CC</td>
<td>Clustering coefficient</td>
</tr>
<tr>
<td>ENM</td>
<td>Ecological niche modeling</td>
</tr>
<tr>
<td>FE</td>
<td>Functional entity</td>
</tr>
<tr>
<td>FN</td>
<td>Fundamental niche</td>
</tr>
<tr>
<td>GCP</td>
<td>Gulf Coastal Plain</td>
</tr>
<tr>
<td>HFS</td>
<td>High faunal similarity</td>
</tr>
<tr>
<td>KPg</td>
<td>End Cretaceous mass extinction</td>
</tr>
<tr>
<td>kyr</td>
<td>Thousands of years</td>
</tr>
<tr>
<td>LDG</td>
<td>Latitudinal diversity gradient</td>
</tr>
<tr>
<td>MAA</td>
<td>Maastrichtian</td>
</tr>
<tr>
<td>MST</td>
<td>Minimum spanning trees</td>
</tr>
<tr>
<td>My</td>
<td>Millions of years</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>PCS</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PNC</td>
<td>Phylogenetic niche conservation</td>
</tr>
<tr>
<td>RN</td>
<td>Realized niche</td>
</tr>
<tr>
<td>SEI</td>
<td>Shannon’s evenness index</td>
</tr>
<tr>
<td>SI</td>
<td>Supplementary information</td>
</tr>
<tr>
<td>SQS</td>
<td>Shareholder quorum subsampling</td>
</tr>
<tr>
<td>SME</td>
<td>Simpsons measure of evenness</td>
</tr>
<tr>
<td>TA</td>
<td>Transcontinental arch</td>
</tr>
<tr>
<td>TH</td>
<td>Threshold</td>
</tr>
<tr>
<td>WIS</td>
<td>Western Interior Seaway</td>
</tr>
</tbody>
</table>
**Figure 1.** Campanian to Maastrichtian biostratigraphic chart showing polarity chorn intervals (Malinverno et al. 2020), biozones (Ogg et al. 2016; Lynds and Slattery 2017), sea level fluctuations (Haq 2014), global stable carbon and oxygen isotope trends (Gale et al. 2020), and stable oxygen and carbon isotope data collated for the Western Interior region (Dwyer, 2019). Ages marked with an asterisk have been added or updated based on Malinverno et al. (2020) from Slattery et al. (2013). WI isotope values modified from unpublished data by Dwyer (2019) which span the R9 regressive interval of the Bearpaw Cyclothem (*D. cheyennenses – B. baculus* ammonite zones).
environmental data, and the common biases which may influence our understanding of these phenomena. Questions that can potentially be investigated to better understand the Earth system in this way include how taxa and the functional entities distribute themselves across space and how species abiotic ecological niches, the primary ecological unit hypothetically underlying these biogeographical patterns, can be accurately and consistently assessed using fossil data. By collating one of the most extensive, high-resolution paleontological and sedimentological databases available to date, this research addresses some aspects of these issues and provides guidance for additional investigations, creating an essential framework from which even greater understanding of the Late Cretaceous and the modern world can be achieved.

Geologic Setting: The Western Interior Seaway and Gulf Coastal Plain

During the late Albian, rising sea levels flooded the North American continent, connecting the Arctic Ocean in the north with the Tethys Sea in the south at the proto-Gulf of Mexico (Williams and Stelek 1975; Kauffman and Caldwell 1993; Roberts and Kirschbaum 1995; Ziegler and Rowley 1998). This epeiric seaway occupied the forearc basins formed by the tectonic movement of the North American Plate west relative to the European Plate, beginning in the Late Triassic (Monger 1993; DeCelles 2004; Miall et al. 2008). These basins were bound on the west by the Cordilleran Fold-Thrust Belt and on the east by the stable North American Craton (Kauffman and Caldwell 1993; DeCelles 2004; Miall et al. 2008). Sedimentary units from the WIS record tectonically influenced subsidence and associated deposition, primarily from erosion of the uplifted fold-thrust belt to the west. Deposits from the GCP, in contrast, represent deposition along the passive but subsiding margin of the North American Continent (Braile et al. 1986). Geologic units from both regions represent a
dense record of marine deposition from approximately 100 to 59.2 Ma (Caldwell, 1974; Slattery et al., 2013).

Across this interval, eustatic and tectonic, regional changes to sea level greatly influenced the paleoceanography and geometry of the WIS and its connection with the GCP, eventually resulting in the diminution and draining of the WIS itself at the end of the Cretaceous (Gill and Cobban 1966; Kauffman 1984; Kauffman and Caldwell 1993; Slattery et al. 2013; Lowery et al. 2018). Oceanic differences between the WIS and GCP may also have influenced the macroevolution and macroecology of marine biota. The WIS had a unique and potentially dynamically shifting set of conditions, including distinct habitat shifts such as moving from siliciclastic to carbonate deposition eastward over its longitudinal extent (Slattery et al. 2013), the presence of a counterclockwise gyre and distinct cool and warm water masses (Fisher et al. 1994; Slingerland et al. 1996; Steel et al. 2012; Elderbak and Leckie 2016; Lowery et al. 2018), and potential periods of oceanic stratification (Cochran et al. 2003; He et al. 2005; Fricke et al. 2010; Petersen et al. 2016; Lowery et al. 2018). Distinct biotic patterns across both space and time have been noted throughout the Campanian and Maastrichtian of the WIS in particular, often associated with abiotic factors such as sea level (Sohl 1967, 1971, 1987; Jeletzky 1971; Kauffman 1973, 1984; Scott and Cobban 1986).

The dense fossil and sedimentological record of the WIS and GCP, their persistence across a critical interval just prior to a major mass extinction event, their broad latitudinal distribution, and the variation in environmental conditions they experienced make these two regions ideal for testing various hypotheses of ecological patterns and processes on geological timescales. Indeed, many previous workers have made great strides in
understanding the paleogeographic, biotic, and oceanic patterns that persisted throughout the
WIS and GCP regions across this interval. Most of these studies are more than thirty years
old, and the addition of novel methods and more extensive datasets make reevaluating these
patterns using modern analytical methods both desirable and valuable. For example, it has
long been assumed – based on investigations primarily conducted in the late 1980s – that the
WIS was characterized by distinct biogeographical subprovinces whose distribution was
closely associated with fluctuations in regional sea level (Kauffman 1984). No dedicated
analysis has reassessed these claims using modern methods or updated taxonomic data.

Few palaeoecological studies, furthermore, focus on studying spatiotemporal patterns
within a relatively “stable” interval of Earth’s history, rather than an period of significant
abiotic change such as a mass extinction interval (i.e., Dunhill et al., 2018; Foster and
Twitchet, 2014). While these studies are relevant to determining how major abiotic
perturbations influence biota, they rarely include long-term, in-depth analysis for background
intervals before or after the mass extinction event itself. Understanding how ecological
patterns present themselves during background intervals, is therefore highly relevant to
establishing baseline assumptions. Since the end Cretaceous mass extinction occurs at the
end of the Maastrichtian, these analyses illuminate the spatial and temporal ecological
patterns just before this event, providing a point of comparison for shifts across the mass
extinction interval.

Lastly, understanding how abiotic and biotic systems interact is useful for predicting
the effects of modern environmental changes, such as warming climates and shifting sea
levels. Unlike most modern analyses, paleontological analyses allow for long-term
assessment of broad-scale patterns during intervals of environmental change. Deep time data
can therefore be used as analogues for current and future conditions, informing modern conservation efforts. These analyses provide information about various aspects of a marine system over approximately 17 million years, informing our understanding of our current ecological system and its potential responses to future environmental change.

Research Chapters and Databases

This dissertation consists of three research studies that focus on fundamental ecological aspects of the WIS and GCP region during the Campanian and Maastrichtian. They utilize an extensive fossil dataset including over 33,000 total occurrences of marine invertebrates, and composed of 1113 different species and 574 genera (Table 1, Figure 2). Major clades represented include bivalves, cephalopods, gastropods, echinoderms, corals, brachiopods, bryozoans, and crustaceans. The complete database characterizes a substantial portion of marine animal taxonomic and functional diversity preserved in the WIS present during the Late Cretaceous and their geographic extent. High taxonomic resolution make these data ideal for future studies into primary ecological characteristics such as paleobiogeographic patterns, functional assemblages, and realized/fundamental niche analyses (Peterson et al. 2011).

The first chapter focuses on reassessing the paleobiogeographic provinces present in the WIS during the Late Cretaceous and their abiotic associations (Purcell et al., 2023). Biotic provinces associated with latitudinal gradients that shift across space relative to major abiotic shifts such as sea level change were hypothesized to exist in the WIS and GCP across the interval (Kauffman 1984). However, only two major provinces were observed using quantitative network analysis, with greater provinciality resulting from sea level fall across
Table 1. Taxonomic database summary

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Unique Genera</th>
<th>Number of Unique Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastropoda</td>
<td>194</td>
<td>376</td>
</tr>
<tr>
<td>Bivalvia</td>
<td>192</td>
<td>395</td>
</tr>
<tr>
<td>Cephalopoda</td>
<td>82</td>
<td>220</td>
</tr>
<tr>
<td>Malacostraca</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>Gymnolaemata</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Echinoidea</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>Anthozoa</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Polychaeta</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Hexanauplia</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Rhynchonellata</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Scaphopoda</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Xiphosura</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Stenolaemata</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Demospongea</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Asteroidea</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Crinoidea</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chromadorea</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Homarus</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lingulata</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Hydrozoa</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Thecostraca</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Maxillipoda</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bryozoa (undifferentiated)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>574</strong></td>
<td><strong>1113</strong></td>
</tr>
</tbody>
</table>

The interval. Conclusions support a WIS biogeographically homogenized, that became increasingly distinct through time from the open-ocean-facing GCP.

The second chapter explores the distinction between functional and taxonomic diversity in the WIS and GCP. Within this analysis, the dataset was also tested for how spatial aggregation of occurrences might influenced results of studies of this ilk. Spatiotemporal patterns of functional diversity were compared with taxonomic diversity under the primary hypothesis that functional diversity would be both spatially and temporally stable relative to
Figure 2. Distribution of taxonomic (brown diamonds), stratigraphic (green circles), and hydrothermal seep (yellow triangles) from the WIS across the study interval (Campanian – Maastrichtian). See Chapters 1 and 4 for details.
taxonomic diversity. This hypothesis was largely supported by the data, though differences in functional assemblages were observed in the WIS relative to the GCP. These patterns were found to reflect specific abiotic factors present in the WIS as compared to the GCP and to support ecological buffering against taxonomic loss during environmentally stressed intervals. Diversity would be both spatially and temporally stable relative to taxonomic diversity. This hypothesis was largely supported by the data, though differences in functional assemblages were observed in the WIS relative to the GCP. These patterns were found to reflect specific abiotic factors present in the WIS as compared to the GCP and to support ecological buffering against taxonomic loss during environmentally stressed intervals.

The final chapter incorporates a high-resolution stratigraphic database from across the WIS to assess potential sources of bias in the application of ecological niche modeling using deep time paleontological data. This chapter specifically tests the influence of temporal resolution on ecological niche quantification and provides a case study test of phylogenetic niche conservation in fossil species as they relate to their genus and the potential utility of using generic-level data as a proxy for species characteristics. It was hypothesized that low temporal resolution negatively influences the fidelity of niche characterization, and that individual species within the same genus with present high niche conservation with each other and with the genus-level niche itself. Temporal resolution was found to negatively correlate with niche characterization, but genus-level data was not found to represent species-level niches well.
Table 2. The number of sedimentary data locality points at each level of biostratigraphic confidence in the twelve biozone intervals. See Chapter 4 for details.

<table>
<thead>
<tr>
<th>Interval</th>
<th>High Confidence</th>
<th>Medium Confidence</th>
<th>Low Confidence</th>
<th>Total Localities</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. birkelundae</em>-H. <em>nebrascensis</em></td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td><em>B. clinolobatus</em></td>
<td>5</td>
<td>19</td>
<td>15</td>
<td>39</td>
</tr>
<tr>
<td><em>B. baculus</em>-B. <em>grandis</em></td>
<td>10</td>
<td>18</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td><em>B. reesidei</em>-B. <em>eliasi</em></td>
<td>28</td>
<td>6</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td><em>B. compressus</em>-B. <em>cuneatus</em></td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td><em>D. cheyennense</em></td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td><em>D. nebrascense</em>-E. <em>jenneyi</em></td>
<td>12</td>
<td>21</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td><em>B. reduncus</em>-B. <em>scotti</em></td>
<td>13</td>
<td>7</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td><em>B. perplexus</em>-B. <em>gregoryensis</em></td>
<td>38</td>
<td>10</td>
<td>19</td>
<td>67</td>
</tr>
<tr>
<td><em>B. macleanii</em>-B. <em>sp.</em> (smooth)</td>
<td>24</td>
<td>17</td>
<td>30</td>
<td>71</td>
</tr>
<tr>
<td><em>B. obtusus</em></td>
<td>25</td>
<td>9</td>
<td>29</td>
<td>63</td>
</tr>
<tr>
<td><em>S. leei</em>-B. <em>sp.</em> (weak flank ribs)</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>59</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>506</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sedimentological database included in the third research chapter will furthermore provide a foundation for exploring the more detailed localized aspects of ecology and abiotic conditions. Stratigraphic data incorporates sedimentary information relevant to environmental conditions present during sedimentation, including grain size, siliciclastic versus carbonate composition, sedimentary structures, bedding thickness, and bioturbation. These data were used to create WIS maps of continuous sedimentary characteristics constrained by paleo-shorelines.
CHAPTER 2: FAUNAL PROVINCIALITY IN THE LATE CRETACEOUS WESTERN INTERIOR SEAWAY USING NETWORK MODELING

Ceara Purcell¹, Louis Scuderi¹, Corinne Myers¹

¹Department of Earth and Planetary Sciences, Northrop Hall, 221 Yale Blvd NE University of New Mexico, Albuquerque, New Mexico 87131

Abstract

The Western Interior Seaway (WIS) was historically divided into latitudinal faunal provinces that were taxonomically distinct from the adjacent Gulf Coastal Plain (GCP) and shift in space due to sea level changes. However, no rigorous quantitative analysis using recent taxonomic updates have reassessed these provinces and their associations. We used network modeling of macro-invertebrate WIS and GCP fauna to test whether biotic provinces existed and to examine their relationships with abiotic change. Results suggest a cohesive WIS unit existed across the Campanian, and distinct WIS and GCP provinces existed in the Maastrichtian. Sea level changes coincided with changes in network metrics. These results indicate that, while the WIS did not contain sub-provinces in the Late Cretaceous, environmental factors influenced faunal associations and their communication over time.

Introduction

The Western Interior Seaway (WIS) and Gulf Coastal Plain (GCP) are characterized by a dense fossil record of marine invertebrates in the latest Cretaceous (~100–66 Ma; Caldwell, 1974; Slattery et al., 2013), spanning 45° latitude, that experienced a wide range of environmental shifts. Fluctuating sea levels (Figure 3C), for example, modified basin geometry and water mass distributions, impacting marine life (e.g., He et al., 2005; Kauffman, 1984; Lowery et al., 2018). A restricted connection between the WIS and the
open ocean affected oceanic conditions relative to the GCP and may have influenced biotic provinces.

Figure 3. A) Previously defined subprovinces and northernmost extent of tropical/subtropical faunas during transgressions, modified from Kauffman (1984). B) Occurrence map of data from this study. C) Global and regional sea levels curves with major transgressive-regressive events. D) Complete database 360 km network with Kauffman’s subprovinces indicated. A modified version of this figure with additional biostratigraphic and isotopic data is available as Appendix A, Figure S30.
Biotic provinces are geographic regions characterized by distinct ecological associations. Previous studies of biotic provinces using fossil materials have: attributed them to major climatic regions (e.g., Kocsis et al., 2021), associated shifting provinces with sea level fluctuations (e.g., Kauffman, 1984), compared spatiotemporal influences on taxonomic association patterns (e.g., Kiel, 2017), and observed changes in provinciality relative to taxonomic loss (e.g., Kocsis et al., 2018).

Quantifying biogeographic patterns can therefore shed light on macroecology over evolutionary time. Kauffman (1984) described three significant biotic incursions during transgressions based on changes to WIS subprovinces (Figure 3A). These subprovinces, determined using percent endemism of mollusk records from 1960-80s, are: (1) Northern Interior; (2) Southern Interior; and (3) Central Interior Subprovince (Kauffman 1984). Another identified faunal province was the Gulf and Atlantic Coastal Plain Subprovince. However, delineation of these paleobiogeographic provinces and their changes through time were based on qualitative assemblages limited by the available fossil data (Kauffman, 1984 and references therein). Analysis of WIS provinciality using current fossil data will improve the validity of these interpretations.

Network modeling analysis of faunal provinces is a novel approach (Kiel 2016, 2017) to quantifying faunal similarity across spatiotemporal units. Using a well-vetted database of over 33,000 fossil occurrences from the WIS and the GCP we used this approach to reevaluate WIS provinciality in the Campanian (CAM) and Maastrichtian (MAA) stages of the Late Cretaceous. While previous studies have used network analysis to explore provincialism in fossil taxa (Kiel, 2016, 2017; Rojas et al., 2021; Muscente et al., 2018; Kocsis et al., 2018, 2021), none have applied the technique to a geochemically unique,
restricted ocean system, characterized by over 100 years of dedicated sampling. This research may also inform general patterns of Earth-Life interaction over long timescales (e.g., the nature of faunal variation through time and space) and serve as a foundation for future WIS/GCP investigations.

**Methods**

Records of marine invertebrates from the CAM and MAA of the WIS and the GCP were compiled from digital databases, including the Paleobiology Database (08/25/21 download) and iDigBio (08/30/21 download), and records from museum collections at the Black Hills Institute, USGS-Denver (Cobban Collection), and the Mackenzie (2007) thesis database (Table S1). Taxa were binned into the Early, Middle, and Late CAM and the Early and Late MAA stages (Figure 3B and S1). Localities were converted to paleocoordinates within a 60 km grid for analysis; nodes with fewer than three unique taxa were removed. The vetted database was analyzed prior to network modeling to determine fundamental sources of taxonomic and spatial bias which should be considered during network interpretations (Appendix A, Table S2, Figure S2-S9).

Faunal provinces were delimited for substages individually and for the complete database (combined substages) using threshold weighted networks (Keil, 2016) in the EDENetworks software (Appendix A, Table S3 & Figure S10-S20; Moalic et al., 2012; Kivelä et al., 2015; Kiel, 2016). Network components which disconnect at and below a network-specific threshold identified by EDENetworks known as the percolation point are interpreted as representing distinct “community” groups (Newman 2012) or faunal provinces. General patterns in network connections across all substages together were assessed using
coarser spatial aggregations of the data and minor network components were assessed for spatiotemporal consistency (Appendix A, Figure S21–S23).

Average network clustering coefficient (CC) values, indicative of network organization ranging from 0 (no cluster) to 1 (fully connected cluster), were compared with a null model of randomized networks to determine if the topology was more or less clustered than a random distribution (Table S4; Kiel, 2016). Link weights, or the degree of dissimilarity between nodes, and betweenness centrality (BC), a measure of the degree to which a node acts as a geographic connection between regions, were averaged by 5° paleolatitudinal bins for comparison (Appendix A, Figure S24-S26). Link weights were also binned by geographic distance to test for correlation between faunal dissimilarity and distance (Appendix A, Table S5 & Figure S27). Given latitudinal overlap between the WIS and GCP, around 35°N, we separated the data by major components and evaluated link similarity between the two regions independently (Appendix A, Table S6 & Figure S28). Sampling bias influence on network communication was assessed by binning average betweenness centrality (BCave), by generic richness as a proxy for sampling effort (Appendix A, Figure S25) and using minimum spanning trees (MST; Appendix A, Figure S29).

Additional explanation of methods is provided in the supplementary materials.

Results and Discussion

WIS vs. GCP Provinces

The presence of a single faunal province in the WIS is supported by all network permutations and subsequent analysis (Figure 4). This province is geographically consistent with the WIS (Appendix A, Figure S21), is maintained at all threshold levels (Appendix A,
Figure 4. Substage and complete database networks at percolation points.
Figure S10-15), does not contain spatiotemporally consistent minor components (Appendix A, Figure S22-23) and is faunally distinct from GCP grid cells (Appendix A, Figure S21). For MAA substages, which contain more GCP fossil occurrences relative to the CAM, a well-supported GCP faunal province was observed (Figure 4, S16-S19). This distinct GCP province is supported by the full database network as well (Appendix A, Figure S20). Network randomization comparisons demonstrated that these results are non-random and likely reflect biogeographic patterns (Table S4; average CC >3 stand. dev. from mean). They do not result from sampling bias based on comparisons with spatial cluster analysis (Appendix A, Figure S5-S9), BC comparisons (Appendix A, Figure S25), and MST (Appendix A, Table S7-S12). Thus, our quantitative analysis does not support the existence of WIS biotic subprovinces but does support a distinct GCP province in the MAA (Kauffman 1984).

The WIS and GCP provinces may have resulted from geochemical and bathymetric changes across the transcontinental arch (TA), which may have acted as a bathymetric high between the regions (He et al., 2005; Lowery et al., 2018 and references therein), rather than by latitudinal factors (e.g., temperature). Geochemical studies have found evidence for non-normal marine conditions in the WIS, including low salinity or brackish conditions (Cochran et al., 2003; Dennis et al., 2013; Fricke et al., 2010) and lower δ18O of seawater than the open ocean (Fricke et al. 2010; Petersen et al. 2016). There is also evidence for stratification within the WIS during the CAM and MAA, produced by mixing water masses (He et al. 2005; Lowery et al. 2018). These factors could have created a habitat barrier between the GCP and WIS, facilitating provincialism. The WIS fauna may have been more tolerant of non-normal conditions, supported by a lack of abundant reef-building and reef-associated
taxa (Gill and Cobban 1966b; Caldwell 1968; Kauffman 1984; Kauffman and Caldwell 1993).

The lack of a latitudinally-defined provinces within the WIS is unsurprising given a flattened latitudinal temperature gradient in the Cretaceous greenhouse (Mannion et al., 2014; Super et al., 2018). However, evidence for different water mass distributions and salinity/temperature gradients have long been associated with latitude and faunal gradients in the WIS (Fisher et al., 1994; Slingerland et al., 1996; Longman et al., 1998; Elderbak and Leckie, 2016; Lowery et al., 2018). During much of the study interval, a cool-water mass circulated south through the WIS from the northern connection with Greenland and northern Europe, interacting with the northward moving warm water mass from the Tethys, forming a counterclockwise gyre (Steel et al., 2012; Lowery et al., 2018). However, no evidence for provinces matching these water bodies was observed in our results. Instead, this gyre could have contributed to faunal homogenization despite ocean stratification or abiotic gradients. Further, the unique geochemical nature of the basin may have encouraged WIS incumbents and generalists to flourish over specialists or invaders. Dataset differences, including improved sampling and the lack of foraminifera in this study, may have hindered observation of Kauffman’s (1984) subprovinces, though this requires further investigation. Indeed, the potential for along-seaway variation within specific WIS faunas, as observed by previous authors (i.e., Sohl, 1971; Jeletzky, 1971, etc.), was not tested by this analysis which tests for discrete clusters of faunal assemblages.

*Decreasing Faunal Similarity and Sea Level Fall*

Network dissimilarity values increase through time (i.e., decreasing similarity), particularly in the WIS province, based on average link values per substage (Table 3); in
contrast, the GCP province shows increasing similarity through the MAA (Table S6). Bathymetric and geochemical changes coincide with these shifts, suggesting a potential relationship. The WIS gyre that may have promoted mixing of water masses, WIS dispersal, and mixing with the GCP (Fisher et al. 1994; Slingerland et al. 1996; Longman et al. 1998; Elderbak and Leckie 2016). However, falling sea levels likely impacted circulation patterns, water mass dynamics, and geochemical and environmental gradients (e.g., non-normal salinity, nutrient load) that would limit WIS migration and thereby insulate existing fauna from outside invasion (Cochran et al. 2003b; He et al. 2005; Fricke et al. 2010; Petersen et al. 2016). Shallowing along the TA may also have created a geographic barrier between the WIS and GCP as early as the Late CAM (Lehman, 1987; Lowery et al., 2018 and references therein).

Table 3. Average/median link weights within substages and between different substages. 95% confidence intervals (CI) of the mean indicated.

<table>
<thead>
<tr>
<th>Substage Link Weights (Mean ± 95% CI/Median)</th>
<th>Lower Camp.</th>
<th>Middle Camp.</th>
<th>Upper Camp.</th>
<th>Lower Maastr.</th>
<th>Upper Maastr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp. (0.78 ± 0.004/0.80)</td>
<td>-</td>
<td>0.83 ± 0.002/0.84</td>
<td>0.85 ± 0.002/0.86</td>
<td>0.87 ± 0.003/0.88</td>
<td>0.96 ± 0.001/0.98</td>
</tr>
<tr>
<td>Middle Camp. (0.79 ± 0.002/0.82)</td>
<td>-</td>
<td>-</td>
<td>0.83 ± 0.001/0.85</td>
<td>0.86 ± 0.002/0.88</td>
<td>0.95 ± 0.001/0.96</td>
</tr>
<tr>
<td>Upper Camp. (0.82 ± 0.002/0.84)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.86 ± 0.002/0.88</td>
<td>0.95 ± 0.001/0.96</td>
</tr>
<tr>
<td>Lower Maastr. (0.83 ± 0.004/0.86)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.93 ± 0.003/0.95</td>
</tr>
<tr>
<td>Upper Maastr. (0.85 ± 0.006/0.95)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Dampened circulation and salinity gradients may have also caused declining faunal similarity within the WIS alone, as evidenced by lower average faunal similarity within each substage network across time (Table S6). Below ~1000 km faunal similarity comparisons show weak correlation between distance and link weight (Figure 5 and S28), indicating only slight decline in similarity over distance within a substage, despite decreasing similarity though time. Sedimentary evidence for tidal circulation influences through at least the Middle and Late CAM (Steel et al. 2012) suggest continued circulation and mixing that could promote homogenization. Distance comparisons for the WIS and GCP components individually produced similar patterns (Figure S28), indicating that these results are not basin specific. Additional study of WIS oceanography is needed to confidently assess the potential influence of Late Campanian oceanographic changes.

![Figure 5](image)

**Figure 5.** Plots of average link weights within geographic distance bins (WGS84 ellipsoid). Bars indicate 95% confidence intervals.
While faunal connectivity within the WIS and between the WIS and GCP decreased over time, MAA average link weights show that GCP faunal similarity increased (Appendix A, Table S6; Figure S21). As an longitudinally broad, open-ocean facing province, the GCP would have experienced normal marine conditions, less latitudinal variation, and the potential for long-distance dispersal, potentially supporting faunal similarity across the MAA by reducing the endemism. Within the WIS province, from the Early CAM to the Early MAA, network connections remained strong between substages across time, indicating weak faunal turnover in the region even as faunal similarity decreased, until the Late MAA when the province disconnects from previous iterations (Figure 2D). The Late MAA disconnect with previous iterations matches expectations of oceanic changes that decrease dispersal and habitat homogenization for WIS taxa (Elderbak & Leckie, 2016; Steel et al., 2012; Fisher et al., 1994; Longman et al., 1998; Slingerland et al., 1996).

**Latitudinal Patterns**

Results do not support latitude-based faunal provinces, despite changes in network metrics across latitudes (Figure 4). However, the region of highest average faunal similarity (HFS) shifts 5° north from the 40-45°N bin to the 45-50°N bin across the R9 regressive event at the end of the Middle CAM (Kauffman & Caldwell, 1993; Figure 2C). Prior to the R9 regression, sea levels were more stable and the HFS region was fixed (Appendix A, Figure S26). This HFS shift appears to reflect biogeographic patterns and is unlikely to be a product of data distributions given that faunal similarity compared to geodesic distance are relatively stable over <1,000 km (Figure 5). The geodesic distance covered by a 5° latitudinal bin in this region is approximately 555 km and the distance covered by two bins is approximately 1,110 km. Thus, similarity begins to strongly decrease over distances greater than 10°
latitude. This suggests regional control on network metrics. If faunal similarity only depended on distance, then similarity should show a uniform pattern across latitude rather than the observed peaks and dips (Appendix A, Figure S25). Therefore, a HFS region that shifts parallel to sea level likely represents a distinct biogeographical component influenced by oceanographic changes.

Similarly, although the region of highest $B_{\text{ave}}$ (indicates highest faunal communication between regions) primarily occupies northern latitudes (40-60°N), it shifts south from the Middle to Late CAM (Appendix A, Figure S25). This region of highest communication may indicate intermediate habitat (Kiel 2016), uniformity of conditions (i.e., water depth), or currents which transport taxa long distances (Lowery et al., 2018). The latter would support larval migration of marine taxa, especially those with long planktonic larval stages (Nickols et al. 2015). However, more specific bathymetric, geochemical, and sedimentological evidence for habitat conditions is sorely needed, but outside the scope of this analysis. The region of highest $B_{\text{ave}}$ may also correspond to Kauffman’s (1984) mixing zone. Despite shifts in highest $B_{\text{ave}}$, all networks show a minor or major peak in the central WIS (45-50°N; Appendix A, Figure S25) corresponding with a region of mixing water masses (Lowery et al. 2018). This supports oceanographic or habitat controls on fauna in the WIS.

**Conclusions**

Network analysis of the Late Cretaceous WIS and GCP supports a single biogeographic province throughout the CAM and an independent GCP province in the MAA with decreasing faunal connectivity through time. This contrasts with Kauffman’s (1984) original division of the region into four “subprovinces.” Decreasing faunal similarity over the
study interval is consistent with oceanographic and geochemical changes that restricted the WIS and exacerbated non-normal marine conditions. Though no overarching relationship between faunal associations and latitude are observed, regional movement of HFS and highest $B_{Ca}^{w}$ suggest environmental changes (i.e., falling sea levels and associated effects) were the primary control on biogeographic connections. We provide further evidence for the utility of network modeling to quantitatively characterize paleobiogeographic trends on evolutionary timescales relative to major environmental shifts, an important analytical tool in modern tests of marine biogeographic change under predicted global change.

Acknowledgements

Thanks to Timothy Nagle-McNaughton, Steffen Kiel, and Xi Gong for helpful discussions and assistance with methods and visualization. Thanks also to Steve Holland and two anonymous reviewers whose comments improved this manuscript. CP acknowledges support from NSF #2021744; CEM acknowledges support from NSF #1924807.
CHAPTER 3: A NETWORK ANALYSIS APPROACH TO UNDERSTANDING TRENDS IN FUNCTIONAL DIVERSITY IN THE LATE CRETACEOUS OF NORTH AMERICA

Ceara Purcell¹, Naveen Balaji², Carlie Pietsch³, Corinne Myers¹

¹Department of Earth and Planetary Sciences, Northrop Hall, 221 Yale Blvd NE University of New Mexico, Albuquerque, New Mexico 87131 U.S.A. Email: cpurcell@unm.edu

²Data Engineer, BBInsight, 1750 East Golf Road, Suite 232, Schaumburg, Illinois 60173 U.S.A. Email: nvnbalji@gmail.com.

³Department of Geology, Duncan Hall, One Washington Square, San Jose University, San Jose, California 95192 U.S.A. Email: carlie.pietsch@sjsu.edu

Abstract.

Functional diversity is a complex aspect of macroecology that may respond to environmental changes differently than taxonomic diversity across space and time. To investigate spatiotemporal patterns of functional diversity over geologic time we utilize paleontological data from the environmentally diverse North American Western Interior Seaway and Gulf Coastal Plain through the Campanian and Maastrichtian geologic stages. Invertebrate taxa are classified into functional entities (FEs) based on their motility, attachment style, tiering level, and feeding strategy and analyzed across four levels of spatial aggregation over five geologic sub-stages. Network models of functional assemblages are also used to assess functional biogeographic patterns. Results indicate that regional functional diversity follows previously observed global trends of stability across geologic intervals and supports high functional redundancy (i.e., packing of genera) within specific FEs, buffing against loss. Unlike present-day taxa, however, the number of Late Cretaceous marine FEs do not decline across a latitudinal temperature gradient. Biogeographic patterns in functional diversity are instead associated with paleo-oceanographic conditions and the
resulting habitat differences between the Western Interior Seaway and Golf Coastal Plain. These analyses provide insight into the spatiotemporal consistency of functional diversity patterns relevant to understanding current, ancient, and future Earth-life dynamics.

**Introduction**

The relationship between functional diversity and environmental change, while relatively poorly understood, is relevant to current, ancient, and future Earth systems. Given a relationship does exist between the two, how abiotic factors influence functional diversity under different circumstances has not been well constrained and functional diversity has previously been found to respond differently across space and time relative to environmental influences (Foster and Twitchett 2014; Dunhill et al. 2018; Edie et al. 2018; Schumm et al. 2019; Pimiento et al. 2020). For example, functional diversity seems to be little affected during mass extinction events despite severe loss of taxonomic diversity (Foster and Twitchett 2014; Dunhill et al. 2018; Edie et al. 2018). This is potentially due to functional redundancy, which is defined as the number of taxa within a given functional entity (FE) (Pimiento et al. 2020). In recent ecosystems functional diversity trends are negatively correlated with the latitudinal diversity gradient (LDG), that is, decreasing at greater latitudes, regardless of functional redundancy (Edie et al. 2018; Schumm et al. 2019; Floyd et al. 2020; Forsyth and Gilbert 2021). Identifying factors that may influence these patterns and how they differ during background intervals (versus mass extinction intervals) in the geologic record are relevant to a better understanding of macroecology on evolutionary timescales. Paleontological data is uniquely able to assess patterns of functional diversity across both space and time (Roopnarine and Angielczyk 2015; Muscente et al. 2018; Foster et al. 2020).
The latest Cretaceous (~ 85 – 66 Ma, Campanian-Maastrichtian) Western Interior Seaway (WIS) and Gulf Coastal Plain (GCP) of North America are characterized by a spatiotemporally dense fossil record of marine invertebrate assemblages living across a 45-degree latitudinal gradient (Caldwell 1974; Kauffman 1984; Slattery et al. 2013). Fluctuating sea level across the interval impacted marine geochemistry, water mass distributions, and biogeography (Figure 6) (e.g., He et al., 2005; Kauffman, 1984; Lowery et al., 2018; Purcell et al., 2023, etc.) (Figure 6). The WIS itself was characterized by an asymmetrical foreland basin geometry with a deep, siliciclastic foredeep to the west, a latitudinally intermittent central forebulge, and a broad, shallow carbonate shelf to the east (Slattery et al. 2013; Minor et al. 2022). Furthermore, variable restriction between the WIS and GCP promoted distinct habitat conditions between the regions, wherein the WIS experienced episodes of a stratified, potentially brackish, and geochemically distinct restricted seaway system and the GCP represented a normal-marine, unrestricted shoreline (Fisher et al. 1994; Slingerland et al. 1996; Longman et al. 1998; Elderbak and Leckie 2016; Lowery et al. 2018; Dwyer 2019). The Late Cretaceous WIS and GCP are therefore excellent regions in which to explore the spatiotemporal patterns of functional diversity and their relationship to macroecological patterns.

The Late Cretaceous faunal record in these regions is characterized by abiotic change and background levels of biotic turnover. Functional entities are globally stable across the end-Cretaceous mass extinction (KPg) itself (e.g., Edie et al., 2018), but none have yet analyzed regional changes in functional diversity during background intervals of the Late Cretaceous. Moreover, latitudinal changes in functional diversity have been poorly studied in the fossil record. Though present-day functional diversity declines as latitude increases (Edie et al.
2018; Schumm et al. 2019), the latest Cretaceous is characterized by a relatively flat latitudinal temperature gradient that may present a more uniform distribution of FEs compared to today (Mannion et al. 2012, 2014; Super et al. 2018). Dramatic oceanographic differences between the GCP and the WIS have additionally been found to correspond with marine provinciality and altered biotic connectivity (Purcell et al. 2023). Regional differences in abiotic factors may therefore result in distinct Late Cretaceous patterns in the structure of functional

**Figure 6.** Campanian to Maastrichtian data locations.
assemblages, or the types of FEs present in a specific time and place (Schumm et al. 2019). Global functional diversity analyses, which have previously been the focus of most paleo-FE studies, may also introduce longitudinal variations that overprint regional patterns (Mannion et al. 2014; Schumm et al. 2019). Generally, the influence of data aggregation across space on functional diversity analyses has not been rigorously tested. By focusing on a single, well-studied region such as the WIS/GCP, we can address these aspects of the biogeography FEs over time.

Thus, we use a well-vetted fossil occurrence database representing approximately 18 million years of an ecologically diverse suite of marine invertebrates, to observe how functional diversity relates to major environmental shifts in space and time within the WIS and adjacent GCP. We hypothesize that 1) functional richness and evenness are stable across space and within a single temporal interval, and across the minor sea level oscillations and environmental changes present during the Campanian and Maastrichtian; 2) the WIS and GCP lack notable biogeographic differences in specific FEs, despite previously observed taxonomic provinciality and habitat differences; 3) functional redundancy is associated with FE stability, both in terms of specific FEs presence and absence and in terms of FE richness and evenness, across space/time, and 4) these patterns are observable at several levels spatial aggregation. By documenting spatiotemporal FE stability we can contribute to a better understanding of how functional diversity connects with FE stability, habitat- and dispersal-based provinciality, and illustrate the influence of spatial resolution on functional diversity analyses. We demonstrate that packing of species and genera into FEs leads to stability of functional guild structure throughout the study interval. A major sea level regression and disconnect of the WIS and GCP leads to paleoenvironmental change and regional extirpation.
of FEs. Thus, our research better illuminates the biogeography of functional diversity and its structure during a background interval that leads into a major taxonomic depletion event: the end-Cretaceous mass extinction (Edie et al. 2018). This work informs both modern and ancient understanding of functional diversity dynamics and its relationship to macroecology.

**Methods**

Fossil occurrence data consists of marine invertebrates from WIS and GCP Campanian and Maastrichtian strata compiled from digital databases, including the Paleobiology Database (08/25/21 download) and iDigBio (08/30/21 download), and records from museum collections at the Black Hills Institute, USGS William A. Cobban Cretaceous Ammonite Collection, and a thesis database (Mackenzie 2007). All data were vetted to remove incomplete, erroneous, or poorly spatially and temporally resolved occurrences and to update taxonomic nomenclature. Stratigraphic ages were vetted to the substage level based on stratigraphic and biozone information or from literature sources (see Purcell et al. 2023). Spatial resolution that could not be confirmed to less than 30 km of uncertainty were removed to avoid spatial ambiguity and records east of -80° longitude and south of 23° latitude were removed to avoid records from outside the WIS/GCP region.

The final dataset of 32,864 fossil occurrences includes a total of 552 genera of bivalves, cephalopods, gastropods, echinoderms, corals, brachiopods, bryozoans, and crustaceans (see Purcell et al. 2023 for additional details and references used in dataset vetting). These groups represent a wide variety of lifestyle modes both within clades and between distinct phylogenetic groups. They are furthermore important components in both modern and ancient shallow marine ecosystems and therefore excellent test subjects for ecological analysis. Taxa were binned into the Early, Middle, and Late Campanian and the
Early and Late Maastrichtian substages based on stratigraphic information. Tests confirming signal over noise in this dataset were previously published in Purcell et al. (2023).

Genera were classified into functional entities (FEs) based on 1) mobility and attachment style, 2) lifestyle mode or tiering level, and 3) feeding strategy following Bambach et al. (2007; see also Bush & Bambach, 2011) (Table 4). Characteristics were assigned based on extant relatives or higher order traits (i.e., Family or Order level traits, see excel file Table S36) (i.e., Family or Order level traits; e.g., Aberhan & Kiessling, 2015; Edie et al., 2018; Foster et al., 2020; Foster & Twitchett, 2014; Sessa et al., 2012). Taxa were grouped into FEs using the mFD package in R (Magneville et al. 2022). Functional diversity metrics were calculated using five levels of spatial aggregation, or units: regional (WIS+GCP, unit size 1), basinal (WIS and GCP, separately, unit size 2), 5° paleolatitudinal bins (unit size 3), 360km² grid cells (unit size 4), and 60km² grid cells (unit size 5). These units assess how functional diversity metrics change across space, test sensitivity to spatial binning, and allow assessment of patterns that may emerge at different spatiotemporal scales of analysis.

Aggregations in 60km² and 360km² grid cells (referred to as “nodes” in Kiel 2017) are intended to capture local paleo-“community” structure, though they do not necessarily scale to present-day ecological communities or homogeneous environments. Unit locations were estimated based on the average location of occurrences they contain, converted to paleocoordinates using the chonosphere package in R (Kocsis and Raja 2020).
Table 4. Table of the FE present in the database and their code values.

<table>
<thead>
<tr>
<th>Mobility level</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile, unattached</td>
<td>MU</td>
<td>Includes actively mobile, freely mobile, slow moving, fast moving, and creeping unattached taxa.</td>
</tr>
<tr>
<td>Facultatively mobile, unattached</td>
<td>FU</td>
<td>Includes facultatively mobile and unattached taxa.</td>
</tr>
<tr>
<td>Facultatively mobile, attached</td>
<td>FA</td>
<td>Includes facultatively mobile and swimming taxa with some form of attachment (byssate, pedicle, or cementing)</td>
</tr>
<tr>
<td>Immobile, unattached</td>
<td>IU</td>
<td>Includes free-lying, boring taxa.</td>
</tr>
<tr>
<td>Immobile, attached</td>
<td>IA</td>
<td>Includes stationary and attached (cemented, epibiont, byssate) taxa.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feeding mechanism</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspension feeding</td>
<td>S</td>
<td>Includes all forms of micro-feeding from the water column (i.e., filter feeding, suspension feeding, etc.). Various diet types are included in this strategy (i.e., carnivore) but it is non-specific.</td>
</tr>
<tr>
<td>Deposit feeding</td>
<td>D</td>
<td>Includes deposit feeding in the subsurface or at the sediment surface and detritivore feeding. Some taxa have mixed deposit/suspension feeding but just classified as deposit feeding.</td>
</tr>
<tr>
<td>Herbivore / grazing</td>
<td>H</td>
<td>Includes all forms of herbivorous feeding, including grazing.</td>
</tr>
<tr>
<td>Carnivore / predatory</td>
<td>C</td>
<td>Includes all macro-carnivorous and predatory feeding strategies. Implies active feeding but includes carnivorous scavenging as well.</td>
</tr>
<tr>
<td>Omnivore</td>
<td>O</td>
<td>Includes feeding strategies that incorporate both carnivorous and herbivorous feeding. If multiple differing feeding modes were attributed to a taxon, including omnivore, omnivore listed.</td>
</tr>
<tr>
<td>Photosymbiotic</td>
<td>P</td>
<td>Includes photosymbiotic taxa.</td>
</tr>
<tr>
<td>Chemosymbiotic</td>
<td>CH</td>
<td>Includes chemosymbiotic taxa, including chemosymbiotic deposit feeders.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tiering Level/Lifestyle Mode</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boring</td>
<td>B</td>
<td>Includes boring taxa and nestlers.</td>
</tr>
<tr>
<td>Epifaunal</td>
<td>E</td>
<td>Includes all taxa living above the sediment-water interface (non-burrowing) that do not have upward mobility in the water column (includes all epifaunal levels, i.e., intermediate epifaunal)</td>
</tr>
<tr>
<td>Semi-infaunal</td>
<td>PI</td>
<td>Includes all taxa which living partially beneath the sediment surface and taxa which both burrow and live epifaunally.</td>
</tr>
<tr>
<td>Infaunal</td>
<td>I</td>
<td>Includes all taxa living below the sediment surface, either shallow, deep, or at uncertain infaunal depths</td>
</tr>
<tr>
<td>Nektonic</td>
<td>N</td>
<td>Includes all taxa living freely in the water column, well above the sediment surface.</td>
</tr>
<tr>
<td>Nekto-benthic</td>
<td>NB</td>
<td>Includes all taxa that live in the water column close to the sediment surface or those with uncertainty as to which level of the water column they occupy.</td>
</tr>
</tbody>
</table>
Generic richness, functional richness ($\alpha_F$; the number of unique FEs), and functional evenness were calculated for each spatiotemporal level (i.e., five time bins at five spatial scales). Functional evenness, or the relative abundance of genera within FEs, was calculated using Simpson’s Measure of Evenness (SME). SME is calculated by normalizing the inverse Simpson’s Diversity Index by the number of total FE present:

$$S = \frac{1}{\sum p_i^2}$$  \hspace{1cm} (EQ 1)

$$Simpson's measure of evenness = \frac{D}{S}$$  \hspace{1cm} (EQ 2)

Where $p_i$ represents the proportion of genera in a given FE and $S$ is the total number of FEs present. Values closer to 0 indicate lower evenness and higher values indicate higher evenness (SME has no upper limit); no value is assigned for units containing a single FE. Similarly, SME cannot be calculated if all FEs contain a single genus. In both cases, units were given SME values of 0. SME is weighted by the abundance or dominance of a specific FE and was chosen because it is less sensitive to taxon-richness and provides an estimate of the generic abundance distribution (Magurran, 2003; see SI for discussion of other metrics tested). Richness and evenness values were compared between temporally and spatially adjacent units by calculating the proportional change in $\alpha_F$ and SME values.

Fundamental sources of bias in the database were tested using several methods. The relationship between generic and functional richness was analyzed by correlating the two variables based on 60km unit aggregates. Bootstrapping of 60km unit $\alpha_F$ values based on the number of unique genera were performed to determine how $\alpha_F$ varies with generic richness ($\alpha_G$). Additionally, SQS subsampling (Alroy 2010) of the FEs in each substage was
performed to find the generic sample size estimates appropriate for capturing the majority of FEs (see SI page 2).

Functional entity assemblages within spatiotemporal units were compared below the regional level using network modeling in the EDENetworks software (Moalic et al. 2012; Kivelä et al. 2015) following methods outlined in Kiel (2017) and Purcell et al. (2023). EDENetworks for presence-absence data utilizes Bray-Curtis dissimilarity values to assign weights to the links between each unit in the network, allowing connections to be viewed at different threshold values of dissimilarity. Betweenness centrality, a measure of the degree to which a unit acts as a geographic connection between regions, was also calculated (Kivelä et al., 2015). Network topology patterns at the percolation point were used to determine if FE assemblages were provincial (i.e., geographically controlled) given the assumption that network components represent distinct biologically-based “community” groups (Newman 2012). The percolation point depicts the threshold at which all components in a network are connected without forming a single component (Kivelä et al. 2015). The specific FEs and associated PBDB paleoenvironmental characteristics of occurrences within components found at and just below percolation were qualitatively assessed (Appendix B, Table S35). To determine if the WIS and GCP differ in $\alpha_F$, functional diversity, and FEs present, the two basins (unit size 2) were analyzed separately using all previously described methods. Differences in the three most common clades (cephalopods, bivalves, and gastropods) were also compared. Other clades made up too few of the occurrences to form useful networks. See supplementary materials for details regarding methods and results (SI, page 23).
Results

Relationship Between Functional and Generic Diversity

Generic and functional richness have a strong positive correlation that loses strength at approximately more than 100 genera (Appendix B, Figure S1). Generic richness ($\alpha_G$), previously considered a proxy for sampling bias in this database (Purcell et al., 2023), is therefore not representative of potential $\alpha_F$ when more than ~100 genera are present. Furthermore, Bootstrap analysis of 60km units indicate standard error of $\alpha_F$ in units has a positive correlation with unit $\alpha_F$ that depreciates above approximately 20 FEs (Appendix B, Figure S3). This suggests that the distribution of genera in FEs becomes more variable (i.e., less even) when more than 20 FEs are present; that is, functional evenness should be more biased in units with < 20 FEs, while also poorly representing potential $\alpha_F$. However, SQS analysis of substages, which was used to determine what level of $\alpha_G$ best captures the majority of common FEs, indicates that $\alpha_G > 16$ in the Campanian and $\alpha_G > 18$ in the Maastrichtian are likely to capture the most common FEs (Appendix B, Table S1). Therefore, spatiotemporal units with fewer than 100 unique genera are under-sampled relative to $\alpha_F$ and functional evenness but units with more than 16 and 18 unique genera may reliably represent FE assemblages (paleo-“communities”) for the Campanian and Maastrichtian, respectively.

All units containing less than 16 or 18 unique genera for the Campanian and Maastrichtian, respectively, were removed given that they are not considered to be representative of either FE assemblages or functional diversity metrics. This includes 67 out of 533 units at 60km (unit size 5), 80 out of 132 units at 360km (unit size 4) aggregated spatial resolution, and seven out of 32 paleolatitude bins (unit size 3). Removal of these
under-sampled units was not found to notably alter results at any level of aggregation. Figures and tables including all units regardless of generic richness are available in the SI.

Substage-level Diversity:

There are a total of 38 FEs in the database (Appendix B, Table S2): 30 each in the Early Campanian and the Middle Campanian, 33 in the Late Campanian, 35 in the Early Maastrichtian, and all 38 in the Late Maastrichtian, representative of 18 taxonomic classes (Table 4 & 5). Functional entities are overwhelmingly assigned at taxonomic levels below class. Regional $\alpha_F$, in unit size 1, is highly stable, rising only slightly (~6% to 10%) across any two substages while generic richness is more variable, rising across the Campanian, before falling between the Late Campanian and Early Maastrichtian and rising through the Late Maastrichtian (Figure 6A, Appendix B, Table S3). Functional evenness (SME) decreases 50% between the Middle and Late Campanian and fluctuates slightly across the other substages (Figure 6B). SME change only corresponds positively with generic richness in the Early to Middle Campanian while $\alpha_F$ is stable (Appendix B, Table S3).

Generic and FE diversity are furthermore disconnected through time. Only three FEs are added between the Middle to Late Campanian, despite the sharp increase in generic richness (81%) (Figure 8). Simultaneously, mobile unattached epifaunal carnivores (MU-E-C in Figure 4; composed of gastropods and crustaceans) rose in dominance from ~8% to ~30% in proportional generic richness, while facultatively mobile unattached infaunal suspension feeders (FU-I-SF in Figure 8; composed of bivalves) fell from 17% to 10% (Appendix B, Table S4). No other FE experienced a change in proportional generic diversity of more than 5% across the interval. The few shifts in generic occupation of FEs caused functional evenness to decline sharply. Following the Late Campanian, MU-E-Cs remain dominant;
immobile attached epifaunal suspension feeders (IA-E-SF in Figure 8) and FU-I-SF, both previously the most diverse FEs, never again reach their former degree of dominance (Appendix B, Table S4). All other FEs have low proportional generic diversity throughout all five time bins (<9%). At any given substage no more than eight FEs are “gained”

![Figure 7. Line plots of substage values for A) generic richness (red, solid line) and functional richness (blue, dotted line) and B) generic richness (red, solid line) and SME (blue, dotted line). The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric. Proportional change in values is denoted as percentages. Camp. = Campanian, Maastr. = Maastrichtian.](image_url)
Table 5. List of taxonomic classes and the FE they represent. See Table S36 for details on genera. Asterisks denote FE s represented by a single class.

<table>
<thead>
<tr>
<th>FE</th>
<th>Class</th>
<th>FE</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA-E-P</td>
<td>Anthozoa</td>
<td>IA-B-SF</td>
<td>Demospongea</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>Anthozoa</td>
<td>IA-E-SF</td>
<td>Demospongea</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>Asteroidea</td>
<td>MU-E-H</td>
<td>Echinidea</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>Bivalvia</td>
<td>MU-E-O</td>
<td>Echinidea</td>
</tr>
<tr>
<td>*FU-I-C</td>
<td>Bivalvia</td>
<td>*MU-PI-D</td>
<td>Echinidea</td>
</tr>
<tr>
<td>*FU-I-CH</td>
<td>Bivalvia</td>
<td>MU-I-D</td>
<td>Echinidea</td>
</tr>
<tr>
<td>*MU-I-CH</td>
<td>Bivalvia</td>
<td>MU-E-C</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*FA-I-D</td>
<td>Bivalvia</td>
<td>MU-PI-C</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*FU-I-D</td>
<td>Bivalvia</td>
<td>MU-I-C</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>Bivalvia</td>
<td>*FU-E-H</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>IA-E-P</td>
<td>Bivalvia</td>
<td>MU-E-H</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*IA-PI-P</td>
<td>Bivalvia</td>
<td>*FU-PI-H</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*IU-PI-P</td>
<td>Bivalvia</td>
<td>*MU-PI-H</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*FU-B-SF</td>
<td>Bivalvia</td>
<td>MU-E-O</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>IA-B-SF</td>
<td>Bivalvia</td>
<td>FA-E-SF</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*IU-B-SF</td>
<td>Bivalvia</td>
<td>FU-E-SF</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>Bivalvia</td>
<td>*MU-E-SF</td>
<td>Gastropoda</td>
</tr>
<tr>
<td>*FU-I-SF</td>
<td>Bivalvia</td>
<td>IA-E-SF</td>
<td>Gymnolaemata</td>
</tr>
<tr>
<td>*IU-I-SF</td>
<td>Bivalvia</td>
<td>IA-E-SF</td>
<td>Hexanauplia</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>Bivalvia</td>
<td>MU-E-O</td>
<td>Homarus</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>Bivalvia</td>
<td>IA-E-SF</td>
<td>Hydrozoa</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>Bivalvia</td>
<td>FA-I-SF</td>
<td>Lingulata</td>
</tr>
<tr>
<td>*IU-E-SF</td>
<td>Bivalvia</td>
<td>MU-E-C</td>
<td>Malacostraca</td>
</tr>
<tr>
<td>*FA-PI-SF</td>
<td>Bivalvia</td>
<td>MU-I-C</td>
<td>Malacostraca</td>
</tr>
<tr>
<td>*IA-PI-SF</td>
<td>Bivalvia</td>
<td>MU-I-D</td>
<td>Malacostraca</td>
</tr>
<tr>
<td>*MU-I-SF</td>
<td>Bivalvia</td>
<td>*MU-E-D</td>
<td>Malacostraca</td>
</tr>
<tr>
<td>*IA-I-SF</td>
<td>Bivalvia</td>
<td>MU-E-O</td>
<td>Malacostraca</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>Bryozoa</td>
<td>MU-E-C</td>
<td>Merostomata</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>Cephalopoda</td>
<td>IA-E-SF</td>
<td>Rhychonellata</td>
</tr>
<tr>
<td>*MU-NB-C</td>
<td>Cephalopoda</td>
<td>MU-I-D</td>
<td>Scaphopoda</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>Cephalopoda</td>
<td>IA-E-SF</td>
<td>Stenolaemata</td>
</tr>
<tr>
<td>*MU-NB-SF</td>
<td>Cephalopoda</td>
<td>IA-E-SF</td>
<td>Thecostraca</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>Crinoidea</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 8. Plot of FE relative generic richness through time. Circle size indicates relative generic richness within each substage, color indicates a proportional increase (red) greater than 0.05, decrease (blue) greater than 0.05, change less than 0.05 (orange) from the previous substage. Grey indicates a value cannot be compared with a previous substage. Green rectangles denote FEs only found in the GCP and red squares denote FEs missing from the Late Maastrichtian WIS. Gastropod silhouettes denote gastropod-type FEs; larger green silhouettes indicate herbivorous FEs absent in the Late Maastrichtian 360km WIS network component. FEs that are present in every 60km unit just below percolation in each substage are denoted with an asterisk (GCP component) and a cross (WIS component). See Table 4 for FE naming key.
(i.e., not observed in the previous interval but currently present) and only two FEs are “lost”
(i.e., previously observed but missing in the current interval). All FEs lost/gained have very
low proportional generic diversity (Figure 8). The two FEs that are unobserved in a substage
but present in the preceding and succeeding substages are facultatively unattached epifaunal
herbivores (FU-E-H) and immobile attached epifaunal photosymbionts (IA-E-P), in the
Middle Campanian and Late Maastrichtian, respectively (Figure 8, Appendix B, Table S4).
While both FEs are among the least diverse FEs within the substage prior to their apparent
extirpation (each represented by a single genus), in the succeeding substage they are not the
least diverse and are represented by these same genera as well as either one (FU-E-H) or two
(IA-E-P) additional genera.

WIS vs GCP Influence

The WIS, analyzed independently (unit size 2), has fairly stable $\alpha_F$ (4% and 7%
increase) in the Campanian, but declines from the Late Campanian through the Maastrichtian
(13-30% decline; Appendix B, Figure S4-S5, Table 6). WIS generic richness parallels this
pattern with greater proportional change. In the GCP generic and functional richness increase
throughout Appendix B, (Figure S6-S7), particularly between the Middle and Late
Campanian, indicating that the GCP data strongly influences richness trends in the overall
database. WIS and GCP SME change inversely to generic richness, even when $\alpha_F$ is stable or
increases (Appendix B, Table S6 and S8).

Of the eight FEs “gained” across the study interval, five are only found in the GCP
(Figure 8): three photosymbionts, one boring suspension feeder, and an attached infaunal
deposit feeder. No photosymbionts and only one boring FE are present in the WIS at any
time bin. Sampling is not a reasonable driver of FE differences between the WIS and GCP.

All WIS

**Table 6.** Table of regional generic richness, functional richness ($\alpha_f$), and functional evenness (SME) values in each substage. Values for the overall database are given first, followed by the values observed in the WIS and then GCP. Camp. = Campanian, Maastr. = Maastrichtian.

<table>
<thead>
<tr>
<th>Substage</th>
<th>Generic Richness</th>
<th>Functional Richness ($\alpha_f$)</th>
<th>Functional Evenness (SME)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Camp.</td>
<td>143 (108/47)</td>
<td>30 (28/14)</td>
<td>0.48 (0.53/0.61)</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>182 (151/65)</td>
<td>30 (29/22)</td>
<td>0.50 (0.51/0.67)</td>
</tr>
<tr>
<td>Late Camp.</td>
<td>329 (199/213)</td>
<td>33 (31/29)</td>
<td>0.25 (0.41/0.21)</td>
</tr>
<tr>
<td>Early Maastr.</td>
<td>300 (119/248)</td>
<td>35 (27/34)</td>
<td>0.28 (0.51/0.26)</td>
</tr>
<tr>
<td>Late Maastr.</td>
<td>323 (42/309)</td>
<td>38 (19/38)</td>
<td>0.23 (0.78/0.22)</td>
</tr>
<tr>
<td>Mean</td>
<td>255.4 (123.8/176.4)</td>
<td>33.2 (26.8/27.4)</td>
<td>0.35 (0.55/0.39)</td>
</tr>
<tr>
<td>Median</td>
<td>300 (119/213)</td>
<td>33 (28/29)</td>
<td>0.28 (0.51/0.26)</td>
</tr>
<tr>
<td>S.D.</td>
<td>86.60 (57.79/115.3)</td>
<td>3.42 (4.60/9.58)</td>
<td>0.13 (0.14/0.23)</td>
</tr>
</tbody>
</table>

Substages have more than adequate sampling to produce representative FE assemblages (i.e., >16 and 18 unique genera for the Campanian and Maastrichtian, respectively), and all but the Late Maastrichtian WIS dataset have adequate generic richness to capture potential $\alpha_f$ (i.e., >100 unique genera). No FEs are completely lost between the Early Campanian and Late Maastrichtian in the GCP, however, several FEs are absent in the Campanian of the GCP where generic richness is frequently <100; this suggests that these absent FEs may be caused by poor sampling (Appendix B, Figure S7; Table 6 and S10). That said, generic richness in all GCP substages is adequate to capture the 20 most common FEs and the FEs with the highest overall proportional generic diversity providing a baseline community structure throughout the study interval.
Class Influences.

Cephalopods represent 85 unique genera and 13,170 fossil occurrences. They are characterized by only four FEs, all of them mobile and unattached, nektonic or nektobenthic, and carnivorous or suspension feeders (Table 5). All four FEs are present throughout the study interval and are fairly even (<16% SME change) despite changes to cephalopod generic richness (5-54% generic richness change; Appendix B, Figure S8-S9, Table S12). Bivalves make up 184 unique genera, 14,085 fossil occurrences, and represent 24 FEs in the database. Bivalve generic and functional richness generally increase across the study interval, but SME values are more stable, indicating even dispersal of additional genera to bivalve FEs (Appendix B, Figure S10-S11). Bivalve SME only changes notably between the Early and Middle Campanian, decreasing by ~23% while generic diversity increases by 33% and $\alpha_F$ increases by 6% (Appendix B, Table S14). Gastropods, which make up 184 unique genera, 4664 fossil occurrences, and represent 11 FEs, are less stable than bivalves or cephalopods (Appendix B, Figure S12-S13). The number of gastropod genera rises sharply between the Middle and Late Campanian (408%), resulting in a sharp increase in $\alpha_F$ (57%), and a clear decline in SME (66%; Appendix B, Figure S12, Table S16).

Functional versus Generic Diversity Across Distance and Spatial Scales: Paleolatitudinal Bins (unit size 3).

When the data are aggregated by paleolatitudinal bins and compared between substages, patterns of proportional change in all metrics are dissimilar from the regional analysis (unit size 1) except between the Middle and Late Campanian (Table 7 and 8; Appendix B, Figure S15). Proportional change across paleolatitudes bins (i.e., LDG comparison) indicates generally decreasing generic and functional richness and increasing
Figure 9. Beeswarm plot of proportional change in A) generic richness, B) functional richness ($\alpha_f$), and C) functional evenness (SME). Values reflect change in metric between adjacent paleolatitude for each substage. Camp. = Campanian, Maastr. = Maastrichtian.
Table 7. Table of paleolatitude bin generic richness, functional richness (αf), and functional evenness (SME) and mean, median, and standard deviation values for each substage. NA value indicate no value measured for that unit in that substage. LC = Early Campanian, MC = Middle Campanian, UM = Late Campanian, LM = Early Maastrichtian, UM = Late Maastrichtian.

<table>
<thead>
<tr>
<th>Generic Richness</th>
<th>Functional Richness (αf)</th>
<th>Functional Evenness (SME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EC MC LC EM LM</td>
<td>EC MC LC EM LM</td>
</tr>
<tr>
<td>30-35°N</td>
<td>11 NA 14 210 238</td>
<td>4 NA 5 32 37</td>
</tr>
<tr>
<td>35-40°N</td>
<td>41 65 142 155 271</td>
<td>14 22 28 30 36</td>
</tr>
<tr>
<td>40-45°N</td>
<td>14 40 129 19 3</td>
<td>7 20 18 10 2</td>
</tr>
<tr>
<td>45-50°N</td>
<td>67 134 149 109 15</td>
<td>22 28 29 27 8</td>
</tr>
<tr>
<td>50-55°N</td>
<td>59 70 119 48 27</td>
<td>22 23 27 16 15</td>
</tr>
<tr>
<td>55-60°N</td>
<td>37 27 87 18 16</td>
<td>17 16 25 9 9</td>
</tr>
<tr>
<td>60-65°N</td>
<td>5 1 7 NA NA</td>
<td>5 1 4 NA NA</td>
</tr>
<tr>
<td>Mean</td>
<td>33 56 92 93 95</td>
<td>13 18 19 21 18</td>
</tr>
<tr>
<td>Median</td>
<td>37 53 119 79 22</td>
<td>14 21 25 22 12</td>
</tr>
<tr>
<td>S.D.</td>
<td>24. 45. 59. 78. 124</td>
<td>7.7 9.4 10. 10. 15.</td>
</tr>
</tbody>
</table>

Table 8. Summary statistics for proportional change in generic richness, functional richness (αf), and functional evenness (SME) for paleolatitude bins across adjacent substages. Camp. = Campanian, Maastr. = Maastrichtian. Values with less than the SQS sample estimate have been removed from these summary statistics.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Richness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.38</td>
<td>1.29</td>
<td>-0.48</td>
</tr>
<tr>
<td>Median</td>
<td>0.39</td>
<td>1.18</td>
<td>-0.60</td>
</tr>
<tr>
<td>SD</td>
<td>0.54</td>
<td>0.93</td>
<td>0.39</td>
</tr>
</tbody>
</table>

| Functional Richness (αf) |                        |                         |                          |
| Mean                   | 0.21                   | 0.19                    | -0.30                    | -0.20                    |
| Median                 | 0.16                   | 0.17                    | -0.41                    | -0.03                    |
| SD                     | 0.28                   | 0.25                    | 0.29                     | 0.44                     |

| Functional Evenness (SME) |                        |                         |                          |
| Mean                   | 0.00                   | -0.37                   | 1.23                     | 0.24                     |
| Median                 | -0.03                  | -0.35                   | 0.66                     | 0.28                     |
| SD                     | 0.12                   | 0.27                    | 2.02                     | 0.41                     |
evenness across paleolatitude from low to high latitudes, except between the 40-45°N to 45-50°N bins where richness increases and evenness decreases in all substages (Figure 9). However, these patterns are considered to be strongly influenced by sampling bias, as will be discussed more below.

360km and 60km Units (unit size 4).

Temporal patterns apparent at the substage and paleolatitudinal bin level are present but dampened at higher spatial resolutions (360km and 60km units; Figure 10, Appendix B, Figures S18-19; Table S30 and S32). Proportional changes in generic and functional richness between 360km and 60km units are not well supported by linear regression but indicate that the two values are poorly correlated across distance in the Campanian and positively correlated (i.e., increasing with increased distance) in the Maastrichtian (Appendix B, Figure S20-S27). Proportional changes in 60km and 360km units SME values across distance are negatively correlated in all substages except in the Early Campanian 360km units where SME is slightly positively correlated with distance (Appendix B, Figure S22 and S26). Thus, SME decreases across distance in most substages and, regardless of spatial resolution, agrees with assumptions about spatial autocorrelation and previous clustering assessments of the database (Purcell et al., 2023). Functional richness only increases with distance in the Maastrichtian, most likely due to the addition of GCP units with high generic richness (Appendix B, Figure S1).
Figure 10. Box and whisker plots of proportional change between substages for generic richness, functional richness ($\alpha_f$), and functional evenness (SME) at the 360km and 60km unit scales. Triangles indicate mean values. Camp. = Campanian, Maastr. = Maastrichtian.
Network models of Functional Assemblages:

Network models indicate spatiotemporal units are well connected overall across the study region, though regional connectivity diminishes with higher spatial resolution (as expected) and is less robust in the Late Maastrichtian. Paleolatitude bin networks (unit size 3) show strong connections between both adjacent and nonadjacent paleolatitudes for all substages, except in bins with poor sampling (i.e., <SQS threshold; Figure 11). 360km and 60km unit networks (unit sizes 4 and 5, respectively) do not separate into distinct components in any substage except the Late Maastrichtian (Appendix B, Figure S28-S29, 12-14). The 360km Middle Campanian network has two geographically overlapping components, represented by a functionally diverse assemblage and a depauperate assemblage of only mobile unattached nektobenthic suspension feeders (MU-NB-SF in Figure 8; Appendix B, Table S33; Figure S28). However, these components are not observed in the 60km Middle Campanian unit network (Figure 12); this suggests that data aggregation influences network modeling results methodologically (e.g., via percolation points) and may not relate directly to biological associations. The 60km Early Maastrichtian network also has distinct geographically overlapping components representing functionally diverse and functionally depauperate assemblages, most likely due to sampling differences between units (Figure 12; Appendix B, Table S34).
Figure 11. Network models across paleolatitude bins. Colored points represent the paleolatitude (cool = higher paleolatitude, warm = lower paleolatitude). Size indicates betweenness centrality. Line color and thickness indicate degree of similarity (thinner/thicker lines = greater similarity). TH = Threshold.
Figure 12. Campanian 60km unit networks of FE assemblages plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Unit size represents Betweenness Centrality. Unit colors indicate paleolatitude as in Fig. 8 (warmer = low paleolatitudes and cooler = higher paleolatitudes). TH = Threshold.
Figure 13. Early Maastrichtian 60km unit networks plotted by FE assemblages (left) and based on paleo-coordinates (right). Unit colors of upper networks indicate paleolatitude (warmer = lower paleolatitudes and cooler = higher paleolatitudes). Lower networks are colored by groups of highly similar units or components. TH = Threshold.
Figure 14. Late Maastrichtian 60km unit networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Unit colors of upper networks indicate paleolatitude (warm = southern; cool = northern). Lower networks are colored by component. TH = Threshold.
In the Late Maastrichtian 360km and 60km unit networks, however, separate components at percolation occupy mainly either the WIS or GCP, indicating provincialism (Figure 14, S29; see Purcell et al. 2023 for more on provincialism). These components represent a functionally diverse GCP assemblage whose units all share five FEs and a functionally less diverse WIS assemblage (Appendix B, Table S33 and S34). When all 360km substage units are modeled together, strong links persist between all regions and substages regardless of distinct components observed in individual substage networks (Appendix B, Figure S30). WIS sampling is less extensive in the Late Maastrichtian than in the GCP, which may explain these differences.

Differentiation in network components based on dominantly siliciclastic and carbonaceous sediments is potentially present in 60km unit networks (Appendix B, Table S35), but a more thorough investigation beyond the scope of this paper would be necessary to determine its prevalence. Minimum Bray-Curtis dissimilarity values between units increase with distance in all substages except in the 360km Early Campanian network (Appendix B, Figure S31-S32), indicating spatial autocorrelation and diminishing FE similarity across distance, as expected. Class-level analysis of 60km networks indicates bivalves and cephalopods form strong, non-provincial network connections (Appendix B, Figure S33-S36) but gastropods form provincial components given their WIS-dominated Campanian distribution and GCP-dominated Maastrichtian distribution (Appendix B, Figure S37-S38). Three of the five FEs shared by all units in the Late Maastrichtian 60km unit network GCP component are represented by gastropods (Figure 8; Appendix B, Table S34). However, there is strong evidence that this is once again a product of sampling bias due to differential
preservation of taxonomically and ecologically informative features (Smith 2001; Smith et al. 2001; Dean et al. 2019).

**Discussion**

Several previous studies looking at functional diversity across major environmental perturbations have suggested that functional ecological signals are more robust to disturbances such as mass extinctions (Foster and Twitchett 2014; Dunhill et al. 2018; Edie et al. 2018; Pimiento et al. 2020), but are not stable across latitudes (Edie et al. 2018; Schumm et al. 2019). Here we looked at the dynamics of functional entities (FEs) during a background (non-mass extinction) interval that experienced environmental changes in the form of sea level oscillations and their effects on ocean circulation in a restricted epicontinental seaway (WIS) as well as the adjacent normal-marine shelf (GCP). Our results indicate that functional diversity is indeed stable through time despite environmental shifts, and moreover, is stable across a 35° latitudinal gradient. Both patterns can be explained through functional redundancy found previously to produce stability despite environmental perturbations (Pimiento et al., 2020). The WIS and GCP are also functionally distinct, a characteristic of oceanic basin environmental differences previously observed by Edie et al. (2018). Our results suggest, overall, that paleontological data in deep time is relevant and useful for spatiotemporal comparisons of functional trends, but that aspects of data aggregation and type, such as class-level diversity should be taken into account.

**Functional Diversity in Classes:**

Analyses of FD trends within specific molluscan classes showed some taxon-specificity, which should be noted for future research as a potential source of bias (e.g., especially among gastropods). In most substages, additional bivalve genera are distributed
into FEs evenly, resulting in <7% SME change despite more dramatic generic and functional richness fluctuations (Appendix B, Figure S11; Table S14). Cephalopods, on the other hand, are functionally stable but dramatic changes to cephalopod generic richness resulted in changes to functional evenness (<16%) that was not necessarily inverse to generic richness as would be expected (Appendix B, Figure S8, Table S12). Finally, gastropod richness and evenness values fluctuated strongly across all substages relative to other mollusks (Appendix B, Figure S13). Gastropod generic diversity increased sharply between the Middle and Late Campanian (408%), resulting in rising $\alpha_F$ (57%) and a clear decline in SME (66%). This parallels changes to database metrics not observed in either cephalopod or bivalve data alone, suggesting that gastropod patterns strongly influenced the overall results. These results show that trends in functional diversity observed within single taxonomic classes may poorly represent other clades or overall paleo-“community” structure. Thus, ideal datasets should include diverse taxonomic sampling to best observe broad paleo-“community” structure through space and time. However, it should be noted that ecological and sampling identification biases in specific clades, such as WIS gastropods, may also skew results. Close attention should therefore be given to the diversity of different higher level taxonomic groups in a dataset. Network models using the different mollusk classes also indicate that ecological differences can influence these results. Bivalves have high evenness, resulting in more consistent representation of FE within high-resolution spatial units that increases network connectivity while gastropods have higher FE diversity ($\alpha_F$) and low evenness, which could generate distinct network components. However, determining the specific causes are beyond the scope of this analyses.
**Functional Stability through Time:**

Our analyses overall indicate temporal stability in functional diversity across background intervals, despite changes to the environment, supporting hypothesis 1. Results show that $\alpha_F$ increases only slightly across the study interval, despite more rigorous (fluctuating) changes to patterns of generic richness. No FEs are lost, and only a few are “gained” between any two substages. FEs gained/lost between individual substages are rare, with low proportional generic richness, throughout their temporal range of existence. Given their rarity, timing of loss/gain may be influenced by taphonomic or sampling biases (e.g., the Signor-Lipps effect; (Signor and Lipps 1982; Nawrot et al. 2018; Dean et al. 2019). Moreover, previous analyses support that any loss of FEs likely reflects (brief) regional extirpation versus a global pattern of extinction, which is also observed here (Foster and Twitchett 2014; Dunhill et al. 2018; Edie et al. 2018; Pimiento et al. 2020). For example, immobile attached epifaunal photosymbionts (IA-E-P) is represented by only the genus *Titanosarcolites* in the Late Campanian, is unobserved in the Early Maastrichtian, potentially indicating extirpation, and is subsequently represented by *Titanosarcolites* as well as both *Dictyaraea* and *Gyropleura* in the Late Maastrichtian. Therefore, IA-E-P is most likely either present in the Early Maastrichtian, though unsampled, or only briefly absent due to the short-lived extirpation of *Titanosarcolites*. Paleolatitude bins and 60/360km units are less functionally stable, but this is likely due to spatial resolution and sampling: reduced geographic bin size often results in fewer than 100 genera/unit, which results in control of $\alpha_F$ and functional evenness by generic richness rather than ecology (Appendix B, Figure S1 and S3). Unsurprisingly, functional evenness is higher and more stable at higher spatial resolutions across time, which demonstrates a limit within spatial resolution to adequately
capture the full range of FEs present, or the distribution of genera within them. Thus, functional diversity trends are not necessarily consistent at all spatial scales, refuting hypothesis 4, though further discussion of the utility of observations made using different spatial resolutions is made below. This is an important finding in a field where higher resolution is typically considered more “real,” with the implication being that there is a specific spatial resolution for plausibly reconstructing paleo-“communities.” Determining whether this “best” resolution is generalizable or case-specific will take continued analyses that include tests of spatial sensitivity across different time periods and habitats.

**A Flat Latitudinal Functional Diversity Gradient:**

Our analysis shows that the Late Cretaceous WIS + GCP region lack a latitudinal functional diversity gradient signal; this confirms hypothesis 1 in part, which predicted that functional diversity is stable across space, and contradicts previous studies of latitudinal functional diversity gradients in the Recent (Edie et al., 2018; Schumm et al., 2019). At face value, both generic and functional richness decline between adjacent paleolatitude bins from low to high latitude but few bins contain sufficient unique genera (i.e., >100) to adequately represent “real” functional diversity (Table 7). The influence of true diversity change cannot be confidently separated from sampling bias. Paleolatitude bin networks are consistently well-connected regardless of degree of paleolatitude bin separation spatially, and both 360km and 60km unit networks lack latitudinally distinct components, indicating similar functional assemblages (Figure 12 & 13). Even taken at face value, paleolatitudinal patterns would indicate diversity peaks in mid-latitudes of the study region, potentially due to sampling bias such as is present in current diversity studies (i.e., Menegotto & Rangel, 2018). Mid-latitude peaks in diversity, even when accounting for sampling bias, have been observed for
Cretaceous terrestrial fauna potentially resulting from landmass distributions (Mannion et al. 2012); these mid-latitude peaks may therefore be representative of shifting WIS basin geometries, as can be observed in the data distribution of this study (Figure 6).

There is, furthermore, poor evidence for a global Late Cretaceous LDG in the literature once sampling bias has been taken into account (Huang et al. 2014; Mannion et al. 2014; Brodie and Mannion 2022). In our WIS+GCP data, functional diversity metrics do not change consistently from south to north among bins with sufficient sampling (>100 GR) and in the Late Campanian, where paleolatitude bins have >84 genera from 35° to 60°N, α_F declines inconsistently by a total of ~12%, losing only three FEs. This contrasts sharply with the dramatic decline in α_F observed by Schumm and colleagues for present-day bivalves across the same 25° gradient (Schumm et al., 2019) (Figure 6), and also conflicts with LDG patterns observed in present-day birds showing decreased α_F at higher latitudes in parallel with generic richness (Edie et al., 2018; Schumm et al., 2019). Changes in functional diversity across latitude furthermore do not form consistent nested subsets of FEs (Table. S24-28), as is observed in the latitudinal gradient of today (Schumm et al., 2019). This suggests that FEs are not lost systematically from some maximum α_F assemblage, but instead the specific FE vary across latitudinal space via turnover.

This suggests that the lack of change in functional diversity across latitude is a distinct pattern of the Late Cretaceous compared to today, or at least one common to Greenhouse climate regimes in Earth history (Huang et al. 2014; Brodie and Mannion 2022). A flat Late Cretaceous latitudinal temperature gradient documented for the study area could have promoted reduced environmental disparity across latitude, thereby fostering functional stability (Huber et al. 1995; Mannion et al. 2012). A flat temperature gradient is supported by
the observation that terrestrial diversity in the Late Cretaceous correlates with landmass distributions rather than climate gradients (Baron, 1989; Poulsen, 2010; Mannion et al., 2012; Nicholson et al., 2016). The latitudinally extensive distribution of the WIS basin and its unique oceanographic conditions, including non-normal marine conditions such as lower salinity and periodic vertical stratification, as well as mixing via counterclockwise gyre throughout the Late Cretaceous (Fisher et al. 1994; Slingerland et al. 1996; Leckie et al. 1998; Longman et al. 1998; Steel et al. 2012; Lowery et al. 2018), may contribute to marine invertebrate taxonomic stability despite 25° of geographic distance. The oceanographic conditions throughout the WIS that lead to a breakdown or lack of development of a latitudinal diversity gradient may also have promoted functional homogenization and marine invertebrate stability in functional diversity.

_Provincialism in Functional Assemblages:_

The WIS and GCP regions are functionally distinct and evolve towards independent provinces as habitat differences become exacerbated through time, conflicting with our second hypothesis of stability between regions. Five FEs are absent in the WIS which are present in the GCP when sampling intensities are sufficient, including photosymbiotic and boring taxa consistent with previous studies that observed few reef-associated taxa in the WIS (Gill and Cobban 1966b; Sohl 1967; Caldwell 1968; Kauffman 1984; Kauffman and Caldwell 1993). While these unoccupied FEs are consistently rare in the GCP, their absence over approximately 18 My suggests that the WIS is less functionally diverse than the GCP, most likely due to differences in their ocean environment (i.e., restricted epicontinental sea versus open-ocean facing shelf environment; Caldwell, 1968; Gill & Cobban, 1966; Kauffman, 1984; Kauffman & Caldwell, 1993; Sohl, 1967; Purcell et al., in press) (Figure 6).
Basinal differences like these have been observed in oceans today as well, resulting in functional turnover and distinct FE assemblages in different regional basins (Schumm et al. 2019). Low salinity conditions in present-day oceans have also been associated with poor development and survival of photosynthetic corals and bivalves (Soo and Todd 2014; Aguilar et al. 2019); the non-normal salinity and stratification of the WIS (Cochran et al. 2003a; Fricke et al. 2010; Petersen et al. 2016) may have thus inhibited the survival of photosymbionts and other reef-associated taxa (Kauffman 1984; Kauffman and Caldwell 1993).

Only the Late Maastrichtian networks show strong basin provincialism (Figure 14 and S29), suggesting higher spatial resolution comparisons only detect geographic FE differences when sea levels fell dramatically in the Maastrichtian. This provincialism is driven by taxonomic differences, particularly among gastropods, between the two regions, consistent with observations by Sohl (1967) among others (Figure 8). Furthermore, most FEs extirpated from the Late Maastrichtian WIS were found there in previous intervals, especially gastropod-type FEs (Sohl 1967) (Appendix B, Figure S5). Of the 11 possible gastropod-type FEs, all are present in at least two earlier substages of the WIS, but only six are present in the Late Maastrichtian WIS, where only one of which is herbivorous (Figure 8; Appendix B, Table S32 and S33). Only gastropods show clear network provincialism when molluscan clades were analyzed independently (Appendix B, Figure S37-S38). Even when considering all taxa together, WIS richness values declined (-13% αF) and SME increased (+23% SME), between the Late Campanian and Early Maastrichtian, and from the Early to Late Maastrichtian following the T9 transgression (-30% αF, +54% SME; Figure 6 and S4). These patterns are consistent with the timing of sea level fall during the Late Campanian that
exacerbated water mass differences and promoted taxonomic provincialism either through habitat change or restrictions to taxon dispersal (Figure 6; Purcell et al., 2023). Increasingly non-normal marine conditions likely stressed WIS taxa and promoted their extirpation (Cochran et al. 2003b; He et al. 2005; Fricke et al. 2010; Petersen et al. 2016). Lack of functional redundancy plausibly made these FEs susceptible to loss while more dominant FEs experienced only generic abundance decline or even increase (Figure 8 & S5). It should be noted, however, that network-based provincialism may also be influenced by sampling biases in gastropods in particular due to their lesser preservation potential and few diagnostic characters (Smith et al., 2001; Smith, 2001; Dean et al., 2019)

Smaller scale differences between the specific FE assemblages present within the study region are most likely also due to habitat differences. Several studies have found that distinct habitats have not only unique levels of functional diversity, but also different collections of specific FEs within different assemblages (de Arruda Almeida et al. 2018; Sulemana et al. 2022; Walsh et al. 2022). In all intervals except the Early Campanian, maximum FE assemblage similarity (i.e., the similarity between the specific FEs present) weakly decreases across distance (Appendix B, Figure S31 and S32). WIS data analyzed independently follows a similar pattern in all substages. That the maximum similarity of FE assemblages at both unit size 4 and 5 resolutions decreases as distances increase, even within the WIS, suggests some spatial control on ecological relationships. Similar assemblages of FEs are more likely to occur close to one another, which suggests that habitat distributions influence functional community structure (Schumm et al. 2019), even within a region with a weak latitudinal temperature change. Future studies that assess the influence of abiotic variables on these patterns will potentially help determine their influence on this dataset, but
are beyond the scope of this paper. These patterns are furthermore not observed at spatial resolutions lower than 360km units (i.e., in latitudinal bins or regionally), thus, functional diversity does not appear stable across space at all scales, partially contradicting hypothesis 1. This may in part be due to the way network models are created using a threshold method. Network connections are always present in nodes with shared FEs, so only a complete lack of similarity will cause the network to break down at high thresholds. As nodes become smaller, they contain fewer FEs, and therefore are either well connected (i.e., potentially containing the same common FEs) or have no similarity (i.e., containing rare, different FEs). This result adds support to the notion that not only does the study of functionally interesting paleo-“communities” have a higher resolution limit, but also a lower resolution limit, which suggests spatial scale sensitivity analysis should be considered in studies like this. Future studies that focus on the influence of spatial unit size and similarity should be conducted to better determine how well networks such as these assess paleo-“communities” and the specific FEs that make them up, but that is beyond the scope of this analysis.

**Packing and Functional Redundancy:**

Spatiotemporal functional stability in these results is well explained based on functional redundancy, supporting hypothesis 3. When generic richness increased across time and space, taxa are preferentially “packed” into existing FEs before adding new FEs, promoting functional redundancy in specific FE categories (Figure 7). For example, between the Middle and Late Campanian, as generic richness increased by ~81% (147 new genera), $\alpha_F$ only increased by ~10% (three new FEs). Across this interval, the most common FE, mobile unattached epifaunal carnivores (MU-E-C in Figure 8), rose in proportional generic diversity from ~8% to ~30% while the majority of other FEs fell in proportional generic diversity.
(Figure 8). High functional redundancy in specific groups diminishes functional evenness, as can be seen throughout our results. Prior studies of functional diversity patterns across space have also found that functional richness increases with area, but evenness decreases, as specific FEs become more dominant (Karadimou et al. 2016). Moreover, this is observed in our data where functional evenness responds inversely to changes in generic richness in almost every comparison across either time or space, indicating preferential packing in specific FEs, such as mobile unattached epifaunal carnivores and facultatively mobile unattached infaunal suspension feeders (MU-E-C and FU-I-SF in Figure 8, respectively); these are generally the most dominant FEs in all substages and experience some of the most significant changes in proportional generic diversity (both positive and negative) (Figure 8).

Differential packing of genera into FEs can be viewed either from the perspective of sampling bias (i.e., poorer sampled units are more even because the most common FEs are likely to be represented by the few genera present, but not likely to have many genera in each FE) or as an ecological phenomenon, wherein the primary ecological niches are filled first, and then those with greater resource availability are packed with more genera as diversity increases. The distinction between these two interpretations is not clear, and further research would be necessary to determine if these patterns represent ecological processes.

Dominance of specific FEs is also consistent through time (Figure 8). Similar trends have been documented within other taxonomic groups, indicating that either some trait characteristics allow for greater subdivision or partitioning, or that some FEs have greater capacity for taxonomic diversity due to greater resource abundance (Halpern and Floeter 2008; Oliveira et al. 2016; Schumm et al. 2019; Pimiento et al. 2020). In this case, mobile epifaunal carnivores, facultatively mobile infaunal suspension feeders, and immobile
attached epifaunal suspension feeders are the most dominant throughout all intervals (MU-E-C, F-I-SF, and IA-E-SF in Figure 8, respectively). These groups may have contributed to greater resource partitioning and/or utilized resources that were more abundant in the WIS/GCP at this time. Among marine groups like bivalves, functional redundancy is most likely caused by greater resource availability rather than subdivision of trait space (Stanley 2008; Schumm et al. 2019). In the case of MU-E-Cs, which are consistently dominant throughout the study interval, taxa would be less limited by spatial constraints since they are mobile and unattached, making prey abundance a more likely limiting factor.

Functional redundancy appears to have buffered most FEs from complete loss even when generic diversity declined. Loss of generic richness by 8.8% between the Late Campanian and Early Maastrichtian is associated with the loss of only one FE: immobile attached epifaunal photosymbionts (IA-E-P in Figure 8). This FE had very low proportional generic richness in the Late Campanian (0.3% of genera) and is spatially restricted to the GCP (Appendix B, Table S4). These results are consistent with patterns observed in mollusks across the Miocene to Pleistocene by Pimiento et al. (2020), wherein functional diversity was sustained despite taxonomic loss by the preferential extinction of functionally redundant hypothesized competitors. Functional evenness increased from the Late Campanian to the Early Maastrichtian (SME proportional increase of ~11%) as the relative generic richness of the most dominant FE (MU-E-C) declined from ~30% to ~26%. No other FE experienced such a large decrease. Thus, whereas functional redundancy in this dominant FE likely buffered it against regional extinction, less dominant FEs were generally more likely to be extirpated. Along these lines, unaffected, less dominant FEs were more likely buffered by reduced resource competition among fewer taxa (Pimiento et al. 2020).
Conclusions

Analysis of Campanian and Maastrichtian Western Interior Seaway (WIS) and Gulf Coastal Plain (GCP) of North American invertebrates indicates stability in functional richness ($\alpha_f$) and instability in functional evenness as additional genera are preferentially packed into a few dominant functional entities (FEs). Additionally, functional redundancy is found to buffer against FE loss across both time and space. Sampling biases these results at finer, and potentially very broad, spatial scales; although comparisons of FE assemblages across space are still viable given that they can capture the influence of habitat variability. The WIS and GCP region lack a present-day-style LDG, potentially caused by a flat latitudinal temperature gradient across these latitudes. Instead, the WIS and GCP regions display distinct FE assemblages and functional provinciality by the Late Maastrichtian which mirrors taxonomic provinciality with a lag time of one substage (Purcell et al. 2023). Provinciality is likely caused by oceanographic differences between the basins, particularly non-normal marine conditions in the WIS which were probably exacerbated by falling sea levels in the Maastrichtian. Non-normal marine conditions stress photosymbiotic taxa, including both corals and bivalves, boring taxa, and FEs occupied by herbivorous gastropods. This is observed herein by a decrease in generic richness or complete loss of these FEs in the WIS. The lag between provinces formed in the Early Maastrichtian for taxa but the Late Maastrichtian for FEs may suggest either a data-based difference between the two analyses, or an ecological pattern wherein functional diversity remained stable for a period of time after taxonomic turnover produced taxonomic provincialism. Functional diversity trends also differ by taxonomic class, discouraging future use of single clades for interpretation of spatiotemporal trends in assemblage (paleo-“community”) structure. That said, overall this
investigation provides evidence for FE paleo-“community” stasis during background intervals not characterized by a global climate upheaval (e.g., mass extinction events). Results also provide additional information regarding the stability of functional diversity and redundancy through changing environmental conditions and at different spatial scales, improving our understanding about long-term and broad-scale functional diversity dynamics for both current and ancient systems.

**Acknowledgements**

CP acknowledges funding support from NSF #2021744; CEM acknowledges support from NSF #1924807, and #1601878. CP would also like to thank the many researchers who have worked to identify and describe the functional characteristics of marine invertebrates, without whom this work would not be possible.
CHAPTER 4: TESTING FUNDAMENTAL NICHE CHARACTERISTICS USING HIGH-RESOLUTION LATE CRETACEOUS WESTERN INTERIOR MARINE INVERTEBRATES

Abstract

Ecological niche models (ENMs) are a common tool used to analyze various aspects of ecology, evolution, biotic distribution, and community structure by correlating environmental variables with species occurrence distributions based on the concept of a realized niche. Similarly, paleoENM analysis is applied to address these same questions across long intervals of Earth’s history and species’ durations. This enables the exploration of fundamental assumptions about the niche itself (e.g., whether it is stable over species’ durations and/or conserved across evolutionary lineages). However, very little analysis has been performed testing the fidelity of these methods under the unique biases in paleontological data. In particular, the influence of temporal aggregation and assumed ecological similitude between individual species and their genera have been poorly explored. Using a high-resolution, geographically widespread, deep time dataset, this paper analyzes the impacts of temporal resolution on paleoENM-based niche prediction, and performs basic tests of phylogenetic niche conservation between sister species of marine bivalves to draw conclusions about appropriate application of paleoENM analysis. The results of these tests indicate that great attention should be paid to environmental data sources, interpolation processes, and ecological interpretation when using fossil data, and that substage-level aggregation is unlikely to produce meaningful results within intervals characterized by highly variable environments. Furthermore, we find that individual species-level niche predictions do not match well that of the generic-level estimation. These inherent biases in fossil data likely give rise to researchers utilizing generalist taxa rather than specialists, resulting in
apparent ecological homogenization. However, our results also indicate that paleoENM is useful for assessing general aspects of ecological shift or differentiation and for analyzing niche abiotic habitat characteristics of functional entities, particularly when primary ecological characteristics and environmental factors are appropriately scaled.

**Introduction**

The ecological niche is a fundamental species’ characteristic that is frequently used to address macroecological and evolutionary questions (Brown 1995; Peterson et al. 2011). Although an ecological “niche” can be defined in many ways, here we use the Hutchinson niche: an n-dimensional hypervolume composed of environmental factors that allow a population to survive and reproduce in a given area (Hutchinson 1957; Peterson 2001, 2011; Soberón and Peterson 2005). Hutchinson divided niches into two forms: (1) fundamental niche (FN) – encompassing all abiotic factors necessary for survival and reproduction, and (2) realized niche (RN) – a subset of the FN wherein the FN is restricted by biotic interactions and dispersal capacity (Soberón and Peterson 2005; Peterson et al. 2011). Many modern studies have attempted to quantify the FN/RN to elucidate ecological and evolutionary concepts, including species’ life histories, biotic interactions, geographic range patterns, invasion potential, biotic response to environmental shifts, and extinction vulnerability (i.e., Adhikari et al., 2015; Benito Garzón et al., 2011; Liu et al., 2022; Lockwood et al., 2013; Peterson, 2003; Planas et al., 2014; Purcell & Stigall, 2021; Valencia-Rodríguez et al., 2021; Varela et al., 2010; etc.). Given the wide variety of species, environments, and hypotheses being tested, each dataset is unique, making the identification of universal or generalizable “rules” of FN/RN interactions with evolutionary or ecological processes difficult to define or quantify comparatively.
The application of these techniques to the deep time fossil record adds additional dimensions, first and foremost the element of time, to using niche estimation to test macroevolutionary and macroecological hypotheses unique to paleontological data. Modern datasets usually involve higher spatial resolution, denser species occurrence data, can incorporate genome-based phylogenetic information, and use well-established climatic variables that can produce detailed estimation of niche dynamics (Peterson et al. 2011). In contrast, ecological niche analyses using deep time datasets have the advantage of using fossil and sedimentological data to compare long-term patterns of FN/RN not achievable in modern datasets (Maguire and Stigall 2009; Dudei and Stigall 2010; Malizia and Stigall 2011; Brame and Stigall 2014; Saupe et al. 2014, 2015, 2019; Stigall 2014, Myers et al. 2015b; Purcell and Stigall 2021). Given that species are known to survive on million-year (Myrs) timescales, the extension of these methods to deep time is imperative to truly understand how niche dynamics relate to evolutionary and ecological trajectories of species and clades.

However, paleo-niche analyses must also consider the loss of spatiotemporal detail inherent in the geologic record (reviewed in Myers et al., 2015). Primarily, the process of sedimentation homogenizes both temporal and, to a lesser degree, spatial information. At best, geologic units may record $10^4$ years of environmental information per centimeter within a single defined sedimentary layer (Schindel, 1980). Fine-scale analysis of ecology and environmental characteristics at less than this resolution is therefore rarely viable. In practice, most niche estimation in deep time relies on temporal aggregation at the geological substage to stage level (i.e., generally 100s-1000s kyrs). Moreover, temporal aggregation increases as the geographic scope of analyses gets larger – for example, when using regional or global
data, which is particularly important for “big data” paleobiological analyses so common today (e.g., Maguire and Stigall 2009; Saupe et al. 2019; Purcell and Stigall 2021). Therefore, while fossil analyses allow for long-term tests of niche dynamics and their relationship to macroevolutionary and macroecological process that are not possible using modern species, by necessity they assume either localized environmental stability across long time intervals or that long-term averages of environmental conditions are consistent with a species’ preferred habitat.

Paleontological data is also biased both by the fossilization process itself which preserves hard-bodied taxa significantly more often than soft-bodied taxa (Kidwell and Flessa 1996), by the incomplete exposure of strata, and the erosion of exposed material (Antell et al. 2024). Spatial uncertainty, present in all datasets, requires that species’ occurrences be spatially thinned to a single occurrence per grid cell (already likely larger than in modern datasets). Given that fossil species’ occurrences, rarely extensive or dense to begin with, are often spatially clumped, spatial thinning typically leads to substantial reduction in occurrence observations (Shcheglovitova and Anderson 2013; Aiello-Lammens et al. 2015). For example, Purcell and Stigall (2021) used a spatial error of 30kms, based on the spatial error margin for the fossil taxa analyzed and linked to 30km geographic grid cell size used. This results in very few fossil species with sufficient distributions to perform robust paleo-niche analysis using popular methods, such as the Maxent algorithm (Phillips et al. 2006), often regarded as greater than five occurrences (Hernandez et al. 2006). To obtain sufficient occurrence numbers after spatial thinning, most paleo-niche analyses are restricted to genus-level comparisons, particularly in deep time (e.g., Brame & Stigall, 2014; Dudei & Stigall, 2010b; Hopkins et al., 2014; Maguire & Stigall, 2009; Malizia & Stigall, 2011b;
Myers et al., 2013; Nurnberg & Aberhan, 2013; Purcell & Stigall, 2021). How well a genus represents species-level ecological traits, however, is still debated (Hendricks et al. 2014).

Unfortunately, very few datasets allow for the high-resolution, densely fossiliferous, and well-documented comparisons necessary to test these issues directly. How well temporally-aggregated geological and paleontological information represents real-world biological phenomenon has yet to be well established. Even assuming that temporal resolution is not a factor in accurate niche estimation (e.g., if FNs are stable over species’ durations), niche characteristics may not be conserved between related species and at the genus-level, potentially resulting in less robust paleo-niche analyses. Given the advantages of using fossil data to assess overall community structure (Fara and Benton 2000; Finnegan and Droser 2008), the relationship between environmental change and ecological response (Purcell and Stigall 2021), and various other macroecological questions (Myers et al. 2015), addressing these issues is vital to the production of accurate scientific study.

Using a high-resolution sedimentary record with regionally correlated Western Interior Seaway (WIS) marine units and a well-vetted paleontological database representing approximately 17 million years of marine life, we tested fundamental aspects of niche estimation in deep time. By comparing niche overlap in a simplified environmental space, these data were used to address three main questions: 1) how temporal resolution influences niche estimation, 2) how well fossil data of phylogenetically related taxa (sister species) present phylogenetic niche conservation, and 3) how well the genus-level niche estimation conserves species-level niche characteristics. The results of these analyses inform the utility of paleo-niche estimation in deep time to test macroevolutionary and macroecological hypotheses. We provide best practices for more accurate paleo-niche estimation and
ultimately support the application of ecological niche modeling in the fossil record (paleoENM).

**Methods:**

Our study used marine invertebrate fossil occurrence data from the North American WIS during the Campanian and Maastrichtian stages of the Late Cretaceous. Occurrence data was compiled from online records downloaded from the Paleobiology Database (08/25/21 download) and iDigBio (08/30/21 download), and records from museum collections at the Black Hills Institute, the USGS-Washington Cobban Collection, and a thesis database (Mackenzie 2007). All data were vetted to remove incomplete, erroneous, or poorly spatiotemporally resolved occurrences and to update taxonomic nomenclature (see Purcell et al. 2023 for additional details and references used in dataset vetting). Stratigraphic ages were vetted to the biozone level based on stratigraphic and biozone information in the database itself or from literature sources (Cobban et al. 2006; Merewether and McKinney 2015), resulting in twelve temporal bins (Figure 15, Table 9). To remove spatial ambiguity, taxonomic occurrences with spatial resolution less than 30 km of uncertainty were removed. To include only WIS data, records east of -80° longitude and south of 31.6° latitude were removed (see Purcell et al., 2023).

The final WIS dataset of 21,156 fossil occurrences includes a total of 279 genera of bivalves, cephalopods, gastropods, echinoderms, corals, brachiopods, bryozoans, and crustaceans (Figure 2). Tests confirming signal over noise in this dataset were previously published in Purcell et al. (2023). In the cleaned database, potential taxa for use in niche estimation were identified by greater than six fossil occurrences within at least one biozone interval. More than six fossil occurrences identified at the species-level was selected as an
Figure 15. Biostratigraphic chart showing collated biozone intervals; shading delineates each interval used in analyses (modified from Purcell et al. 2023).
arbitrarily threshold because it is slightly greater than the minimum value recommended for creating ecological niche models (ENMs) using the popular Maxent algorithm (Hernandez et al. 2006), while still providing a relatively relaxed threshold for sparse fossil data. Selected taxa were spatially thinned to the 30 km geographic grid cell size determined by allowed spatial uncertainty in occurrences. After spatial thinning, only the *Inoceramus* genus had sufficient species-level occurrences for phylogenetic niche conservation tests; in this case, seven *Inoceramus* species had six or more occurrences in at least one time interval. These were also evaluated with three “comparison taxa” species from the *Ctena, Lucina*, and *Chlamys* genera that had sufficient occurrences in most of the same biozone intervals (Table 9). All ten bivalve species were used for analysis of temporal sensitivity.
Table 9. Summary of the number of species occurrences in each biozone interval. Note that some biozone intervals were not used for analysis, due to a lack of adequate species occurrence information.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INGROUP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus barabini</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus convexus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus sagensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus sublaevis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus oblongus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus saskatchewanensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Inoceramus azerbaydjanensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>COMPARISON TAXA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamys nebrascensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ctena imbricatula</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lucina subundata</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to estimate environmental niches, 11 spatially continuous environmental layers were reconstructed based on sedimentary proxies, distance from shore, and distance from hydrothermal seep deposits (following Myers et al 2015). Stratigraphic data used to characterize environmental proxy variables were collected across WIS localities from literature sources, including theses and dissertations, published journal articles, and field guides (see SI for references). The sedimentary environmental proxy variables included: percent grain sizes in siliciclastic rocks (ranging from mud to coarse sand), percent...
limestone, stratigraphic bedding style, bedding thickness, and degree of bioturbation (Table S28). Environmental layer values were averaged at individual localities based on three levels of temporal resolution: biozone, substage—Middle and Late Campanian, and stage—Maastrichtian only. The Point Statistics tool and the Inverse Distance Weighting method in ArcGIS Pro was used to interpolate values at locations without sedimentary data and create a spatially continuous, smoothed set of raster layers with 30 km resolution. Raster layers for the distance-from-shore variable were created using paleo-shorelines published in previous works (Gill and Cobban 1973; Cobban et al. 1994; Roberts and Kirschbaum 1995; Slattery et al. 2013) and updated based on fossil localities and stratigraphic information from the current database. The distance-from-hydrothermal-seeps variable was based on the locations of individual or clusters of seeps published in various literature sources (see SI for details on sedimentary data collection, aggregation, and interpolation). Environmental layers were then intersected with outcrop masks to restrict continuous variables only to areas of potential fossil occurrence, thus reducing model overfitting. These were created using relevant Late Cretaceous outcrop polygons buffered an additional 30 km to allow for spatial error in the dataset (see SI for references). Outcrop masks were clipped to strandline boundaries for the appropriate interval and then used to clip raster data for each variable (Figure 16). This resulted in environmental layer characterization for the twelve biozone-level intervals, five substages (early, middle, and late Campanian, and the early and late Maastrichtian), and the Maastrichtian stage. Based on the taxa selected for analysis, only seven biozone intervals, the early and late Campanian, the early Maastrichtian, and the Maastrichtian were utilized in the following analyses.
Figure 16. Example interpolated maps within the twelve biozone intervals, representing percent mud. Cool colors indicate low percentage and warmer colors indicate high percentage of mud. Shorelines are denoted with black. Note that limitations in sedimentary data collected resulted in unmapped portions of the basin in some intervals.
Niche overlap was assessed using comparisons of observed occurrences within a simplified two-dimensional environmental principal components space (PCS) using the ecospat package in R (Broennimann et al. 2012; Di Cola et al. 2017). Niche overlap is a metric describing the degree of similarity between two niches in environmental space (e-space). When two temporal bins were compared, only shared, non-correlated environmental proxy variables were used (Appendix C, Figure S2 and S3, Table S7). Niche overlap is quantified using two metrics of similarity, a version of the Hellinger distance, known as the I statistic, and Schoener’s (1968) D statistic (Warren et al., 2008). These metrics range from 0 to 1 (indicating zero to complete overlap, respectively) and were calculated by comparing smoothed kernel density functions representing taxonomic occupancy for taxon in each PCS grid cell (Broennimann et al. 2012; Di Cola et al. 2017). Both metrics produce similar results, though Schoener’s D has a greater ecological implication than the I statistic (see Warren et al., 2008 for discussion). In this analysis both metrics were calculated for comparison, but given the similarity of their results, only the D statistic is discussed in detail.

Three other metrics of niche change were also calculated: niche stability, expansion, and unfilling (Di Cola et al. 2017). Niche stability is the amount of niche space within the PCS which is maintained or conserved between the two taxa being compared (or between the same taxon in two intervals). Niche unfilling and expansion are used to determine how a niche has changed either by taxa moving away from previously occupied e-space or expanding into new e-space, respectively. Both metrics are interpreted here as instability, given that these analyses do not assess alteration of the niche across time or space. Calculations of niche overlap, stability, unfilling, and expansion were performed using both the niche equivalency and niche similarity tests in ecospat to determine if values were greater
or less than would be expected at random (Di Cola et al. 2017). The niche equivalency test is used to determine if two distributions are more or less identical than the overlap of a random set of simulated distributions created using pooled data. The niche similarity test determines if the distributions are more or less similar than the overlap of a randomized subset of the data within analogous e-space (Di Cola et al. 2017). Niche equivalency is a very strict test of niche overlap, whereas niche similarity is a more lenient test and, in general, niche equivalency is rarely observed in nature (Warren et al. 2010; Aguirre-Gutiérrez et al. 2015).

Data sensitivity to temporal resolution was analyzed by comparing niche overlap and stability between the highest resolution temporal bins (biozones) with lower resolution bins (substage and stage) for all ten bivalve species, resulting in a total of twenty-one comparisons. It was hypothesized that niche overlap and stability would fail to be statistically high between different temporal bin resolutions. Phylogenetic niche conservation (PNC) was analyzed by comparing niche overlap and stability between the species of *Inoceramus* bivalves (*I. barabini, I. convexus, I. sagensis, I. sublaevis, I. saskatchewanensis, I. oblongus, and I. azerbaydjanensis*). These analyses were contrasted with comparisons between *Inoceramus* species and three non-*Inoceramus* bivalve species (*Ctena imbricatula, Lucina subundata, and Chlamys nebrascensis*); in total, twenty-eight PNC comparisons were tested. It was hypothesized that niche overlap and stability would be statistically higher, and higher on average, between *Inoceramus* species than between *Inoceramus* species and non-*Inoceramus* taxa. Finally, niche overlap and stability between *Inoceramus* species and the genus-level *Inoceramus* niche estimation were analyzed to assess how well the genus represents the environmental niche space of its component species, resulting in nine comparisons. Partitioning of the genus-level niche by individual species was visualized in R
using convex hulls in a two-dimensional PCS of shared, non-correlated environmental variables within a single biozone.

**Results**

*Temporal bin Size Comparison*

Table 10 provides niche overlap results for taxon comparisons at variable temporal resolution. Of the twenty-one temporal comparisons, three (~14.3%) had statistically more equivalent niche overlap, and only 7 (~33%) had statistically high overlap similarity (see bolded D-values and stability scores in Table 10, S9, and S11). Niche stability overall in these comparisons averaged 72% (Table 11 and S8). Only five of the 21 comparisons (~24%) had statistically higher niche stability values (Table 10). Unfilling and expansion were statistically lower than expected at random four and five times, respectively, in niche similarity tests (Appendix C, Table S11). A weak positive correlation between the interval duration ratio (biozone to substage and stage) and niche overlap furthermore indicates that as temporal aggregation increases niche overlap values decrease, as predicted (Figure 17). Statistically low niche values were calculated but given that their biological significance cannot be well constrained, they are not included in these results (see SI for details).

Non-analogous environmental conditions were common between temporal bins of different resolution (Appendix C, Figure S4). Furthermore, niche distributions were disjointed within PCS at the substage level for two of the eleven taxa-intervals assessed (Appendix C, Figure S4). Only one of the biozone-level niches had a disjointed distribution. These results suggest that non-analogous conditions are introduced by aggregating variables at lower temporal resolutions and the estimation of niche space occupancy for taxa within these environments is artifactually modified through this process. Furthermore, results of
overlap comparisons are somewhat unique to the species and intervals compared, suggesting that lower temporal resolution in these analyses do not produce consistent results across time interval sizes.

**Table 10.** Results (p-value, alpha=0.5) of greater niche similarity tests for temporal resolution comparisons. Bolded values indicate statistical significance at the p-value < 0.05 level.

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. nebrascensis</em></td>
<td><em>B. baculus</em>- <em>B. grandis</em> to lower Maastrichtian</td>
<td>0.001*</td>
<td>0.005*</td>
</tr>
<tr>
<td><em>C. nebrascensis</em></td>
<td><em>B. baculus</em>- <em>B. grandis</em> to Maastrichtian</td>
<td>0.105</td>
<td>0.118</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. compressus</em>- <em>B. cuneatus</em> to upper Campanian</td>
<td>0.133</td>
<td>0.166</td>
</tr>
<tr>
<td><em>I. azerbaydjanensis</em></td>
<td><em>B. maclearni</em>- <em>B. sp. (smooth)</em> to middle Campanian</td>
<td>0.165</td>
<td>0.161</td>
</tr>
<tr>
<td><em>C. imbricatula</em></td>
<td><em>B. perplexus</em>- <em>B. gregoryensis</em> to middle Campanian</td>
<td>0.002*</td>
<td>0.191</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus</em>- <em>B. scotti</em> to middle Campanian</td>
<td>0.243</td>
<td>0.171</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus</em>- <em>B. scotti</em> to middle Campanian</td>
<td>0.105</td>
<td>0.003*</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus</em>- <em>B. scotti</em> to middle Campanian</td>
<td>0.028*</td>
<td>0.024*</td>
</tr>
<tr>
<td><em>C. imbricatula</em></td>
<td><em>B. reduncus</em>- <em>B. scotti</em> to upper Campanian</td>
<td>0.102</td>
<td>0.239</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.034</td>
<td>0.080</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.113</td>
<td>0.205</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.014*</td>
<td>0.191</td>
</tr>
<tr>
<td><em>C. imbricatula</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.085</td>
<td>0.168</td>
</tr>
<tr>
<td><em>L. subundata</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.337</td>
<td>0.139</td>
</tr>
<tr>
<td><em>C. nebrascensis</em></td>
<td><em>B. reesidei</em>- <em>B. eliasi</em> to upper Campanian</td>
<td>0.040*</td>
<td>0.021*</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.001*</td>
<td>0.022*</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.059</td>
<td>0.133</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.003*</td>
<td>0.126</td>
</tr>
<tr>
<td><em>C. imbricatula</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.093</td>
<td>0.153</td>
</tr>
<tr>
<td><em>L. subundata</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.076</td>
<td>0.133</td>
</tr>
<tr>
<td><em>C. nebrascensis</em></td>
<td><em>D. nebrascense</em>- <em>E. jenneyi</em> to upper Campanian</td>
<td>0.003*</td>
<td>0.016*</td>
</tr>
</tbody>
</table>
Figure 17. Correlation plot of duration ratio to mean niche overlap (D value).
Table 11. Summary of niche overlap and stability results. Sbst. D = duration of the substage/stage; b. D. = duration of biozone interval; D Ratio = ratio of the biozone duration; N = number of comparisons.

<table>
<thead>
<tr>
<th>Intervals Compared</th>
<th>sbst. D</th>
<th>b. D.</th>
<th>D Ratio</th>
<th>N</th>
<th>D-value Mean</th>
<th>Median</th>
<th>Stability Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Bin Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maastrichtian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>6.1</td>
<td>1.8</td>
<td>0.3</td>
<td>1</td>
<td>0.38</td>
<td>0.38</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Lower Maastrichtian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. baculus-B. grandis to lower Maastrichtian</td>
<td>2</td>
<td>1.8</td>
<td>0.9</td>
<td>1</td>
<td>0.78</td>
<td>0.78</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Upper Campanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reesidei-B. eliasi to upper Campanian</td>
<td>4.1</td>
<td>1.4</td>
<td>0.3</td>
<td>6</td>
<td>0.41</td>
<td>0.45</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td>B. compressus-B. cuneatus to upper Campanian</td>
<td>4.1</td>
<td>0.7</td>
<td>0.2</td>
<td>1</td>
<td>0.43</td>
<td>0.43</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>D. nebrascense-E. jenneyi to upper Campanian</td>
<td>4.1</td>
<td>1.7</td>
<td>0.4</td>
<td>6</td>
<td>0.45</td>
<td>0.42</td>
<td>0.73</td>
<td>0.77</td>
</tr>
<tr>
<td>B. reduncus-B. scotti to upper Campanian</td>
<td>4.1</td>
<td>1.3</td>
<td>0.3</td>
<td>4</td>
<td>0.38</td>
<td>0.37</td>
<td>0.82</td>
<td>0.95</td>
</tr>
<tr>
<td>Middle Campanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. perplexus-B. gregoryensis to middle Campanian</td>
<td>4.3</td>
<td>1.4</td>
<td>0.3</td>
<td>1</td>
<td>0.60</td>
<td>0.60</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>B. maclearni-B. sp. (smooth) to middle Campanian</td>
<td>4.3</td>
<td>1.4</td>
<td>0.3</td>
<td>1</td>
<td>0.09</td>
<td>0.09</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>TOTAL</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
<td>0.43</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>Phylogenetic Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reesidei-B. eliasi</td>
<td>-</td>
<td>1.4</td>
<td>-</td>
<td>3</td>
<td>0.67</td>
<td>0.68</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>D. nebrascense-E. jenneyi</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>3</td>
<td>0.33</td>
<td>0.10</td>
<td>0.52</td>
<td>0.41</td>
</tr>
<tr>
<td>B. reduncus-B. scotti</td>
<td>-</td>
<td>1.3</td>
<td>-</td>
<td>3</td>
<td>0.58</td>
<td>0.52</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>0.54</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>Comparison taxa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reesidei-B. eliasi</td>
<td>-</td>
<td>1.4</td>
<td>-</td>
<td>6</td>
<td>0.65</td>
<td>0.63</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>B. compressus-B. cuneatus</td>
<td>-</td>
<td>0.7</td>
<td>-</td>
<td>1</td>
<td>0.29</td>
<td>0.29</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>D. nebrascense-E. jenneyi</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>9</td>
<td>0.48</td>
<td>0.57</td>
<td>0.76</td>
<td>0.83</td>
</tr>
<tr>
<td>B. reduncus-B. scotti</td>
<td>-</td>
<td>1.3</td>
<td>-</td>
<td>3</td>
<td>0.36</td>
<td>0.38</td>
<td>0.50</td>
<td>0.53</td>
</tr>
<tr>
<td>TOTAL</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>0.57</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>Niche Partitioning Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inoceramus Species to Genus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reesidei-B. eliasi</td>
<td>-</td>
<td>1.4</td>
<td>-</td>
<td>3</td>
<td>0.45</td>
<td>0.46</td>
<td>0.63</td>
<td>0.67</td>
</tr>
<tr>
<td>D. nebrascense-E. jenneyi</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>3</td>
<td>0.39</td>
<td>0.55</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>B. reduncus-B. scotti</td>
<td>-</td>
<td>1.3</td>
<td>-</td>
<td>3</td>
<td>0.71</td>
<td>0.72</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>0.48</td>
<td>0.66</td>
<td>0.67</td>
</tr>
</tbody>
</table>
**Phylogenetic Niche Conservation Comparisons**

Only *Inoceramus oblongus* vs. *I. sagensis* comparison in the *B. reesidei-B. eliasi* interval had statistically greater niche equivalence than would be expected at random (Appendix C, Table S14). Four of the nine *Inoceramus* species (~44%) showed greater niche similarity than would be expected at random (Table 12 and Appendix C, Table S16). Only one species comparison (*I. barabini* vs. *I. sagensis* in the *B. reesidei-B. eliasi* biozone interval) had greater niche stability than would be expected at random using the niche similarity test. Overlap plots support these results (Figure 18 and S5).

Comparisons of the *Inoceramus* species with the three non-*Inoceramus* species indicate that statistically high overlap and stability is only slightly more common between related species than unrelated ones based on the niche similarity test (Table 12). When using the niche equivalency test, seven of the nineteen non-*Inoceramus* to *Inoceramus* species comparisons (~26%) had greater niche equivalence than would be expected at random and three of the nineteen comparisons (~16%) had greater stability and lower niche expansion than would be expected at random (Appendix C, Table S19). Two non-*Inoceramus* comparisons showed statistically greater unfilling based on the niche equivalency test (Appendix C, Table S20). Seven of the nineteen comparisons (~37%) had statistically high D-values, high stability, and low expansion, and four had statistically low unfilling when using niche similarity tests (Appendix C, Table S21). Therefore, the proportion of species with statistically similar niches is slightly higher between related taxa (Appendix C, Table S16), but statistical equivalence was more common between unrelated taxa. Mean niche stability for the non-*Inoceramus* to *Inoceramus* species comparisons (0.76) was higher than within *Inoceramus* species comparisons (0.62; Appendix C, Table 11). Mean expansion was
lower between within *Inoceramus* comparisons relative to non-*Inoceramus* to *Inoceramus* comparisons (0.24 versus 0.38, respectively), and mean unfilling was approximately the same (0.32 and 0.34 respectively; Tables S18 and S13). Plots of niche comparisons for non-*Inoceramus* taxa support these results (Figures 19 and S6). Surprisingly, these analyses suggest that stability and overlap was higher on average between evolutionarily more distant taxa vs. between sister taxa. Statistical assessments furthermore do not indicate that niche equivalency was higher between related taxa than unrelated taxa, and proportions of statistically high overlap and stability based on niche similarity were not distinct.

**Figure 18.** Example overlap plots of *Inoceramus* species against each other and against *Ctena imbricatula* during the *B. reduncus-B. scotti* biozone interval. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the second listed species has expanded beyond the niche of the first listed species), and green represents regions of niche unfilling (i.e., where the second listed species niche has moved away from space occupied by the first listed species). Red lines represent the extent of environmental space realized within the interval and the black arrow indicates the shift of the niche centroid between the two species. See Figures S5 and S6 for all species overlap plots.
Table 12. Results (p-value, alpha=0.5) of greater niche similarity tests for phylogenetic comparisons. Bold text and asterisks indicate significant results.

<table>
<thead>
<tr>
<th>Phylogenetic Conservation Comparisons</th>
<th>Species Compared</th>
<th>Interval</th>
<th>D value</th>
<th>Stability (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ingroup Comparisons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. convexus vs. I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.112</td>
<td>0.381</td>
<td></td>
</tr>
<tr>
<td>I. sublaevis vs. I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.221</td>
<td>0.353</td>
<td></td>
</tr>
<tr>
<td>I. convexus vs. I. sublaevis</td>
<td>B. reduncus-B. scotti</td>
<td>0.431</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>I. convexus vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.429</td>
<td>0.209</td>
<td></td>
</tr>
<tr>
<td>I. convexus vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.431</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>I. barabini vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td><strong>0.001</strong></td>
<td>0.251</td>
<td></td>
</tr>
<tr>
<td>I. barabini vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.015</strong></td>
<td><strong>0.026</strong></td>
<td></td>
</tr>
<tr>
<td>I. barabini vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.038</strong></td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td>I. oblongus vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.048</strong></td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td><strong>Comparison taxa Comparisons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>B. reduncus-B. scotti</td>
<td>0.411</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>B. reduncus-B. scotti</td>
<td>0.306</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.112</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td><strong>0.001</strong></td>
<td><strong>0.001</strong></td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.398</td>
<td>0.305</td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.356</td>
<td>0.626</td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td><strong>0.001</strong></td>
<td><strong>0.001</strong></td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.449</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td><strong>0.015</strong></td>
<td><strong>0.045</strong></td>
<td></td>
</tr>
<tr>
<td>L. subundata vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.101</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td>L. subundata vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.501</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>L. subundata vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td><strong>0.028</strong></td>
<td>0.448</td>
<td></td>
</tr>
<tr>
<td>L. subundata vs. I. sagensis</td>
<td>B. compressus-B. cuneatus</td>
<td>0.265</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.006</strong></td>
<td>0.098</td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.067</td>
<td><strong>0.045</strong></td>
<td></td>
</tr>
<tr>
<td>C. nebrascensis vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.101</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.017</strong></td>
<td>0.172</td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.044</strong></td>
<td><strong>0.035</strong></td>
<td></td>
</tr>
<tr>
<td>C. imbricatula vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.076</td>
<td><strong>0.018</strong></td>
<td></td>
</tr>
</tbody>
</table>
Species to Genus Comparisons (Niche Partitioning)

Only *I. barabini* in the *B. reesidei-B. eliasi* interval had a statistically greater high niche overlap when compared with the genus-level niche based on niche similarity tests, as well as greater than random niche stability and lower than random niche expansion (Tables 13 and S26). In the *D. nebrascense-E. jenneyi* interval, *I. barabini* also had lower niche unfilling than would be expected at random, but no other comparison of overlap or stability was statistically high (Appendix C, Table S26). In the *D. nebrascense-E. jenneyi* interval, *I. sublaevis* had higher than expected niche unfilling based on the niche equivalency test (Appendix C, Table S25). No other values were statistically significant based on the niche equivalency test.

Mean niche stability for all species to genus comparisons was approximately equal to the mean stability observed among *Inoceramus* species themselves, 52% (Table 11). Unfilling was notably very low for species to genus comparisons, ranging from 0% to 24% and averaging only 5% (Appendix C, Table S23). Niche expansion was more notable, ranging from 9% to 68% and averaging 34% (Appendix C, Table S23). Niche expansion is expected to be higher and niche unfilling extremely low for species to genus comparisons since the genus niche (always analyzed as the “secondary” niche) is expected to be larger. Our results are consistent with these expectations and suggest that niche stability and overlap were not high overall between species and the genus. In cases where the species niche did not overlap with the genus niche (unfilling), it is plausible that spatial thinning removed relevant occurrence points of specific species from the genus, resulting in a broader species’ ecological extent from the genus-level data. This is observed for *I. sublaevis* in the *D.*
**Figure 19.** Convex hull comparisons of the *Inoceramus* genus and *Inoceramus* species niches in PCS for three biozone intervals. Polygons represent occupied environmental spaces for each taxa within PC environmental space. Dark grey points represent environmental data observed in the interval. Colors indicate *Inoceramus* species (green = *I. barabini*, light blue = *I. convexus*, pink = *I. sagensis*, dark blue = *I. oblongus*, orange = *I. sublaevis*, yellow = *I. saskatchewanensis*); grey = *Inoceramus* genus estimation.
nebrascense-E. jenneyi comparison with Inoceramus, which had statistically high unfilling based on niche similarity test results, caused by a gap in the Inoceramus niche distribution, presumably from spatial thinning (Appendix C, Figure S7). Principal component comparisons of multiple Inoceramus species to the genus niche using convex hulls visualizes the structure of species niches within the generic niche (Figure 19). These plots show species’ niche partitioning to some degree within the genus-level niche, but also that overlap is common between species.

Table 13. Results (p-value, alpha=0.5) of greater niche similarity tests for phylogenetic comparisons. Bold text and asterisks indicate significant results.

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>Stability (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti</td>
<td>0.114</td>
<td>0.225</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti</td>
<td>0.165</td>
<td>0.144</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.060</td>
<td>0.145</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenney</td>
<td>0.292</td>
<td>0.202</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenney</td>
<td>0.137</td>
<td>0.347</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenney</td>
<td>0.152</td>
<td>0.170</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td>0.033*</td>
<td>0.041*</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.206</td>
<td>0.089</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.212</td>
<td>0.076</td>
</tr>
</tbody>
</table>

Discussion

The Effects of Temporal Aggregation in Paleo-niche Analyses

The temporal resolution used for deep-time analysis is typically constrained by logistical factors such as geologic age constraints, inadequate taxonomic abundances, and issues with regional correlation. The combination of these factors frequently forces researchers to aggregate sedimentary data at the lower resolutions (substage or higher) when analyzing paleo-niche parameters (e.g., Maguire and Stigall 2009; Hopkins et al. 2014; Saupe et al. 2019; Purcell and Stigall 2021). Higher-resolution studies can be achieved but generally
involve investigating less geographically extensive regions, such as a single, well-sampled sedimentary basin with well correlated high-resolution stratigraphic data (e.g., Brame & Stigall, 2014; Dudei & Stigall, 2010; Malizia & Stigall, 2011). Generally, these studies are forced to assume that data homogenization will increase analytical noise, but not modify the overall signal of niche patterns, if they exist (e.g., Purcell & Stigall, 2021). Here we specifically tested for the influence of temporal aggregation and found that niche characteristics are not well conserved at lower temporal resolutions, although temporal aggregation does not typically introduce unique niche e-space occupation.

Across the Late Cretaceous, temporal biozone intervals ranging in duration from ~0.7 Myrs (B. compressus-B. cuneatus) to ~1.8 Myrs (D. nebrascense-E. jenneyi) were compared with longer intervals lasting ~2 Myrs (early Maastrichtian) to ~4.3 Myrs (late Campanian). We were also able to compare the B. baculus-B. grandis biozone interval with the Maastrichtian stage overall, which represents ~6.1 Myrs of deposition. Across these comparisons, differences in temporal duration were found to correlate with decreased niche overlap (Figure 17). Lower temporal resolution of the original data therefore did not well-characterize the niche of species at comparatively higher resolutions. Within these temporal comparisons, species niches had high overlap (mean D-value = 0.43) and stability (mean = 72%) but values were often statistically insignificant either for the niche equivalency or similarity tests (Table 9). Only ~33% of the taxa analyzed had statistically high D-values of overlap and only ~24% had statistically high stability.

Furthermore, map interpolations at different temporal resolutions show notable differences between biozone-level bins within most environmental proxy variables (Appendix C, Figure S2). For example, percent mud is high (greater than 50%) only in the
northeastern portion of the basin and lower (less than 30%) in the rest of the basin during the
D. nebrascense-E. jenneyi interval. During the succeeding D. cheyennense interval, however,
percent mud is relatively high (greater than 30%) in the majority of the southern and easter
portions of the basin (Figure 16). These patterns are in turn distinct from the distribution of
mud within the other two biozones that make up the Late Campanian. When aggregated at
the Late Campanian substage level, the regions with highest percent mud are in the northern
portion of the basin, within two distinct “patches” (Appendix C, Figure S2). Even in higher-
resolution studies using fossil data that are limited to shorter duration aggregations, such as
0.5 Myrs temporal bins, environmental variables are aggregated across intervals potentially
much longer than the duration of the species occupation (i.e., Dudei and Stigall, 2010). Basin
geometry is also highly variable across the study interval, with marine areas shifting rapidly
with changing sea levels (both eustatic and tectonic), particularly across the Late Campanian
and Maastrichtian (Figure 16) (Cobban et al., 1994; Gill & Cobban, 1973; Roberts &
Kirschbaum, 1995; Slattery et al., 2013). This almost certainly leads to a distortion of
environmental averages within regions that shift from marine to non-marine within a single
interval. Changes to shorelines are likely occurring even within biozone intervals themselves,
though these alternations are too high-resolution to detect in this dataset (and perhaps any).
Thus, aggregating data at a higher geochronological level necessarily homogenizes and
distorts otherwise localized or regional characteristics. This results not only in a loss of
spatiotemporal detail, but also potentially erroneous spatial correlations between shorter-
lived taxa and long-term environmental averages.

While changes in shoreline are not ubiquitous to all regions or intervals of Earth
history, periods of environmental change are common targets for deep-time paleo-niche
analysis. Consequently, researchers must pay close attention to the influence that temporal aggregation has on the interpretation of their results, and not assume homogenization will represent niche characteristics accurately. In all paleoecological analyses, the highest temporal resolution available should be used whenever possible to avoid over-generalizing variables, and changes to broader paleoceanography and data distributions should be carefully considered before attempting or interpreting these kinds of analyses. Paleo-niche analysis can provide useful information provided the resolution of the data matches the question being asked, however. For example, analyses that are conducted to assess how species shift their niche space through time in response to changes in temperature can constrain their area of analyses to consistently marine environments that lack dramatically shifting shoreline.

**Implications for Testing Phylogenetic Niche Conservation in Deep Time**

Phylogenetic niche conservation is a debated biotic pattern relevant to both ecological and evolutionary trends (Crisp and Cook 2012; Pyron et al. 2015). Related taxa are generally considered to have more ecologically similar environmental requirements than unrelated taxa, given the fact that they share more recent ancestral traits (Hadly et al. 2009; Nürnberg and Aberhan 2013). Indeed, PNC has been observed in both allopatric species (Peterson et al. 1999) and in sympatric species (Lovette and Hochachka 2006), indicating that both physical separation of populations and genetic divergence can result in conservation of fundamental ecological traits. If PNC is widespread over macroevolutionary timescales, then it becomes a primary justification supporting the application of genus-level niche traits in paleo-niche modeling and other deep time applications (Dudei and Stigall 2010; Malizia and Stigall 2011; Nürnberg and Aberhan 2013; Brame and Stigall 2014; Hendricks et al. 2014; Purcell
and Stigall 2021). Our secondary hypothesis predicted niche conservatism between sister species for this reason. However, niche comparisons between unrelated taxa produced higher average stability than between related taxa (76% and 62%, respectively) and very similar overlap values (50% and 52%, respectively). The proportion of comparisons with statistically greater niche stability was lower for within-genus species comparisons than species comparisons across different genera (~11% and 37%, respectively), and statistically high overlap was only slightly more common for within-genus versus between-genus species comparisons (~44% and 37%, respectively). These results reject our hypothesis of greater niche conservation between related versus among less closely related taxa and suggest that either PNC is not present within the taxa considered, or substantial biases prevent PNC from being observed. A lack of PNC may be, and was here, observed as within-genus species niche partitioning, which is generally regarded as important to promote speciation (Cavender-Bares et al. 2004; Graham et al. 2004); it is also likely that evidence for PNC is taxon- and/or spatiotemporal scale-specific. Differentiating niche characteristics may furthermore simply not be observable at a given level of analysis or method of niche estimation (Lovette and Hochachka 2006), caused by environmental variable biases and/or the fundamental characteristics in the taxa selected for analyses.

Specifically, low spatial and temporal resolution, including potentially biozone-level resolutions, and the broad-scale distribution of the taxa analyzed likely culminates in an over-generalization of environmental characteristics and a bias towards preservation of generalist taxa. Deep-time environmental variables are especially prone to over-generalization in these kinds of analyses, since they are by necessity time-averaged, autocorrelated proxies (Appendix C, Figure S3). For example, oxygen levels are critical to habitat restrictions of
marine biota (Deutsch et al. 2015), but oxygen levels can only be approximated in sedimentary units by a combination of variables such as sedimentary structures (indicating energy level and water depth) and bioturbation, which is broadly associated with oxygen levels sufficient to allow for activity below the sediment-water interface (Leszczyński 1991). However, neither sedimentary structure nor degree of bioturbation are characterized in sufficient detail in most datasets to capture much more than “oxic-to-some-degree” vs. “anoxic” at each specific location, often due to insufficient preservation and/or the idiosyncrasies of documentation. Generalizations like these culminate in a simplification of real-world patterns such that higher resolution niche differentiation may be difficult to observe, resulting in high niche similarity. Furthermore, other aspects of species’ realized niches, such as biotic interactions, cannot be well-characterized in most cases, even using more recent datasets (Peterson et al. 2011).

These results furthermore indicate that the current method of approximating environmental characteristics using continuous sedimentary variable proxies may be less useful than manually classifying relevant habitat regions on a more categorical level. For example, anoxic and oxygen-rich marine habitats can be interpreted directly by a researcher based on a collection of sedimentological variables and mapped accordingly. While these variable maps may also include generalizations, they have the potential to more accurately represent cumulative environmental variables that are directly relevant to taxonomic distributions in a way that cannot be easily captured using sedimentary data alone. Niche models created using ground truthed categorical habitats such as these may therefore produce more definitive results when comparing niche characteristics between taxa or across time, and should be tested in future analyses.
In addition to the issues with environmental generalization, taxonomic preservation and sampling bias in the fossil record further complicate the resolution issue, since only a very small subset of a population (even across a long temporal interval) is fossilized, collected, and documented (Kidwell and Flessa 1996). This means that it is challenging to collect species occurrences across their entire geographic and temporal range. Niche studies on modern animals and plants have shown that not using the full biogeographic range of species’ distributions can obscure accurate niche estimation, particularly when using ENM algorithms (i.e., Owens et al. 2013). Furthermore, when conducting deep-time niche analysis, these occurrences are further reduced by spatial thinning, which again promotes analysis of common, widespread, and long-lived species rather than taxa with patchy or localized distributions (whether those be real or artifactual). It is commonly assumed that wide-ranging taxa have large niches and are therefore more ecologically generalized (Kammer et al. 1997; Slatyer et al. 2013). Even under ideal circumstances with high-resolution, modern data, generalist taxa have been found to have similar ecological characteristics to one another (Denelle et al. 2020). Given that subtle environmental differences are unlikely to be preserved when using deep-time environmental proxy data, ecological distinctions between generalist taxa occupying similar overall habitats are also unlikely to be observed. The high similarity between both related and unrelated species observed here supports this interpretation. Given that this analysis includes some of the highest resolution and best-vetted data available for paleo-niche analysis, these results suggest that PNC is unlikely to be easily distinguished in fossil data unless higher resolution datasets with a greater biogeographic diversity of taxa and their environments can be achieved. Instead, fossil data is more applicable to analyzing broad-scale changes in niche parameters within generalist species.
Within-species partitioning of generic niches

In paleontological studies, the genus-level, or even above, is often used for ecological analyses since because (1) genus-level data is so much more abundant and (2) genus-level patterns are considered a decent proxy of species-level biotic processes (Hadly et al. 2009; Malizia and Stigall 2011; Nürnberg and Aberhan 2013; Brame and Stigall 2014; Tong et al. 2021). Others argue that the species is the fundamental biological unit of analysis (Vrba 1980, 1984, 1989), and that higher taxonomic levels do not inherently capture ecologically relevant biotic responses (Hendricks et al. 2014). For example, while a genus’ niche might show expansion across time due to a few species on the extremities of the genus niche moving into new abiotic space, individual species that do not occupy marginal regions of the genus’ niche may instead be experiencing substantial reduction of their niche space occupation in a way that cannot be observed at the genus-level (Hendricks et al. 2014; Purcell and Stigall 2021). This can be visualized here with convex hull reconstructions (Figure 19). In the *B. reduncus-B. scotti* interval, for example, the extremities of the first principal component axis of the genus-level niche are not occupied by the three species analyzed directly, and any change to these species across this axis could not be observed at the genus-level (Figure 19). Furthermore, our results of niche comparisons between *Inoceramus* species and the *Inoceramus* genus-level niche estimation suggest that the genus-level niche does not represent individual species well statistically (Table 11). Average niche overlap for species-level analysis compared to the genus-level is only 0.52, and only one species (*I. barabini*) had statistically high overlap and stability based on the niche similarity test in only one biozone interval of two (Table 13). This contradicts several studies using fossil data that have found similar patterns of niche overlap and conservation between both species and genera (e.g., Hadly et al. 2009; Malizia and Stigall 2011; Nürnberg and Aberhan
2013; Brame and Stigall 2014). How time influences a genus-to-species niche relationship is beyond the scope of this analysis but given the poor overlap in this study between individual species and the genus-level analysis, it is very possible that genus-level paleo-niche patterns reflect ecological phenomenon at the species-level.

Conversely, if the genus is considered a functional entity, rather than an evolutionary entity, its utility is vastly increased. Firstly, there are distinct advantages to utilizing higher-level taxonomic information for addressing biological and ecological traits in general. Obviously, using the genus greatly expands the utility of sparse datasets because it allows researchers to incorporate fossil data that has only be identified to the genus-level. Additionally, studies have found that higher taxonomic levels may present distinct biotic responses relevant to life history traits (Hadly et al. 2009; Smith et al. 2019; Tong et al. 2021). More importantly for using paleontological data for niche analyses, genera classifications are typically based on morphological similarity, which is directly linked in most cases with the functional traits relevant to an organism’s lifestyle (i.e., Weller 1949). While this often means that genera cannot be considered evolutionarily entities, since they may represent convergent morphologies rather than phylogenetically related taxa, this principle enables researchers to classify genera into functional ecological entities that occupy a specific ecological role in their community structure and assess their temporal and spatial distribution relative to abiotic factors on a macroecological level (Foster and Twitchett 2014; Dunhill et al. 2018; Edie et al. 2018).

Therefore, if we consider a genus niche to represent a functional entity rather than an evolutionary entity, the issues related to interpreting the genus relative to the species niche are largely resolved. Analyses using the genus niche can instead enhance our understanding
of how functional entities are distributed relative to fundamental abiotic characteristics and how they respond across space and time. The actual species-level changes occurring in these analyses may not be clear, depending on the genus and number of species included within it, but any genus-level alteration may be indicative of taxonomic or ecological instability (Purcell and Stigall 2021). For example, contraction of a genus niche requires either species extinction or contraction of species’ niches (Figure 20). Depending on the degree of overlap between species within the niche (i.e., PNC), these alterations must furthermore be simultaneous across the same interval of comparison. Ignoring the specific species-level changes however, this pattern would indicate an overall ecological responses by the functional entity represented by the genus, and can lead to further, more detailed analyses related to abiotic pressures and ecological processes. Only genus niche stability, in fact, may be entirely nonspecific about the species niche patterns (Figure 20C), while still indicating that either niche characteristics of the functional entity are being maintained through stability at the species level or through compensation for any alteration by individual species by the others within the genus. Therefore, though the genus niche is not representative of individual species ecological niches, it is in fact highly relevant to understanding functional diversity patterns at a broad scale.
Figure 20. Examples of niche dynamics through time at the species and genus-level, which can be interpreted as representing the niche of an ecological entity. The genus-level niche is in black, species niches are colored.

Conclusions

This dataset represents the highest-resolution, most geographically extensive, and thoroughly cleaned deep-time dataset used for niche comparisons to date. We find that paleo-niche analyses should be considered individualistically and wholistically to avoid over-generalizing biological phenomenon. The size of a temporal bin, particularly during intervals with substantial changes to paleoceanography and paleogeography, can strongly influence the outcome of niche estimation in deep time, and accuracy diminishes with environmental data aggregated over longer intervals. High niche similarity values resulting from such low-resolution analysis are therefore potentially misleading, given that they likely result from the over-generalization of already biased environmental proxy data and the use of geographically
widespread, generalist taxa. The case study comparing genus-level niche estimation to the species-level did not find that the generic-based niche represented individual species’ niches well, despite the fact that this genus is composed of abundant, generalist species with relatively high niche overlap between them. This suggests that the structure of species’ paleo-niches within genera is an exciting research area in need of more and specific analysis. Furthermore, if the genus is considered as a functional entity, genus-level niche analyses may open up an extremely useful avenue for further research into how abiotic factors across space and time influence the fundamental functional characteristics of ecological communities.
CHAPTER 5: CONCLUSION

The Western Interior Seaway has been studied for over 200 years (Everhart, 2017), resulting in the collection of vast numbers of fossils and geologic data all collected with specific, and not necessarily complementary research goals. The Western Interior Seaway persisted from approximately the late Jurassic to the early Paleocene (approximately 100 Ma) and is one of the best preserved and studied sedimentary basins in the world (Miall et al., 2008). Not only is the WIS well studied as a whole, but the digital age has provided a previously unprecedented venue for cataloguing and disseminating paleontological and sedimentological data. The evolution of quantitative methods applicable to fossil data has also greatly advanced, and computer software has made exploring “big data” questions related to Earth history very accessible such that researchers can investigate large biostratigraphic regions like the WIS with ease relative to the methods available to previous researchers. Geologic data, recording aspects of the interacting biotic and abiotic aspects of the Earth system, can now more than ever before be used to address questions related to how the Earth functions in the past and evaluate underlying principles that may explain how it will progress in the future.

This expanding scientific horizon requires that we carefully assess the match between data uncertainty and the questions we are asking, but also the methods we are using to answer them. Limitations in paleontological data, while somewhat mitigated by our ever-growing datasets and increased understanding about biases, still exist and must be taken into account. Previously established paradigms regarding biotic patterns and processes, while not inherently wrong, require reassessment using updated information and methods to support their accuracy. For example, the analyses documented here indicate that biotic subprovinces
were not an established characteristic of the WIS, as previously established by Kauffman (1980). In short, the fundamental relationships between biotic and abiotic data that paleontologists and other researchers explore should be carefully considered not just within the confines of a priori factors, but also within the limits of what paleontological and geological data can achieve using current methods. One goal of this dissertation is to explicitly address some of these issues using a high-resolution, well establish dataset from an extensively studied, geographically extensive region; thus testing potential biases in as comprehensive a dataset as exists for deep time marine ecosystems.

The first research chapter (Chapter 2) uses relatively novel quantitative methods to reassess a long-established assumption regarding the paleobiogeography of the WIS, and the relationship between taxonomic distributions and abiotic factors. This chapter finds that the current dataset is incompatible with previous conclusions regarding WIS provinciality (Kauffman, 1980), supporting further research into the biogeographic structure of continental-sized regions in the fossil record. The second research chapter (Chapter 3) takes these methods further, addressing a popular topic in paleontology today: that functional diversity patterns result from ecological processes unique from taxonomic diversity, and tests the influence of spatial constraints on paleobiogeographic analytical results. This study compares functional diversity patterns with major abiotic shifts present at the same time at different spatial scales and contrasts these patterns with modern patterns. The last research chapter (Chapter 4) explores the validity of ecological niche modeling using deep-time datasets by comparing the influence that temporal resolution has on niche analyses and assessing the conservation of niche characteristics between related specific taxa and their genus. It furthermore provides an extensive and detailed stratigraphic dataset that
compliments the taxonomic data used in previous chapters and lays the groundwork for using stratigraphic data to create sedimentological maps of the WIS at the biostratigraphic level. This dataset will contribute significantly to continued studies linking paleobiogeographical and paleoenvironmental patterns in the WIS.

This research, while presenting novel data related to the biotic and abiotic system present in the WIS, are first and foremost linked to testing how methods shape research results (and consequent interpretations). They are meant to highlight the importance of considering data from multiple angles and scales, while also taking biases into account whenever feasible. The Late Cretaceous WIS represents one of the best studied, most vetted, and extensively documented datasets in the world, allowing researchers to not only explore highly relevant questions related to abiotic and biotic interactions, but also to establish principles for paleontological research moving forward. It is my hope that, regardless of whether the conclusions of this research remain eternally accurate, that these data and the methods used to explore them can create a foundation for future research, improving our understanding of the Earth system and its exciting history as a whole.
REFERENCES


Denelle, P., C. Violle, and F. Munoz. 2020: Generalist plants are more competitive and more functionally similar to each other than specialist plants: insights from network analyses. Journal of Biogeography 47:1922–1933.


APPENDIX A: Supplementary Materials for “Faunal provinciality in the Late Cretaceous Western Interior Seaway using a network modeling approach” by Purcell et al.:

Appendix A-1: Detailed Methods and Results

Database Vetting:

Original downloads from the PBDB and iDigBio were configured to include all taxa from the Late Cretaceous and then vetted to remove incomplete, erroneous, or poorly spatially and temporally resolved data. All genera denoted with “?”, “c.f.”, “aff.” and taxa not identified to the genus-level were removed to avoid occurrences with insufficient taxonomic resolution. All taxonomic names were vetted using peer-reviewed literature to identify taxonomic updates, synonymies, and to correct any spelling errors (Appendix B provides all references used to vet taxonomic designations). Stratigraphic ages were updated and vetted to geologic substage temporal resolution based on stratigraphic and biozone information included in the original database file, primary literature sources cited for each occurrence, and the WIS stratigraphic database of Myers et al. (2015). Occurrence records with spatial resolution > 30km of uncertainty were removed to avoid spatially ambiguous data, and records south of 23° latitude and east of -80° longitude were removed to avoid records from outside the WIS and GCP regions. The resulting geographic range of the dataset extends from 23° to 54° N and 83° to 118° W (Figure S1).

The final vetted database contained 33,168 fossil occurrences (Table S1; Figure S1) representing 574 unique genera and 1113 unique species of marine macroinvertebrates including bivalves, ammonites, gastropods, echinoderms, corals, brachiopods, polychaetes, and crustaceans. These groups represent a wide variety of lifestyle modes both within clades and between distinct phylogenetic groups. They are furthermore important components in both modern and ancient shallow marine ecosystems (Bush and Bambach 2011) and therefore allow comparison to modern systems.
Figure S1. Map of fossil occurrence data information. Point colors indicate substage.
Database Evaluation:

Prior to network modeling, the vetted database was analyzed to assess for taphonomic and sampling bias (Vilhena and Smith 2013). This included comparisons of taxonomic richness with sampling bias, or the number of fossil occurrences per geographic unit. Since taxonomic information is categorical and autocorrelation of this data type is complex and frequently includes duplicated locations, alternative methods for analyzing general sampling bias spatial autocorrelation were utilized. This was achieved by creating a 60km grid in ArcGIS Pro using the Grid Index Tool (ESRI, 2021) that was overlain on the data. This grid size was chosen to account for uncertainty in fossil location information, while still maintaining localized community ranges that would be consistent across geographic space. An alternative 30km grid size at our occurrence resolution was tested and produced primarily singleton grid cells (i.e., a grid cell with only 1 fossil occurrence) and was therefore uninformative for our analysis of fossil communities. The genus- and species-level data were summarized by number of fossil occurrences per 60km grid cell as well as the number of unique genera and species per grid cell, though only genus-level data were analyzed using network modeling (see Appendix C for R code).

The data, once aggregated, was found to be highly right-skewed (Figure S2 and S3) and to have a strong correlation between the number of fossil occurrences in a grid cell (i.e., sampling) and the taxonomic richness (Figure S4). Thus, the data is non-normal and there is a strong positive relationship between taxonomic richness and the number of occurrences sampled in a given 60-km pixel, or node, herein regarded as sampling effort. This correlation is used in subsequent analyses, including MST assessments, to determine the influence that sampling bias has on network outputs by allowing for indirect comparisons between the number of fossils present in a node as indicated by generic richness, and network values or patterns.
Figure S2. Histograms and Q-Q plots of genus-level occurrence counts and richness for the data aggregated in 60-km grid cells. The data is highly right-skewed and deviates greatly from the trend line, indicating that it is non-normal.
Figure S3. Histograms and Q-Q plots of species-level occurrence counts and richness for the data aggregated in 60-km grid cells. The data is highly right-skewed and deviates greatly from the trend line, indicating that it is non-normal.
Figure S4. Scatter plots and log-transformed plots of taxonomic richness and the number of occurrences for the data aggregated in 60-km grid cells. The data is strongly correlated prior to log₁₀ normalization. Once normalized, the correlation is smaller though still high, and the trend line fits with much smaller residual error.

The nature of the geographic distribution and autocorrelation embedded in the dataset was analyzed with ArcGIS Pro (Spatial Autocorrelation (Global Moran's I) (Spatial Statistics) tool) and
the Cluster and Outlier Analysis (Anselin Local Moran's I (Spatial Statistics) tool; ESRI, 2021) to
determine the degree of local and global spatial autocorrelation of genus and species-level richness
and total number of fossils collected at each location. Global and local spatial autocorrelation of these
data was then analyzed using the Moran’s I statistic and local Anselin Moran’s I statistic,
respectively, in ArcGIS Pro (ESRI, 2021). Overall, results of spatial autocorrelation analyses (both
global and local) indicate high spatial clustering within the WIS and GCP with some high-low and
low-high outliers based on region (Figures S4-S8). Visual comparisons between Anselin Moran’s I
with network results, however, did not indicate that any obvious parallels between autocorrelation
patterns of clustering/outliers with either link weights or betweenness centrality. These results
indicate that spatial bias is not a primary control on general network topology.

The Global Moran’s I analysis indicates that all substages are statistically clustered (p-values
ranging from 0 to 0.078), with only the distribution of species-level occurrence counts in the Lower
Campanian falling just outside the 95% confidence interval. All z-scores for the tests are positive,
indicating clustering rather than dispersal. These results suggest that the data overall is highly
clustered globally when summarized based on the 60km grid, both in terms of taxonomic richness and
sampling effort. The results of the Anselin Local Moran’s I analysis indicates similar clustering at
local scales (Figure S4-S8). For each substage, many summarized 60km grid points have statistically
high clustering of either high values with other high values, or low values with other low values for
both unique taxa and total taxa occurrence counts.
Table S2: Results of Global Moran’s I analysis. Distance threshold values were calculated automatically to ensure that each occurrence had at least one neighbor and have been rounded up to the nearest km.

<table>
<thead>
<tr>
<th>Substage</th>
<th>Number Unique Genera/Grid Cell</th>
<th>Number Genus Occ/Grid Cell</th>
<th>Number Unique Species/Grid Cell</th>
<th>Distance Threshold (nearest km)</th>
<th>Number Species Occ/Grid Cell</th>
<th>Distance Threshold (nearest km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low CAM</td>
<td>Moran’s I p: 0.003 Moran’s I z: 2.997</td>
<td>Moran’s I p: 0.006 Moran’s I z: 2.724</td>
<td>Moran’s I p: 0.006 Moran’s I z: 2.757</td>
<td>Distance Threshold (nearest km)</td>
<td>Moran’s I p: 0.078 Moran’s I z: 1.762</td>
<td>Distance Threshold (nearest km)</td>
</tr>
<tr>
<td>Mid CAM</td>
<td>Moran’s I p: 0.000 Moran’s I z: 14.046</td>
<td>Moran’s I p: 0.000 Moran’s I z: 8.393</td>
<td>Moran’s I p: 0.000 Moran’s I z: 10.389</td>
<td>Distance Threshold (nearest km)</td>
<td>Moran’s I p: 0.000 Moran’s I z: 7.006</td>
<td>Distance Threshold (nearest km)</td>
</tr>
<tr>
<td>Up CAM</td>
<td>Moran’s I p: 0.000 Moran’s I z: 5.577</td>
<td>Moran’s I p: 0.000 Moran’s I z: 6.468</td>
<td>Moran’s I p: 0.000 Moran’s I z: 3.662</td>
<td>Distance Threshold (nearest km)</td>
<td>Moran’s I p: 0.000 Moran’s I z: 4.428</td>
<td>Distance Threshold (nearest km)</td>
</tr>
<tr>
<td>Low MAA</td>
<td>Moran’s I p: 0.000 Moran’s I z: 8.280</td>
<td>Moran’s I p: 0.000 Moran’s I z: 5.163</td>
<td>Moran’s I p: 0.000 Moran’s I z: 7.604</td>
<td>Distance Threshold (nearest km)</td>
<td>Moran’s I p: 0.000 Moran’s I z: 5.019</td>
<td>Distance Threshold (nearest km)</td>
</tr>
<tr>
<td>Up MAA</td>
<td>Moran’s I p: 0.000 Moran’s I z: 10.486</td>
<td>Moran’s I p: 0.000 Moran’s I z: 5.786</td>
<td>Moran’s I p: 0.000 Moran’s I z: 13.393</td>
<td>Distance Threshold (nearest km)</td>
<td>Moran’s I p: 0.000 Moran’s I z: 7.144</td>
<td>Distance Threshold (nearest km)</td>
</tr>
</tbody>
</table>

Local spatial relationships related to sampling and taxonomic richness were assessed using the Cluster and Outlier Analysis (Anselin Local Moran’s I) (Spatial Statistics) tool in ArcGIS Pro. Anselin Local Moran’s I analysis determines if the values of a particular feature are more or less similar to its neighbors than would be expected at random using a 95% confidence interval. The results can indicate spatial outliers and local clusters that are not assessed using the global statistic. Overall, Anselin Local Moran’s I analysis indicates that the data from the Campanian substages are concentrated in the WIS region rather than the GCP based both on the number of unique taxa (i.e., taxonomic richness) and the total fossil occurrence counts (i.e., sampling count). These analyses also strongly suggest that the distribution of genera/species richness and sampling counts are not random but show high clustering. Analyses of the Campanian substages in general show that the dominance of high-high clusters for richness shifts from the more northern grid cells of the WIS in the Lower Campanian, to the more central portion of the WIS in the Middle and Upper Campanian. The Maastrichtian substages, on the other hand, indicate that higher richness and sampling are concentrated in the GCP and that the WIS is dominated by low-low clustering for all analysis.

It also appears that clustering/outlier patterns for sampling counts parallel that of taxonomic richness, except in the Middle Campanian substage where the generic richness and the total genus
sampling count results appear inverse. These results indicate that, despite the fact that the southern portion of the WIS (Colorado and southern Wyoming) has clusters of low sampling, this area is dominated by high-high clusters of generic richness. Similarly, the northern portion of the WIS (Montana and Canada) is dominated by low-low clusters of generic richness, despite the fact that it is dominated by high sampling. This suggests that the few occurrence records in the southern region of the WIS that do exist have a disproportionately high genus richness, and the reverse for the northern region. Interestingly, this pattern does not hold when the data are analyzed based on species for the Middle Campanian. Details for each substage are listed below.

Visual comparisons of networks relative to Aneslin Local Moran’s I analysis indicate that spatial distribution of unique genera counts within grid cells did produce noticeable bias in the substage networks (Figure S5 – S9). In the Lower Campanian, the northern portion of the WIS, which is dominated by high-high clusters of unique genera counts had overall weaker faunal associations and stronger links in the southern portion where low-low clusters dominate. In the Middle Campanian, this pattern is reversed: the southern portion of the basin, dominated by high-high clusters, maintains relatively stronger links than the northern portion, which is characterized by low-low clusters. The Upper Campanian network has the strongest links between regions with low-low clustering in the northwest and east-central regions, but relatively strong faunal associations were found across the majority of the WIS, connecting the southwestern, northwestern, and east-central portions (see Figure S10, S12, S14, S16, and S18). The Lower Maastrichtian network maintains its strongest links between the southern and east-central portions of the WIS, which are characterized by low-low clustering, but relatively strong links are maintained across the whole basin and are weakest in the north where few clusters or outliers were observed. The Lower Maastrichtian network also presents strong links within the GCP between grid cells along the eastern margin, where high-high clusters dominate. Finally, the Upper Maastrichtian network is characterized by strong links both in the WIS, where low-low clusters dominate, and in the GCP, particularly along the eastern margin
where high-high clusters are the most common. The faunal associations are somewhat stronger in the WIS during this substage than they are in the GCP.

*Lower Campanian:* Data from the Lower Campanian summarized by the 60-km grid cells indicate that the WIS has clusters of both high-high and low-low counts of unique genera (Figure S5). Clusters of high-high unique genera are mostly found in central Wyoming and Montana, while low-low clusters are mostly distributed in northern New Mexico and southern Utah, as well as Alberta, Canada. Outliers of low unique genera counts surrounded by high (low-high outliers) are also fairly common and distributed across Montana and Wyoming. Only three grid cells were found to be statistical outliers of high unique genera counts surrounded by low (high-low outliers), and these are located in southern Utah and northern New Mexico. The majority of grid cells in Utah, Colorado, and the eastern regions of Montana and Wyoming were not statistically significant clusters or outliers regarding unique genera counts. The GCP region had no significant clusters or outliers for the Lower Campanian data. Patterns found using the summarized number of unique species were very similar to those of the unique genera, except that the number of high-high clusters was far fewer in central Wyoming and Montana and that a greater number of high-low outliers were found in the western Colorado. One grid cell was determined to be a high-low outlier for the number of unique species in the GCP of Alabama. The data summarized by the number of total genera indicate that there are both high-high clusters and low-high outliers distributed throughout the central WIS, primarily in Wyoming and northern Utah. Low-low clusters of genus occurrence counts are located primarily in Alberta and northern New Mexico. Only two grid cells were determined to be high-low outliers, located in Alabama and northern New Mexico. Patterns for the summarized number of species occurrence counts were very similar to those of genus occurrence counts, except that the number of high-high clusters and low-high outliers in central WIS were far fewer and that four total grid cells were identified as high-low outliers in northern New Mexico and Colorado.
Figure S5. Map results of Anselin Local Moran’s I analysis showing the distribution of clustering and outliers for number of unique genera summarized using the 60km grid for the Lower Campanian.
Middle Campanian: Data from the Middle Campanian summarized by the 60-km grid cells indicate that the northern WIS in Montana is dominated by low-low clusters with few high-low outliers and that the central WIS in Colorado, Utah, and Wyoming is dominated by both high-high clusters and low-high outliers for both the number of unique genera and species (Figure S6). Grid cells located in New Mexico, eastern Wyoming, North Dakota, and most of the GCP were predominantly non-significant, excluding a few low-low clusters. When the data was analyzed based on the number of species occurrences, a very similar pattern emerges, except that the GCP has a much higher number of low-low clusters. When analyzed based on the number of genus occurrences, however, the distribution of clusters and outliers is almost completely opposite that of the number of species occurrences. For data summarized by genus occurrences, the Montana region is dominated by high-high clusters with a single cell that represents a low-high outlier, the central region of the WIS is dominated by both low-low clusters and high-low outliers, and grid cells in the GCP region were almost all found to be low-low clusters.
Figure S6. Map results of Anselin Local Moran’s I analysis showing the distribution of clustering and outliers for number of unique genera summarized using the 60km grid for the Middle Campanian.
**Upper Campanian:** Data from the Upper Campanian summarized by the 60-km grid cells indicate that the northern WIS in Montana, Alberta, and Saskatchewan is dominated by low-low clusters with few high-low outliers located in Montana and that the central WIS in Colorado and southern Wyoming is dominated by both high-high clusters and low-high outliers for both the number of unique genera and the number of unique species (Figure S7). Grid cells located in New Mexico, eastern Wyoming, North Dakota, Nebraska, South Dakota, and most of the GCP were predominantly non-significant. However, the data analyzed based on unique genera indicate a concentration of low-low clusters in South Dakota, northern Nebraska, and North Dakota that is not found when analyzed based on unique species counts. There are also several grid cells identified as low-high outliers across the GCP for unique species, and one high-low outlier for unique genera counts in eastern Texas. The distribution of significant clusters and outliers identified when the data was analyzed based on both the total number of genus and species count is extremely similar to those found for the number of unique genera and species in the Upper Campanian.
Figure S7. Map results of Anselin Local Moran’s I analysis showing the distribution of clustering and outliers for number of unique genera summarized using the 60km grid for the Upper Campanian.
*Lower Maastrichtian:* Data from the Lower Maastrichtian summarized by the 60-km grid cells indicate that the South Dakota and southern Colorado regions are dominated by low-low clusters of unique genera and unique species, but that the majority of the WIS region has no significant clustering or outliers (Figure S8). The GCP region, however, is dominated by low-low clustering in northern Mexico and Arkansas for the number of unique genera and mix of high-high clusters with low-high outliers in the east. This pattern holds for the GCP region when analyzed based on the number of unique species, but with only a single low-low cluster grid cell in northern Mexico. Two grid cells were identified as high-low outliers for both the unique genus and species analysis, one in eastern Montana and another in eastern Texas. The distribution of clusters and outliers for analysis performed based on the total species and genus occurrence counts were very similar to that of the number of unique genera overall, except that Colorado does not have significant low-low cluster grid cells while Montana and southern Canada regions do.
Figure S8. Map results of Anselin Local Moran’s I analysis showing the distribution of clustering and outliers for number of unique genera summarized using the 60km grid for the Lower Maastrichtian.
Upper Maastrichtian: Data from the Upper Maastrichtian summarized by the 60-km grid cells indicate that the WIS is dominated by low-low clusters of unique genera and unique species, and that the western portion of the GCP, including Texas and northern Mexico have no significant outliers or clusters (Figure S9). The eastern GCP region, however, is dominated by low-low clustering, particularly in Alabama and Georgia, and low-high outliers for both the number of unique genera and species. This general pattern holds when analyzed based on both genus and species occurrence counts, excepting that the WIS region in North and South Dakota are predominantly non-significant.
Figure S9. Map results of Anselin Local Moran’s I analysis showing the distribution of clustering and outliers for number of unique genera summarized using the 60km grid for the Upper Maastrichtian.
Network Analysis:

Faunal occurrence records of genera were aggregated into “paleogeographic analysis units” or nodes based on the 60km-grid that was created in ArcGIS Pro using the Grid Index tool with cell sizes set to 60x60km (see above section on data evaluation). Aggregated data in the form of a presence-absence matrix were input into the EDENetworks version 2018 software package (Kivelä et al. 2015). Nodes with fewer than three genera were removed from the analysis to allow for better comparisons of assemblages and to avoid singletons (i.e., Kiel, 2017). Weighted distances between nodes, or links, were calculated based on the Bray–Curtis dissimilarity index, which calculates how similar two nodes are based on the number of shared taxa,

\[
D(A, B) = \frac{2 \sum_{i=1}^{k} \min(A_i, B_i)}{\sum_{i=1}^{k}(A_i + B_i)}
\]

(EQ 1)

where A and B are two nodes, calculated over a vector with length K where every i\(^{th}\) element represents the presence or absence of a specific taxa. This calculation is computationally identical to the Sorensen or Dice coefficient (Somerfield 2008). The index value ranges from 0, indicating that the two nodes share identical taxa, to 1, indicating that they share none. R Code used to configure the database into tables that could be input into EDENetworks can be found in Appendix C.

All data locations were converted to paleo coordinates using the R package chronosphere (Kocsis and Raja 2020) to reproject the data back to their original positions and relative distances. Faunal provinces were delimited using these distance weight measures in a thresholding network approach (Moalic et al. 2012; Kivelä et al. 2015; Kiel 2016) where the threshold represents the maximum weight (level of dissimilarity) of the links used to construct the network; weights that fall above this value (i.e., those with greater faunal dissimilarity) were removed from the network (Kivelä et al. 2015). Thus, at lower thresholds the network is more fractured (more and smaller network components) and at higher thresholds the network is more connected (fewer and larger components).
The point at which all components are connected without forming a unitary component is referred to as the percolation point (Kivelä et al. 2015). Components which become disconnected at thresholds at and below the percolation point can be interpreted as representing distinct “community” groups (Newman 2012) and therefore can be used to draw conclusions about the distribution of faunal provinces.

Network models were created using EDENetworks for all substages individually and for the complete database, which combines all substages together. Networks were produced between the 0.99 threshold and 0.40 threshold to analyze the consistency of topological patterns at intervals of 0.05, and an additional network was produced for each percolation point (Table S3). General patterns in network connection across all substages for the complete database were also visually assessed by using a coarser spatial aggregation of the data, by a 360km grid aggregation (see below). This aggregation produced lower resolution comparison with fewer nodes but allowed for easier visualization of network connections across both space and time simultaneously.

Networks and their figures were produced for each substage between the 0.99 threshold and 0.40 threshold to analyze the consistency of topological patterns at intervals of 0.05, and an additional network was produced for each percolation point (Table S3). The complete database, containing the data from all substages, was only analyzed at thresholds between 0.76 and 0.40 due to computational time constraints. Topological patterns were generally consistent at all thresholds for all networks, and no distinct minor components or subprovinces become more obvious within the primary WIS and GCP provinces at lower thresholds (Figure S10-S20).

Results of the network analysis indicate a single faunal province, representing the WIS throughout the Campanian, and two primary faunal provinces in the Maastrichtian (one circumscribing the WIS and the other delineating the GCP) (Figure S16-S19); the two provinces were also observed in the complete database network (Figure S20). The lack of a GCP province in the Campanian (Figure S10-S15) is probably driven by a lack of GCP data; only in the Maastrichtian
when GCP data is more abundant does the GCP province become clear. Topological patterns were generally consistent at all thresholds for all networks, and no distinct minor components become more obvious within the primary WIS and GCP provinces at lower thresholds.
Table S3. Summary table of network analysis results for all substages and complete database including summary values for individual threshold analysis.

<table>
<thead>
<tr>
<th></th>
<th>Low CAM</th>
<th>Mid CAM</th>
<th>Up CAM</th>
<th>Low MAA</th>
<th>Up MAA</th>
<th>Complete DB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
<td>75</td>
<td>109</td>
<td>138</td>
<td>68</td>
<td>50</td>
<td>440</td>
</tr>
<tr>
<td><strong>Distance Measure</strong></td>
<td>Jaccard</td>
<td>Jaccard</td>
<td>Jaccard</td>
<td>Jaccard</td>
<td>Jaccard</td>
<td>Jaccard</td>
</tr>
<tr>
<td><strong>Avg Distance</strong></td>
<td>0.79</td>
<td>0.8</td>
<td>0.83</td>
<td>0.84</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>Min Distance</strong></td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Max Distance</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>PERCOLATION POINT</strong></td>
<td>0.61</td>
<td>0.57</td>
<td>0.5</td>
<td>0.8</td>
<td>0.86</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>367</td>
<td>175</td>
<td>174</td>
<td>674</td>
<td>447</td>
<td>16965</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>9.79</td>
<td>3.21</td>
<td>2.52</td>
<td>19.82</td>
<td>17.88</td>
<td>77.11</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>33</td>
<td>11</td>
<td>17</td>
<td>41</td>
<td>29</td>
<td>119</td>
</tr>
<tr>
<td><strong>Avg Clustering (&lt;c&gt;)</strong></td>
<td>0.41</td>
<td>0.25</td>
<td>0.18</td>
<td>0.68</td>
<td>0.8</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.51</td>
<td>0.48</td>
<td>0.44</td>
<td>0.69</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>18</td>
<td>25</td>
<td>74</td>
<td>3</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>57</td>
<td>84</td>
<td>61</td>
<td>66</td>
<td>44</td>
<td>415</td>
</tr>
<tr>
<td><strong>MINIMUM SPANNING TREE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>74</td>
<td>108</td>
<td>137</td>
<td>67</td>
<td>49</td>
<td>439</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>1.97</td>
<td>1.98</td>
<td>1.99</td>
<td>1.97</td>
<td>1.96</td>
<td>2</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td><strong>Avg Clustering (&lt;c&gt;)</strong></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.49</td>
<td>0.5</td>
<td>0.52</td>
<td>0.56</td>
<td>0.58</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>connected</td>
<td>connected</td>
<td>connected</td>
<td>connected</td>
<td>connected</td>
<td>connected</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>57</td>
<td>84</td>
<td>61</td>
<td>66</td>
<td>44</td>
<td>415</td>
</tr>
<tr>
<td><strong>THRESHOLD 0.99</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>2322</td>
<td>5741</td>
<td>8761</td>
<td>2066</td>
<td>987</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>61.92</td>
<td>105.34</td>
<td>126.97</td>
<td>60.76</td>
<td>39.48</td>
<td>-</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>71</td>
<td>108</td>
<td>135</td>
<td>65</td>
<td>48</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Clustering (&lt;c&gt;)</strong></td>
<td>0.93</td>
<td>0.98</td>
<td>0.97</td>
<td>0.94</td>
<td>0.89</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.75</td>
<td>0.8</td>
<td>0.81</td>
<td>0.82</td>
<td>0.84</td>
<td>-</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>3</td>
<td>connected</td>
<td>2</td>
<td>2</td>
<td>connected</td>
<td>-</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>73</td>
<td>137</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>THRESHOLD 0.95</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>2299</td>
<td>5492</td>
<td>8235</td>
<td>1832</td>
<td>606</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>61.31</td>
<td>100.77</td>
<td>119.35</td>
<td>53.88</td>
<td>24.24</td>
<td>-</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>70</td>
<td>108</td>
<td>133</td>
<td>65</td>
<td>38</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Clustering (&lt;c&gt;)</strong></td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>0.91</td>
<td>0.87</td>
<td>-</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.75</td>
<td>0.79</td>
<td>0.8</td>
<td>0.8</td>
<td>0.75</td>
<td>-</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>3</td>
<td>connected</td>
<td>2</td>
<td>3</td>
<td>connected</td>
<td>-</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>73</td>
<td>137</td>
<td>66</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>THRESHOLD 0.90</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>2042</td>
<td>4731</td>
<td>6901</td>
<td>1505</td>
<td>497</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Avg Degree (&lt;k&gt;)</td>
<td>Max Degree (kmax)</td>
<td>Avg Clustering (&lt;c&gt;)</td>
<td>Avg Distance (&lt;d&gt;)</td>
<td>Connectivity (components)</td>
<td>Connectivity (largest)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td>54.45</td>
<td>67</td>
<td>0.88</td>
<td>0.72</td>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>86.81</td>
<td>102</td>
<td>0.89</td>
<td>0.77</td>
<td>2</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>100.01</td>
<td>126</td>
<td>0.85</td>
<td>0.78</td>
<td>3</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>44.26</td>
<td>58</td>
<td>0.83</td>
<td>0.78</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>19.88</td>
<td>29</td>
<td>0.85</td>
<td>0.71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>THRESHOLD 0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1684</td>
<td>3651</td>
<td>4958</td>
<td>1070</td>
<td>431</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.91</td>
<td>66.99</td>
<td>71.86</td>
<td>31.47</td>
<td>17.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>97</td>
<td>109</td>
<td>53</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>0.79</td>
<td>0.74</td>
<td>0.75</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.69</td>
<td>0.73</td>
<td>0.74</td>
<td>0.74</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>106</td>
<td>135</td>
<td>66</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>THRESHOLD 0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1400</td>
<td>2711</td>
<td>3586</td>
<td>674</td>
<td>357</td>
<td>26242</td>
</tr>
<tr>
<td></td>
<td>37.33</td>
<td>49.74</td>
<td>51.97</td>
<td>19.82</td>
<td>14.28</td>
<td>119.28</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>84</td>
<td>91</td>
<td>41</td>
<td>26</td>
<td>268</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.7</td>
<td>0.68</td>
<td>0.68</td>
<td>0.73</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.66</td>
<td>0.7</td>
<td>0.71</td>
<td>0.69</td>
<td>0.66</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>104</td>
<td>135</td>
<td>66</td>
<td>30</td>
<td>423</td>
</tr>
<tr>
<td>THRESHOLD 0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1086</td>
<td>18.51</td>
<td>2333</td>
<td>506</td>
<td>300</td>
<td>16253</td>
</tr>
<tr>
<td></td>
<td>28.96</td>
<td>33.96</td>
<td>33.81</td>
<td>14.88</td>
<td>12</td>
<td>73.88</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>65</td>
<td>71</td>
<td>35</td>
<td>25</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.62</td>
<td>0.58</td>
<td>0.63</td>
<td>0.68</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>103</td>
<td>128</td>
<td>59</td>
<td>28</td>
<td>377</td>
</tr>
<tr>
<td>THRESHOLD 0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>789</td>
<td>1111</td>
<td>1347</td>
<td>309</td>
<td>215</td>
<td>9009</td>
</tr>
<tr>
<td></td>
<td>21.04</td>
<td>20.39</td>
<td>19.52</td>
<td>9.09</td>
<td>8.6</td>
<td>40.95</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>42</td>
<td>53</td>
<td>22</td>
<td>23</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.51</td>
<td>0.46</td>
<td>0.54</td>
<td>0.59</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.62</td>
<td>0.62</td>
<td>0.61</td>
<td>0.6</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>6</td>
<td>14</td>
<td>8</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>102</td>
<td>125</td>
<td>56</td>
<td>27</td>
<td>336</td>
</tr>
<tr>
<td>THRESHOLD 0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>523</td>
<td>579</td>
<td>710</td>
<td>189</td>
<td>145</td>
<td>4737</td>
</tr>
<tr>
<td></td>
<td>13.95</td>
<td>10.62</td>
<td>10.29</td>
<td>5.56</td>
<td>5.8</td>
<td>21.53</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>24</td>
<td>33</td>
<td>17</td>
<td>18</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.43</td>
<td>0.38</td>
<td>0.41</td>
<td>0.51</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>0.57</td>
<td>0.56</td>
<td>0.57</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Threshold 0.60</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>304</td>
<td>274</td>
<td>352</td>
<td>79</td>
<td>97</td>
<td>2310</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>8.11</td>
<td>5.03</td>
<td>5.1</td>
<td>2.32</td>
<td>3.88</td>
<td>10.5</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>32</td>
<td>16</td>
<td>21</td>
<td>10</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td><strong>Avg Clustering (c&lt;&gt;)</strong></td>
<td>0.36</td>
<td>0.29</td>
<td>0.28</td>
<td>0.26</td>
<td>0.41</td>
<td>0.34</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.494</td>
<td>0.52</td>
<td>0.5</td>
<td>0.52</td>
<td>0.52</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>24</td>
<td>15</td>
<td>43</td>
<td>25</td>
<td>18</td>
<td>88</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>48</td>
<td>94</td>
<td>92</td>
<td>40</td>
<td>22</td>
<td>317</td>
</tr>
<tr>
<td><strong>Threshold 0.55</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>203</td>
<td>142</td>
<td>195</td>
<td>51</td>
<td>55</td>
<td>1372</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>5.41</td>
<td>2.61</td>
<td>2.83</td>
<td>1.5</td>
<td>2.2</td>
<td>6.24</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>19</td>
<td>11</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td><strong>Avg Clustering (c&lt;&gt;)</strong></td>
<td>0.33</td>
<td>0.2</td>
<td>0.2</td>
<td>0.16</td>
<td>0.3</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.45</td>
<td>0.46</td>
<td>0.45</td>
<td>0.46</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>27</td>
<td>36</td>
<td>64</td>
<td>37</td>
<td>25</td>
<td>142</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>47</td>
<td>54</td>
<td>73</td>
<td>16</td>
<td>17</td>
<td>261</td>
</tr>
<tr>
<td><strong>Threshold 0.50</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>181</td>
<td>112</td>
<td>174</td>
<td>42</td>
<td>33</td>
<td>1199</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>4.83</td>
<td>2.06</td>
<td>2.52</td>
<td>1.24</td>
<td>1.32</td>
<td>5.45</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>18</td>
<td>11</td>
<td>17</td>
<td>9</td>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td><strong>Avg Clustering (c&lt;&gt;)</strong></td>
<td>0.31</td>
<td>0.17</td>
<td>0.18</td>
<td>0.1</td>
<td>0.27</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.43</td>
<td>0.45</td>
<td>0.44</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>29</td>
<td>47</td>
<td>74</td>
<td>43</td>
<td>34</td>
<td>195</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>46</td>
<td>51</td>
<td>61</td>
<td>15</td>
<td>10</td>
<td>207</td>
</tr>
<tr>
<td><strong>Threshold 0.45</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>80</td>
<td>42</td>
<td>77</td>
<td>12</td>
<td>5</td>
<td>433</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>2.13</td>
<td>0.77</td>
<td>1.12</td>
<td>0.35</td>
<td>0.2</td>
<td>1.97</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>12</td>
<td>6</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td><strong>Avg Clustering (c&lt;&gt;)</strong></td>
<td>0.23</td>
<td>0.07</td>
<td>0.09</td>
<td>0.03</td>
<td>0</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.34</td>
<td>0.26</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>45</td>
<td>74</td>
<td>95</td>
<td>57</td>
<td>45</td>
<td>281</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>30</td>
<td>24</td>
<td>35</td>
<td>5</td>
<td>3</td>
<td>128</td>
</tr>
<tr>
<td><strong>Threshold 0.40</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>58</td>
<td>31</td>
<td>47</td>
<td>10</td>
<td>5</td>
<td>303</td>
</tr>
<tr>
<td><strong>Avg Degree (&lt;k&gt;)</strong></td>
<td>1.55</td>
<td>0.57</td>
<td>0.68</td>
<td>0.29</td>
<td>0.2</td>
<td>1.38</td>
</tr>
<tr>
<td><strong>Max Degree (kmax)</strong></td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td><strong>Avg Clustering (c&lt;&gt;)</strong></td>
<td>0.22</td>
<td>0.05</td>
<td>0.08</td>
<td>0.04</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Avg Distance (&lt;d&gt;)</strong></td>
<td>0.33</td>
<td>0.34</td>
<td>0.32</td>
<td>0.32</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Connectivity (components)</strong></td>
<td>47</td>
<td>83</td>
<td>108</td>
<td>59</td>
<td>45</td>
<td>311</td>
</tr>
<tr>
<td><strong>Connectivity (largest)</strong></td>
<td>28</td>
<td>18</td>
<td>31</td>
<td>3</td>
<td>3</td>
<td>108</td>
</tr>
</tbody>
</table>
Figure S10. Set of network figures for the Lower Campanian. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WIS, where the network is densely interconnected without substructure. Only one major component in WI. All other interconnections not included in this major component have less than 2 nodes each. No configuration seemed to make any subprovinces more visible, so the Minimum Spanning Tree configuration were used to plot the network figures. GCP region is not well connected at or below the percolation threshold. High betweenness centrality values in WIS. No obvious connection between low generic richness and weak linkages.
Figure S11. Set of network figures for the Lower Campanian with nodes placed based on geographic coordinates. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WIS, where the network is densely interconnected without substructure. Only one major component in WI with weak connections between WI and GCP nodes. GCP region is not well connected at or below the percolation threshold. Northern sites were also poorly connected and become less so with lower thresholds. High betweenness centrality values in WIS and strong links between southeast, central-east, and southwest nodes of the WI.
Figure S12. Set of network figures for the Middle Campanian. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WIS, where nodes interconnected without substructure. Only one major component in WIS and all other components have less than 2 nodes each. No configuration seemed to make any subprovinces more visible, so used Minimum Spanning Tree configuration to plot figures. High betweenness centrality values in the WIS. At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region. No obvious connection between low generic richness and weak linkages. In fact, strongest links appear to be between those sites with less than 5 genera.
Figure S13. Set of network figures for the Middle Campanian with nodes placed based on geographic coordinates. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WIS, where nodes interconnected without substructure. Only one major component in WIS and all other components have less than 2 nodes each. GCP region is not well connected at all. At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region, particularly in the south, south-western, and eastern nodes.
Figure S14. Set of network figures for the Upper Campanian. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WI, where nodes are interconnected without substructure. Only one major component in WIS and any others have less than 3 nodes each. No configuration seemed to make any subprovinces more visible, so used Minimum Spanning Tree configuration to plot figures. At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region. No obvious connection between low generic richness and weak linkages. In fact, strongest links appear to be between those sites with less than 10 genera.
Figure S15. Set of network figures for the Upper Campanian with nodes placed based on geographic coordinates. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. No obvious structure except in WI, where nodes are interconnected without substructure. GCP region is not well connected at all (only one weak connection between the WIS and GCP nodes). At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region, particularly between and northwest to eastern nodes.
**Figure S16.** Set of network figures for the Lower Maastrichtian. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. Two major components in the WIS and GCP regions. One weak connection between GCP site (isolated otherwise) and WIS. No configuration seemed to make any subprovinces more visible, so used Minimum Spanning Tree configuration to plot figures. GCP region on the east side maintains its connections throughout thresholds. High betweenness centrality values in WIS, mostly at lower thresholds, not at percolation point. At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region. No obvious connection between low generic richness and weak linkages. In fact, strongest links appear to be between those sites with less than 10 genera.
Figure S17. Set of network figures for the Lower Maastrichtian with nodes placed based on geographic coordinates. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. One weak connection between GCP site and WIS below the percolation point. GCP region on the east side maintains its connections throughout thresholds. High betweenness centrality values in WIS. At lower thresholds, linkages break down somewhat in WIS but still strong connections across whole region, though they are strongest between the southern, central-east, and central nodes.
**Figure S18.** Set of network figures for the Upper Maastrichtian. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. Two major components, the WIS and GCP regions, connected by one weak connection between the WIS and GCP that is gone by 0.80 threshold. No configuration seemed to make any subprovinces more visible in WI, so used Minimum Spanning Tree configuration to plot figures. GCP region has strong connections across region. High betweenness centrality values in GCP, mostly at lower thresholds, not at percolation point. At very low thresholds (0.5), linkages break down somewhat in GCP and WIS but still strong connections across both regions, particularly in eastern side of GCP. No obvious connection between low generic richness and weak linkages. In fact, some of strongest links appear to be between those sites with less than 5 genera in WIS.
Figure S19. Set of network figures for the Upper Maastrichtian with nodes placed based on geographic coordinates. Node sizes indicate betweenness centrality, color denotes the number of unique genus occurrences present at the site, and link weights between nodes denoted by line color and thickness. Two major components, the WIS and GCP regions, connected by one weak connection between the WIS and GCP that is gone by 0.80 threshold. GCP region has strong connections across region. High betweenness centrality values in GCP, mostly at lower thresholds, not at percolation point. At very low thresholds (0.50), linkages break down somewhat in GCP and WIS but still strong connections within both regions, particularly in eastern side of GCP.
Figure S20. Set of network figures for the entire database. Node sizes indicate betweenness centrality, color denotes node substage, and link weights between nodes denoted by line color and thickness. Only 2 major components, not divided otherwise, one is the WIS region and the other is GCP region (mostly MAA). The WIS and GCP regions connected mainly by Upper Campanian sites. Same age sites tend to be more closely connected to each other than different age sites on Minimum Spanning Tree. Topology maintained throughout thresholds.
Network Randomization Comparisons

Clustering coefficient values, which describe the ratio of the number existing connections between a node and its neighbors to the maximum possible such connections, were calculated and then averaged over the network as a whole (Saramäki et al. 2007). Clustering coefficients for networks created at the percolation point were compared with a null model of randomized networks that contained the same number of nodes and links to determine if the network was more or less clustered than would be expected from a random distribution (Kiel, 2016; Table S4). Average clustering coefficient values that are distinct from random indicate substructure in the network, suggesting that the network represents aspects of the real-world system, such as community groups or faunal provinces, that differ in assemblage along distinct network components (Newman 2012).

Each clustering coefficient was compared with a null model of randomized networks that contain the same number of nodes and links to determine if the network is more or less clustered than would be expected at random (Kiel 2016). Networks were randomized in R using the igraph and tnet packages (Csárdi & Nepusz, 2006; Opsahl, 2009; see Appendix C for R codes). Randomization for each substage network were run 1000 times, rewiring the same numbers of nodes and links while maintaining link weight distributions. Average clustering coefficient distributions for the randomized networks were then compared with the clustering coefficient of the original network (Table S4).

Results indicate that all network clustering coefficients are greater or less than three standard deviations from the mean, and therefore highly distinct from random.
Table S4. Results of randomization analysis to compare clustering coefficients. 1000 randomized networks created for each network using the same numbers of nodes and links (with maintained link weights) rewired.

<table>
<thead>
<tr>
<th>Networks</th>
<th>Mean Randomized &lt;CC&gt;</th>
<th>1st S.D. Rand. &lt;CC&gt;</th>
<th>2nd S.D. Rand. &lt;CC&gt;</th>
<th>3rd S.D. Rand. &lt;CC&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Campanian</td>
<td>0.531</td>
<td>0.395</td>
<td>0.404</td>
<td>0.414</td>
</tr>
<tr>
<td>Middle Campanian</td>
<td>0.307</td>
<td>0.065</td>
<td>0.079</td>
<td>0.093</td>
</tr>
<tr>
<td>Upper Campanian</td>
<td>0.440</td>
<td>0.171</td>
<td>0.189</td>
<td>0.206</td>
</tr>
<tr>
<td>Lower Maastrichtian</td>
<td>0.660</td>
<td>0.498</td>
<td>0.503</td>
<td>0.507</td>
</tr>
<tr>
<td>Upper Maastrichtian</td>
<td>0.923</td>
<td>0.595</td>
<td>0.601</td>
<td>0.606</td>
</tr>
<tr>
<td>Complete Database</td>
<td>0.569</td>
<td>0.368</td>
<td>0.369</td>
<td>0.370</td>
</tr>
</tbody>
</table>
Complete 360km² Grid Cell Aggregation Network

We also spatially aggregated the data to a 360km grid cell size, both to check for sensitivity and to allow for easier spatiotemporal visualization of network connections. This analysis was performed using the same methods as the 60km²-aggregation using instead a 360km² grid index layer created in ArcGIS Pro using the Grid Index Tool (ESRI, 2021). The 360km aggregation produced lower resolution networks that compared fewer nodes. Furthermore, networks were only created at the percolation point for the composite database as a whole rather than for individual substages and at different thresholds to save time on computation (and following the consistency of network topology patterns found in using the 60km resolution nodes, above).

Results indicate that the patterns observed at the higher resolution 60km-grid-aggregation were maintained at this courser aggregation. The WIS and GCP components were still distinct from one another in the Maastrichtian. One advantage of coarser aggregation is that it allowed for easier visual comparisons across substages when the nodes were mapped geographically (Figure S21). The WIS component is well connected across all substages excluding the Late Maastrichtian, and connections are common between non-consecutive substages from the Middle Campanian through the Early Maastrichtian, though node links are most common between consecutive substages.

These patterns are consistent with patterns results from the 60km-grid aggregation, and furthermore suggest that the WIS was a distinct faunal province that persisted for a significant period of time. During the Late Maastrichtian, the WIS component was still well interconnected and distinct from the GCP component, but it does not share strong connections with the earlier WIS components (i.e., the three Campanian substages or the Lower Maastrichtian). This suggests turnover of fauna in the region while maintaining internal spatial homogeneity within the Late Maastrichtian itself. These patterns support interpretations discussed in the main text, that sea levels fell towards the end of the interval, potentially altering climate and circulation patterns, and altering faunal associations in the Late Maastrichtian.
Figure S21. Network maps of the composite network created using a 360km² grid aggregation. Note the well-connected WIS component throughout all substages excluding the Upper Maastrichtian.
**Detailed Subcomponent Analysis**

To determine if much smaller but spatially consistent subcomponents (potential communities at lower spatial scale than “subprovinces”) existed in any of the substage networks, identified subcomponents containing more than four interconnected nodes were analyzed at thresholds below the percolation point. A limit of at least four nodes was chosen arbitrarily for this analysis; few subcomponents were observed in any network with fewer than 4 connected nodes. These subcomponents were mapped using the ggplot2 package (Wickham 2016) in R to determine if they represent consistent geographic subprovinces of highly connected nodes. R code for this analysis can be found in Appendix C. Only the Middle Campanian and the Late Maastrichtian were found to contain subcomponents with greater than four interconnected nodes, at threshold levels below the percolation points. Figure S22 and S23 display the major and subcomponents from these networks plotted on a U.S. map and coded by color.

Manual subcomponent analysis did not find any subcomponents in the WIS that were geographically or temporally stable. In the Middle Campanian, the secondary component observed is not consistent through all thresholds, instead only representing the same nodes between the 0.45 and 0.40 thresholds. The third component observed at the 0.45 threshold is furthermore not observed at any other threshold analyzed here. **These results indicate that these components do not represent consistent network features, but are instead lesser faunal associations that shift with changing thresholds.** It is possible, however, that the secondary component observed between the 0.45 and 0.40 thresholds in the WIS represents a significant faunal association, perhaps related to previously described biogeographic features like Kauffman (1984)’s “endemic center” or Central Interior Province. The geographic position of this subcomponent does correspond roughly with both units based on Kauffman’s (1984) map (Fig. 7, p. 286). Further work may be able to elucidate if the faunal assemblages in this region are indeed related to a distinct endemic center or a weak subprovince in the WIS but cannot be conclusively determined here. In the Lower Maastrichtian, the secondary
component observed in the 0.75 and 0.60 thresholds indicates that the GCP region, along its eastern margin, is characterized by a consistent subcomponent observed in the overall network topology. However, this component breaks down at the 0.50 threshold. Thus, no other subcomponents are observed in the network that could correspond with subprovinces.
Figure S22. Maps of major and minor components (containing >3 interconnected nodes) from network analysis of the Middle Campanian for thresholds 0.40, 0.45, and 0.50. All other node components containing less than 4 nodes were not mapped. The major component for all thresholds covers the majority of the WIS region while the secondary component occupies the middle WI at all threshold levels. However, the secondary component observed at the 0.50 does not consist of the same nodes as those of the 0.45 and 0.40 thresholds. This indicates that a consistent subcomponent only emerges at the 0.45 threshold while the secondary component observed in the 0.50 threshold breaks down into smaller pieces. The third subcomponent is only observed at the 0.45 threshold, indicating that it is a portion of the major component that becomes distinct at the 0.45 threshold level before breaking down into smaller pieces at lower thresholds.
Figure S23. Maps of major and minor components (containing >3 nodes) from network analysis of the Lower Maastrichtian for thresholds 0.50, 0.60, and 0.75. All other node components containing less than 4 nodes were not mapped. The major component for all thresholds covers the majority of the WIS region. The secondary component is located in the eastern GCP from the 0.75 to 0.60 thresholds, but shifts to the WIS region at the 0.50 threshold level as the GCP region subcomponent breaks down into ones with less than 4 interconnected nodes. This indicates that the secondary component is fairly consistent and represents the eastern GCP region for most threshold levels, but no other consistent subcomponents exist within the network.
Average Betweenness Centrality:

Within each network, the betweenness centrality (BC) metric was measured for each node, describing the importance of that node in connecting other nodes through a shortest path (Kivelä et al. 2015). Betweenness centrality of a node is calculated as

\[
BC(i) = \sum_{s\neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}
\]

Where \(\sigma_{st}\) is the number of shortest paths connecting nodes s and t and \(\sigma_{st}(i)\) is the number of those paths which pass through node i. This metric indicates the degree to which a node/region/province acts as potential geographic or phylogenetic connection between otherwise dissimilar regions/provinces. Nodes were binned into 5-degree paleolatitudinal bins to compare average BC (BCave) across latitudinal space.

It should be noted that, given that the GCP does not contain as many latitudinal bins as the WIS (only existing in the study area between approximately 25-37°N), latitudinal binning comparisons across the GCP are inappropriate at this scale of analysis. Further, there is some latitudinal overlap between the southernmost WIS and the northernmost GCP, around 35°N, due to the Mississippi Embayment which may influence these results. Since we found that generic richness was correlated strongly with sampling effort (Figure S4) we assessed its impact on network communication by comparing BCave values calculated using generic richness bins (Figure S25). Generic richness values did not have a strong relationship with BCave values, indicating that generic richness and, by that proxy, sampling effort did not strongly control BCave. Had sampling effort and generic richness correlated strongly with BCave, it would be indicative of sampling bias.

Betweenness centrality varied between substages and at different threshold levels for each substage (Figure S24). The values of BCave were consistently higher at the percolation threshold for all substages than at all thresholds below the percolation points, indicating that the number of shortest paths running through nodes was particularly high in the network at the percolation point. Higher
$B_{ave}$ values at thresholds above the percolation point are present but not common, and most substages show a decrease in $B_{ave}$ at thresholds above the percolation point (Figure S24). These results suggest that nodes within each substage network are particularly important for forming connections between other nodes at the percolation point. However, it should be noted that betweenness centrality is greatly dependent on network density, or the number of links between nodes, and therefore the level of influence a set of nodes has on network connectivity can only be directly compared within the same threshold level for a substage. This enables comparisons of BC within substage thresholds based on different latitude bins and binned numbers of unique fossil occurrences (Figure S25).
Figure S24. Average betweenness centrality values for the individual substage networks at different threshold levels at and below their percolation points. The highest value of BC was recorded at the percolation point for the Middle Campanian (0.57), at 156.6. This value suggests that the network was highly structured. All other substages have values less than 40, and all substages show a general decrease in BC values at thresholds below their percolation points excluding the Lower Maastrichtian. The Lower Maastrichtian has a relatively stable BC value between its percolation point (0.80) and 0.60, with some variation, indicating that structure remains relatively high in these networks up until the 0.55 threshold level. Line color and symbol indicates substage, dashed lines are interpolated across several threshold levels.
Figure S25 shows the results of BC_{ave} for each substage at different thresholds based on latitudinal bins and sampling count bins. Average BC values in the Campanian substages are highest in the mid- to northern latitudes (40-60°N; Figure 5), consistent with higher sampling in the WIS vs. GCP during these intervals. However, at the percolation point for each substage, the latitudinal bin with the highest BC_{ave} shifts south through time. In the Lower and Middle Campanian, the highest BC_{ave} is found in the 50-55°N and 55-60°N bins, respectively, but in the Upper Campanian the highest BC_{ave} value was observed in the 40-45°N bin. This pattern is consistent but dampened at other thresholds, except in the Upper Campanian where the pattern is amplified at thresholds just above the percolation point (Figure S24). The highest average BC value for the Lower Maastrichtian substage is in the 35-40°N latitude bin; this shifts in the Upper Maastrichtian to highest average BC values in the 45–50°N bin with a second highest value in the 35–40°N bin. The fact that the second-highest BC_{ave} value in the Lower Maastrichtian exists in the 35-40°N bin is probably a product of increased sampling effort for that latitude, and suggests network patterns are somewhat influenced by sampling distributions in geographic space. Overall, patterns observed at the percolation point for each substage were paralleled but dampened at other thresholds, except for the Upper Campanian and both Maastrichtian substages where these trends were amplified but not substantially distinct. These patterns indicate that network topology is consistent at different thresholds.

Average BC binned by generic richness (i.e., the number of unique genera per 60-km grid cell) (vs. generic richness) do not indicate that sampling bias impacts the ability for nodes to act as communicating links between other areas. If greater sampling bias and therefore higher generic richness greatly influenced network communication, we would expect there to be a positive relationship between BC_{ave} and the number of unique genera. However, nodes with lower richness appear to have higher general BC_{ave} than nodes with greater richness for most substages and their thresholds, though this is not a universal trend and specifics depend on threshold level as well as substage. At thresholds below the percolation point, peaks in BC_{ave} for richness bins are diminished,
excluding the Lower Maastrichtian substage. **Average BC of generic richness bins therefore suggests no strong correlation exists between the number of unique taxa and a node’s ability to communicate information within the network (Figure S25).**

Overall, these results indicate that lower numbers of unique taxa do not diminish a node’s ability for be positioned along the shortest connecting paths between other nodes, communicating faunal information succinctly, and may in fact result in increasing the likelihood for these positions. Linkage patterns within individual substage Minimum Spanning Tree (MST) furthermore indicate that, while associations the number of unique genera in a node are non-random, they are not dominated by links between nodes of the same bin of unique genera counts (see more on MST below). This suggests that the number of unique genera within a grid cell has some influence on network topology, but that it is not the primary deciding factor. Nodes with fewer unique genera are probably characterized by highly common taxa, making them important for connecting nodes with more distinct and rarer genera. This interpretation corresponds well with the results of MST assessments, since nodes with a lower number of unique genera appear to form stronger links than those with greater number of unique genera. This may indicate that the nodes with fewer unique taxa that are common form strong links to one another and form links with nodes containing a more unique genera, while nodes with more and rarer taxa differ more, forming weaker links. **Thus, sampling bias was not found to strongly impact network connections.**
Figure S25. Average BC values for the different substages binned by latitude and by the number of fossil occurrences aggregated into each grid cell or node.
Average Link Weights through Time and by Latitude:

A comparison of average/median network link weights with latitude revealed significant variability within each substage and no clear trend across substages (Figure S26, A and B; Table S5), suggesting that faunal similarity does not vary consistently with latitude through time or space. In other words, network similarity neither increases or decreases across latitude in any interval. Generally, the lowest average similarity was observed in the 35-40°N bin. Highest similarity shifts north from the 40-45°N bins in the Lower and Middle Campanian to 45-50°N in the Upper Campanian; similarly, the Lower Maastrichtian showed the highest average similarity at 30-35°N, shifting to 45-50°N in the Upper Maastrichtian. Faunal similarity is overall greater between nodes within a latitudinal bin and those nodes without and higher between nodes in the WIS bins (40-50°N) to those without than the GCP (Figure S26, C and D).
Figure S26. Plots of average and median weights (Bray–Curtis dissimilarity index) within bins of latitude.
Distance Comparisons:

The relationship between network links and geographic distance was analyzed to determine how distance impacts faunal similarity across the study region. Under the null hypothesis it is expected that, as distance increases, link dissimilarity values or weights will increase as well given an assumption that faunal similarity is primarily spatially controlled. Distance values were calculated for the completely connected networks (Threshold = 0.99) for each substage in order to assess all faunal links based on paleocoordinates using the geodesic distance calculation from the geosphere package in R (Hijmans, 2021; code in Appendix C). Link weights between different nodes were binned by 50- to 500-kms and used to calculate average and median weight values. Networks were also analyzed by binning link values into 5-degree paleolatitudinal bins as mentioned previously to compare link weight change across latitude.

Comparisons between geographic distance and faunal similarity determined that the similarity of nodes is greater in general when the two nodes are closer together, but weakens significantly at ~1000 km distance. At very great distances (i.e., >2000 km), the correlation similarity loses power in most substages (Figure S27). Results show little increase in weight with distance until the 1000-1500 km bin at which weights increase sharply. Average network link weight did not consistently continue to increase beyond the 2000 – 2500 km bin range. Overlap in confidence intervals indicate that most substages do not have significantly different similarity values. However, mean similarity notably decreases from the Lower to the Middle Campanian (Figure S27 A). Median values showed similar trends, with generally higher average values in each bin.

To test for differences in similarity across distance within the different basins, the WIS and GCP components identified by the networks were also analyzed individually, following the same procedure as use for overall basin link weight averages (see section Average Dissimilarity by Network-Identified Components below). Patterns observed in the collective database were largely maintained in the WIS and GCP analyzed individually, but distances greater than ~2000 km were not
frequently observed (Figure S29). In the WIS, only the Lower Maastrichtian is distinctly less similar than other substages across most distance bins (Figure S29 A). In the GCP, confidence interval values in almost all distance bins overlap, suggesting that no significant difference in similarity is observed (Figure S29 C).
Figure S27. Plots of average (A) and median (B) weights (Bray–Curtis dissimilarity index) within bins of geographic distance (geodesic distance using the WGS84 ellipsoid). 95% Confidence Intervals indicated by vertical bars.
Figure S28. Plots of average (A) and median (B) WIS weights (Bray–Curtis dissimilarity index) and average (C) and median (D) GCP weights within bins of geographic distance (geodesic distance using the WGS84 ellipsoid). 95% Confidence Intervals indicated by vertical bars.
Table S5. Table of mean/median weights (Bray–Curtis dissimilarity index) binned by latitude for each substage. NA values indicate that the bin lacked occupied grid cells.

<table>
<thead>
<tr>
<th>Degree Bins</th>
<th>Low CAM</th>
<th>Mid CAM</th>
<th>Up CAM</th>
<th>Low MAA</th>
<th>Up MAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35</td>
<td>0.88/0.94</td>
<td>0.70/0.70</td>
<td>NA</td>
<td>0.51/0.51</td>
<td>NA</td>
</tr>
<tr>
<td>35-40</td>
<td>1.0/1.0</td>
<td>0.91/1.0</td>
<td>0.93/0.96</td>
<td>0.89/0.96</td>
<td>0.79/0.80</td>
</tr>
<tr>
<td>40-45</td>
<td>0.65/0.67</td>
<td>0.75/0.78</td>
<td>0.95/1.0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>45-50</td>
<td>0.70/0.70</td>
<td>0.77/0.78</td>
<td>0.75/0.76</td>
<td>0.81/0.83</td>
<td>0.47/0.50</td>
</tr>
<tr>
<td>50-55</td>
<td>0.75/0.75</td>
<td>0.77/0.79</td>
<td>0.79/0.81</td>
<td>0.76/0.78</td>
<td>0.75/0.76</td>
</tr>
<tr>
<td>55-60</td>
<td>0.85/0.85</td>
<td>NA</td>
<td>0.83/0.84</td>
<td>0.65/0.65</td>
<td>NA</td>
</tr>
<tr>
<td>60-65</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Total</td>
<td>0.78/0.80</td>
<td>0.79/0.82</td>
<td>0.82/0.84</td>
<td>0.83/0.86</td>
<td>0.85/0.95</td>
</tr>
</tbody>
</table>
Average Dissimilarity by Network-Identified Components:

The GCP region is defined as the submerged continent below the TA (~ northern TX) and extends east to overlap in latitude the southern WIS in the Mississippi Embayment, which might render latitude-based comparisons of the region highly uninformative. It is furthermore possible that the WIS and GCP provinces experienced different faunal association patterns across time which cannot be observed when the two are assessed in the same network. To test if network patterns differ between the WIS and GCP basins, we examined each provincial component individually. We separated the WIS from the GCP based on network components identified just below the percolation point for each substage (when components become disassociated from one another), which are geographically associated with the WIS and GCP in R. Faunal dissimilarity values represented by the links between nodes present in these two components were then used to calculate average faunal dissimilarity for both the WIS and GCP at each substage. WIS and GCP component values were averaged to include all links (i.e., a threshold of 100) but based on network-identified components at a level just below the percolation point for each, when the two components are disconnected completely. The GCP component was only present and therefore only calculated for in the Maastrichtian substages.

Results of average link weights for the WIS and GCP as distinct component not restricted to latitude show that WIS average faunal similarity was stable during the Campanian then decreased in the Maastrichtian while similarity increased within the GCP over the Maastrichtian (Table S6).
Table S6. Results of dissimilarity comparisons between network identified components representing the WIS and GCP in each substage. The threshold level used to identified nodes within each component are listed, as well as average dissimilarity values for links between identified nodes in the fully connected network for each substage.

<table>
<thead>
<tr>
<th>Substage</th>
<th>Threshold of ID’ed components</th>
<th>WIS Avg Dissimilarity</th>
<th>GCP Avg Dissimilarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>0.60</td>
<td>0.726</td>
<td>NA</td>
</tr>
<tr>
<td>Middle Camp.</td>
<td>0.55</td>
<td>0.771</td>
<td>NA</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>0.45</td>
<td>0.779</td>
<td>NA</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>0.75</td>
<td>0.810</td>
<td>0.919</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>0.85</td>
<td>0.882</td>
<td>0.816</td>
</tr>
</tbody>
</table>
Minimum Spanning Trees:

Beyond the previous analysis binning BCave by generic richness demonstrating that sampling bias did not control network component results (i.e., identified WIS and GCP faunal provinces), we tested for sampling bias by comparing the influence of generic richness on network topology using minimum spanning trees (MST). MST are subnetworks that connect nodes using the minimum link weights possible while preventing loops (forcing the network to produce only branching trees where branches cannot reconnect via looping), therefore representing fundamental aspects of network topology (Graham and Hell 1985; Mareš 2008). The topology of MSTs test sampling bias by calculating the degree of clustering between nodes containing similar generic richness. If sampling bias is the primary factor controlling topology, MSTs should show extremely high clustering between nodes of the same richness bin, a lower relative proportion of links between adjacent bins, and virtually no links between non-adjacent bins. Essentially, there should be a high dominance of links between nodes of the same richness level, were richness (i.e., sampling bias) the primary factor controlling how similar nodes are to one another.

Results of MST link proportion comparisons for substages (Figure S29, A-E) indicate that a high and relatively even proportion of links exist between nodes with the same or similar generic richness, and very low proportions of non-adjacent bin links (Tables S9-S13). This result suggests that while sampling bias influenced network topology, it was not the primary factor delineating network connections. No MST shows a majority of clustering between nodes of the same richness bin with a lower proportion between adjacent bins and virtually none between non-adjacent bins, excluding perhaps the Upper Maastrichtian. Therefore, sampling bias is unlikely to have strongly influenced network topology in all substages, except perhaps the Upper Maastrichtian. The MST for the complete database indicates that most substages are dominated by within-interval connections, rather than randomly or geographically sharing links with other intervals (Figure S29, F). This is expected based on an assumption of temporal autocorrelation and supports age-defined
communities as a primary control on topology. Substages are also linked with adjacent substages, rather than to non-adjacent ones, consistent with the expectation that communities will become less similar through time (Table S14).
Figure S29. Minimum spanning trees for each substage and the collective network. For substages (A-E), node color indicates the number of unique fossil genera present in each node, node size indicates betweenness centrality, and link size and color denotes the weight between two nodes (Bray–Curtis dissimilarity index). See Figures S10-S20 for legends regarding nodes and edges. Nodes for the complete database (F) are colored based on the substage they belong to. Nodes are labeled with their grid cell ID.
Table S7. List of the number of links between unique fossil occurrence count bins in the minimum spanning tree of the Lower Campanian. There are substantially more links between nodes of the same bin count and similar bin counts (adjacent bins) than between non-adjacent bin counts.

<table>
<thead>
<tr>
<th>Lower Campanian – Minimum Spanning Tree Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Link</strong></td>
</tr>
<tr>
<td>Same Bin:</td>
</tr>
<tr>
<td>Adjacent Bin</td>
</tr>
<tr>
<td>Non-Adjacent Bin</td>
</tr>
</tbody>
</table>

Table S8. List of the number of links between unique fossil occurrence count bins in the minimum spanning tree of the Middle Campanian. There are substantially more links between nodes of the same bin count and similar bin counts (adjacent bins) than between non-adjacent bin counts.

<table>
<thead>
<tr>
<th>Middle Campanian – Minimum Spanning Tree Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Link</strong></td>
</tr>
<tr>
<td>Same Bin:</td>
</tr>
<tr>
<td>Adjacent Bin</td>
</tr>
<tr>
<td>Non-Adjacent Bin</td>
</tr>
</tbody>
</table>

Table S9. List of the number of links between unique fossil occurrence count bins in the minimum spanning tree of the Upper Campanian. There are substantially more links between nodes of the same bin count and similar bin counts (adjacent bins) than between non-adjacent bin counts.

<table>
<thead>
<tr>
<th>Upper Campanian – Minimum Spanning Tree Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Link</strong></td>
</tr>
<tr>
<td>Same Bin:</td>
</tr>
<tr>
<td>Adjacent Bin</td>
</tr>
<tr>
<td>Non-Adjacent Bin</td>
</tr>
</tbody>
</table>

Table S10. List of the number of links between unique fossil occurrence count bins in the minimum spanning tree of the Lower Maastrichtian. There are substantially more links between nodes of the same bin count and similar bin counts (adjacent bins) than between non-adjacent bin counts.

<table>
<thead>
<tr>
<th>Lower Maastrichtian – Minimum Spanning Tree Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Link</strong></td>
</tr>
<tr>
<td>Same Bin:</td>
</tr>
<tr>
<td>Adjacent Bin</td>
</tr>
<tr>
<td>Non-Adjacent Bin</td>
</tr>
</tbody>
</table>
Table S11. List of the number of links between unique fossil occurrence count bins in the minimum spanning tree of the Upper Maastrichtian. There are substantially more links between nodes of the same bin count and similar bin counts (adjacent bins) than between non-adjacent bin counts.

<table>
<thead>
<tr>
<th>Type of Link</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same Bin:</td>
<td>31</td>
<td>63.27%</td>
</tr>
<tr>
<td>Adjacent Bin</td>
<td>7</td>
<td>14.29%</td>
</tr>
<tr>
<td>Non-Adjacent Bin</td>
<td>11</td>
<td>22.45%</td>
</tr>
</tbody>
</table>

Table S12. List of the number of links between substages in the minimum spanning tree of the complete database. There are substantially more links between nodes of the same substage, less than half as many between adjacent substages, and less than a third between non-adjacent substages.

<table>
<thead>
<tr>
<th>Type of Link</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same Substage:</td>
<td>270</td>
<td>61.36%</td>
</tr>
<tr>
<td>Adjacent Substages:</td>
<td>109</td>
<td>24.77%</td>
</tr>
<tr>
<td>Non-Adjacent Substages:</td>
<td>61</td>
<td>13.86%</td>
</tr>
</tbody>
</table>
Extended Biostratigraphic Chart (Extension of Figure 1):

Figure S30. Campanian to Maastrichtian biostratigraphic chart showing polarity chon intervals (Malinverno et al. 2020), biozones (Ogg et al. 2016; Lynds and Slattery 2017), sea level fluctuations (Haq 2014), global stable carbon and oxygen isotope trends (Gale et al. 2020), and stable oxygen and carbon isotope data collated for the Western Interior region (Dwyer, 2019). Ages marked with an asterisk have been added or updated based on Malinverno et al. (2020) from Slattery et al. (2013). WI isotope values modified from unpublished data by Dwyer (2019) which span the R9 regressive interval of the Bearpaw Cyclothem (D. cheyennenses – B. baculus ammonite zones).
SI References:


Dwyer, C.H., 2019, Constraining the oxygen values of the Late Cretaceous Western Interior Seaway using marine bivalves [Thesis]: University of New Mexico, 1–92 p.


Kocsis, Á.T., and Raja, N.B., 2020, chronosphere: Earth system history variables:


Appendix A-2. Taxonomic References List (See Table S13)

The following list of references were used to vet and update taxonomic nomenclature for the fossil data included in Table S1 (excel file). Table S15 lists the original names provided by the primary databases (i.e., iDigBio, PBDB, etc.), the updated name, if one is necessary, the reference used to check the name, and any additional notes. Taxa were removed if they were not marine invertebrates, identified down to at least the genus level, or if they included “c.f.”, “aff.”, or “?” following the genus name. Most references were found using basic google searches, but others were researched in specific available volumes (i.e., Akers and Akers, 2002a).

Taxonomic References


Aberhan, M., 2010, Jurassic bivalves: Taxonomic opinions:


Alstine, J.B. van, 1974, Paleontology of brackish-water faunas in two tongues of the Cannonball Formation (Paleocene, Danian), Slope and Golden Valley counties, southwestern North Dakota [Master’s Thesis]: University of North Dakota.


Boyle, C.B., 1893, Catalogue and Bibliography of North American Mesozoic Invertebrata.:

Brown, L.G., and Neville, B.D., 2015, Catalog of the recent taxa of the families Epitoniidae and Nystiellidae (Mollusca: Gastropoda) with a bibliography of the descriptive and systematic literature : Zootaxa, v. 3907.


Clark, W.B., and Twitchell, M.W., 1915, The Mesozoic and Cenozoic Echinodermata of the United States:


191


Cobban, W.A., 1962a, Baculites from the Lower Part of the Pierre Shale and Equivalent Rocks in the Western Interior: https://about.jstor.org/terms.


Cobban, W.A., 1977b, Characteristic Marine molluscan Fossils from the Dakota Sandstone and Intertongued Mancos Shale, West-Central New Mexico: Geological Survey Professional Paper 1009:


Cobban, W.A., 1961, The Ammonite Family Binneyitidae Reeside in the Western Interior of the United States Author:


Cooke, C.W., 1953, American Upper Cretaceous Echinoidea.


Correa, M.L., Freiwald, A., Hall-Spencer, J., and Taviani, M., 2005, Distribution and habitats of Acesta excavata (Bivalvia: Limidae) with new data on its shell ultrastructure, in Freiwald A. and

Coryell, H.N., and Salmon, E.S., 1934, A molluscan faunule from the Pierre Formation in Eastern Montana.: 


Dockery III, D.T., 1993, The Streptoneuran gastropods, exclusive of the Stenoglossa, of the Coffee Sand (Campanian) of Northeastern Mississippi.: 


Dunker, W., 1845, Diagnoses Molluscorum quorundam novorum, quae ex itinere ad oras Africae occidentalis reportavit cl. G. Tams, Med. Dr.: Zeitschrift für Malakozoologie.,


Elias, M.K., 1933, Cephalopods of Pierre Formation of Wallace County, Kansas, and Adjacent Area: The University of Kansas Science Bulletin, v. XXI.


Harbison, A., 1945, Upper Cretaceous Mollusks of the Lower Ripley Formation near Dumas, Mississippi:.


Jäger, M., and Jäger, M., 2012, Sabellids and serpulids (Polychaeta sedentaria) from the type Maastrichtian, the Netherlands and Belgium.


Kafanov, A.I., and Ogasawara, K., 2005, Cenozoic molluscan (Bivavia) cenozone of Hokkaido, Northern Japan.:  


Kosnik, M.A., 2002, personal opinion:


Marcou, J.B., 1885, A list of the Mesozoic and Cenozoic types in the Collections of the U.S. National Museum:


Meek, F.B., 1873, Preliminary paleontological report, consisting of lists and descriptions of fossils, with remarks on the ages of the rocks in which they were found, etc.: U.S. Geological Survey of the Territories (Hayden), Sixth Annual Report, p. 429–518.

Meek, F.B., and Hayden, F. v, 1857, Description of new Species and Genera of Fossils, collected by Dr. F. V. Hayden in Nebraska Territory, under the direction of Lieut. G. K. Warren, U.S. Topographical Engineer; with some remarks on the Tertiary and Cretaceous formations of the north-west, and the parallelism of the latter with those of other portions of the United States and Territories: Proceedings of the Academy of Natural Sciences of Philadelphia, v. 9, p. 117–148.

Meek, F.B., and Hayden, F. v, 1861, Descriptions of new Lower Silurian, (Primordial), Jurassic, Cretaceous, and Tertiary fossils, collected in Nebraska, by the exploring expedition under the command of Capt. Wm. F. Reynolds, U.S. Top. Engineers with some remarks on the rocks from which they were obtained: Academy of Natural Sciences of Philadelphia Proceedings, v. 13, p. 415–447.


Mitchell, S.F., 2010b, Revision of three large species of Barrettia from Jamaica Taxonomy of Rudist Bivalves (Bivalvia: Hippuritida) View project Geology and geological evolution of Jamaica View project Revision of three large species of Barrettia from Jamaica: https://www.researchgate.net/publication/233924620.


Morton, S.G., 1834, Synopsis of the organic remains of the Cretaceous group of the United States: Illustrations by nineteen plates to which is added an appendix, containing a tabular view of the Tertiary fossils hitherto discovered in North America: Philadelphia, Key and Biddle, 1–88 p.


Powell, J.D., 1963, Turonian (Cretaceous) Ammonites from Northeastern Chihuahua, Mexico.


Rathbun, M.J., 1935, Fossil Crustacea of the Atlantic and Gulf Coastal Plain: Geological Society of America Special Papers,.


Sari, B., Kandemir, R., Özer, S., Walaszczyk, I., Görmüş, M., Demircan, H., and Yılmaz, C., 2014, Upper Campanian calciclastic turbidite sequences from the Hacimehmet area (eastern Pontides, NE


Schweizerbart, E., 1894, New Yearbook of Mineralogy, Geology and Paleontology, Part 2:


Sealey, P.L., and Lucas, S.G., 2011, Upper Cretaceous (Turonian) ammonites from the Carlile Member and Reference are of the Juana Lopez Member of the Mancos Shale, Eastern side of the San Juan Basin, La Ventana, Sandoval County, New Mexico: New Mexico Museum of Natural History and Science, Bulletin, v. 53, p. 370-379.


Sidwell, R., 1932, New Species from the Colorado Group, Cretaceous, in South Central Wyoming.: 


Sohl, N.F., and Koch, C.F., 1984, Upper Cretaceous (Maestrichtian) Larger Invertebrate Fossils form the Haustator bilara Assemblage Zone in the West Gulf Coastal Plain: Open File Report 84-687.: 


Sowerby, J. de C., 1829, The mineral conchology of Great Britain; or coloured figures and descriptions of those remains of testaceous animals or shells, which have been preserved at various times and depths in the Earth: London, privately published, v. VI.


Stephenson, L.W., 1954, Additions to the Fauna of the Raritan Formation (Cenomanian) of New Jersey: Geological Survey Professional Paper 264-B.: 


Stephenson, L.W., 1941, The larger invertebrate fossils of the Navarro Group of Texas: University of Texas Publication number 4101, p. 1–641.


Twenhofel, W., 1924, State Geological Survey of Kansas Bulletin 9: The Geology and Invertebrate Paleontology of the Comanchean and “Dakota” Formations of Kansas:


Walaszczyk, I., 2004, Inoceramids and inoceramid biostratigraphy of the Upper Campanian to basal Maastrichtian of the Middle Vistula River section, central Poland.


White, C.A., 1884, A Review of the Fossil Ostreidae of North America; and a comparison of the fossil with the living forms:


White, C.A., 1889, Bulletin of the United States Geological Survey No. 51: On Invertebrate Fossils from the Pacific Coast:


Wick, S.L., 2021, Paleontological inventory of Paleozoic, Late Mesozoic, and Cenozoic plant, invertebrate, and vertebrate fossil species from Big Bend National Park, Texas, USA - over a century of paleontological discovery: Zitteliana, v. 95, p. 95–134, doi:10.3897/zitteliana.95.73026.


Appendix A-3. R Code used for Analysis

The following scripts were used in R to perform analyses.

*Code for Comparing Generic Richness and Occurrence Number (Sampling Bias)*

```r
library(ggplot2)
library(ggmap)
library(tidyr)
library(sf)
library(mapview)
library(raster)
library(rgdal)
library(dismo)
library(XML)
library(maps)
library(mFD)
library(reshape2)
library(vegan)
library(dplyr)
library(tidyr)

setwd("C:/Users/ceara/Documents/Province Project/Occurrence Data/VettedCombinedDB")

# read in the data table (includes the name of grid cell based on 60km grid)
data <- read.csv("R_Config_for_EN_OUTPUT/Vetted_Substage_Joined_60.csv", header=TRUE)

#### RECONFIGURE ENTIRE DB (same basic code as used to configure into EDENETWORKS format) ####

colnames(data)

# simplify data to simple dataframes that can be used in EDENETWORKS analyses
simplify_data <- function(data) {
  data1 <- data.frame(data)
  data2 <- data1[,c(1:2,8:9,19,25:29)]
  return(data2)
}

trunk_data <- simplify_data(data)
ncol(trunk_data)
colnames(trunk_data)
nrow(trunk_data)

# create column collating the grid cell and the age
trunk_data$Abrev_Age <- with(trunk_data, ifelse(Substage_from_Zone.Mbr == "CAM (low)",
  'LC', ifelse(Substage_from_Zone.Mbr == "CAM (mid)",
    'MC', ifelse(Substage_from_Zone.Mbr == "CAM (up)",
      'UC', ifelse(Substage_from_Zone.Mbr == "MAA (low)",
        'LM', 'UM'))))

head(trunk_data)
```

head(trunk_data)
nrow(trunk_data)

# replace the grid cell name column with the grid cell name and abbreviated age collated
trunk_data$PageName <- paste(trunk_data$PageName,trunk_data$Abrev_Age,sep="")
head(trunk_data)
colnames(trunk_data)
nrow(trunk_data)

# Add the faunal province information
FP_list <- read.csv("FaunalProv_AttributesList.csv")
FP_list <- FP_list[,1:3] # shorted to just the Province and Grid name columns
nrow(FP_list)
unique(FP_list$Province)

FP_list$Color <- with(FP_list, ifelse(Province == "NIP", 'mediumblue', ifelse(Province == "SIP", 'gold', ifelse(Province == "CIP", 'darkolivegreen', ifelse(Province == "GCP", 'chocolate', ifelse(Province == "EC", 'darkmagenta', 'grey'))))))

FP_list$ColorAge <- with(FP_list, ifelse(Age == "LC", 'darkblue', ifelse(Age == "MC", 'olivedrab', ifelse(Age == "UC", 'yellow', ifelse(Age == "LM", 'darkorange', 'darkred')))))

# Check for difference in the two datasets
pn <- trunk_data[,10]
fp <- FP_list[,2]
setdiff(pn, fp)

#merge the two (table join by Grid cell name)
trunk_data <- merge(trunk_data,FP_list,by="PageName")
colnames(trunk_data)
head(trunk_data)
nrow(trunk_data) #check that there are the same # occ still

# subset just the spp-level occ data into
trunk_data_spp <- subset(trunk_data, Updated_Sp != "")
head(trunk_data_spp)
nrow(trunk_data_spp)
colnames(trunk_data)

# Get the number of fossil occurrence present for each grid cell
trunk_gen_occnum <- as.data.frame(table(trunk_data$PageName))

trunk_spp_occnum <- as.data.frame(table(trunk_data_spp$PageName))

trunk_gen_occnum <- data.frame(trunk_gen_occnum[-,1], row.names = trunk_gen_occnum[,1])

224
trunk_spp_occnum <- data.frame(trunk_spp_occnum[, -1], row.names = trunk_spp_occnum[, 1])

# function to transform into an abundance matrix of locations with genus names
substg_gen_grid <- function(data){
data1 <- as.data.frame(subset(data)[, c(1, 4)]) # subset out just the genus name and location information
colnames(data1) <- c("PageName", "Genus_Name") # Give the columns names
data2 <- dcast(data1, PageName~Genus_Name, length) # transform into a pres-abs matrix based on Grid cells
}

trunk_matrix_gen <- substg_gen_grid(trunk_data)

# function to transform into an abundance matrix of locations with species names
substg_spp_grid <- function(data){
data1 <- as.data.frame(subset(data)[, c(1, 8)]) # subset out just the species name and location information
colnames(data1) <- c("PageName", "Species_Name") # Give the columns names
data2 <- dcast(data1, PageName~Species_Name, length) # transform into a pres-abs matrix based on Grid cells
}

trunk_matrix_spp <- substg_spp_grid(trunk_data_spp)

# function to configure the matrix into the correct format for EDENETWORKS and add the grid cell info
config_matrix <- function(data){
  base1 <- (merge(data, lookup, by = 'LocationKey', all.X = TRUE, all.y = FALSE)) # only use if trying to make column of latlong using unique latlong key
  x <- nrow(data)
data$SampleKey <- seq(1:x)
a <- ncol(data)
b <- a - 1
  base2 <- data[, c(1, a, 2:b)] # This part would need to change if using lat/long as unique ID
  base3 <- data.frame(base2[, -1], row.names = base2[, 1]) # make the first column with pagename the index
}

trunk_matrix_gen_final <- config_matrix(trunk_matrix_gen)
head(trunk_matrix_gen_final)
trunk_matrix_gen_final[4,]

trunk_matrix_spp_final <- config_matrix(trunk_matrix_spp)

#MATRIX OF TOTAL DB BASED ON SPP NAME (For figuring out faunal provinces... not for final analysis)
#trunk_matrix_TOTAL_final <- config_matrix(trunk_matrix_TOTAL)

pres_ab <- function(data){
  numeric_cols <- vapply(data, is.numeric, logical(1)) # make all values 1 if not 0
data[numeric_cols] <- as.integer(data[numeric_cols] != 0)
data
}
trunk_matrix_gen_final_pres <- pres_ab(trunk_matrix_gen_final)
trunk_matrix_spp_final_pres <- pres_ab(trunk_matrix_spp_final)

# get the richness per grid cell count by summing across the rows
#For genera
ncol(trunk_matrix_gen_final_pres)
numb_gen <- data.frame(rowSums(trunk_matrix_gen_final_pres[,2:575])) # sum rows to get total # gen
trunk_matrix_gen_final_pres_pre <- trunk_matrix_gen_final_pres # create new matrix name to use
trunk_matrix_gen_final_pres_pre$nGen <- numb_gen # Add column of total occ number
trunk_gen_rich <- trunk_matrix_gen_final_pres_pre[,c(576)]
colnames(trunk_gen_rich)

#For species
ncol(trunk_matrix_spp_final_pres)
numb_spp <- data.frame(rowSums(trunk_matrix_spp_final_pres[,2:1114])) # sum rows to get total # occ
trunk_matrix_spp_final_pres_pre <- trunk_matrix_spp_final # create new matrix name to use
trunk_matrix_spp_final_pres_pre$nSpp <- numb_spp # Add column of total occ number
trunk_spp_rich <- trunk_matrix_spp_final_pres_pre[,c(1115)]
colnames(trunk_spp_rich)

# Combine the tables
genus_data <- cbind(trunk_gen_occnum,trunk_gen_rich)
colnames(genus_data) <- c("Occ count","Genus Richness")
species_data <- cbind(trunk_spp_occnum,trunk_spp_rich)
colnames(species_data) <- c("Occ count","Species Richness")

# Look at the correlation and distributions of the genus-level data

pdf(file = "DB_Map_Visualize/genus_histograms.pdf")
hist(genus_data[,1],xlab='Genus Occurrence Count',main='Genus Occurrence Frequency')
hist(genus_data[,2],xlab='Genus Richness',main='Genus Richness Frequency')
dev.off()

# perform Shapiro test for normality
shapiro.test(genus_data[,1]) # very much not normal
shapiro.test(genus_data[,2]) # very much not normal

# Q-Q plots to look at normality

pdf(file = "DB_Map_Visualize/genus_QQplots.pdf")
ggqqplot(genus_data[,1], ylab = "Genus Occurrence Count")
ggqqplot(genus_data[,2], ylab = "Genus Richness")
dev.off()

# conduct spearman's rho correlation (non-parametric because the data is highly skewed)
cor(genus_data$'Occ count', genus_data$'Genus Richness', method = "spearman", use = "complete.obs")
# result: 0.9327709 >> highly correlated

cor.test(genus_data$'Occ count', genus_data$'Genus Richness', method = "spearman", exact=FALSE)
# p-value is highly significant: < 2.2e-16

## Correlation is highly significant

# Look at a correlation plot of the data
library("ggpubr")

pdf(file = "DB_Map_Visualize/genus_scatter_trend.pdf")

ggscatter(genus_data, x = "Occ count", y = "Genus Richness",
add = "reg.line", conf.int = TRUE,
cor.coef = TRUE, cor.method = "spearman",
xlab = "Number of Occurrences", ylab = "Genus Richness")

dev.off()

# create linear model for the correlation
lm.model = lm(genus_data$'Occ count' ~ genus_data$'Genus Richness', data = genus_data)

summary(lm.model)

# create a log transformed model
lm_log.model = lm(log1p(genus_data$'Occ count') ~ log1p(genus_data$'Genus Richness'), data = genus_data)

summary(lm_log.model)
str(summary(lm_log.model))
names(summary(lm_log.model))

# Results show that there is a very high correlation between sampling effort (number of fossil occurrences collected)
# and the genus richness observed.

pdf(file = "DB_Map_Visualize/genus_logtransformed_data.pdf")

# perform log10 transformation on both sets of data and graph
ggplot(data = genus_data, aes(x = genus_data$'Occ count', y = genus_data$'Genus Richness')) + geom_point() + scale_x_log10() + scale_y_log10() + xlab("Log10 transformed Number of Occurrences") + ylab("Log10 transformed Genus Richness") + ggtitle("Log Transfromed Richness vs. Occurrence Count") + geom_smooth(method=lm, level=0.99) +
annotate('text',x = 10, y = 100,label=paste('R-squared: ', summary(lm_log.model)$r.squared))

dev.off()

# Look at the correlation and distributions of the species-level data

pdf(file = "DB_Map_Visualize/species_histograms.pdf")

hist(species_data[,1],xlab='Species Occurrence Count',main='Species Occurrence Frequency')
hist(species_data[,2],xlab='Species Richness',main='Species Richness Frequency')

dev.off()

# perform Shapiro test for normality

shapiro.test(species_data[,1]) # very much not normal
shapiro.test(species_data[,2]) # very much not normal

# Q-Q plots to look at normality

pdf(file = "DB_Map_Visualize/species_QQplots.pdf")

ggqqplot(species_data[,1], ylab = "Species Occurrence Count")
ggqqplot(species_data[,2], ylab = "Species Richness")

dev.off()  

# Conduct spearman rho correlation (non-parametric because the data is highly skewed)

cor(species_data$`Occ count`, species_data$`Species Richness`, method = "spearman", use = "complete.obs")
# result: 0.9353668

cor.test(species_data$`Occ count`, species_data$`Species Richness`, method = "spearman", exact=FALSE)
# p-value is highly significant: < 2.2e-16

## Correlation is highly significant

# Look at a correlation plot of the data

library("ggpubr")

pdf(file = "DB_Map_Visualize/species_scatter_trend.pdf")

ggscatter(species_data, x = "Occ count", y = "Species Richness", add = "reg.line", conf.int = TRUE, cor.coef = TRUE, cor.method = "spearman", xlab = "Number of Occurrences", ylab = "Species Richness")

dev.off()

# Create linear model for the correlation
lm.model_spp = lm(species_data$'Occ count' ~ species_data$'Species Richness', data = species_data)
summary(lm.model_spp)

# create a log transformed model
lm_log.model_spp = lm(log1p(species_data$'Occ count') ~ log1p(species_data$'Species Richness'), data = species_data)
summary(lm_log.model_spp)

# Results show that there is a very high correlation between sampling effort (number of fossil occurrences collected)
# and the species richness observed.

pdf(file = "DB_Map_Visualize/species_logtransformed_data.pdf")

# perform log10 transformation on both sets of data and graph
ggplot(data = species_data, aes(x = species_data$'Occ count', y = species_data$'Species Richness')) +
  geom_point() +
  scale_x_log10() + scale_y_log10() +
  xlab("Log10 transformed Number of Occurrences") +
  ylab("Log10 transformed Species Richness") +
  ggtitle("Log Transformed Richness vs. Occurrence Count") +
  geom_smooth(method=lm, level=0.99) +
  annotate('text',x = 10, y = 100,label=paste('R-squared:',summary(lm_log.model_spp)$r.squared))
dev.off()
Code to configure Occurrence Database into format for EDENetworks Analysis

```r
## Load necessary packages
library(ggplot2)
library(ggmap)
library(tidyverse)
library(sf)
library(mapview)
library(raster)
library(rgdal)
library(dismo)
library(XML)
library(maps)
library(mFD)
library(reshape2)
library(vegan)
library(dplyr)
library(tidyr)
# set working directory
setwd("C:/Users/ceara/Documents/Province Project/Occurrence Data/VettedCombinedDB")

##### RECONFIGURE THE FULL DB FILE BASED ON LOCATION AND GRID INFO #####
# read in the csv file of the DB that has only the substage-level occ data
data <- read.csv('Vetted_Combined_Genus-Spp-lvl_SubstageOnly_sansACP-lowLAT.csv')
# add column with the lat long information as a unique location ID
data$LocationKey <- paste(data$Lat,"",data$Lon)
# add a column with the genus and spp names concatenated
data$SppName <- paste(data$Updated_Genus," ",data$Updated_Sp)
# add lat and long columns
data$Latitude <- paste(data$lat)
data$Longitude <- paste(data$lon)
head(data)
nrow(data)
lattr <- data[3:4]
head(lattr)
nrow(unique(lattr))
nrow(data.frame(unique(data$LocationKey)))

# Read in grids made in ArcGIS pro
grid_60km <- st_read("SpatialGrids/grid_60km.shp")
plot(grid_60km$geometry,xlim=c(-118,-
68),ylim=c(28,53),col="blue",pch=19,xlab="Longitude",ylab="Latitude")
points(data$Long,data$Lat,col="red",pch=19,xlab="Longitude",ylab="Latitude") # add occ pts

#transform database into shapefile
sf_data <- st_as_sf(data, coords = c("Long", "Lat"), crs = 4326)
```
# function to spatially join DB with grids
spatial_join <- function(shapefile_data, grid) {
  data1 <- st_join(shapefile_data, left=TRUE, grid["PageName"])
  return(data1)
}

# spatially join Grid cell Page Names to data table
join60_data <- spatial_join(sf_data, grid_60km)
join60_data <- as.data.frame(join60_data)

ncolnames(join60_data)
nrow(data.frame(unique(join60_data$Updated_Genus)))
nrow(data.frame(unique(join60_data$SppName)))

# add column name to last column for geometry
names(join60_data)[length(names(join60_data))] <- "geometry2"
colnames(join60_data)
nrow(join60_data)

# write csv file with joined info
write.csv(join60_data, file="R_Config_for_EN_OUTPUT/Vetted_Substage_Joined_60.csv")

### Get summary information about the dataset as a whole ###
class_genus <- join60_data[,c(14,8)]
class_species <- subset(join60_data, Updated_Sp != """)
class_species <- class_species[,c(14,26)]

nb_gen_class <- count(class_genus) # get the frequency of a genus with class ID
nb_spp_class <- count(class_species) # get the frequency of a species with class ID
nb_gen <- nrow(data.frame(unique(class_genus$Updated_Genus))) # get the number of unique genera
nb_spp <- nrow(data.frame(unique(class_species$SppName))) # get the number of unique species
nb_classes <- nrow(data.frame(unique(class_genus$HigherTax_Class))) # get number of occ for each unique class-level
nb_occ <- nrow(join60_data) # get number of fossil occurrences

write.csv(unique(class_genus), "R_Config_for_EN_OUTPUT/class_unique_genus.csv", row.names=FALSE)
write.csv(unique(class_species), "R_Config_for_EN_OUTPUT/class_unique_species.csv", row.names=FALSE)

### RECONFIGURE ENTIRE DB TO USE IN EDENETWORKS ###

# simplify data to simple dataframes that can be used in EDENETWORKS analyses
simplify_data <- function(data) {
  data1 <- data.frame(data)
  data2 <- data1[,c(1:2,8:9,19,25:29)]
  return(data2)
}
trunk60_data <- simplify_data(join60_data)
nrow(trunk60_data)
colnames(trunk60_data)

# create column collating the grid cell and the age
trunk60_data$Abrev_Age <- with(trunk60_data, ifelse(Substage_from_Zone.Mbr == "CAM (low)", 'LC', ifelse(Substage_from_Zone.Mbr == "CAM (mid)", 'MC', ifelse(Substage_from_Zone.Mbr == "CAM (up)", 'UC', ifelse(Substage_from_Zone.Mbr == "MAA (low)", 'LM', 'UM'))))

head(trunk60_data)
nrow(trunk60_data)

# replace the grid cell name column with the grid cell name and abbreviated age collated
trunk60_data$PageName <- paste(trunk60_data$PageName, trunk60_data$Abrev_Age, sep="")
head(trunk60_data)
colnames(trunk60_data)
nrow(trunk60_data)

# Add the faunal province information
FP_list <- read.csv("FaunalProv_AttributesList.csv")
FP_list <- FP_list[,1:3]  # shortened to just the Province and Grid name columns
nrow(FP_list)
unique(FP_list$Province)

FP_list$Color <- with(FP_list, ifelse(Province == "NIP", 'mediumblue', ifelse(Province == "SIP", 'gold', ifelse(Province == "CIP", 'darkolivegreen', ifelse(Province == "GCP", 'chocolate', ifelse(Province == "EC", 'darkmagenta', 'grey')))))))

FP_list$ColorAge <- with(FP_list, ifelse(Age == "LC", 'darkblue', ifelse(Age == "MC", 'olivedrab', ifelse(Age == "UC", 'yellow', ifelse(Age == "LM", 'darkorange', 'darkred')))))))

# Check for difference in the two datasets
pn <- trunk60_data[,10]
fp <- FP_list[,2]
setdiff(pn, fp)

# merge the two (table join by Grid cell name)
trunk60_data <- merge(trunk60_data, FP_list, by="PageName")
colnames(trunk60_data)
head(trunk60_data)
nrow(trunk60_data)  # check that there are the same # occ still

# subset just the spp-level occ data into
trunk60_data_spp <- subset(trunk60_data, Updated_Sp != "")
head(trunk60_data_spp)
nrow(trunk60_data_spp)

colnames(trunk60_data)

# function to transform into an abundance matrix of locations with genus names
substg_gen_grid <- function(data){
data1 <- as.data.frame(subset(data)[,c(1,4)]) # subset out just the genus name and location information
colnames(data1) <- c("PageName","Genus_Name") # Give the columns names
data2 <- dcast(data1, PageName~Genus_Name, length) # transform into a pres-abs matrix based on Grid cells
}

trunk60_matrix_gen <- substg_gen_grid(trunk60_data)
head(trunk60_matrix_gen)

# function to transform into an abundance matrix of locations with species names
substg_spp_grid <- function(data){
data1 <- as.data.frame(subset(data)[,c(1,8)]) # subset out just the species name and location information
colnames(data1) <- c("PageName","Species_Name") # Give the columns names
data2 <- dcast(data1, PageName~Species_Name, length) # transform into a pres-abs matrix based on Grid cells
}

trunk60_matrix_spp <- substg_spp_grid(trunk60_data_spp)
head(trunk60_matrix_spp)

#GET MATRIX OF ALL DB BASED ON SPP NAME (For figuring out faunal provinces... not for final analysis)
#trunk60_matrix_TOTAL <- substg_spp_grid(trunk60_data)

# function to configure the matrix into the correct format for EDENETWORKS and add the grid cell info
config_matrix <- function(data){
#base1 <- (merge(data, lookup, by = 'LocationKey',all.X=TRUE, all.y=FALSE)) # only use if trying to make column of latlong using unique latlong key
x <- nrow(data)
data$SampleKey <- seq(1:x)
a <- ncol(data)
b <- a - 1
base2 <- data[,c(1,a,2:b)] # This part would need to change if using lat/long as unique ID
base3 <- data.frame(base2[,1], row.names = base2[,1]) # make the first column with pagename the index
}

trunk60_matrix_gen_final <- config_matrix(trunk60_matrix_gen)

trunk60_matrix_spp_final <- config_matrix(trunk60_matrix_spp)

#MATRIX OF TOTAL DB BASED ON SPP NAME (For figuring out faunal provinces... not for final analysis)
pres_ab <- function(data){
  numeric_cols <- vapply(data, is.numeric, logical(1)) # make all values 1 if not 0
  data[numeric_cols] <- as.integer(data[numeric_cols] != 0)
  data
}

trunk60_matrix_gen_final <- pres_ab(trunk60_matrix_gen_final)
trunk60_matrix_spp_final <- pres_ab(trunk60_matrix_spp_final)

#MATRIX OF TOTAL DB BASED ON SPP NAME (For figuring out faunal provinces... not for final analysis)
#trunk60_matrix_TOTAL_final <- pres_ab(trunk60_matrix_TOTAL_final)

# write csv and txt files of the matrix for the whole dataset
write.csv(trunk60_matrix_gen_final,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_gen_final.csv")
#write.table(trunk60_matrix_gen_final,file="R_Config_for_EN_OUTPUT/trunk60_matrix_gen_final.txt",col.names = TRUE)
write.csv(trunk60_matrix_spp_final,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_spp_final.csv")

# (For figuring out faunal provinces... not for final analysis)
#write.csv(trunk60_matrix_TOTAL_final,file="R_Config_for_EN_OUTPUT/trunk60_matrix_TOTAL_final.csv")

## Create matrix with grid cells containing less than 3 occ removed

#For genera
ncol(trunk60_matrix_gen_final)
numb_occ_gen <- data.frame(rowSums(trunk60_matrix_gen_final[,2:575])) # sum rows to get total # occ
trunk60_matrix_gen_final_pre <- trunk60_matrix_gen_final # create new matrix name to use
trunk60_matrix_gen_final_pre$nOccGen <- numb_occ_gen # Add column of total occ number
colnames(trunk60_matrix_gen_final_pre)
trunk60_matrix_gen_great3 <- subset(trunk60_matrix_gen_final_pre, nOccGen >= 3) # subset out all grid
cells with greater 3 occ
ncol(trunk60_matrix_gen_great3)
trunk60_matrix_gen_great3 <- trunk60_matrix_gen_great3[,1:575] # remove occ number column
colnames(trunk60_matrix_gen_great3)

# export new csv and txt files of the dataset with grids containing <3 occ removed
write.csv(trunk60_matrix_gen_great3,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_gen_great3.csv")
# write.table(trunk60_matrix_gen_great3,file="R_Config_for_EN_OUTPUT/trunk60_matrix_gen_great3.txt",col.names = TRUE)
# for species
ncol(trunk60_matrix_spp_final)
numb_occ_spp <- as.data.frame(rowSums(trunk60_matrix_spp_final[,2:575])) # sum rows to get total # occ
trunk60_matrix_spp_final_pre <- trunk60_matrix_spp_final # create new matrix name to use
trunk60_matrix_spp_final_pre$nOccSpp <- numb_occ_spp # Add column of total occ number
colnames(trunk60_matrix_spp_final_pre)
trunk60_matrix_spp_great3 <- subset(trunk60_matrix_spp_final_pre, nOccSpp >= 3) # subset out all grid
cells with greater 3 occ
ncol(trunk60_matrix_spp_great3)
trunk60_matrix_spp_great3 <- trunk60_matrix_spp_great3[,1:1114] # remove occ number column
colnames(trunk60_matrix_spp_great3)

# export new csv and txt files of the dataset with grids containing <3 occ removed
write.csv(trunk60_matrix_spp_great3, file =
"R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_spp_great3.csv")
write.table(trunk60_matrix_gen_great3, file =
"R_Config_for_EN_OUTPUT/trunk60_matrix_gen_great3.txt", col.names = TRUE)

#### Get attribute files for EDENetwork analysis ####

# function for creating an attribute table of averaged coor for each occupied grid cell (attribute data can be
# added to this function)
coor_atrib <- function(data){
  loc <- data[,c(1,9:10)]
  loc$Latitude <- as.numeric(as.character(loc$Latitude))# get just lat/long and grid cell names to get
  loc$Longitude <- as.numeric(as.character(loc$Longitude)) # get just lat/long and grid cell names to get
  avg_lat <- aggregate( Latitude ~ PageName, loc, mean ) # create a matrix of lat lon avg for grid cells
  avg_lon <- aggregate( Longitude ~ PageName, loc, mean ) # create a matrix of lat lon avg for grid cells
  avg_loc <- merge(avg_lat,avg_lon)
  colnames(avg_loc) <- c(“node_label”, “y”, “x”)
  avg_loc
}

colnames(trunk60_data)
nrow(trunk60_data)

# Get the attribute table of just the locations for genera level
trunk60_gen_avg_loc <- coor_atrib(trunk60_data)
head(trunk60_gen_avg_loc)
nrow(trunk60_gen_avg_loc)

# Get list of all grid cells with greater/less than 2 dev of median occ count?? Trying different methods
colnames(trunk60_matrix_gen_final_pre)
lastcol <- ncol(trunk60_matrix_gen_final_pre)
grid_gen_occ_count <- trunk60_matrix_gen_final_pre[,c(lastcol)]
colnames(grid_gen_occ_count) <- "Numb_Occ" # get a list of the total # occ in each grid cell
head(grid_gen_occ_count)
summary(grid_gen_occ_count) # check the mean, median, etc for the list
typeof(grid_gen_occ_count)
grid_gen_occ_count_calc <- as.numeric(unlist(grid_gen_occ_count)) # convert to numeric
hist(grid_gen_occ_count_calc) # values are highly right skewed... most have less than 20 occ
med_ab_dev <- mad((grid_gen_occ_count_calc)) # calculate the median absolute deviation of the data
### The MAD is too high, even without being multiplied by 2 like SD, can't really go below median with it
sd(grid_gen_occ_count_calc)
### The SD is too high too... can't subtract SD from mean either to get low lying values
quantile(grid_gen_occ_count_calc) # get quantile values
### Quantiles show that 1st quantile is 2. Might be worth using but very low... try median?

## Going to try to just give color attributes based on below median and below 2nd quantile but will color based on bins... no <3 occ
grid_gen_occ_count$OccColor <- with(grid_gen_occ_count, ifelse(Numb_Occ <= 3, 'black', ifelse(Numb_Occ <= 5,'grey', ifelse(Numb_Occ <= 10, 'blue', ifelse(Numb_Occ <= 15,'springgreen', ifelse(Numb_Occ <= 20,'gold',ifelse(Numb_Occ <= 25,'indianred','purple')))))))
grid_gen_occ_count <- cbind(node_label = rownames(grid_gen_occ_count), grid_gen_occ_count) # make the index the first row (Grid cell names)
rownames(grid_gen_occ_count) <- 1:nrow(grid_gen_occ_count) # create a new index

#Add a column of age and province to the attributes table
nrow(trunk60_data)
atrib_gen <- unique(trunk60_data[,c(1,12:15)])
colnames(atrib_gen) <- c("node_label","Province","Age","ProvColor","AgeColor")
nrow(atrib_gen)
head(atrib_gen)

#merge with table of less than median and 2nd quantile value colors
trunk60_gen_atrib <- merge(grid_gen_occ_count,atrib_gen,by="node_label")

#Add age level info to the loc table
loc_age_merge <- trunk60_gen_atrib[,c(1,5)]
head(loc_age_merge)
trunk60_gen_avg_loc <- merge(trunk60_gen_avg_loc,loc_age_merge,by="node_label")

write.csv(trunk60_gen_avg_loc,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_gen_avg_loc.csv",row.names=FALSE)
write.csv(trunk60_gen_atrib,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_gen_atrib.csv",row.names=FALSE)

# Get the attribute table of just the locations for species level
trunk60_spp_avg_loc <- coor_atrib(trunk60_data_spp)
head(trunk60_spp_avg_loc)
nrow(trunk60_spp_avg_loc)

# Get list of all grid cells with greater/less than 2 dev of median occ count?? Trying different methods
colnames(trunk60_matrix_spp_final_pre)
lastcol <- ncol(trunk60_matrix_spp_final_pre)
grid_spp_occ_count <- trunk60_matrix_spp_final_pre[,c(lastcol)]
colnames(grid_spp_occ_count) <- "Numb_Occ" # get a list of the total # occ in each grid cell
head(grid_spp_occ_count)
summary(grid_spp_occ_count) # check the mean, median, etc for the list

## Going to try to just give color attributes based on below median and below 2nd quantile

grid_spp_occ_count <- cbind(node_label = rownames(grid_spp_occ_count), grid_spp_occ_count) # make the index the first row (Grid cell names)
rownames(grid_spp_occ_count) <- 1:nrow(grid_spp_occ_count) # create a new index

# Add a column of age and province to the attributes table
colnames(trunk60_data_spp)
atrib_spp <- unique(trunk60_data_spp[,c(1,12:15)])
colnames(atrib_spp) <- c("node_label","Province","Age","ProvColor","AgeColor")
head(atrib_spp)

# Merge the two tables together
trunk60_spp_atrib <- merge(grid_spp_occ_count,atrib_spp, by="node_label",all.y=TRUE)
head(trunk60_spp_atrib)

# Add age level info to the loc table
loc_age_merge <- trunk60_spp_atrib[,c(1,5)]
head(loc_age_merge)
trunk60_spp_avg_loc <- merge(trunk60_spp_avg_loc,loc_age_merge,by="node_label")

write.csv(trunk60_spp_avg_loc,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_spp_avg_loc.csv",row.names=FALSE)
write.csv(trunk60_spp_atrib,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_spp_atrib.csv",row.names=FALSE)

### Make a attributes table for the matrix with grids containing <3 occ removed (this one is messier)
# Reconfigure the matrix without <3 occ grid cells into a simple list of grid names
trunk60_matrix_gen_great3_GridNames <- cbind(node_label = rownames(trunk60_matrix_gen_great3), trunk60_matrix_gen_great3) # make the index the first row (Grid cell names)
rownames(trunk60_matrix_gen_great3_GridNames) <- 1:nrow(trunk60_matrix_gen_great3) # create a new index
trunk60_matrix_gen_great3_GridNames <- data.frame(trunk60_matrix_gen_great3_GridNames[,1]) # remove all rows except the grid cell names
colnames(trunk60_matrix_gen_great3_GridNames) <- "node_label"
nrow(trunk60_matrix_gen_great3_GridNames)
head(trunk60_matrix_gen_great3_GridNames)

# merge the simplified list of occupied grid cells from the matrix which has had <3 occ grids removed with the full atrib table
trunk60_matrix_gen_great3_atrib <- merge(trunk60_matrix_gen_great3_GridNames, trunk60_gen_atrib, all.x=TRUE)
nrow(trunk60_matrix_gen_great3_atrib)
head(trunk60_matrix_gen_great3_atrib)

trunk60_matrix_gen_great3_avgloc <-
mmerge(trunk60_matrix_gen_great3_GridNames,trunk60_gen_avg_loc,all.x=TRUE)

write.csv(trunk60_matrix_gen_great3_atrib,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_gen_great3_atrib.csv",row.names=FALSE)
write.csv(trunk60_matrix_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_gen_great3_avgloc.csv",row.names=FALSE)

# Reconfigure the matrix without <3 occ grid cells into a simple list of grid names for species
trunk60_matrix_spp_great3_GridNames <- cbind(node_label = rownames(trunk60_matrix_spp_great3),
trunk60_matrix_spp_great3) # make the index the first row (Grid cell names)
rownames(trunk60_matrix_spp_great3_GridNames) <- 1:nrow(trunk60_matrix_spp_great3) # create a new index

trunk60_matrix_spp_great3_GridNames <- data.frame(trunk60_matrix_spp_great3_GridNames[,1]) # remove all rows except the grid cell names
colnames(trunk60_matrix_spp_great3_GridNames) <- "node_label"
nrow(trunk60_matrix_spp_great3_GridNames)
head(trunk60_matrix_spp_great3_GridNames)

# merge the simplified list of occupied grid cells from the matrix which has had <3 occ grids removed with the full atrib table
trunk60_matrix_spp_great3_atrib <-
mmerge(trunk60_matrix_spp_great3_GridNames,trunk60_spp_atrib,all.x=TRUE)
nrow(trunk60_matrix_spp_great3_atrib)
head(trunk60_matrix_spp_great3_atrib)

trunk60_matrix_spp_great3_avgloc <-
mmerge(trunk60_matrix_spp_great3_GridNames,trunk60_spp_avg_loc,all.x=TRUE)

write.csv(trunk60_matrix_spp_great3_atrib,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_spp_great3_atrib.csv",row.names=FALSE)
write.csv(trunk60_matrix_spp_great3_avgloc,file="R_Config_for_EN_OUTPUT/CompleteDB/trunk60_matrix_spp_great3_avgloc.csv",row.names=FALSE)

### RECONFIGURE INTO SIMPLIFIED SUBSTAGES TO USE IN EDENETWORKS ###

## SUBSETTING INTO SUBSTAGE BINS ##

# Subset the genus-level data based on substages
CAM_low <- subset(trunk60_data,Substage_from_Zone.Mbr == "CAM (low)")
CAM_mid <- subset(trunk60_data,Substage_from_Zone.Mbr == "CAM (mid)")
CAM_up <- subset(trunk60_data,Substage_from_Zone.Mbr == "CAM (up)")
MAA_low <- subset(trunk60_data,Substage_from_Zone.Mbr == "MAA (low)")
MAA_up <- subset(trunk60_data,Substage_from_Zone.Mbr == "MAA (up)")

# check the number of genus-level occ for each stage
nrow(CAM_low) # 1866
nrow(CAM_mid) # 8403
nrow(CAM_up) # 8564
nrow(MAA_low) # 6480
nrow(MAA_up) # 7855

# write csv of subset genus-level data
write.csv(CAM_low, file= "R_Config_for_EN_OUTPUT/Low CAM/CAM_low_genus_all.csv")
write.csv(CAM_mid, file= "R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_genus_all.csv")
write.csv(CAM_up, file= "R_Config_for_EN_OUTPUT/Up CAM/CAM_up_genus_all.csv")
write.csv(MAA_low, file= "R_Config_for_EN_OUTPUT/Low MAA/MAA_low_genus_all.csv")
write.csv(MAA_up, file= "R_Config_for_EN_OUTPUT/Up MAA/MAA_up_genus_all.csv")

# Subset the spp-level data based on substages
CAM_low_spp <- subset(trunk60_data_spp,Substage_from_Zone.Mbr == "CAM (low)")
CAM_mid_spp <- subset(trunk60_data_spp,Substage_from_Zone.Mbr == "CAM (mid)")
CAM_up_spp <- subset(trunk60_data_spp,Substage_from_Zone.Mbr == "CAM (up)")
MAA_low_spp <- subset(trunk60_data_spp,Substage_from_Zone.Mbr == "MAA (low)")
MAA_up_spp <- subset(trunk60_data_spp,Substage_from_Zone.Mbr == "MAA (up)")

# Check number of spp occ in each substage (have not removed duplicates)
nrow(CAM_low_spp) # 896
nrow(CAM_mid_spp) # 4856
nrow(CAM_up_spp) # 5282
nrow(MAA_low_spp) # 4012
nrow(MAA_up_spp) # 4866

# write csv of subset spp-level data
write.csv(CAM_low_spp, file= "R_Config_for_EN_OUTPUT/Low CAM/CAM_low_species_all.csv")
write.csv(CAM_mid_spp, file= "R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_species_all.csv")
write.csv(CAM_up_spp, file= "R_Config_for_EN_OUTPUT/Up CAM/CAM_up_species_all.csv")
write.csv(MAA_low_spp, file= "R_Config_for_EN_OUTPUT/Low MAA/MAA_low_species_all.csv")
write.csv(MAA_up_spp, file= "R_Config_for_EN_OUTPUT/Up MAA/MAA_up_species_all.csv")

#### CREATE MATRIX OF GENUS LEVEL OCC INFORMATION FOR EDENETWORKS ANALYSIS ####

# Original attempted made unique lat/long into the ID locations but that doesn't work, so instead
# I'm going to make the GridCell name the unique location and give a collection ID that is
# meaningless just so the configuration matches what it's supposed to in EDENETWORKS. Any code
# that isn't used anymore based on that original attempt has been left but commented out.

# create lookup table of unique location ID (lat, long) to use as "sites" in EN
# lookup <- seq(1:6000)

# function to transform into an abundance matrix of locations with genus names
substr_gen_grid <- function(data){
  data1 <- as.data.frame(subset(data[,c(1,4)]) # subset out just the genus name and location information
colnames(data1) <- c("PageName","Genus_Name") # Give the columns names
data2 <- dcast(data1, PageName~Genus_Name, length) # transform into a pres-abs matrix based on Grid cells
colnames(CAM_low)

CAM_low_matrix_gen <- substg_gen_grid(CAM_low)
CAM_mid_matrix_gen <- substg_gen_grid(CAM_mid)
CAM_up_matrix_gen <- substg_gen_grid(CAM_up)
MAA_low_matrix_gen <- substg_gen_grid(MAA_low)
MAA_up_matrix_gen <- substg_gen_grid(MAA_up)

# find number of genera in each substage matrix
ncol(CAM_low_matrix_gen) # 148
ncol(CAM_mid_matrix_gen) # 188
ncol(CAM_up_matrix_gen) # 344
ncol(MAA_low_matrix_gen) # 305
ncol(MAA_up_matrix_gen) # 331

# function to configure the matrix into the correct format for EDENETWORKS and add the grid cell info
config_matrix <- function(data){
  #base1 <- (merge(data, lookup, by = 'LocationKey', all.X=TRUE, all.y=FALSE)) # only use if trying to
  # make column of latlong using unique latlong key
  x <- nrow(data)
  data$SampleKey <- seq(1:x)
  a <- ncol(data)
  b <- a - 1
  base2 <- data[,c(1,a,2:b)] # This part would need to change if using lat/long as unique ID
  base3 <- data.frame(base2[,-1], row.names = base2[,1]) # make the first column with pagename the
  index
}

# create matrix in correct config with grid cell names include as first
CAM_low_matrix_gen_final <- config_matrix(CAM_low_matrix_gen)
CAM_mid_matrix_gen_final <- config_matrix(CAM_mid_matrix_gen)
CAM_up_matrix_gen_final <- config_matrix(CAM_up_matrix_gen)
MAA_low_matrix_gen_final <- config_matrix(MAA_low_matrix_gen)
MAA_up_matrix_gen_final <- config_matrix(MAA_up_matrix_gen)

#function to change from abundance to pres-abs matrix
pres_ab <- function(data){
  numeric_cols <- vapply(data, is.numeric, logical(1))
  data[numeric_cols] <- as.integer(data[numeric_cols] != 0)
  data
}

CAM_low_matrix_gen_final <- pres_ab(CAM_low_matrix_gen_final)
CAM_mid_matrix_gen_final <- pres_ab(CAM_mid_matrix_gen_final)
CAM_up_matrix_gen_final <- pres_ab(CAM_up_matrix_gen_final)
MAA_low_matrix_gen_final <- pres_ab(MAA_low_matrix_gen_final)
MAA_up_matrix_gen_final <- pres_ab(MAA_up_matrix_gen_final)

# check the number of rows (# of unique locations w/ genera) for each substage
# The number of unique locations is significantly larger for the Mid and Up CAM

```
write.csv(CAM_low_matrix_gen_final,file= "R_Config_for_EN_OUTPUT/LowCAM/CAM_low_matrix_gen_final.csv")
write.csv(CAM_mid_matrix_gen_final,file= "R_Config_for_EN_OUTPUT/MidCAM/CAM_mid_matrix_gen_final.csv")
write.csv(CAM_up_matrix_gen_final,file= "R_Config_for_EN_OUTPUT/UpCAM/CAM_up_matrix_gen_final.csv")
write.csv(MAA_low_matrix_gen_final,file= "R_Config_for_EN_OUTPUT/LowMAA/MAA_low_matrix_gen_final.csv")
write.csv(MAA_up_matrix_gen_final,file= "R_Config_for_EN_OUTPUT/UpMAA/MAA_up_matrix_gen_final.csv")
```

## Create matrix with grid cells containing less than 3 occ removed for each individual substage

### Lower CAM: For genera
```
ncol(CAM_low_matrix_gen_final)
CAM_low_numb_occ_gen <- data.frame(rowSums(CAM_low_matrix_gen_final[,2:148])) # sum rows to get total # occ
CAM_low_matrix_gen_final_pre <- CAM_low_matrix_gen_final # create new matrix name to use
CAM_low_matrix_gen_final_pre$nOccGen <- CAM_low_numb_occ_gen # Add column of total occ number
colnames(CAM_low_matrix_gen_final_pre)
CAM_low_matrix_gen_great3 <- subset(CAM_low_matrix_gen_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(CAM_low_matrix_gen_great3)
CAM_low_matrix_gen_great3 <- CAM_low_matrix_gen_great3[,1:148] # remove occ number column
colnames(CAM_low_matrix_gen_great3)
```

### Middle CAM: For genera
```
ncol(CAM_mid_matrix_gen_final)
CAM_mid_numb_occ_gen <- data.frame(rowSums(CAM_mid_matrix_gen_final[,2:188])) # sum rows to get total # occ
CAM_mid_matrix_gen_final_pre <- CAM_mid_matrix_gen_final # create new matrix name to use
CAM_mid_matrix_gen_final_pre$nOccGen <- CAM_mid_numb_occ_gen # Add column of total occ number
colnames(CAM_mid_matrix_gen_final_pre)
```
CAM_mid_matrix_gen_great3 <- subset(CAM_mid_matrix_gen_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
col(CAM_mid_matrix_gen_great3)
CAM_mid_matrix_gen_great3 <- CAM_mid_matrix_gen_great3[,1:188] # remove occ number column
colnames(CAM_mid_matrix_gen_great3)

#Upper CAM: For genera
ncol(CAM_up_matrix_gen_final)
CAM_up_numb_occ_gen <- data.frame(rowSums(CAM_up_matrix_gen_final[,2:344])) # sum rows to get total # occ
CAM_up_matrix_gen_final_pre <- CAM_up_matrix_gen_final # create new matrix name to use
CAM_up_matrix_gen_final_pre$nOccGen <- CAM_up_numb_occ_gen # Add column of total occ number
colnames(CAM_up_matrix_gen_final_pre)
CAM_up_matrix_gen_great3 <- subset(CAM_up_matrix_gen_final_pre, nOcc >= 3) # subset out all grid cells with greater 3 occ
col(CAM_up_matrix_gen_great3)
CAM_up_matrix_gen_great3 <- CAM_up_matrix_gen_great3[,1:344] # remove occ number column
colnames(CAM_up_matrix_gen_great3)

#Lower MAA: For genera
ncol(MAA_low_matrix_gen_final)
MAA_low_numb_occ_gen <- data.frame(rowSums(MAA_low_matrix_gen_final[,2:305])) # sum rows to get total # occ
MAA_low_matrix_gen_final_pre <- MAA_low_matrix_gen_final # create new matrix name to use
MAA_low_matrix_gen_final_pre$nOccGen <- MAA_low_numb_occ_gen # Add column of total occ number
colnames(MAA_low_matrix_gen_final_pre)
MAA_low_matrix_gen_great3 <- subset(MAA_low_matrix_gen_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
col(CAM_up_matrix_gen_great3)
MAA_low_matrix_gen_great3 <- MAA_low_matrix_gen_great3[,1:305] # remove occ number column
colnames(MAA_low_matrix_gen_great3)

#Upper MAA: For genera
ncol(MAA_up_matrix_gen_final)
MAA_up_numb_occ_gen <- data.frame(rowSums(MAA_up_matrix_gen_final[,2:331])) # sum rows to get total # occ
MAA_up_matrix_gen_final_pre <- MAA_up_matrix_gen_final # create new matrix name to use
MAA_up_matrix_gen_final_pre$nOccGen <- MAA_up_numb_occ_gen # Add column of total occ number
colnames(MAA_up_matrix_gen_final_pre)
MAA_up_matrix_gen_great3 <- subset(MAA_up_matrix_gen_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
col(MAM_up_matrix_gen_great3)
MAA_up_matrix_gen_great3 <- MAA_up_matrix_gen_great3[,1:331] # remove occ number column
colnames(MAA_up_matrix_gen_great3)

# export new csv files of the dataset with grids containing <3 occ removed
write.csv(CAM_low_matrix_gen_great3, file= "R_Config_for_EN_OUTPUT/Low_CAM/CAM_low_matrix_gen_great3.csv")
write.csv(CAM_mid_matrix_gen_great3, file= "R_Config_for_EN_OUTPUT/Mid  
CAM/CAM_mid_matrix_gen_great3.csv")
write.csv(CAM_up_matrix_gen_great3, file= "R_Config_for_EN_OUTPUT/Up  
CAM/CAM_up_matrix_gen_great3.csv")
write.csv(MAA_low_matrix_gen_great3, file= "R_Config_for_EN_OUTPUT/Low  
MAA/MAA_low_matrix_gen_great3.csv")
write.csv(MAA_up_matrix_gen_great3, file= "R_Config_for_EN_OUTPUT/Up  
MAA/MAA_up_matrix_gen_great3.csv")

#### Get attribute files for EDENetwork analysis for each Substage ####

# subset out the different substages for genera
CAM_low_gen_atrib <- subset(trunk60_gen_atrib, Age == "LC")
CAM_mid_gen_atrib <- subset(trunk60_gen_atrib, Age == "MC")
CAM_up_gen_atrib <- subset(trunk60_gen_atrib, Age == "UC")
MAA_low_gen_atrib <- subset(trunk60_gen_atrib, Age == "LM")
MAA_up_gen_atrib <- subset(trunk60_gen_atrib, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_gen_atrib, file="R_Config_for_EN_OUTPUT/Low  
CAM/CAM_low_gen_atrib.csv", row.names=FALSE)
write.csv(CAM_mid_gen_atrib, file="R_Config_for_EN_OUTPUT/Mid  
CAM/CAM_mid_gen_atrib.csv", row.names=FALSE)
write.csv(CAM_up_gen_atrib, file="R_Config_for_EN_OUTPUT/Up  
CAM/CAM_up_gen_atrib.csv", row.names=FALSE)
write.csv(MAA_low_gen_atrib, file="R_Config_for_EN_OUTPUT/Low  
MAA/MAA_low_gen_atrib.csv", row.names=FALSE)
write.csv(MAA_up_gen_atrib, file="R_Config_for_EN_OUTPUT/Up  
MAA/MAA_up_gen_atrib.csv", row.names=FALSE)

### Make a attributes table for the matrix with grids containing <3 occ removed for substages

# subset out the different substages for genera of greater than 3 occ
CAM_low_gen_great3_atrib <- subset(trunk60_matrix_gen_great3_atrib, Age == "LC")
CAM_mid_gen_great3_atrib <- subset(trunk60_matrix_gen_great3_atrib, Age == "MC")
CAM_up_gen_great3_atrib <- subset(trunk60_matrix_gen_great3_atrib, Age == "UC")
MAA_low_gen_great3_atrib <- subset(trunk60_matrix_gen_great3_atrib, Age == "LM")
MAA_up_gen_great3_atrib <- subset(trunk60_matrix_gen_great3_atrib, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_gen_great3_atrib, file="R_Config_for_EN_OUTPUT/Low  
CAM/CAM_low_gen_great3_atrib.csv", row.names=FALSE)
write.csv(CAM_mid_gen_great3_atrib, file="R_Config_for_EN_OUTPUT/Mid  
CAM/CAM_mid_gen_great3_atrib.csv", row.names=FALSE)
write.csv(CAM_up_gen_great3_atrib, file="R_Config_for_EN_OUTPUT/Up  
CAM/CAM_up_gen_great3_atrib.csv", row.names=FALSE)
write.csv(MAA_low_gen_great3_atrib, file="R_Config_for_EN_OUTPUT/Low  
MAA/MAA_low_gen_great3_atrib.csv", row.names=FALSE)
write.csv(MAA_up_gen_great3_atrib, file="R_Config_for_EN_OUTPUT/Up  
MAA/MAA_up_gen_great3_atrib.csv", row.names=FALSE)
### Get average coordinates files for EDENetwork analysis for each Substage ###

# subset out the different substages for genera
CAM_low_gen_avg_loc <- subset(trunk60_gen_avg_loc, Age == "LC")
CAM_mid_gen_avg_loc <- subset(trunk60_gen_avg_loc, Age == "MC")
CAM_up_gen_avg_loc <- subset(trunk60_gen_avg_loc, Age == "UC")
MAA_low_gen_avg_loc <- subset(trunk60_gen_avg_loc, Age == "LM")
MAA_up_gen_avg_loc <- subset(trunk60_gen_avg_loc, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_gen_avg_loc,file="R_Config_for_EN_OUTPUT/LowCAM/CAM_low_gen_avg_loc.csv",row.names=FALSE)
write.csv(CAM_mid_gen_avg_loc,file="R_Config_for_EN_OUTPUT/MidCAM/CAM_mid_gen_avg_loc.csv",row.names=FALSE)
write.csv(CAM_up_gen_avg_loc,file="R_Config_for_EN_OUTPUT/UpCAM/CAM_up_gen_avg_loc.csv",row.names=FALSE)
write.csv(MAA_low_gen_avg_loc,file="R_Config_for_EN_OUTPUT/LowMAA/MAA_low_gen_avg_loc.csv",row.names=FALSE)
write.csv(MAA_up_gen_avg_loc,file="R_Config_for_EN_OUTPUT/UpMAA/MAA_up_gen_avg_loc.csv",row.names=FALSE)

### Make a average coordinates table for the matrix with grids containing <3 occ removed for substages

# subset out the different substages for genus of greater than 3 occ
CAM_low_gen_great3_avgloc <- subset(trunk60_matrix_gen_great3_avgloc, Age == "LC")
CAM_mid_gen_great3_avgloc <- subset(trunk60_matrix_gen_great3_avgloc, Age == "MC")
CAM_up_gen_great3_avgloc <- subset(trunk60_matrix_gen_great3_avgloc, Age == "UC")
MAA_low_gen_great3_avgloc <- subset(trunk60_matrix_gen_great3_avgloc, Age == "LM")
MAA_up_gen_great3_avgloc <- subset(trunk60_matrix_gen_great3_avgloc, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/LowCAM/CAM_low_gen_great3_avgloc.csv",row.names=FALSE)
write.csv(CAM_mid_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/MidCAM/CAM_mid_gen_great3_avgloc.csv",row.names=FALSE)
write.csv(CAM_up_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/UpCAM/CAM_up_gen_great3_avgloc.csv",row.names=FALSE)
write.csv(MAA_low_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/LowMAA/MAA_low_gen_great3_avgloc.csv",row.names=FALSE)
write.csv(MAA_up_gen_great3_avgloc,file="R_Config_for_EN_OUTPUT/UpMAA/MAA_up_gen_great3_avgloc.csv",row.names=FALSE)

##### CREATE MATRIX OF SPECIES LEVEL OCC INFORMATION FOR EDENETWORKS ANALYSIS #####

# create lookup table of unique location ID (lat,long) to use as "sites" in EN
lookup <- unique(join60_data[,c(25,27)])

# function to transform into a pres-abs matrix of locations with genus names
substg_spp_grid <- function(data) {
  data1 <- as.data.frame(subset(data[,c(1,8)])) # subset out just the genus name and location information
colnames(data1) <- c("PageName","Species_Name") # Give the columns names
data2 <- dcast(data1, PageName~Species_Name, length) # transform into a pres-abs matrix
}

colnames(CAM_low_spp)

CAM_low_matrix_spp <- substg_spp_grid(CAM_low_spp)
CAM_mid_matrix_spp <- substg_spp_grid(CAM_mid_spp)
CAM_up_matrix_spp <- substg_spp_grid(CAM_up_spp)
MAA_low_matrix_spp <- substg_spp_grid(MAA_low_spp)
MAA_up_matrix_spp <- substg_spp_grid(MAA_up_spp)

# find number of spp in each substage matrix
ncol(CAM_low_matrix_spp) # 135
ncol(CAM_mid_matrix_spp) # 211
ncol(CAM_up_matrix_spp) # 586
ncol(MAA_low_matrix_spp) # 381
ncol(MAA_up_matrix_spp) # 452

# function to configure the SPP-level matrix into the correct format for EDENETWORKS and add the
grid cell info
config_matrix_spp <- function(data){
  # base1 <- (merge(data, lookup, by = 'LocationKey',all.X=TRUE, all.y=FALSE))
x <- nrow(data)
data$SampleKey <-seq(1:x)
a <- ncol(data)
b <- a - 1
base2 <- data[,c(1,a,2:b)] # This part would need to change if using lat/long as unique ID
base3 <- data.frame(base2[-1], row.names = base2[,1]) # make the first column with pagename the
  index
}

# create matrix in correct config with grid cell names include as first
CAM_low_matrix_spp_final <- config_matrix_spp(CAM_low_matrix_spp)
CAM_mid_matrix_spp_final <- config_matrix_spp(CAM_mid_matrix_spp)
CAM_up_matrix_spp_final <- config_matrix_spp(CAM_up_matrix_spp)
MAA_low_matrix_spp_final <- config_matrix_spp(MAA_low_matrix_spp)
MAA_up_matrix_spp_final <- config_matrix_spp(MAA_up_matrix_spp)

head(CAM_low_matrix_spp_final)

#function to change from abundance to pres-abs matrix
pres_ab <- function(data){
  numeric_cols <- vapply(data, is.numeric, logical(1))
data[numeric_cols] <- as.integer(data[numeric_cols] != 0)
data
}

# create pres-abs matrix from the abundance matrix
CAM_low_matrix_spp_final <- pres_ab(CAM_low_matrix_spp_final)
CAM_mid_matrix_spp_final <- pres_ab(CAM_mid_matrix_spp_final)
CAM_up_matrix_spp_final <- pres_ab(CAM_up_matrix_spp_final)
MAA_low_matrix_spp_final <- pres_ab(MAA_low_matrix_spp_final)
MAA_up_matrix_spp_final <- pres_ab(MAA_up_matrix_spp_final)

# check number of spp in each substage matrix
ncol(CAM_low_matrix_spp) # 135
ncol(CAM_mid_matrix_spp) # 211
ncol(CAM_up_matrix_spp) # 586
ncol(MAA_low_matrix_spp) # 381
ncol(MAA_up_matrix_spp) # 452

# check the number of rows (# of unique locations w/ species) for each substage
nrow(CAM_low_matrix_spp_final) # 107
nrow(CAM_mid_matrix_spp_final) # 156
nrow(CAM_up_matrix_spp_final) # 188
nrow(MAA_low_matrix_spp_final) # 116
nrow(MAA_up_matrix_spp_final) # 73

# The number of unique locations is significantly larger for the Mid and Up CAM

# write csv of matrices for each substage
write.csv(CAM_low_matrix_spp_final, file= "R_Config_for_EN_OUTPUT/Low
CAM/CAM_low_matrix_spp_final.csv")
write.csv(CAM_mid_matrix_spp_final, file= "R_Config_for_EN_OUTPUT/Mid
CAM/CAM_mid_matrix_spp_final.csv")
write.csv(CAM_up_matrix_spp_final, file= "R_Config_for_EN_OUTPUT/Up
CAM/CAM_up_matrix_spp_final.csv")
write.csv(MAA_low_matrix_spp_final, file= "R_Config_for_EN_OUTPUT/Low
MAA/MAA_low_matrix_spp_final.csv")
write.csv(MAA_up_matrix_spp_final, file= "R_Config_for_EN_OUTPUT/Up
MAA/MAA_up_matrix_spp_final.csv")

## Create matrix with grid cells containing less than 3 occ removed for each individual substage

#Lower CAM: For species
ncol(CAM_low_matrix_spp_final)
CAM_low_numb_occ_spp <- data.frame(rowSums(CAM_low_matrix_spp_final[,2:135])) # sum rows to get total # occ
CAM_low_matrix_spp_final_pre <- CAM_low_matrix_spp_final # create new matrix name to use
CAM_low_matrix_spp_final_pre$nOccGen <- CAM_low_numb_occ_spp # Add column of total occ number
colnames(CAM_low_matrix_spp_final_pre)
CAM_low_matrix_spp_great3 <- subset(CAM_low_matrix_spp_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(CAM_low_matrix_spp_great3)
CAM_low_matrix_spp_great3 <- CAM_low_matrix_spp_great3[,1:135] # remove occ number column
colnames(CAM_low_matrix_spp_great3)

#Middle CAM: For species
ncol(CAM_mid_matrix_spp_final)
CAM_mid_numb_occ_spp <- data.frame(rowSums(CAM_mid_matrix_spp_final[,2:211])) # sum rows to get total # occ
CAM_mid_matrix_spp_final_pre <- CAM_mid_matrix_spp_final # create new matrix name to use
CAM_mid_matrix_spp_final_pre$nOccGen <- CAM_mid_numb_occ_spp # Add column of total occ number
colnames(CAM_mid_matrix_spp_final_pre)
CAM_mid_matrix_spp_great3 <- subset(CAM_mid_matrix_spp_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(CAM_mid_matrix_spp_great3)
CAM_mid_matrix_spp_great3 <- CAM_mid_matrix_spp_great3[,1:211] # remove occ number column
colnames(CAM_mid_matrix_spp_great3)

#Upper CAM: For species
ncol(CAM_up_matrix_spp_final)
CAM_up_numb_occ_spp <- data.frame(rowSums(CAM_up_matrix_spp_final[,2:586])) # sum rows to get total # occ
CAM_up_matrix_spp_final_pre <- CAM_up_matrix_spp_final # create new matrix name to use
CAM_up_matrix_spp_final_pre$nOccGen <- CAM_up_numb_occ_spp # Add column of total occ number
colnames(CAM_up_matrix_spp_final_pre)
CAM_up_matrix_spp_great3 <- subset(CAM_up_matrix_spp_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(CAM_up_matrix_spp_great3)
CAM_up_matrix_spp_great3 <- CAM_up_matrix_spp_great3[,1:586] # remove occ number column
colnames(CAM_up_matrix_spp_great3)

#Lower MAA: For species
ncol(MAA_low_matrix_spp_final)
MAA_low_numb_occ_spp <- data.frame(rowSums(MAA_low_matrix_spp_final[,2:381])) # sum rows to get total # occ
MAA_low_matrix_spp_final_pre <- MAA_low_matrix_spp_final # create new matrix name to use
MAA_low_matrix_spp_final_pre$nOccGen <- MAA_low_numb_occ_spp # Add column of total occ number
colnames(MAA_low_matrix_spp_final_pre)
MAA_low_matrix_spp_great3 <- subset(MAA_low_matrix_spp_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(MAA_low_matrix_spp_great3)
MAA_low_matrix_spp_great3 <- MAA_low_matrix_spp_great3[,1:382] # remove occ number column
colnames(MAA_low_matrix_spp_great3)

#Upper MAA: For species
ncol(MAA_up_matrix_spp_final)
MAA_up_numb_occ_spp <- data.frame(rowSums(MAA_up_matrix_spp_final[,2:452])) # sum rows to get total # occ
MAA_up_matrix_spp_final_pre <- MAA_up_matrix_spp_final # create new matrix name to use
MAA_up_matrix_spp_final_pre$nOccGen <- MAA_up_numb_occ_spp # Add column of total occ number
colnames(MAA_up_matrix_spp_final_pre)
MAA_up_matrix_spp_great3 <- subset(MAA_up_matrix_spp_final_pre, nOccGen >= 3) # subset out all grid cells with greater 3 occ
ncol(MAA_up_matrix_spp_great3)
MAA_up_matrix_spp_great3 <- MAA_up_matrix_spp_great3[,1:452] # remove occ number column
colnames(MAA_up_matrix_spp_great3)

# export new csv of the dataset with grids containing <3 occ removed for each substage and spp
write.csv(CAM_low_matrix_spp_great3,file= "R_Config_for_EN_OUTPUT/Low
CAM/CAM_low_matrix_spp_great3.csv")
write.csv(CAM_mid_matrix_spp_great3,file= "R_Config_for_EN_OUTPUT/Mid
CAM/CAM_mid_matrix_spp_great3.csv")
write.csv(CAM_up_matrix_spp_great3,file= "R_Config_for_EN_OUTPUT/Up
CAM/CAM_up_matrix_spp_great3.csv")
write.csv(MAA_low_matrix_spp_great3,file= "R_Config_for_EN_OUTPUT/Low
MAA/MAA_low_matrix_spp_great3.csv")
write.csv(MAA_up_matrix_spp_great3,file= "R_Config_for_EN_OUTPUT/Up
MAA/MAA_up_matrix_spp_great3.csv")

### Get attribute files for EDENetwork analysis for each Substage ###

# subset out the different substages for species
CAM_low_spp_atrib <- subset(trunk60_spp_atrib, Age == "LC")
CAM_mid_spp_atrib <- subset(trunk60_spp_atrib, Age == "MC")
CAM_up_spp_atrib <- subset(trunk60_spp_atrib, Age == "UC")
MAA_low_spp_atrib <- subset(trunk60_spp_atrib, Age == "LM")
MAA_up_spp_atrib <- subset(trunk60_spp_atrib, Age == "UM")

#write csv files of attributes tables for substages at species level
write.csv(CAM_low_spp_atrib,file="R_Config_for_EN_OUTPUT/Low
CAM/CAM_low_spp_atrib.csv",row.names=FALSE)
write.csv(CAM_mid_spp_atrib,file="R_Config_for_EN_OUTPUT/Mid
CAM/CAM_mid_spp_atrib.csv",row.names=FALSE)
write.csv(CAM_up_spp_atrib,file="R_Config_for_EN_OUTPUT/Up
CAM/CAM_up_spp_atrib.csv",row.names=FALSE)
write.csv(MAA_low_spp_atrib,file="R_Config_for_EN_OUTPUT/Low
MAA/MAA_low_spp_atrib.csv",row.names=FALSE)
write.csv(MAA_up_spp_atrib,file="R_Config_for_EN_OUTPUT/Up
MAA/MAA_up_spp_atrib.csv",row.names=FALSE)

### Make attributes tables of substages with only >3 occ

# subset out the different substages for species
CAM_low_spp_great3_atrib <- subset(trunk60_matrix_spp_great3_atrib, Age == "LC")
CAM_mid_spp_great3_atrib <- subset(trunk60_matrix_spp_great3_atrib, Age == "MC")
CAM_up_spp_great3_atrib <- subset(trunk60_matrix_spp_great3_atrib, Age == "UC")
MAA_low_spp_great3_atrib <- subset(trunk60_matrix_spp_great3_atrib, Age == "LM")
MAA_up_spp_great3_atrib <- subset(trunk60_matrix_spp_great3_atrib, Age == "UM")

#write csv files of attributes tables for substages at species level
write.csv(CAM_low_spp_great3_atrib,file="R_Config_for_EN_OUTPUT/Low
CAM/CAM_low_spp_great3_atrib.csv",row.names=FALSE)
write.csv(CAM_mid_spp_great3_atrib,file="R_Config_for_EN_OUTPUT/Mid
CAM/CAM_mid_spp_great3_atrib.csv",row.names=FALSE)
write.csv(CAM_up_spp_great3_atrib, file="R_Config_for_EN_OUTPUT/Up CAM/CAM_up_spp_great3_atrib.csv", row.names=FALSE)
write.csv(MAA_low_spp_great3_atrib, file="R_Config_for_EN_OUTPUT/Low MAA/MAA_low_spp_great3_atrib.csv", row.names=FALSE)
write.csv(MAA_up_spp_great3_atrib, file="R_Config_for_EN_OUTPUT/Up MAA/MAA_up_spp_great3_atrib.csv", row.names=FALSE)

#### Get average coordinates files for EDENetwork analysis for each Substage species level ####

# subset out the different substages for genera
CAM_low_spp_avg_loc <- subset(trunk60_spp_avg_loc, Age == "LC")
CAM_mid_spp_avg_loc <- subset(trunk60_spp_avg_loc, Age == "MC")
CAM_up_spp_avg_loc <- subset(trunk60_spp_avg_loc, Age == "UC")
MAA_low_spp_avg_loc <- subset(trunk60_spp_avg_loc, Age == "LM")
MAA_up_spp_avg_loc <- subset(trunk60_spp_avg_loc, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_spp_avg_loc, file="R_Config_for_EN_OUTPUT/Low CAM/CAM_low_spp_avg_loc.csv", row.names=FALSE)
write.csv(CAM_mid_spp_avg_loc, file="R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_spp_avg_loc.csv", row.names=FALSE)
write.csv(CAM_up_spp_avg_loc, file="R_Config_for_EN_OUTPUT/Up CAM/CAM_up_spp_avg_loc.csv", row.names=FALSE)
write.csv(MAA_low_spp_avg_loc, file="R_Config_for_EN_OUTPUT/Low MAA/MAA_low_spp_avg_loc.csv", row.names=FALSE)
write.csv(MAA_up_spp_avg_loc, file="R_Config_for_EN_OUTPUT/Up MAA/MAA_up_gen_spp_loc.csv", row.names=FALSE)

### Make a average coordinates table for the matrix with grids containing <3 occ removed for substages species level

# subset out the different coordinates table for genera of greater than 3 occ
CAM_low_spp_great3_avgloc <- subset(trunk60_matrix_spp_great3_avgloc, Age == "LC")
CAM_mid_spp_great3_avgloc <- subset(trunk60_matrix_spp_great3_avgloc, Age == "MC")
CAM_up_spp_great3_avgloc <- subset(trunk60_matrix_spp_great3_avgloc, Age == "UC")
MAA_low_spp_great3_avgloc <- subset(trunk60_matrix_spp_great3_avgloc, Age == "LM")
MAA_up_spp_great3_avgloc <- subset(trunk60_matrix_spp_great3_avgloc, Age == "UM")

#write csv files of attributes tables for substages at genus level
write.csv(CAM_low_spp_great3_avgloc, file="R_Config_for_EN_OUTPUT/Low CAM/CAM_low_spp_great3_avgloc.csv", row.names=FALSE)
write.csv(CAM_mid_spp_great3_avgloc, file="R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_spp_great3_avgloc.csv", row.names=FALSE)
write.csv(CAM_up_spp_great3_avgloc, file="R_Config_for_EN_OUTPUT/Up CAM/CAM_up_spp_great3_avgloc.csv", row.names=FALSE)
write.csv(MAA_low_spp_great3_avgloc, file="R_Config_for_EN_OUTPUT/Low MAA/MAA_low_spp_great3_avgloc.csv", row.names=FALSE)
write.csv(MAA_up_spp_great3_avgloc, file="R_Config_for_EN_OUTPUT/Up MAA/MAA_up_spp_great3_avgloc.csv", row.names=FALSE)

#### CREATE SUMMARY TABLE OF THE NUMBER OF SPP/GEN AND UNIQUE LOCATIONS IN EACH SUBSTAGE ####
Substage <- c("Low CAM", "Mid CAM", "Up CAM", "Low MAA", "Up MAA")

UGenLoc <- c(nrow(CAM_low_matrix_gen_final), nrow(CAM_mid_matrix_gen_final),
              nrow(CAM_up_matrix_gen_final), nrow(MAA_low_matrix_gen_final),
              nrow(MAA_up_matrix_gen_final))

GenGrid <- c(nrow(data.frame(unique(CAM_low_matrix_gen_final$PageName))),
             nrow(data.frame(unique(CAM_mid_matrix_gen_final$PageName))),
             nrow(data.frame(unique(CAM_up_matrix_gen_final$PageName))),
             nrow(data.frame(unique(MAA_low_matrix_gen_final$PageName))),
             nrow(data.frame(unique(MAA_up_matrix_gen_final$PageName))))

GenOcc <- c(nrow(CAM_low),
             nrow(CAM_mid),
             nrow(CAM_up),
             nrow(MAA_low),
             nrow(MAA_up))

Genera <- c(ncol(CAM_low_matrix_gen_final), ncol(CAM_mid_matrix_gen_final),
             ncol(CAM_up_matrix_gen_final), ncol(MAA_low_matrix_gen_final),
             ncol(MAA_up_matrix_gen_final)) - 2

USppLoc <- c(nrow(CAM_low_matrix_spp_final), nrow(CAM_mid_matrix_spp_final),
              nrow(CAM_up_matrix_spp_final), nrow(MAA_low_matrix_spp_final),
              nrow(MAA_up_matrix_spp_final))

SppGrid <- c(nrow(data.frame(unique(CAM_low_matrix_spp_final$PageName))),
              nrow(data.frame(unique(CAM_mid_matrix_spp_final$PageName))),
              nrow(data.frame(unique(CAM_up_matrix_spp_final$PageName))),
              nrow(data.frame(unique(MAA_low_matrix_spp_final$PageName))),
              nrow(data.frame(unique(MAA_up_matrix_spp_final$PageName))))

SppOcc <- c(nrow(CAM_low_spp),
            nrow(CAM_mid_spp),
            nrow(CAM_up_spp),
            nrow(MAA_low_spp),
            nrow(MAA_up_spp))

Species <- c(ncol(CAM_low_matrix_spp_final), ncol(CAM_mid_matrix_spp_final),
              ncol(CAM_up_matrix_spp_final), ncol(MAA_low_matrix_spp_final),
              ncol(MAA_up_matrix_spp_final)) - 2

summary <- as.data.frame(cbind(Substage, GenOcc, UGenLoc, GenGrid, Genera, SppOcc, USppLoc,
                                 SppGrid,
                                 Species))

colnames(summary) <- c("Substage", "TotGenOcc", "UniqueGenLocations", "GenGridPixels", "#Genera",
                     "TotSppOcc", "UniqueSppLocations", "SppGridPixels", "#Species")

summary
write.csv(summary, file = "R_Config_for_EN_OUTPUT/SubstageSummary.csv")
The output shows that there are significantly more unique locations in the Mid and Up CAM but not a much higher number of occupied grid cells. There are far fewer total gen/spp occ points in the Low CAM, but not a much lower number of unique spp and genera relative to other substages. The Up MAA has the fewest total grid cells occupied, despite having a relatively high number of total spp/gen occ points and unique spp/gen.
Code for Conducting Network Randomization Comparisons

setwd("C:/Users/ceara/Documents/Province Project/Occurrence Data/VettedCombinedDB")

#### Lower Campanian ####

# read in the network information data from the pajek file

library(igraph)
library(tnet)

low_cam_percol_net <- read.graph('EDENetwork_outputs/LowCAM/CAM_low_matrix_gen_great3_thresholded_at_061.net', format = "pajek")

typeof(low_cam_percol_net)

cc_low_cam_percol_net <- transitivity(low_cam_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be
# numeric, not grid names)

low_cam_percol_net2 <- cbind(get.edgelist(low_cam_percol_net, names=FALSE),
                     E(low_cam_percol_net)$weight)

# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(low_cam_percol_net))
low_cam_percol_net2 <- symmetrise_w(low_cam_percol_net2)

# make sure conforms to tnet format
low_cam_percol_net2 <- as.tnet(low_cam_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates
# identical connection but only 2 present...)

# Create igraph object
low_cam_percol_net3 <- tnet_igraph(low_cam_percol_net2,type="weighted one-mode tnet")

# Get summary statistics
summary(low_cam_percol_net3)

# check that the cc is still close
cc_low_cam_percol_net2 <- transitivity(low_cam_percol_net3)

### loss of the 2 identical links maybe changed, but not much...

## Plot the network

# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(low_cam_percol_net3)
plot(low_cam_percol_net3, main="Observed Network", layout=layout,
vertex.label=1:length(low_cam_percol_net3), edge.width=E(low_cam_percol_net3)$weight,
edge.label=E(low_cam_percol_net3)$weight)

# Create random network from original network

# create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
  rg_reshuffling_w(x, option="links")
}

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
low_cam_percol_1000rand <- replicate(n, fun(low_cam_percol_net2), simplify=FALSE)

# turn into igraph objects
low_cam_percol_1000rand_igraph <- lapply(low_cam_percol_1000rand, function(x)
  tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_low_cam_percol_1000rand <- lapply(low_cam_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_low_cam_percol_1000rand <- mean(unlist(cc_low_cam_percol_1000rand))

# Calculate standad deviation of the random dist
cc_sd_low_cam_percol_1000rand <- sd(unlist(cc_low_cam_percol_1000rand))

# look at random nw cc distributions
cc_dist_low_cam_percol_1000rand <- (unlist(cc_low_cam_percol_1000rand))
hist(cc_dist_low_cam_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05

t.test(cc_low_cam_percol_net2, mu = cc_avg_low_cam_percol_1000rand, alternative = 'less')

#### Middle Campanian ####
mid_cam_percol_net <- read.graph('EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_057.net', format = "pajek")
typeof(mid_cam_percol_net)
cc_mid_cam_percol_net <- transitivity(mid_cam_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be numeric, not grid names)
mid_cam_percol_net2 <- cbind(get.edgelist(mid_cam_percol_net, names=FALSE),
                           E(mid_cam_percol_net)$weight)

# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(mid_cam_percol_net))
   mid_cam_percol_net2 <- symmetrise_w(mid_cam_percol_net2)

# make sure conforms to tnet format
mid_cam_percol_net2 <- as.tnet(mid_cam_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates identical connection but only 2 present...)
# Create igraph object
mid_cam_percol_net3 <- tnet_igraph(mid_cam_percol_net2, type="weighted one-mode tnet")

# Get summary statistics
summary(mid_cam_percol_net3)

# check that the cc is still close
cc_mid_cam_percol_net2 <- transitivity(mid_cam_percol_net3)
### exactly the same!

## Plot the network

# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(mid_cam_percol_net3)

plot(mid_cam_percol_net3, main="Observed Network", layout=layout,
     vertex.label=1:length(mid_cam_percol_net3), edge.width=E(mid_cam_percol_net3)$weight,
     edge.label=E(mid_cam_percol_net3)$weight)

# Create random network from original network

# layout <- layout.fruchterman.reingold(mid_cam_percol_net3)

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
mid_cam_percol_1000rand <- replicate(n, fun(mid_cam_percol_net2), simplify=FALSE)
# turn into igraph objects
mid_cam_percol_1000rand_igraph <- lapply(mid_cam_percol_1000rand, function(x)
    tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_mid_cam_percol_1000rand <- lapply(mid_cam_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_mid_cam_percol_1000rand <- mean(unlist(cc_mid_cam_percol_1000rand))

# Calculate standard deviation of the random dist
cc_sd_mid_cam_percol_1000rand <- sd(unlist(cc_mid_cam_percol_1000rand))

# look at random nw cc distributions
cc_dist_mid_cam_percol_1000rand <- (unlist(cc_mid_cam_percol_1000rand))
hist(cc_dist_mid_cam_percol_1000rand)

### Run statistical test on the real cc using random networks

### H0: cc[real] = mean of cc[rand]
### H1: cc[real] != mean of cc[rand]
### alpha = 0.05

t.test(cc_mid_cam_percol_net2, mu = cc_avg_mid_cam_percol_1000rand, alternative = 'less')

### Middle Campanian ###

mid_cam_percol_net <- read.graph('EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_057.net', format = "pajek")
typeof(mid_cam_percol_net)

cc_mid_cam_percol_net <- transitivity(mid_cam_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be numeric, not grid names)
mid_cam_percol_net2 <- cbind(get.edgelist(mid_cam_percol_net, names=FALSE),
E(mid_cam_percol_net)$weight)

# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(mid_cam_percol_net))
    mid_cam_percol_net2 <- symmetrise_w(mid_cam_percol_net2)
# make sure conforms to tnet format
mid_cam_percol_net2 <- as.tnet(mid_cam_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates identical connection but only 2 present...)
# Create igraph object
mid_cam_percol_net3 <- tnet_igraph(mid_cam_percol_net2,type="weighted one-mode tnet")

# Get summary statistics
summary(mid_cam_percol_net3)

# check that the cc is still close
cc_mid_cam_percol_net2 <- transitivity(mid_cam_percol_net3)
### exactly the same!

## Plot the network
# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(mid_cam_percol_net3)

plot(mid_cam_percol_net3, main="Observed Network", layout=layout,
vertex.label=1:length(mid_cam_percol_net3), edge.width=E(mid_cam_percol_net3)$weight,
edge.label=E(mid_cam_percol_net3)$weight)

# Create random network from original network

# create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
  rgreshuffling_w(x, option="links")
}

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
mid_cam_percol_1000rand <- replicate(n, fun(mid_cam_percol_net2), simplify=FALSE)

# turn into igraph objects
mid_cam_percol_1000rand_igraph <- lapply(mid_cam_percol_1000rand, function(x)
  tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_mid_cam_percol_1000rand <- lapply(mid_cam_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_mid_cam_percol_1000rand <- mean(unlist(cc_mid_cam_percol_1000rand))

# Calculate standad deviation of the random dist
cc_sd_mid_cam_percol_1000rand <- sd(unlist(cc_mid_cam_percol_1000rand))
# look at random nw cc distributions
cc_dist_mid_cam_percol_1000rand <- unlist(cc_mid_cam_percol_1000rand)
hist(cc_dist_mid_cam_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05
t.test(cc_mid_cam_percol_net2, mu = cc_avg_mid_cam_percol_1000rand, alternative = 'less')

#### Upper Campanian ####

up_cam_percol_net <- read.graph('EDENetwork_outputs/UpCAM/CAM_up_matrix_gen_great3_thresholded_at_050.net', format = "pajek")
typeof(up_cam_percol_net)

cc_up_cam_percol_net <- transitivity(up_cam_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be
# numeric, not grid names)
up_cam_percol_net2 <- cbind(get.edgelist(up_cam_percol_net, names=FALSE),
                          E(up_cam_percol_net)$weight)

# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(up_cam_percol_net))
  up_cam_percol_net2 <- symmetrise_w(up_cam_percol_net2)

# make sure conforms to tnet format
up_cam_percol_net2 <- as.tnet(up_cam_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates
# identical connection but only 2 present...)
# Create igraph object
up_cam_percol_net3 <- tnet_igraph(up_cam_percol_net2,type="weighted one-mode tnet")

# Get summary statistics
summary(up_cam_percol_net3)

# check that the cc is still close
cc_up_cam_percol_net2 <- transitivity(up_cam_percol_net3)
### slightly lower, loss of 3 identical connections (weights = 0) but not very different

## Plot the network
# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(up_cam_percol_net3)

plot(up_cam_percol_net3, main="Observed Network", layout=layout,
    vertex.label=1:length(up_cam_percol_net3), edge.width=E(up_cam_percol_net3)$weight,
    edge.label=E(up_cam_percol_net3)$weight)

# Create random network from original network

# create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
    rg_reshuffling_w(x, option="links")
}

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
up_cam_percol_1000rand <- replicate(n, fun(up_cam_percol_net2), simplify=FALSE)

# turn into igraph objects
up_cam_percol_1000rand_igraph <- lapply(up_cam_percol_1000rand, function(x)
    tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_up_cam_percol_1000rand <- lapply(up_cam_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_up_cam_percol_1000rand <- mean(unlist(cc_up_cam_percol_1000rand))

# Calculate standard deviation of the random dist
cc_sd_up_cam_percol_1000rand <- sd(unlist(cc_up_cam_percol_1000rand))

# look at random nw cc distributions
cc_dist_up_cam_percol_1000rand <- (unlist(cc_up_cam_percol_1000rand))
hist(cc_dist_up_cam_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05

t.test(cc_up_cam_percol_net2, mu = cc_avg_up_cam_percol_1000rand, alternative = 'less')

#### Lower Maastrichtian ####
low_maa_percol_net <- read.graph('EDENetwork_outputs/Low
MAA/maa_low_matrix_gen_great3_thresholded_at_080.net', format = "pajek")

typeof(low_maa_percol_net)

cc_low_maa_percol_net <- transitivity(low_maa_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be
numeric, not grid names)
low_maa_percol_net2 <- cbind(get.edgelist(low_maa_percol_net, names=FALSE),
E(low_maa_percol_net)$weight)

# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(low_maa_percol_net))
low_maa_percol_net2 <- symmetrise_w(low_maa_percol_net2)

# make sure conforms to tnet format
low_maa_percol_net2 <- as.tnet(low_maa_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates
identical connection but only 2 present...)
# Create igraph object
low_maa_percol_net3 <- tnet_igraph(low_maa_percol_net2, type="weighted one-mode tnet")

# Get summary statistics
summary(low_maa_percol_net3)

# check that the cc is still close
cc_low_maa_percol_net2 <- transitivity(low_maa_percol_net3)
### slightly lower, loss of 1 identical connections (wights = 0) but not very different

## Plot the network

# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(low_maa_percol_net3)

plot(low_maa_percol_net3, main="Observed Network", layout=layout,
vertex.label=1:length(low_maa_percol_net3), edge.width=E(low_maa_percol_net3)$weight,
edge.label=E(low_maa_percol_net3)$weight)

# Create random network from original network

# create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
  rg_reshuffling_w(x, option="links")
}
# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
low_maa_percol_1000rand <- replicate(n, fun(low_maa_percol_net2), simplify=FALSE)

# turn into igraph objects
low_maa_percol_1000rand_igraph <- lapply(low_maa_percol_1000rand, function(x) tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_low_maa_percol_1000rand <- lapply(low_maa_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_low_maa_percol_1000rand <- mean(unlist(cc_low_maa_percol_1000rand))

# Calculate standard deviation of the random dist
cc_sd_low_maa_percol_1000rand <- sd(unlist(cc_low_maa_percol_1000rand))

# look at random nw cc distributions
cc_dist_low_maa_percol_1000rand <- (unlist(cc_low_maa_percol_1000rand))
hist(cc_dist_low_maa_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05
t.test(cc_low_maa_percol_net2, mu = cc_avg_low_maa_percol_1000rand, alternative = 'less')

### Upper Maastrichtian ###

up_maa_percol_net <- read.graph('EDENetwork_outputs/Up/MAA/maa_up_matrix_gen_great3_thresholded_at_086.net', format = "pajek")
typeof(up_maa_percol_net)

cc_up_maa_percol_net <- transitivity(up_maa_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be numeric, not grid names)
# run this since the network is undirected, duplicates ties to justify that there is no direction
if(!is.directed(up_maa_percol_net))
    up_maa_percol_net2 <- symmetrise_w(up_maa_percol_net2)

# make sure conforms to tnet format
up_maa_percol_net2 <- as.tnet(up_maa_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates identical connection but only 2 present...)
# Create igraph object
up_maa_percol_net3 <- tnet_igraph(up_maa_percol_net2,type="weighted one-mode tnet")

# Get summary statistics
summary(up_maa_percol_net3)

# check that the cc is still close
cc_up_maa_percol_net2 <- transitivity(up_maa_percol_net3)
### slightly lower, loss of 1 identical connections (weights = 0) but not very different

## Plot the network

# Create random network from original network
# Create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
    rg_reshuffling_w(x, option="links")
}

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
up_maa_percol_1000rand <- replicate(n, fun(up_maa_percol_net2), simplify=FALSE)

# turn into igraph objects
up_maa_percol_1000rand_igraph <- lapply(up_maa_percol_1000rand, function(x)
    tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_up_maa_percol_1000rand <- lapply(up_maa_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_up_maa_percol_1000rand <- mean(unlist(cc_up_maa_percol_1000rand))

# Calculate standard deviation of the random dist
cc_sd_up_maa_percol_1000rand <- sd(unlist(cc_up_maa_percol_1000rand))

# look at random nw cc distributions
cc_dist_up_maa_percol_1000rand <- (unlist(cc_up_maa_percol_1000rand))
hist(cc_dist_up_maa_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05

t.test(cc_up_maa_percol_net2, mu = cc_avg_up_maa_percol_1000rand, alternative = 'less')

##### Complete DB #####

comp_db_percol_net <-
    read.graph("EDENetowrk_outputs/CompleteDB/trunk60_matrix_gen_great3_thresholded_at_076.net", format = "pajek")

typeof(comp_db_percol_net)

cc_comp_db_percol_net <- transitivity(comp_db_percol_net)

# convert to a i,j,w weighted edgelist (NOTE: Could also do this by getting distance matrix and then
# converting to NW using graph.adjacency() then convert using get.data.frame())

### Get the network data into the right format for tnet to calculate random nws (need the nodes to be
# numeric, not grid names)
comp_db_percol_net2 <- cbind(get.edgelist(comp_db_percol_net, names=FALSE),
                        E(comp_db_percol_net)$weight)

# run this since the network is undirected, dcomplicates ties to justify that there is no direction
if(!is.directed(comp_db_percol_net))
    comp_db_percol_net2 <- symmetrise_w(comp_db_percol_net2)

# make sure conforms to tnet format
comp_db_percol_net2 <- as.tnet(comp_db_percol_net2, type="weighted one-mode tnet")

### Convert from tnet format to igraph (this should remove any zero weights, which essentially indicates
# identical connection but only 2 present...)
# Create igraph object
comp_db_percol_net3 <- tnet_igraph(comp_db_percol_net2, type="weighted one-mode tnet")
# Get summary statistics
summary(comp_db_percol_net3)

# check that the cc is still close
cc_comp_db_percol_net2 <- transitivity(comp_db_percol_net3)
### slightly lower, loss of 16 identical connections (weights = 0) but not very different

## Plot the network

# Get layout of observed network and plot it... need to look into this more...
layout <- layout.fruchterman.reingold(comp_db_percol_net3)

plot(comp_db_percol_net3, main="Observed Network", layout=layout,
vertex.label=1:length(comp_db_percol_net3), edge.width=E(comp_db_percol_net3)$weight,
edge.label=E(comp_db_percol_net3)$weight)

# Create random network from original network

# create function to reshuffle by links (reshuffles weighted links while maintaining negree)
fun <- function(x) {
  rg_reshuffling_w(x, option="links")
}

# set number of times to replicate
n = 1000

# run analysis to create array with different randomized networks
comp_db_percol_1000rand <- replicate(n, fun(comp_db_percol_net2), simplify=FALSE)

# turn into igraph objects
comp_db_percol_1000rand_igraph <- lapply(comp_db_percol_1000rand, function(x)
  tnet_igraph((x),type="weighted one-mode tnet"))

# run cc analysis
cc_comp_db_percol_1000rand <- lapply(comp_db_percol_1000rand_igraph,function(x) transitivity(x))

# get random networks cc average
cc_avg_comp_db_percol_1000rand <- mean(unlist(cc_comp_db_percol_1000rand))

# Calculate standad deviation of the random dist
cc_sd_comp_db_percol_1000rand <- sd(unlist(cc_comp_db_percol_1000rand))

# look at random nw cc distributions
cc_dist_comp_db_percol_1000rand <- (unlist(cc_comp_db_percol_1000rand))
hist(cc_dist_comp_db_percol_1000rand)

### Run statistical test on the real cc using random networks

# H0: cc[real] = mean of cc[rand]
# H1: cc[real] != mean of cc[rand]
# alpha = 0.05

t.test(cc_comp_db_percol_net2, mu = cc_avg_comp_db_percol_1000rand, alternative = 'less')

### Take randomized results (mean and sd) and put in table for comparison with real cc

```r
rand_cc_means <-
c(cc_avg_low_cam_percol_1000rand, cc_avg_mid_cam_percol_1000rand, cc_avg_up_cam_percol_1000rand, cc_avg_low_maa_percol_1000rand, cc_avg_up_maa_percol_1000rand, cc_avg_comp_db_percol_1000rand)
```

```r
rand_cc_sd <-
c(cc_sd_low_cam_percol_1000rand, cc_sd_mid_cam_percol_1000rand, cc_sd_up_cam_percol_1000rand, cc_sd_low_maa_percol_1000rand, cc_sd_up_maa_percol_1000rand, cc_sd_comp_db_percol_1000rand)
```

```r
rand_cc_1sd <- (rand_cc_sd) + rand_cc_means
```

```r
rand_cc_2sd <- (rand_cc_sd * 2) + rand_cc_means
```

```r
rand_cc_3sd <- (rand_cc_sd * 3) + rand_cc_means
```

```r
cc_means <-
c(cc_low_cam_percol_net2, cc_mid_cam_percol_net2, cc_up_cam_percol_net2, cc_low_maa_percol_net2, cc_up_maa_percol_net2, cc_comp_db_percol_net2)
```

```r
c_table <- cbind(cc_means, rand_cc_means, rand_cc_1sd, rand_cc_2sd, rand_cc_3sd)
colnames(c_table) <- c("Networks <CC>", "Mean Rand. <CC>", "1 S.D. Rand. <CC>", "2 S.D. Rand. <CC>", "3 S.D. Rand. <CC>")
rownames(c_table) <- c("Lower Camp.", "Middle Camp.", "Upper Camp.", "Lower Maastr.", "Upper Maastr.", "Complete Database")
```

```r
c_table
```

write.csv(c_table, file="NWR_Outputs/rnw_dist_results_table_1000.csv")

x <- 1:5
set.seed(49)
mat1 <- do.call(rbind, lapply(1:10, function(y) sample(x, 3)))
lapply(1:10, sample(x, 2))

x <- seq(1, 10)

r <- array(dim = c(10, 3, 1))

for (i in 1:1){
\[
\begin{align*}
\text{set.seed}(1) \\
\text{gen_mat} & \leftarrow \text{function}(x) \text{ matrix}(c(1, 1, 1, x + \text{rnorm}(1)), \text{nrow} = 2) \\
\text{gen_mat}(5) \\
\text{n} & \leftarrow 10 \\
\text{# give 1 to gen_mat n-times} \\
\text{lapply(rep}(1, n), \text{gen_mat}) \\
\text{try} & \leftarrow \text{replicate}(1, r) \\
\text{try}[1] \\
\text{ri} & \leftarrow \text{tnet_igraph}(r, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{r1} & \leftarrow \text{tnet_igraph}(r11, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{r2} & \leftarrow \text{tnet_igraph}(r22, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{r3} & \leftarrow \text{tnet_igraph}(r33, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{r4} & \leftarrow \text{tnet_igraph}(r44, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{r5} & \leftarrow \text{tnet_igraph}(r55, \text{type} = \text{"weighted one-mode tnet"}) \\
\text{transitivity}(\text{ri}) \\
\text{transitivity}(\text{r1}) \\
\text{transitivity}(\text{r2}) \\
\text{transitivity}(\text{r3}) \\
\text{transitivity}(\text{r4}) \\
\text{transitivity}(\text{r5}) \\
\text{gg} & \leftarrow \text{get.data.frame}(g) \\
\text{sum}(\text{gg_final}$weight != 0) \\
\text{nrow}(gg) \\
\text{gg_names} & \leftarrow \text{data.frame}(\text{unique}(gg[,1])) \\
\text{colnames}(\text{gg_names}) & \leftarrow \text{"names"} \\
\text{gg_names2} & \leftarrow \text{data.frame}(\text{unique}(gg[,2])) \\
\text{colnames}(\text{gg_names2}) & \leftarrow \text{"names"} \\
\text{gg_names_list} & \leftarrow \text{unique}(\text{rbind}(\text{gg_names}, \text{gg_names2})) \\
\text{gg_names_list} & \leftarrow \text{cbind}(\text{newColName} = \text{rownames}(\text{gg_names_list}), \text{gg_names_list}) \\
\text{rownames}(\text{gg_names_list}) & \leftarrow 1:\text{nrow}(\text{gg_names_list}) \\
\text{gg_names_list}$to & \leftarrow \text{gg_names_list}[2] \\
\text{colnames}(\text{gg_names_list}) & \leftarrow \text{c}(\text{"ID"}, \text{"from"}, \text{"to"}) \\
\text{gg_mergefrom} & \leftarrow \text{merge}(gg, gg_names_list, \text{by} = \text{"from"}, \text{all.x} = \text{TRUE})
\end{align*}
\]
colnames(gg_mergefrom) <- c("from","to","weight","ID","to_y")

gg_mergeto <- merge(gg_mergefrom,gg_names_list,by="to",all.x=TRUE)

head(gg_mergeto)

gg_final <- gg_mergeto[,c(4,6,3)]

h <- subset(gg_final, weight != 0)

colnames(h) <- c("i","j","w")

rnw1 <- rg_reshuffling_w(h, option="links")

rnw1_igraph <- tnet_igraph(rnw,type="weighted one-mode tnet",directed =FALSE)

betweenness(rnw_igraph)

# Look at distribution of weights in the network
low_cam_dist <- read.table("EDENetwork_outputs/Low CAM/low_CAM_distanceMatrix.txt", header = FALSE)

low_cam_names <- read.table("EDENetwork_outputs/Low CAM/low_CAM_distanceMatrix_names.txt",header = FALSE)

summary(low_cam_names)

colnames(low_cam_names) <- "Node"

nrow(low_cam_names)
nrow(low_cam_dist)


colnames(low_cam_dist) <- c(low_cam_names$Node)

rownames(low_cam_dist) <- c(low_cam_names$Node)

low_cam_dist_long <- melt(low_cam_dist)

#make into a numeric table

try <- matrix(data = NA, nrow = dim(low_cam_dist)[1], ncol = dim(low_cam_dist)[2])

colnames(try) <- rownames(low_cam_dist)

rownames(try) <- colnames(low_cam_dist)

for (i in 1:dim(low_cam_dist)[2]) {
  try[,i] <- c(as.numeric(low_cam_dist[[i]]))
}

try

hist(try)

### Node weights not normally distributed, so probably definitely need to resample links from the existing networks, rather than creating them from a dist
Code for Manually Comparing Potential Network Subcomponents

## Code to assess the number and attributes of network components, or clusters of interconnected
nodes, in the substage networks.
## Notes are included throughout which clarify results and interpretations.

setwd("C:/Users/ceara/Documents/Province Project/Occurrence Data/VettedCombinedDB")

#### Lower Campanian ####

# read in text files for different thresholds

cam_low_0.61th <- readLines("EDENetwork_outputs/Low
CAM/CAM_low_matrix_gen_great3_thresholded_at_061.txt")
cam_low_0.55th <- readLines("EDENetwork_outputs/Low
CAM/CAM_low_matrix_gen_great3_thresholded_at_055.txt")
cam_low_0.50th <- readLines("EDENetwork_outputs/Low
CAM/CAM_low_matrix_gen_great3_thresholded_at_050.txt")
cam_low_0.45th <- readLines("EDENetwork_outputs/Low
CAM/CAM_low_matrix_gen_great3_thresholded_at_045.txt")
cam_low_0.40th <- readLines("EDENetwork_outputs/Low
CAM/CAM_low_matrix_gen_great3_thresholded_at_040.txt")

# Check which th have components consisting of multiple nodes

cam_low_0.61th # one major component, all others less than 2 nodes each
cam_low_0.55th # one major component, all others less than 2 nodes each
cam_low_0.50th # one major component, all others less than 2 nodes each
cam_low_0.45th # one major component, all others less than 2 nodes each
cam_low_0.40th # one major component, all others less than 2 nodes each

## Do not need to run further visualization of these components, all in WI for LOW CAM

#### Middle Campanian ####

# read in text files for different thresholds

cam_mid_0.57th <- readLines("EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_057.txt")
cam_mid_0.50th <- readLines("EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_050.txt")
cam_mid_0.45th <- readLines("EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_045.txt")
cam_mid_0.40th <- readLines("EDENetwork_outputs/Mid
CAM/CAM_mid_matrix_gen_great3_thresholded_at_040.txt")

# Check which th have components consisting of multiple nodes

cam_mid_0.57th # one major component, all others less than 2 nodes each
cam_mid_0.50th # one major component, one minor with 6, all others less than 3 nodes each
cam_mid_0.45th # one major component, one minor with 6, one with 4, all others less than 3 nodes each

cam_mid_0.40th # one major component, one minor with 5, all others less than 3 nodes each

## Should map to check location of minor components within WI

#### Upper Campanian ####

# read in text files for different thresholds

cam_up_0.50th <- readLines("EDENetwork_outputs/Up CAM/CAM_up_matrix_gen_great3_thresholded_at_050.txt")
cam_up_0.45th <- readLines("EDENetwork_outputs/Up CAM/CAM_up_matrix_gen_great3_thresholded_at_045.txt")
cam_up_0.40th <- readLines("EDENetwork_outputs/Up CAM/CAM_up_matrix_gen_great3_thresholded_at_040.txt")

# Check which th have components consisting of multiple nodes

cam_up_0.50th # one major component, all others less than 3 nodes each

cam_up_0.45th # one major component, all others less than 3 nodes each

cam_up_0.40th # one major component, all others less than 1 nodes each

## Do not need to run further visualization of these components

#### Lower Maastrichtian ####

# read in text files for different thresholds

maa_low_0.80th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_080.txt")
maa_low_0.75th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_075.txt")
maa_low_0.70th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_070.txt")
maa_low_0.65th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_065.txt")
maa_low_0.60th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_060.txt")
maa_low_0.55th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_055.txt")
maa_low_0.50th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_050.txt")
maa_low_0.45th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_045.txt")
maa_low_0.40th <- readLines("EDENetwork_outputs/Low MAA/MAA_low_matrix_gen_great3_thresholded_at_040.txt")

# Check which th have components consisting of multiple nodes
maa_low_0.80th # one major component, all others less than 1 nodes each
maa_low_0.75th # one major component, one with 7, all others less than 1 nodes each
maa_low_0.70th # one major component, one with 7, all others less than 1 nodes each
maa_low_0.65th # one major component, one with 4, all others less than 2 nodes each
maa_low_0.60th # one major component, one with 4, all others less than 2 nodes each
maa_low_0.55th # one major component, 2 with 7, 1 with 4, all others less than 2 nodes each
maa_low_0.50th # one major component, one with 5, all others less than 3 nodes each
maa_low_0.45th # one component with 5, one with 4, all others less than 2 nodes each
maa_low_0.40th # all component less than 3 nodes each

## Map out 0.75, 0.60, and 0.50 to look at component locations

#### Upper Maastrichtian ####

# read in text files for different thresholds
maa_up_0.86th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_086.txt")
maa_up_0.80th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_080.txt")
maa_up_0.75th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_075.txt")
maa_up_0.70th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_070.txt")
maa_up_0.65th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_065.txt")
maa_up_0.60th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_060.txt")
maa_up_0.55th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_055.txt")
maa_up_0.50th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_050.txt")
maa_up_0.45th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_045.txt")
maa_up_0.40th <- readLines("EDENetwork_outputs/Up MAA/MAA_up_matrix_gen_great3_thresholded_at_040.txt")

# Check which th have components consisting of multiple nodes

maa_up_0.86th # one major component, all others less than 2 nodes
maa_up_0.80th # one major component, 2ndary component with 13 nodes, all others less than 1 nodes each
maa_up_0.75th # one major component, 2ndary component with 13 nodes, all others less than 1 nodes each
maa_up_0.70th # one major component, 2ndary component with 13 nodes, all others less than 1 nodes each
maa_up_0.65th # one major component, 2ndary component with 12 nodes, all others less than 1 nodes each
maa_up_0.60th # one major component, 2ndary component with 12 nodes, all others less than 1 nodes each
maa_up_0.55th # one major component, 2ndary component with 10 nodes, all others less than 1 nodes each
maa_up_0.50th # one major component, 2ndary component with 8 nodes, all others less than 1 nodes each
maa_up_0.45th # all components less than 3 nodes each
maa_up_0.40th # all component less than 3 nodes each

#### Middle Campanian Components for 0.50 to 0.40 thresholds

# Extract component info and map to visualize:

typeof(cam_mid_0.50th) ## Contains: one major component, one minor with 6, all others less than 3 nodes each >> NEED TO LOOK AT FIRST 2 COMPONENTS

# split the character string into lists of component information
cam_mid_0.50th_split <- split(cam_mid_0.50th, cumsum( grepl("^---", cam_mid_0.50th)))
cam_mid_0.50th_split$"0"

# create individual list objects of each relevant component
cam_mid_0.50th_split_1st <- cam_mid_0.50th_split$"0"
cam_mid_0.50th_split_2nd <- cam_mid_0.50th_split$"1"
typeof(cam_mid_0.50th_split_2nd)

# get length of each component (number of rows)
nrow_1st <- length(cam_mid_0.50th_split_1st)
nrow_2nd <- length(cam_mid_0.50th_split_2nd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all components after 1st have extra leading row)
cam_mid_0.50th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
cam_mid_0.50th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into vectors

# 1st component
for (i in 3:length(cam_mid_0.50th_split_1st)){
    cam_mid_0.50th_split_1st_matrix[i-2,] <- unlist(strsplit(cam_mid_0.50th_split_1st[i], "\t"))
}

# 2nd component
for (i in 4:length(cam_mid_0.50th_split_2nd)){
    cam_mid_0.50th_split_2nd_matrix[i-3,] <- unlist(strsplit(cam_mid_0.50th_split_2nd[i], "\t"))
}
# add column names to the matrices
colnames(cam_mid_0.50th_split_1st_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ","Age","betweenness","location","OccColor","AgeColor","ProvColor")
colnames(cam_mid_0.50th_split_2nd_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ","Age","betweenness","location","OccColor","AgeColor","ProvColor")

# make the matrices into data frames
.cam_mid_0.50th_split_1st_df <- as.data.frame(cam_mid_0.50th_split_1st_matrix)
cam_mid_0.50th_split_2nd_df <- as.data.frame(cam_mid_0.50th_split_2nd_matrix)

typeof(cam_mid_0.50th_split_1st_df)

# read in the lat/long file for the substage
.cam_mid_loc <- read.csv("R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_gen_great3_avgloc.csv")
colnames(cam_mid_loc) <- c("Node","y","x","Age")

# merge the dfs together to get the location information with the components
.cam_mid_0.50th_split_1st_merge <- merge(cam_mid_0.50th_split_1st_df,cam_mid_loc,by="Node")
cam_mid_0.50th_split_2nd_merge <- merge(cam_mid_0.50th_split_2nd_df,cam_mid_loc,by="Node")

# simplify to just the x,y coord for each component
cam_mid_0.50th_split_1st_loc <- cam_mid_0.50th_split_1st_merge[,c(12,13)]
cam_mid_0.50th_split_2nd_loc <- cam_mid_0.50th_split_2nd_merge[,c(12,13)]

# plot the map of the different components

library("ggplot2")
theme_set(theme_bw())
library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")

world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map
ggplot(data = world) +
geom_sf(size=0.1) +
xlab("Longitude") + ylab("Latitude") +
annotation_scale(location = "bl", width_hint = 0.5) +
annotation_north_arrow(location = "bl", which_north = "true",
height = unit(0.4,"in"), width = unit(0.3,"in"),
pad_x = unit(0.1, "in"), pad_y = unit(0.3, "in"),
style = north_arrow_minimal) +
annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
   fontface = "italic", color = "grey22", size = 3) +
theme(panel.grid.major = element_line(color = "grey",
   size = 0.3,
   linetype = 2)) +
geom_point(data = cam_mid_0.50th_split_1st_loc, aes(x = x, y = y), size = 3,
   shape = 21, fill = "darkred") +
geom_point(data = cam_mid_0.50th_split_2nd_loc, aes(x = x, y = y), size = 3,
   shape = 21, fill = "lightblue") +
coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/cam_mid_0.50th.pdf")

# Second TH level:

### Contains: one major component, one minor with 6, one with 4, all others less than 3 nodes each >>>

Plot 3 Components

# split the character string into lists of component information

cam_mid_0.45th_split <- split(cam_mid_0.45th, cumsum( grepl("^---", cam_mid_0.45th)))
cam_mid_0.45th_split$"0"

# create individual list objects of each relevant component

cam_mid_0.45th_split_1st <- cam_mid_0.45th_split$"0"
cam_mid_0.45th_split_2nd <- cam_mid_0.45th_split$"1"
cam_mid_0.45th_split_3rd <- cam_mid_0.45th_split$"2"

typeof(cam_mid_0.45th_split_2nd)

# get length of each component (number of rows)
nrow_1st <- length(cam_mid_0.45th_split_1st)
nrow_2nd <- length(cam_mid_0.45th_split_2nd)
nrow_3rd <- length(cam_mid_0.45th_split_3rd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all
# components after 1st have extra leading row)
cam_mid_0.45th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
cam_mid_0.45th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)
cam_mid_0.45th_split_3rd_matrix <- matrix(data = NA, nrow=nrow_3rd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into
# vectors

# 1st component
for (i in 3:length(cam_mid_0.45th_split_1st)){
   cam_mid_0.45th_split_1st_matrix[i-2,] <- unlist(strsplit(cam_mid_0.45th_split_1st[i], "\t"))
}
# 2nd component
for (i in 4:length(cam_mid_0.45th_split_2nd)){
    cam_mid_0.45th_split_2nd_matrix[i-3,] <- unlist(strsplit(cam_mid_0.45th_split_2nd[i], "\t"))
}

# 3rd component
for (i in 4:length(cam_mid_0.45th_split_3rd)){
    cam_mid_0.45th_split_3rd_matrix[i-3,] <- unlist(strsplit(cam_mid_0.45th_split_3rd[i], "\t"))
}

# add column names to the matrices
colnames(cam_mid_0.45th_split_1st_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ","Age","betweenness","location","OccColor","AgeColor","ProvColor")
colnames(cam_mid_0.45th_split_2nd_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ","Age","betweenness","location","OccColor","AgeColor","ProvColor")
colnames(cam_mid_0.45th_split_3rd_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ","Age","betweenness","location","OccColor","AgeColor","ProvColor")

# make the matrices into data frames
.cam_mid_0.45th_split_1st_df <- as.data.frame(cam_mid_0.45th_split_1st_matrix)
cam_mid_0.45th_split_2nd_df <- as.data.frame(cam_mid_0.45th_split_2nd_matrix)
cam_mid_0.45th_split_3rd_df <- as.data.frame(cam_mid_0.45th_split_3rd_matrix)

typeof(cam_mid_0.45th_split_1st_df)

# read in the lat/long file for the substage
cam_mid_loc <- read.csv("R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_gen_great3_avgloc.csv")
colnames(cam_mid_loc) <- c("Node","y","x","Age")

# merge the dfs together to get the location information with the components
cam_mid_0.45th_split_1st_merge <- merge(cam_mid_0.45th_split_1st_df,cam_mid_loc,by="Node")
cam_mid_0.45th_split_2nd_merge <- merge(cam_mid_0.45th_split_2nd_df,cam_mid_loc,by="Node")
cam_mid_0.45th_split_3rd_merge <- merge(cam_mid_0.45th_split_3rd_df,cam_mid_loc,by="Node")

# simplify to just the x,y coord for each component
cam_mid_0.45th_split_1st_loc <- cam_mid_0.45th_split_1st_merge[,c(12,13)]
cam_mid_0.45th_split_2nd_loc <- cam_mid_0.45th_split_2nd_merge[,c(12,13)]
cam_mid_0.45th_split_3rd_loc <- cam_mid_0.45th_split_3rd_merge[,c(12,13)]

# plot the map of the different components
library("ggplot2")
theme_set(theme_bw())
library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")
world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map
ggplot(data = world) +
  geom_sf(size=0.1) +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.5) +
  annotation_north_arrow(location = "bl", which_north = "true",
    height = unit(0.4,"in"), width = unit(0.3,"in"),
    pad_x = unit(0.1, "in"), pad_y = unit(0.3, "in"),
    style = north_arrow_minimal) +
  annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
    fontface = "italic", color = "grey22", size = 3) +
  theme(panel.grid.major = element_line(color = "grey",
    size = 0.3,
    linetype = 2)) +
  geom_point(data = cam_mid_0.45th_split_1st_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "darkred") +
  geom_point(data = cam_mid_0.45th_split_2nd_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "lightblue") +
  geom_point(data = cam_mid_0.45th_split_3rd_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "limegreen") +
  coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/cam_mid_0.45th.pdf")

### NOTE: though the 2nd component points appear to overlap in the same geographic region as in the
0.50 TH level in both maps, they do not
# represent the same nodes. So, it isn't a consistent component through time... Weird that it overlaps
so well spatially but based on different nodes.
# The 3rd component here also doesn't correspond with the 2nd component from the 0.50 TH.

# Third TH level:

### Contains: one major component, one minor with 5, all others less than 3 nodes each >>> NEED TO
PLOT 2 COMPONENTS ONLY

# split the character string into lists of component information
cam_mid_0.40th_split <- split(cam_mid_0.40th, cumsum( grepl("^--", cam_mid_0.40th)))
cam_mid_0.40th_split$"0"

# create individual list objects of each relevant component
cam_mid_0.40th_split_1st <- cam_mid_0.40th_split$"0"
cam_mid_0.40th_split_2nd <- cam_mid_0.40th_split$"1"

typeof(cam_mid_0.40th_split_2nd)
# get length of each component (number of rows)
nrow_1st <- length(cam_mid_0.40th_split_1st)
nrow_2nd <- length(cam_mid_0.40th_split_2nd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all components after 1st have extra leading row)
cam_mid_0.40th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
cam_mid_0.40th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into vectors

# 1st component
for (i in 3:length(cam_mid_0.40th_split_1st)){
cam_mid_0.40th_split_1st_matrix[i-2,] <- unlist(strsplit(cam_mid_0.40th_split_1st[i], "\t"))
}

# 2nd component
for (i in 4:length(cam_mid_0.40th_split_2nd)){
cam_mid_0.40th_split_2nd_matrix[i-3,] <- unlist(strsplit(cam_mid_0.40th_split_2nd[i], "\t"))
}

# add column names to the matrices
colnames(cam_mid_0.40th_split_1st_matrix) <- c("Node","Degree","Clustering", 'Province', 'Numb_Occ', 'Age', 'betweenness', 'location', 'OccColor', 'AgeColor', 'ProvColor')
colnames(cam_mid_0.40th_split_2nd_matrix) <- c("Node","Degree","Clustering", 'Province', 'Numb_Occ', 'Age', 'betweenness', 'location', 'OccColor', 'AgeColor', 'ProvColor')

# make the matrices into data frames
cam_mid_0.40th_split_1st_df <- as.data.frame(cam_mid_0.40th_split_1st_matrix)
cam_mid_0.40th_split_2nd_df <- as.data.frame(cam_mid_0.40th_split_2nd_matrix) # REPRESENTS SAME NODES AS THE 2ND COMPONENT IN THE 0.45 TH (but one less)
typeof(cam_mid_0.40th_split_1st_df)

# read in the lat/long file for the substage
cam_mid_loc <- read.csv('R_Config_for_EN_OUTPUT/Mid CAM/CAM_mid_gen_great3_avgloc.csv')
colnames(cam_mid_loc) <- c("Node","y","x","Age")

# merge the dfs together to get the location information with the components
cam_mid_0.40th_split_1st_merge <- merge(cam_mid_0.40th_split_1st_df,cam_mid_loc,by="Node")
cam_mid_0.40th_split_2nd_merge <- merge(cam_mid_0.40th_split_2nd_df,cam_mid_loc,by="Node")

# simplify to just the x,y coord for each component
cam_mid_0.40th_split_1st_loc <- cam_mid_0.40th_split_1st_merge[,c(12,13)]
cam_mid_0.40th_split_2nd_loc <- cam_mid_0.40th_split_2nd_merge[,c(12,13)]
# plot the map of the different components

library("ggplot2")
theme_set(theme_bw())
library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")

world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map

ggplot(data = world) +
geom_sf(size=0.1) +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.5) +
  annotation_north_arrow(location = "bl", which_north = "true",
                         height = unit(0.4,"in"), width = unit(0.3,"in"),
                         pad_x = unit(0.1, "in"), pad_y = unit(0.3, "in"),
                         style = north_arrow_minimal) +
  annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
           fontface = "italic", color = "grey22", size = 3) +
  theme(panel.grid.major = element_line(color = "grey",
                                        size = 0.3,
                                        linetype = 2)) +
  geom_point(data = cam_mid_0.40th_split_1st_loc, aes(x = x, y = y), size = 3,
             shape = 21, fill = "darkred") +
  geom_point(data = cam_mid_0.40th_split_2nd_loc, aes(x = x, y = y), size = 3,
             shape = 21, fill = "lightblue") +
  coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/cam_mid_0.40th.pdf")

### NOTE: Secondary component in the 0.40 TH same as that of the 0.45 TH, so consistent between the two. Again, not the same as the 2nd Component from # the 0.50 TH, however.

#### Lower Maastrichtian Components for 0.75, 0.60, and 0.50 thresholds

# Extract component info and map to visualize:

typeof(maa_low_0.50th)

### Contains: one major component, one with 5, all others less than 3 nodes each >>> PLOT 2 COMPONENTS
# split the character string into lists of component information
maa_low_0.50th_split <- split(maa_low_0.50th, cumsum( grepl("^---", maa_low_0.50th)))
maa_low_0.50th_split$"0"

# create individual list objects of each relevant component
maa_low_0.50th_split_1st <- maa_low_0.50th_split$"0"
maa_low_0.50th_split_2nd <- maa_low_0.50th_split$"1"

typeof(maa_low_0.50th_split_2nd)

# get length of each component (number of rows)
nrow_1st <- length(maa_low_0.50th_split_1st)
nrow_2nd <- length(maa_low_0.50th_split_2nd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all components after 1st have extra leading row)
maa_low_0.50th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
maa_low_0.50th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into vectors

# 1st component
for (i in 3:length(maa_low_0.50th_split_1st)){
  maa_low_0.50th_split_1st_matrix[i-2,] <- unlist(strsplit(maa_low_0.50th_split_1st[i], "\t"))
}

# 2nd component
for (i in 4:length(maa_low_0.50th_split_2nd)){
  maa_low_0.50th_split_2nd_matrix[i-3,] <- unlist(strsplit(maa_low_0.50th_split_2nd[i], "\t"))
}

# add column names to the matrices
colnames(maa_low_0.50th_split_1st_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ",
'Age', 'betweenness', 'location', 'OccColor', 'AgeColor', 'ProvColor')
colnames(maa_low_0.50th_split_2nd_matrix) <- c("Node","Degree","Clustering","Province",
'Numb_Occ', 'Age', 'betweenness', 'location', 'OccColor', 'AgeColor', 'ProvColor')

# make the matrices into data frames
maa_low_0.50th_split_1st_df <- as.data.frame(maa_low_0.50th_split_1st_matrix)
maa_low_0.50th_split_2nd_df <- as.data.frame(maa_low_0.50th_split_2nd_matrix)
typeof(maa_low_0.50th_split_1st_df)

# read in the lat/long file for the substage
maa_low_loc <- read.csv("R_Config_for_EN_OUTPUT/low maa/maa_low_gen_great3_avgloc.csv")
colnames(maa_low_loc) <- c("Node","y","x","Age")
# merge the dfs together to get the location information with the components
maa_low_0.50th_split_1st_merge <- merge(maa_low_0.50th_split_1st_df, maa_low_loc, by = "Node")
maa_low_0.50th_split_2nd_merge <- merge(maa_low_0.50th_split_2nd_df, maa_low_loc, by = "Node")

# simplify to just the x,y coord for each component
maa_low_0.50th_split_1st_loc <- maa_low_0.50th_split_1st_merge[, c(12, 13)]
maa_low_0.50th_split_2nd_loc <- maa_low_0.50th_split_2nd_merge[, c(12, 13)]

# plot the map of the different components

library("ggplot2")
theme_set(theme_bw())
library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")

world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map
ggplot(data = world) +
  geom_sf(size = 0.1) +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.5) +
  annotation_north_arrow(location = "bl", which_north = "true",
    height = unit(0.4, "in"), width = unit(0.3, "in"),
    pad_x = unit(0.1, "in"), pad_y = unit(0.3, "in"),
    style = north_arrow_minimal) +
  annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
    fontface = "italic", color = "grey22", size = 3) +
  theme(panel.grid.major = element_line(color = "grey",
    size = 0.3, linetype = 2)) +
  geom_point(data = maa_low_0.50th_split_1st_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "darkred") +
  geom_point(data = maa_low_0.50th_split_2nd_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "lightblue") +
  coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/maa_low_0.50th.pdf")

# Second TH level:

### Contains: one major component, one with 4, all others less than 2 nodes each >>> PLOT 2
COMPONENTS
# split the character string into lists of component information
maa_low_0.60th_split <- split(maa_low_0.60th, cumsum( grepl("^---", maa_low_0.60th)))
maa_low_0.60th_split$"0"

# create individual list objects of each relevant component
maa_low_0.60th_split_1st <- maa_low_0.60th_split$"0"
maa_low_0.60th_split_2nd <- maa_low_0.60th_split$"1"

typeof(maa_low_0.60th_split_2nd)

# get length of each component (number of rows)
nrow_1st <- length(maa_low_0.60th_split_1st)
nrow_2nd <- length(maa_low_0.60th_split_2nd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all components after 1st have extra leading row)
maa_low_0.60th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
maa_low_0.60th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into vectors

# 1st component
for (i in 3:length(maa_low_0.60th_split_1st)){
    maa_low_0.60th_split_1st_matrix[i-2,] <- unlist(strsplit(maa_low_0.60th_split_1st[i], "\t"))
}

# 2nd component
for (i in 4:length(maa_low_0.60th_split_2nd)){
    maa_low_0.60th_split_2nd_matrix[i-3,] <- unlist(strsplit(maa_low_0.60th_split_2nd[i], "\t"))
}

# add column names to the matrices
colnames(maa_low_0.60th_split_1st_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ",
'Age','betweenness','location','OccColor','AgeColor','ProvColor')
colnames(maa_low_0.60th_split_2nd_matrix) <- c("Node","Degree","Clustering","Province",
'Numb_Occ','Age','betweenness','location','OccColor','AgeColor','ProvColor')

# make the matrices into data frames
maa_low_0.60th_split_1st_df <- as.data.frame(maa_low_0.60th_split_1st_matrix)
maa_low_0.60th_split_2nd_df <- as.data.frame(maa_low_0.60th_split_2nd_matrix) # NOTE: Not the same component as the 2nd component for the 0.50 TH level

typeof(maa_low_0.60th_split_1st_df)

# read in the lat/long file for the substage
maa_low_loc <- read.csv("R_Config_for_EN_OUTPUT/low maa/maa_low_gen_great3_avgloc.csv")
colnames(maa_low_loc) <- c("Node","y","x","Age")

# merge the dfs together to get the location information with the components
maa_low_0.60th_split_1st_merge <- merge(maa_low_0.60th_split_1st_df,maa_low_loc,by="Node")
maa_low_0.60th_split_2nd_merge <- merge(maa_low_0.60th_split_2nd_df,maa_low_loc,by="Node")

# simplify to just the x,y coord for each component
maa_low_0.60th_split_1st_loc <- maa_low_0.60th_split_1st_merge[,c(12,13)]
maa_low_0.60th_split_2nd_loc <- maa_low_0.60th_split_2nd_merge[,c(12,13)]

# plot the map of the different components

library("ggplot2")
theme_set(theme_bw())
library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")

world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map

ggplot(data = world) +
geom_sf(size=0.1) +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.5) +
  annotation_north_arrow(location = "bl", which_north = "true",
    height = unit(0.4,"in"), width = unit(0.3,"in"),
    pad_x = unit(0.1, "in"), pad_y = unit(0.3, "in"),
    style = north_arrow_minimal) +
  annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
    fontface = "italic", color = "grey22", size = 3) +
  theme(panel.grid.major = element_line(color = "grey",
    size = 0.3,
    linetype = 2)) +
  geom_point(data = maa_low_0.60th_split_1st_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "darkred") +
  geom_point(data = maa_low_0.60th_split_2nd_loc, aes(x = x, y = y), size = 3,
    shape = 21, fill = "lightblue") +
  coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/maa_low_0.60th.pdf")

# Third TH level:
### Contains: one major component, one with 7, all others less than 1 nodes each

**COMPONENTS**

# split the character string into lists of component information
maa_low_0.75th_split <- split(maa_low_0.75th, cumsum(grepl("^---", maa_low_0.75th)))
maa_low_0.75th_split$"0"

# create individual list objects of each relevant component
maa_low_0.75th_split_1st <- maa_low_0.75th_split$"0"
maa_low_0.75th_split_2nd <- maa_low_0.75th_split$"1"

typeof(maa_low_0.75th_split_2nd)

# get length of each component (number of rows)
nrow_1st <- length(maa_low_0.75th_split_1st)
nrow_2nd <- length(maa_low_0.75th_split_2nd)

# create an empty matrix to put the first component in (use length of rows for dimensions, remember all
components after 1st have extra leading row)
maa_low_0.75th_split_1st_matrix <- matrix(data = NA, nrow=nrow_1st-2,ncol=11)
maa_low_0.75th_split_2nd_matrix <- matrix(data = NA, nrow=nrow_2nd-3,ncol=11)

# split the string information from the first component in the list by "\t" and unlist to put values into
vectors

# 1st component
for (i in 3:length(maa_low_0.75th_split_1st)){
  maa_low_0.75th_split_1st_matrix[i-2,] <- unlist(strsplit(maa_low_0.75th_split_1st[i], "\t"))
}

# 2nd component
for (i in 4:length(maa_low_0.75th_split_2nd)){
  maa_low_0.75th_split_2nd_matrix[i-3,] <- unlist(strsplit(maa_low_0.75th_split_2nd[i], "\t"))
}

# add column names to the matrices
colnames(maa_low_0.75th_split_1st_matrix) <- c("Node","Degree","Clustering","Province","Numb_Occ",
'Age','betweenness','location','OccColor','AgeColor','ProvColor')
colnames(maa_low_0.75th_split_2nd_matrix) <- c("Node","Degree","Clustering","Province",
'Numb_Occ','Age','betweenness','location','OccColor','AgeColor','ProvColor')

# make the matrices into data frames
maa_low_0.75th_split_1st_df <- as.data.frame(maa_low_0.75th_split_1st_matrix)
maa_low_0.75th_split_2nd_df <- as.data.frame(maa_low_0.75th_split_2nd_matrix) # NOTE: Same 2nd
component as the 0.60 TH level but with more nodes

typeof(maa_low_0.75th_split_1st_df)
# read in the lat/long file for the substage
maa_low_loc <- read.csv('R_Config_for_EN_OUTPUT/low maa/maa_low_gen_great3_avgloc.csv')
colnames(maa_low_loc) <- c("Node","y","x","Age")

# merge the dfs together to get the location information with the components
maa_low_0.75th_split_1st_merge <- merge(maa_low_0.75th_split_1st_df,maa_low_loc,by="Node")
maa_low_0.75th_split_2nd_merge <- merge(maa_low_0.75th_split_2nd_df,maa_low_loc,by="Node")

# simplify to just the x,y coord for each component
maa_low_0.75th_split_1st_loc <- maa_low_0.75th_split_1st_merge[,c(12,13)]
maa_low_0.75th_split_2nd_loc <- maa_low_0.75th_split_2nd_merge[,c(12,13)]

# plot the map of the different components
library("ggplot2")
theme_set(theme_bw())

library("sf")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")

world <- ne_countries(scale = "medium", returnclass = "sf") # get sf info for whole world from rnaturalearth
class(world)

# Plotting the map
ggplot(data = world) +
  geom_sf(size=0.1) +
  xlab("Longitude") + ylab("Latitude") +
  annotation_scale(location = "bl", width_hint = 0.5) +
  annotation_north_arrow(location = "bl", which_north = "true",
                         height = unit(0.4,"in"), width = unit(0.3,"in"),
                         pad_x = unit(0.1,"in"), pad_y = unit(0.3,"in"),
                         style = north_arrow_minimal) +
  annotate(geom = "text", x = -90, y = 26, label = "Gulf of Mexico",
            fontface = "italic", color = "grey22", size = 3) +
  theme(panel.grid.major = element_line(color = "grey",
                                           size = 0.3, linetype = 2)) +
  geom_point(data = maa_low_0.75th_split_1st_loc, aes(x = x, y = y), size = 3,
             shape = 21, fill = "darkred") +
  geom_point(data = maa_low_0.75th_split_2nd_loc, aes(x = x, y = y), size = 3,
             shape = 21, fill = "lightblue") +
  coord_sf(xlim = c(-120, -75), ylim = c(20, 60), expand = FALSE)
ggsave("ComponentAnalysis/maa_low_0.75th.pdf")
Maps show that the 2nd component for TH levels 0.75-0.60 are the same, located in the eastern side of the GCP. The secondary component in the 0.50 TH, however, is not the same and occupies the middle WIS (assumably it is a result of the WIS major component breaking down at lower THs). At TH below the 0.60 TH, it appears that the GCP breaks down completely to less than 4 connected node components.
Code for Calculating Average Network Link Weights across Distance Bins

setwd("C:/Users/ceara/Documents/Province Project/Occurrence Data/VettedCombinedDB")

# Calculalte geographic distance for paleo-coords between 5deg lats
x0 <- c(-100.0,35.0)
y0 <- c(-100.0,40.0)
x  <- c(-100.0,40.0)
y  <- c(-100.0,45.0)
x1  <- c(-100.0,45.0)
y1  <- c(-100.0,50.0)
x2  <- c(-100.0,50.0)
y2  <- c(-100.0,55.0)

555*2

library(geosphere) # load geosphere package

# load the chronosphere package to run this code. This should allow for reconstructions of plat and plong:
# https://www.evolv-ed.net/post/chronosphere-paleomap/chronosphere-paleomap/
library(chronosphere)
citation("geosphere")

##### Lower Campanian #####

# read in the lat/long file for the substage
cam_low_loc <- read.csv('R_Config_for_EN_OUTPUT/Low CAM/CAM_low_gen_great3_avgloc.csv')

# get just the lat long locations from the table
.cam_low_simple_loc <- cam_low_loc[,c(3,2)]

# NOTE: Remember that these are average locations based on the fossil occ found in each grid cell,
# and therefore already approximations
# which may be biased by not converting the fossil locations to paleocoordinates originally...

# Get the map information for this substage (age chosen is very approximate for this substage because limited options)
cam_low_maps <- reconstruct("plates", age=80) # get paleo plate info
cam_low_coast <- reconstruct("coastlines", age=80) # get coastline info

# Reconstruct paleocoord using age of rough age of substage (approximate, will use the same age for M Cam because limited options...)
cam_low_ploc <- reconstruct(cam_low_simple_loc, age=c(80))

# get min and max paleolong
cam_low_minlong <- min(cam_low_ploc[,1]) - 10
cam_low_maxlong <- max(cam_low_ploc[,1]) + 10

# get min and max paleolat
cam_low_minlat <- min(cam_low_ploc[,2]) - 15
cam_low_maxlat <- max(cam_low_ploc[,2]) + 15

# plot the maps and points in paleo positions to get an idea about their layout
par(mfrow = c(1,1))
plot(cam_low_maps, col = 'grey', border = NA, xlim = c(cam_low_minlong,cam_low_maxlong), ylim = c(cam_low_minlat,cam_low_maxlat))
lines(cam_low_coast, col = "black", xlim = c(cam_low_minlong,cam_low_maxlong), ylim = c(cam_low_maxlat,cam_low_minlat))
points(cam_low_ploc, pch=3, col="red")

# get number of unique locations to use for matrix dimensions
low_cam_count <- nrow(cam_low_loc)

# create empty dataframe to put the geographic distances into (use the names of the different grid cells as row and column names)
low_cam_geodist_matrix <- data.frame(matrix(ncol = low_cam_count, nrow = low_cam_count))
colnames(low_cam_geodist_matrix) <- cam_low_loc$node_label # give the matrix column names based on original grid names
rownames(low_cam_geodist_matrix) <- cam_low_loc$node_label # give the matrix row names based on original grid names

# run for loop to get distances using geodesic distance calc
for (i in 1:nrow(cam_low_ploc)){
  low_cam_geodist_matrix[i,] <- distGeo(cam_low_simple_loc[i,], cam_low_simple_loc, a=6378137, f=1/298.257223563)
}

# results are in meters of distance!!!

# check the dimensions and summary info for the matrix
ncol(low_cam_geodist_matrix)
head(low_cam_geodist_matrix)

max(low_cam_geodist_matrix)
min(low_cam_geodist_matrix)
mean(as.matrix(low_cam_geodist_matrix), na.rm = TRUE)
median(as.matrix(low_cam_geodist_matrix), na.rm = TRUE)

# make the matrix into a lower triangle (to make following calc less cumbersome and remove duplicates)
low_cam_geodist_matrix[lower.tri(low_cam_geodist_matrix, diag = TRUE)] <- NA
low_cam_geodist_matrix
# create a sequence of numbers to use for making bins (these bins were chosen to try to capture variation in values while minimizing bin number)
seq <- c(0,50000, 100000, 200000, 400000, 600000, 1000000, 1500000, 2000000, 2500000, 3000000, 3500000)

# read in the Jaccard dist matrix created using EDENetworks
low_cam_dist <- read.table("EDENetwork_outputs/Low CAM/low_CAM_distanceMatrix.txt", header = FALSE) # jaccard values matrix

low_cam_names <- read.table("EDENetwork_outputs/Low CAM/low_CAM_distanceMatrix_names.txt",header = FALSE) # node names vector
colnames(low_cam_names) <- "Node"

# give the rows and columns names based on the node names file
colnames(low_cam_dist) <- c(low_cam_names$Node)
rownames(low_cam_dist) <- c(low_cam_names$Node)

# check that the two matrices have an identical arrangement
a <- colnames(low_cam_geodist_matrix)
b <- colnames(low_cam_dist)
a == b # They do, yay!

#### Calculate the average and mean jaccard weights for the different bins

## 0-50 km ##

# get index positions of all values in geographic distance matrix within 0-50 km
low_cam_geodist_0_50 = which(low_cam_geodist_matrix <50000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_cam_geodist_0_50_jaccard <-
low_cam_dist[c(low_cam_geodist_0_50[,1]),c(low_cam_geodist_0_50[,2])]

# get the average and median jaccard weight values for the 0-50 km bin
low_cam_geodist_0_50_avg <- mean(as.matrix(low_cam_geodist_0_50_jaccard), na.rm = TRUE)
low_cam_geodist_0_50_med <- median(as.matrix(low_cam_geodist_0_50_jaccard), na.rm = TRUE)

## 50-100 km ##

# get index positions of all values in geographic distance matrix within 50-100 km
low_cam_geodist_50_100 = which(low_cam_geodist_matrix >50000 & low_cam_geodist_matrix <100000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_cam_geodist_50_100_jaccard <-
low_cam_dist[c(low_cam_geodist_50_100[,1]),c(low_cam_geodist_50_100[,2])]

# get the average and median jaccard weight values for the 50-100 km bin
low_cam_geodist_50_100_avg <- mean(as.matrix(low_cam_geodist_50_100_jaccard), na.rm = TRUE)
low_cam_geodist_50_100_med <- median(as.matrix(low_cam_geodist_50_100_jaccard), na.rm = TRUE)
## 100-200 km ##

# get index positions of all values in geographic distance matrix within 100-200 km
low_cam_geodist_100_200 = which(low_cam_geodist_matrix > 100000 & low_cam_geodist_matrix < 200000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_100_200_jaccard <-
low_cam_dist[c(low_cam_geodist_100_200[1,], low_cam_geodist_100_200[2,])]

# get the average and median jaccard weight values for the 100-200 km bin
low_cam_geodist_100_200_avg <- mean(as.matrix(low_cam_geodist_100_200_jaccard), na.rm = TRUE)
low_cam_geodist_100_200_med <- median(as.matrix(low_cam_geodist_100_200_jaccard), na.rm = TRUE)

## 200-400 km ##

# get index positions of all values in geographic distance matrix within 200-400 km
low_cam_geodist_200_400 = which(low_cam_geodist_matrix > 200000 & low_cam_geodist_matrix < 400000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_200_400_jaccard <-
low_cam_dist[c(low_cam_geodist_200_400[1,], low_cam_geodist_200_400[2,])]

# get the average and median jaccard weight values for the 200-400 km bin
low_cam_geodist_200_400_avg <- mean(as.matrix(low_cam_geodist_200_400_jaccard), na.rm = TRUE)
low_cam_geodist_200_400_med <- median(as.matrix(low_cam_geodist_200_400_jaccard), na.rm = TRUE)

## 400-600 km ##

# get index positions of all values in geographic distance matrix within 400-600 km
low_cam_geodist_400_600 = which(low_cam_geodist_matrix > 400000 & low_cam_geodist_matrix < 600000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_400_600_jaccard <-
low_cam_dist[c(low_cam_geodist_400_600[1,], low_cam_geodist_400_600[2,])]

# get the average and median jaccard weight values for the 400-600 km bin
low_cam_geodist_400_600_avg <- mean(as.matrix(low_cam_geodist_400_600_jaccard), na.rm = TRUE)
low_cam_geodist_400_600_med <- median(as.matrix(low_cam_geodist_400_600_jaccard), na.rm = TRUE)

## 600-1000 km ##

# get index positions of all values in geographic distance matrix within 600-1000 km
low_cam_geodist_600_1000 = which(low_cam_geodist_matrix > 600000 & low_cam_geodist_matrix < 1000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_600_1000_jaccard <-
low_cam_dist[c(low_cam_geodist_600_1000[1,], low_cam_geodist_600_1000[2,])]

## 1000-2000 km ##

# get index positions of all values in geographic distance matrix within 1000-2000 km
low_cam_geodist_1000_2000 = which(low_cam_geodist_matrix > 1000000 & low_cam_geodist_matrix < 2000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_1000_2000_jaccard <-
low_cam_dist[c(low_cam_geodist_1000_2000[1,], low_cam_geodist_1000_2000[2,])]

# get the average and median jaccard weight values for the 1000-2000 km bin
low_cam_geodist_1000_2000_avg <- mean(as.matrix(low_cam_geodist_1000_2000_jaccard), na.rm = TRUE)
low_cam_geodist_1000_2000_med <- median(as.matrix(low_cam_geodist_1000_2000_jaccard), na.rm = TRUE)
# get the average and median jaccard weight values for the 600-1000 km bin
low_cam_geodist_600_1000_avg <- mean(as.matrix(low_cam_geodist_600_1000_jaccard), na.rm = TRUE)
low_cam_geodist_600_1000_med <- median(as.matrix(low_cam_geodist_600_1000_jaccard), na.rm = TRUE)

## 1000-1500 km ##

# get index positions of all values in geographic distance matrix within 1000-1500 km
low_cam_geodist_1000_1500 = which(low_cam_geodist_matrix >1000000 & low_cam_geodist_matrix < 1500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_cam_geodist_1000_1500_jaccard <-
low_cam_dist[c(low_cam_geodist_1000_1500[,1]),c(low_cam_geodist_1000_1500[,2])]

# get the average and median jaccard weight values for the 1000-1500 km bin
low_cam_geodist_1000_1500_avg <- mean(as.matrix(low_cam_geodist_1000_1500_jaccard), na.rm = TRUE)
low_cam_geodist_1000_1500_med <- median(as.matrix(low_cam_geodist_1000_1500_jaccard), na.rm = TRUE)

## 1500-2000 km ##

# get index positions of all values in geographic distance matrix within 1500-2000 km
low_cam_geodist_1500_2000 = which(low_cam_geodist_matrix >1500000 & low_cam_geodist_matrix < 2000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_cam_geodist_1500_2000_jaccard <-
low_cam_dist[c(low_cam_geodist_1500_2000[,1]),c(low_cam_geodist_1500_2000[,2])]

# get the average and median jaccard weight values for the 1500-2000 km bin
low_cam_geodist_1500_2000_avg <- mean(as.matrix(low_cam_geodist_1500_2000_jaccard), na.rm = TRUE)
low_cam_geodist_1500_2000_med <- median(as.matrix(low_cam_geodist_1500_2000_jaccard), na.rm = TRUE)

## 2000-2500 km ##

# get index positions of all values in geographic distance matrix within 2000-2500 km
low_cam_geodist_2000_2500 = which(low_cam_geodist_matrix >2000000 & low_cam_geodist_matrix < 2500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_cam_geodist_2000_2500_jaccard <-
low_cam_dist[c(low_cam_geodist_2000_2500[,1]),c(low_cam_geodist_2000_2500[,2])]

# get the average and median jaccard weight values for the 2000-2500 km bin
low_cam_geodist_2000_2500_avg <- mean(as.matrix(low_cam_geodist_2000_2500_jaccard), na.rm = TRUE)
low_cam_geodist_2000_2500_med <- median(as.matrix(low_cam_geodist_2000_2500_jaccard), na.rm = TRUE)
## 2500-3000 km ##

# get index positions of all values in geographic distance matrix within 2500-2500 km
low_cam_geodist_2500_3000 = which(low_cam_geodist_matrix > 2500000 & low_cam_geodist_matrix < 3000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_2500_3000_jaccard <- low_cam_dist[c(low_cam_geodist_2500_3000[,1]),c(low_cam_geodist_2500_3000[,2])]

# get the average and median jaccard weight values for the 2500-2500 km bin
low_cam_geodist_2500_3000_avg <- mean(as.matrix(low_cam_geodist_2500_3000_jaccard), na.rm = TRUE)
low_cam_geodist_2500_3000_med <- median(as.matrix(low_cam_geodist_2500_3000_jaccard), na.rm = TRUE)

## 3000-3500 km ##

# get index positions of all values in geographic distance matrix within 3000-3500 km
low_cam_geodist_3000_3500 = which(low_cam_geodist_matrix > 3000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
low_cam_geodist_3000_3500_jaccard <- low_cam_dist[c(low_cam_geodist_3000_3500[,1]),c(low_cam_geodist_3000_3500[,2])]

# get the average and median jaccard weight values for the 3000-3500 km bin
low_cam_geodist_3000_3500_avg <- mean(as.matrix(low_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
low_cam_geodist_3000_3500_med <- median(as.matrix(low_cam_geodist_3000_3500_jaccard), na.rm = TRUE)

low_cam_freq_geodist_bins <- c(nrow(low_cam_geodist_0_50),
    nrow(low_cam_geodist_50_100),
    nrow(low_cam_geodist_100_200),
    nrow(low_cam_geodist_200_400),
    nrow(low_cam_geodist_400_600),
    nrow(low_cam_geodist_600_1000),
    nrow(low_cam_geodist_1000_1500),
    nrow(low_cam_geodist_1500_2000),
    nrow(low_cam_geodist_2000_2500),
    nrow(low_cam_geodist_2500_3000))

low_cam_geodist_avg_bins <- c(low_cam_geodist_0_50_avg,
    low_cam_geodist_50_100_avg,
    low_cam_geodist_100_200_avg,
    low_cam_geodist_200_400_avg,
    low_cam_geodist_400_600_avg,
    low_cam_geodist_600_1000_avg,
    low_cam_geodist_1000_1500_avg,
    low_cam_geodist_1500_2000_avg,
    low_cam_geodist_2000_2500_avg,
    low_cam_geodist_2500_3000_avg,
low_cam_geodist_3000_3500_avg)

low_cam_geodist_med_bins <- c(low_cam_geodist_0_50_med,
                              low_cam_geodist_50_100_med,
                              low_cam_geodist_100_200_med,
                              low_cam_geodist_200_400_med,
                              low_cam_geodist_400_600_med,
                              low_cam_geodist_600_1000_med,
                              low_cam_geodist_1000_1500_med,
                              low_cam_geodist_1500_2000_med,
                              low_cam_geodist_2000_2500_med,
                              low_cam_geodist_2500_3000_med,
                              low_cam_geodist_3000_3500_med)

low_cam_geodist_bins_table <- rbind(low_cam_geodist_avg_bins, low_cam_geodist_med_bins, low_cam_freq_geodist_bins)

colnames(low_cam_geodist_bins_table) <- c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km","600-1000 km",
                                           "1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km")

rownames(low_cam_geodist_bins_table) <- c("Mean Jaccard Weight","Median Jaccard Weight","Numb. Sites in Bin")

low_cam_geodist_bins_table

write.csv(low_cam_geodist_bins_table, file = 'WeightAnalysis/low_cam_geodist_bins_table.csv')

### Plot the lines ######

plot(low_cam_geodist_bins_table[1,],col="blue", pch=16, axes=FALSE, ann=FALSE)
axis(1, at=1:11, lab=c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km",
                      "600-1000 km","1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km"))
axis(2, at=c(0.7,0.75,0.8,0.85,0.9,0.95,1))
box()
title(xlab="Distance Bins (km)")
title(ylab="Bray–Curtis Dissimilarity Index Weight ")
title(main="Average Weight by Geographic Distance")

##### Middle Campanian ######

# read in the lat/long file for the substage

# get just the lat long locations from the table

cam_mid_simple_loc <- cam_mid_loc[,c(3,2)]

# NOTE: Remember that these are average locations based on the fossil occ found in each grid cell, and therefore already approximations
# which may be biased by not converting the fossil locations to paleocoordinates originally...
# Get the map information for this substage (age chosen is very approximate for this substage because limited options)
cam_mid_maps <- reconstruct("plates", age=80) # get paleo plate info
cam_mid_coast <- reconstruct("coastlines", age=80) # get coastline info

# Reconstruct paleocoord using age of rough age of substage (approximate, will use the same age for M Cam because limited options...)
cam_mid_ploc <- reconstruct(cam_mid_simple_loc, age=c(80))

# get min and max paleolong
cam_mid_minlong <- min(cam_mid_ploc[,1]) - 10
.cam_mid_maxlong <- max(cam_mid_ploc[,1]) + 10

# get min and max paleolat
.cam_mid_minlat <- min(cam_mid_ploc[,2]) - 15
.cam_mid_maxlat <- max(cam_mid_ploc[,2]) + 15

# plot the maps and points in paleo positions to get an idea about their layout
par(mfrow = c(1,1))
plot(cam_mid_maps, col = 'grey', border = NA, xlim = c(cam_mid_minlong,cam_mid_maxlong), ylim = c(cam_mid_minlat,cam_mid_maxlat))
lines(cam_mid_coast, col = "black", xlim = c(cam_mid_minlong,cam_mid_maxlong), ylim = c(cam_mid_maxlat,cam_mid_minlat))
points(cam_mid_ploc, pch=3, col="red")

# get number of unique locations to use for matrix dimensions
mid_cam_count <- nrow(cam_mid_loc)

# create empty dataframe to put the geographic distances into (use the names of the different grid cells as row and column names)
mid_cam_geodist_matrix <- data.frame(matrix(ncol = mid_cam_count, nrow = mid_cam_count))
colnames(mid_cam_geodist_matrix) <- cam_mid_loc$node_label # give the matrix column names based on original grid names
rownames(mid_cam_geodist_matrix) <- cam_mid_loc$node_label # give the matrix row names based on original grid names

# run for loop to get distances using geodesic distance calc
for (i in 1:nrow(cam_mid_ploc)){
  mid_cam_geodist_matrix[i,] <- distGeo(cam_mid_simple_loc[i,],cam_mid_simple_loc, a=6378137, f=1/298.257223563)
}

# results are in meters of distance!!!

# check the dimensions and summary info for the matrix
ncol(mid_cam_geodist_matrix)
head(mid_cam_geodist_matrix)

max(mid_cam_geodist_matrix)
min(mid_cam_geodist_matrix)
mean(as.matrix(mid_cam_geodist_matrix), na.rm = TRUE)
median(as.matrix(mid_cam_geodist_matrix), na.rm = TRUE)

# make the matrix into a mider triangle (to make folmiding calc less cumbersome and remove duplicates)
mid_cam_geodist_matrix[lower.tri(mid_cam_geodist_matrix, diag = TRUE)] <- NA
mid_cam_geodist_matrix

# create a sequence of numbers to use for making bins (these bins were chosen to try to capture variation in values while minimizing bin number)
seq <- c(0,50000, 100000, 200000, 400000, 600000, 1000000, 1500000, 2000000, 2500000, 3000000, 3500000)

# read in the Jaccard dist matrix created using EDENetworks
mid_cam_dist <- read.table("EDENetwork_outputs/mid CAM/CAM_mid_Matrix_distance.txt", header = FALSE) # jaccard values matrix
mid_cam_names <- read.table("EDENetwork_outputs/mid CAM/CAM_mid_Matrix_distance_names.txt",header = FALSE) # node names vector
colnames(mid_cam_names) <- "Node"

# give the rows and columns names based on the node names file
colnames(mid_cam_dist) <- c(mid_cam_names$Node)
rownames(mid_cam_dist) <- c(mid_cam_names$Node)

# check that the two matrices have an identical arrangement
a <- colnames(mid_cam_geodist_matrix)
b <- colnames(mid_cam_dist)
a == b # They do, yay!

#### Calculate the average and mean jaccard weights for the different bins

## 0-50 km ##

# get index positions of all values in geographic distance matrix within 0-50 km
mid_cam_geodist_0_50 = which(mid_cam_geodist_matrix <50000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_0_50_jaccard <- mid_cam_dist[cbind(mid_cam_geodist_0_50[,1]),cbind(mid_cam_geodist_0_50[,2])]

# get the average and median jaccard weight values for the 0-50 km bin
mid_cam_geodist_0_50_avg <- mean(as.matrix(mid_cam_geodist_0_50_jaccard), na.rm = TRUE)
mid_cam_geodist_0_50_med <- median(as.matrix(mid_cam_geodist_0_50_jaccard), na.rm = TRUE)

## 50-100 km ##

# get index positions of all values in geographic distance matrix within 50-100 km
mid_cam_geodist_50_100 = which(mid_cam_geodist_matrix >50000 & mid_cam_geodist_matrix < 1000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_50_100_jaccard <-
mid_cam_dist[c(mid_cam_geodist_50_100[,1]),c(mid_cam_geodist_50_100[,2])]

# get the average and median jaccard weight values for the 50-100 km bin
mid_cam_geodist_50_100_avg <- mean(as.matrix(mid_cam_geodist_50_100_jaccard), na.rm = TRUE)
mid_cam_geodist_50_100_med <- median(as.matrix(mid_cam_geodist_50_100_jaccard), na.rm = TRUE)

## 100-200 km ##

# get index positions of all values in geographic distance matrix within 100-200 km
mid_cam_geodist_100_200 = which(mid_cam_geodist_matrix >100000 & mid_cam_geodist_matrix < 200000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_100_200_jaccard <-
mid_cam_dist[c(mid_cam_geodist_100_200[,1]),c(mid_cam_geodist_100_200[,2])]

# get the average and median jaccard weight values for the 100-200 km bin
mid_cam_geodist_100_200_avg <- mean(as.matrix(mid_cam_geodist_100_200_jaccard), na.rm = TRUE)
mid_cam_geodist_100_200_med <- median(as.matrix(mid_cam_geodist_100_200_jaccard), na.rm = TRUE)

## 200-400 km ##

# get index positions of all values in geographic distance matrix within 200-400 km
mid_cam_geodist_200_400 = which(mid_cam_geodist_matrix >200000 & mid_cam_geodist_matrix < 400000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_200_400_jaccard <-
mid_cam_dist[c(mid_cam_geodist_200_400[,1]),c(mid_cam_geodist_200_400[,2])]

# get the average and median jaccard weight values for the 200-400 km bin
mid_cam_geodist_200_400_avg <- mean(as.matrix(mid_cam_geodist_200_400_jaccard), na.rm = TRUE)
mid_cam_geodist_200_400_med <- median(as.matrix(mid_cam_geodist_200_400_jaccard), na.rm = TRUE)

## 400-600 km ##

# get index positions of all values in geographic distance matrix within 400-600 km
mid_cam_geodist_400_600 = which(mid_cam_geodist_matrix >400000 & mid_cam_geodist_matrix < 600000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_400_600_jaccard <-
mid_cam_dist[c(mid_cam_geodist_400_600[,1]),c(mid_cam_geodist_400_600[,2])]

# get the average and median jaccard weight values for the 400-600 km bin
mid_cam_geodist_400_600_avg <- mean(as.matrix(mid_cam_geodist_400_600_jaccard), na.rm = TRUE)
mid_cam_geodist_400_600_med <- median(as.matrix(mid_cam_geodist_400_600_jaccard), na.rm = TRUE)

### 600-1000 km ###
# get index positions of all values in geographic distance matrix within 600-1000 km
mid_cam_geodist_600_1000 = which(mid_cam_geodist_matrix > 600000 & mid_cam_geodist_matrix < 1000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_600_1000_jaccard <- mid_cam_dist[c(mid_cam_geodist_600_1000[,1]), c(mid_cam_geodist_600_1000[,2])]
# get the average and median jaccard weight values for the 600-1000 km bin
mid_cam_geodist_600_1000_avg <- mean(as.matrix(mid_cam_geodist_600_1000_jaccard), na.rm = TRUE)
mid_cam_geodist_600_1000_med <- median(as.matrix(mid_cam_geodist_600_1000_jaccard), na.rm = TRUE)

### 1000-1500 km ###
# get index positions of all values in geographic distance matrix within 1000-1500 km
mid_cam_geodist_1000_1500 = which(mid_cam_geodist_matrix > 1000000 & mid_cam_geodist_matrix < 1500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_1000_1500_jaccard <- mid_cam_dist[c(mid_cam_geodist_1000_1500[,1]), c(mid_cam_geodist_1000_1500[,2])]
# get the average and median jaccard weight values for the 1000-1500 km bin
mid_cam_geodist_1000_1500_avg <- mean(as.matrix(mid_cam_geodist_1000_1500_jaccard), na.rm = TRUE)
mid_cam_geodist_1000_1500_med <- median(as.matrix(mid_cam_geodist_1000_1500_jaccard), na.rm = TRUE)

### 1500-2000 km ###
# get index positions of all values in geographic distance matrix within 1500-2000 km
mid_cam_geodist_1500_2000 = which(mid_cam_geodist_matrix > 1500000 & mid_cam_geodist_matrix < 2000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_1500_2000_jaccard <- mid_cam_dist[c(mid_cam_geodist_1500_2000[,1]), c(mid_cam_geodist_1500_2000[,2])]
# get the average and median jaccard weight values for the 1500-2000 km bin
mid_cam_geodist_1500_2000_avg <- mean(as.matrix(mid_cam_geodist_1500_2000_jaccard), na.rm = TRUE)
mid_cam_geodist_1500_2000_med <- median(as.matrix(mid_cam_geodist_1500_2000_jaccard), na.rm = TRUE)

## 2000-2500 km ##
# get index positions of all values in geographic distance matrix within 2000-2500 km
mid_cam_geodist_2000_2500 = which(mid_cam_geodist_matrix > 2000000 & mid_cam_geodist_matrix < 2500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_2000_2500_jaccard <- mid_cam_dist[c(mid_cam_geodist_2000_2500[,1]),c(mid_cam_geodist_2000_2500[,2])]
# get the average and median jaccard weight values for the 2000-2500 km bin
mid_cam_geodist_2000_2500_avg <- mean(as.matrix(mid_cam_geodist_2000_2500_jaccard), na.rm = TRUE)
mid_cam_geodist_2000_2500_med <- median(as.matrix(mid_cam_geodist_2000_2500_jaccard), na.rm = TRUE)

## 2500-3000 km ##
# get index positions of all values in geographic distance matrix within 2500-2500 km
mid_cam_geodist_2500_3000 = which(mid_cam_geodist_matrix > 2500000 & mid_cam_geodist_matrix < 3000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_2500_3000_jaccard <- mid_cam_dist[c(mid_cam_geodist_2500_3000[,1]),c(mid_cam_geodist_2500_3000[,2])]
# get the average and median jaccard weight values for the 2500-2500 km bin
mid_cam_geodist_2500_3000_avg <- mean(as.matrix(mid_cam_geodist_2500_3000_jaccard), na.rm = TRUE)
mid_cam_geodist_2500_3000_med <- median(as.matrix(mid_cam_geodist_2500_3000_jaccard), na.rm = TRUE)

## 3000-3500 km ##
# get index positions of all values in geographic distance matrix within 3000-3500 km
mid_cam_geodist_3000_3500 = which(mid_cam_geodist_matrix > 3000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
mid_cam_geodist_3000_3500_jaccard <- mid_cam_dist[c(mid_cam_geodist_3000_3500[,1]),c(mid_cam_geodist_3000_3500[,2])]
# get the average and median jaccard weight values for the 3000-3500 km bin
mid_cam_geodist_3000_3500_avg <- mean(as.matrix(mid_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
mid_cam_geodist_3000_3500_med <- median(as.matrix(mid_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
mid_cam_freq_geodist_bins <- c(nrow(mid_cam_geodist_0_50),
nrow(mid_cam_geodist_50_100),
nrow(mid_cam_geodist_100_200),
nrow(mid_cam_geodist_200_400),
nrow(mid_cam_geodist_400_600),
nrow(mid_cam_geodist_600_1000),
nrow(mid_cam_geodist_1000_1500),
nrow(mid_cam_geodist_1500_2000),
nrow(mid_cam_geodist_2000_2500),
nrow(mid_cam_geodist_2500_3000),
nrow(mid_cam_geodist_3000_3500))

mid_cam_geodist_avg_bins <- c(mid_cam_geodist_0_50_avg,
mid_cam_geodist_50_100_avg,
mid_cam_geodist_100_200_avg,
mid_cam_geodist_200_400_avg,
mid_cam_geodist_400_600_avg,
mid_cam_geodist_600_1000_avg,
mid_cam_geodist_1000_1500_avg,
mid_cam_geodist_1500_2000_avg,
mid_cam_geodist_2000_2500_avg,
mid_cam_geodist_2500_3000_avg,
mid_cam_geodist_3000_3500_avg)

mid_cam_geodist_med_bins <- c(mid_cam_geodist_0_50_med,
mid_cam_geodist_50_100_med,
mid_cam_geodist_100_200_med,
mid_cam_geodist_200_400_med,
mid_cam_geodist_400_600_med,
mid_cam_geodist_600_1000_med,
mid_cam_geodist_1000_1500_med,
mid_cam_geodist_1500_2000_med,
mid_cam_geodist_2000_2500_med,
mid_cam_geodist_2500_3000_med,
mid_cam_geodist_3000_3500_med)

mid_cam_geodist_bins_table <-
  rbind(mid_cam_geodist_avg_bins,mid_cam_geodist_med_bins,mid_cam_freq_geodist_bins)

colnames(mid_cam_geodist_bins_table) <- c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km","600-1000 km",
    "1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km")
rownames(mid_cam_geodist_bins_table) <- c("Mean Jaccard Weight","Median Jaccard Weight","Numb. Sites in Bin")

mid_cam_geodist_bins_table

write.csv(mid_cam_geodist_bins_table, file = 'WeightAnalysis/mid_cam_geodist_bins_table.csv')

### Plot the lines ######
plot(low_cam_geodist_bins_table[1,,]col="blue",pch=16,axes=FALSE, ann=FALSE)
points(mid_cam_geodist_bins_table[1,,]col="red",pch=17)
axis(1, at=1:11, lab=c("0-50 ","50-100","100-200","200-400","400-600","600-1000",
"1000-1500","1500-2000","2000-2500","2500-3000","3000-3500"))
axis(2, at=c(0.7,0.75,0.80,0.85,0.90,0.95,1))
box()
title(xlab="Distance Bins (km)")
title(ylab="Bray–Curtis Dissimilarity Index Weight ")
title(main="Average Weight by Geographic Distance")

##### Upper Campanian #####

# read in the lat/long file for the substage
cam_up_loc <- read.csv('R_Config_for_EN_OUTPUT/up CAM/CAM_up_gen_great3_avgloc.csv')

# get just the lat long locations from the table
cam_up_simple_loc <- cam_up_loc[,c(3,2)]

# NOTE: Remember that these are average locations based on the fossil occ found in each grid cell, and therefor already approximations
# which may be biased by not converting the fossil locations to paleocoordinates originally...

# Get the map information for this substage (age chosen is very approximate for this substage because limited options)
cam_up_maps <- reconstruct("plates", age=75) # get paleo plate info
cam_up_coast <- reconstruct("coastlines", age=75) # get coastline info

# Reconstruct paleocoord using age of rough age of substage (approximate)
cam_up_ploc <- reconstruct(cam_up_simple_loc, age=c(75))

# get min and max paleolong
cam_up_minlong <- min(cam_up_ploc[,1]) - 10
cam_up_maxlong <- max(cam_up_ploc[,1]) + 10

# get min and max paleolat
cam_up_minlat <- min(cam_up_ploc[,2]) - 15
cam_up_maxlat <- max(cam_up_ploc[,2]) + 15

# plot the maps and points in paleo positions to get an idea about their layout
par(mfrow = c(1,1))
plot(cam_up_maps, col = 'grey', border = NA, xlim = c(cam_up_minlong,cam_up_maxlong), ylim = c(cam_up_minlat,cam_up_maxlat))
lines(cam_up_coast, col = "black", xlim = c(cam_up_minlat,cam_up_maxlat), ylim = c(cam_up_maxlat,cam_up_minlat))
points(cam_up_ploc, pch=3, col="red")

# get number of unique locations to use for matrix dimensions
up_cam_count <- nrow(cam_up_loc)
# create empty dataframe to put the geographic distances into (use the names of the different grid cells as row and column names)
up_cam_geodist_matrix <- data.frame(matrix(ncol = up_cam_count, nrow = up_cam_count))
colnames(up_cam_geodist_matrix) <- cam_up_loc$node_label # give the matrix column names based on original grid names
rownames(up_cam_geodist_matrix) <- cam_up_loc$node_label # give the matrix row names based on original grid names

# run for loop to get distances using geodesic distance calc
for (i in 1:nrow(cam_up_ploc)){
  up_cam_geodist_matrix[i,] <- distGeo(cam_up_simple_loc[i,],cam_up_simple_loc, a=6378137, f=1/298.257223563)
}

# results are in meters of distance!!!

# check the dimensions and summary info for the matrix
ncol(up_cam_geodist_matrix)
head(up_cam_geodist_matrix)

max(up_cam_geodist_matrix)
min(up_cam_geodist_matrix)
mean(as.matrix(up_cam_geodist_matrix), na.rm = TRUE)
median(as.matrix(up_cam_geodist_matrix), na.rm = TRUE)

# make the matrix into an upper triangle (to make foluping calc less cumbersome and remove duplicates)
up_cam_geodist_matrix[lower.tri(up_cam_geodist_matrix, diag = TRUE)] <- NA
up_cam_geodist_matrix

# create a sequence of numbers to use for making bins (these bins were chosen to try to capture variation in values while minimizing bin number)
seq <- c(0,50000, 100000, 200000, 400000, 600000, 1000000, 1500000, 2000000, 2500000, 3000000, 3500000)

# read in the Jaccard dist matrix created using EDENetworks
up_cam_dist <- read.table("EDENetwork_outputs/Up CAM/UCam_distance.txt", header = FALSE) # jaccard values matrix
up_cam_names <- read.table("EDENetwork_outputs/Up CAM/UCam_distance_names.txt",header = FALSE) # node names vector

# give the rows and columns names based on the node names file
rownames(up_cam_dist) <- c(up_cam_names$Node)
colnames(up_cam_dist) <- c(up_cam_names$Node)

# check that the two matrices have an identical arrangement
a <- colnames(up_cam_geodist_matrix)
b <- colnames(up_cam_dist)
a == b # They do, yay!
#### Calculate the average and mean jaccard weights for the different bins

## 0-50 km ##

```r
# get index positions of all values in geographic distance matrix within 0-50 km
up_cam_geodist_0_50 = which(up_cam_geodist_matrix < 50000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_0_50_jaccard <- up_cam_dist[c(up_cam_geodist_0_50[,1]),c(up_cam_geodist_0_50[,2])]

# get the average and median jaccard weight values for the 0-50 km bin
up_cam_geodist_0_50_avg <- mean(as.matrix(up_cam_geodist_0_50_jaccard), na.rm = TRUE)
up_cam_geodist_0_50_med <- median(as.matrix(up_cam_geodist_0_50_jaccard), na.rm = TRUE)
```

## 50-100 km ##

```r
# get index positions of all values in geographic distance matrix within 50-100 km
up_cam_geodist_50_100 = which(up_cam_geodist_matrix > 50000 & up_cam_geodist_matrix < 100000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_50_100_jaccard <- up_cam_dist[c(up_cam_geodist_50_100[,1]),c(up_cam_geodist_50_100[,2])]

# get the average and median jaccard weight values for the 50-100 km bin
up_cam_geodist_50_100_avg <- mean(as.matrix(up_cam_geodist_50_100_jaccard), na.rm = TRUE)
up_cam_geodist_50_100_med <- median(as.matrix(up_cam_geodist_50_100_jaccard), na.rm = TRUE)
```

## 100-200 km ##

```r
# get index positions of all values in geographic distance matrix within 100-200 km
up_cam_geodist_100_200 = which(up_cam_geodist_matrix > 100000 & up_cam_geodist_matrix < 200000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_100_200_jaccard <- up_cam_dist[c(up_cam_geodist_100_200[,1]),c(up_cam_geodist_100_200[,2])]

# get the average and median jaccard weight values for the 100-200 km bin
up_cam_geodist_100_200_avg <- mean(as.matrix(up_cam_geodist_100_200_jaccard), na.rm = TRUE)
up_cam_geodist_100_200_med <- median(as.matrix(up_cam_geodist_100_200_jaccard), na.rm = TRUE)
```

## 200-400 km ##

```r
# get index positions of all values in geographic distance matrix within 200-400 km
up_cam_geodist_200_400 = which(up_cam_geodist_matrix > 200000 & up_cam_geodist_matrix < 400000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_200_400_jaccard <- up_cam_dist[c(up_cam_geodist_200_400[,1]),c(up_cam_geodist_200_400[,2])]
```
# get the average and median jaccard weight values for the 200-400 km bin
up_cam_geodist_200_400_avg <- mean(as.matrix(up_cam_geodist_200_400_jaccard), na.rm = TRUE)
up_cam_geodist_200_400_med <- median(as.matrix(up_cam_geodist_200_400_jaccard), na.rm = TRUE)

## 400-600 km ##

# get index positions of all values in geographic distance matrix within 400-600 km
up_cam_geodist_400_600 = which(up_cam_geodist_matrix >400000 & up_cam_geodist_matrix <
600000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_400_600_jaccard <-
up_cam_dist[c(up_cam_geodist_400_600[,1]),c(up_cam_geodist_400_600[,2])]

# get the average and median jaccard weight values for the 400-600 km bin
up_cam_geodist_400_600_avg <- mean(as.matrix(up_cam_geodist_400_600_jaccard), na.rm = TRUE)
up_cam_geodist_400_600_med <- median(as.matrix(up_cam_geodist_400_600_jaccard), na.rm = TRUE)

## 600-1000 km ##

# get index positions of all values in geographic distance matrix within 600-1000 km
up_cam_geodist_600_1000 = which(up_cam_geodist_matrix >600000 & up_cam_geodist_matrix <
1000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_600_1000_jaccard <-
up_cam_dist[c(up_cam_geodist_600_1000[,1]),c(up_cam_geodist_600_1000[,2])]

# get the average and median jaccard weight values for the 600-1000 km bin
up_cam_geodist_600_1000_avg <- mean(as.matrix(up_cam_geodist_600_1000_jaccard), na.rm = TRUE)
up_cam_geodist_600_1000_med <- median(as.matrix(up_cam_geodist_600_1000_jaccard), na.rm = TRUE)

## 1000-1500 km ##

# get index positions of all values in geographic distance matrix within 1000-1500 km
up_cam_geodist_1000_1500 = which(up_cam_geodist_matrix >1000000 & up_cam_geodist_matrix <
1500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_cam_geodist_1000_1500_jaccard <-
up_cam_dist[c(up_cam_geodist_1000_1500[,1]),c(up_cam_geodist_1000_1500[,2])]

# get the average and median jaccard weight values for the 1000-1500 km bin
up_cam_geodist_1000_1500_avg <- mean(as.matrix(up_cam_geodist_1000_1500_jaccard), na.rm =
TRUE)
up_cam_geodist_1000_1500_med <- median(as.matrix(up_cam_geodist_1000_1500_jaccard), na.rm =
TRUE)

## 1500-2000 km ##
# get index positions of all values in geographic distance matrix within 1500-2000 km
up_cam_geodist_1500_2000 = which(up_cam_geodist_matrix >1500000 & up_cam_geodist_matrix < 2000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
up_cam_geodist_1500_2000_jaccard <- up_cam_dist[c(up_cam_geodist_1500_2000[,1]),c(up_cam_geodist_1500_2000[,2])]

# get the average and median jaccard weight values for the 1500-2000 km bin
up_cam_geodist_1500_2000_avg <- mean(as.matrix(up_cam_geodist_1500_2000_jaccard), na.rm = TRUE)
up_cam_geodist_1500_2000_med <- median(as.matrix(up_cam_geodist_1500_2000_jaccard), na.rm = TRUE)

## 2000-2500 km ##

# get index positions of all values in geographic distance matrix within 2000-2500 km
up_cam_geodist_2000_2500 = which(up_cam_geodist_matrix >2000000 & up_cam_geodist_matrix < 2500000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
up_cam_geodist_2000_2500_jaccard <- up_cam_dist[c(up_cam_geodist_2000_2500[,1]),c(up_cam_geodist_2000_2500[,2])]

# get the average and median jaccard weight values for the 2000-2500 km bin
up_cam_geodist_2000_2500_avg <- mean(as.matrix(up_cam_geodist_2000_2500_jaccard), na.rm = TRUE)
up_cam_geodist_2000_2500_med <- median(as.matrix(up_cam_geodist_2000_2500_jaccard), na.rm = TRUE)

## 2500-3000 km ##

# get index positions of all values in geographic distance matrix within 2500-2500 km
up_cam_geodist_2500_3000 = which(up_cam_geodist_matrix >2500000 & up_cam_geodist_matrix < 3000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
up_cam_geodist_2500_3000_jaccard <- up_cam_dist[c(up_cam_geodist_2500_3000[,1]),c(up_cam_geodist_2500_3000[,2])]

# get the average and median jaccard weight values for the 2500-3000 km bin
up_cam_geodist_2500_3000_avg <- mean(as.matrix(up_cam_geodist_2500_3000_jaccard), na.rm = TRUE)
up_cam_geodist_2500_3000_med <- median(as.matrix(up_cam_geodist_2500_3000_jaccard), na.rm = TRUE)

## 3000-3500 km ##

# get index positions of all values in geographic distance matrix within 3000-3500 km
up_cam_geodist_3000_3500 = which(up_cam_geodist_matrix >3000000, arr.ind = TRUE)

# extract corresponding values from the jaccard distance matrix
up_cam_geodist_3000_3500_jaccard <- up_cam_dist[c(up_cam_geodist_3000_3500[,1]),c(up_cam_geodist_3000_3500[,2])]

# get the average and median jaccard weight values for the 3000-3500 km bin
up_cam_geodist_3000_3500_avg <- mean(as.matrix(up_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
up_cam_geodist_3000_3500_med <- median(as.matrix(up_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
# get the average and median jaccard weight values for the 3000-3500 km bin
up_cam_geodist_3000_3500_avg <- mean(as.matrix(up_cam_geodist_3000_3500_jaccard), na.rm = TRUE)
up_cam_geodist_3000_3500_med <- median(as.matrix(up_cam_geodist_3000_3500_jaccard), na.rm = TRUE)

up_cam_freq_geodist_bins <- c(nrow(up_cam_geodist_0_50),
nrow(up_cam_geodist_50_100),
nrow(up_cam_geodist_100_200),
nrow(up_cam_geodist_200_400),
nrow(up_cam_geodist_400_600),
nrow(up_cam_geodist_600_1000),
nrow(up_cam_geodist_1000_1500),
nrow(up_cam_geodist_1500_2000),
nrow(up_cam_geodist_2000_2500),
nrow(up_cam_geodist_2500_3000),
nrow(up_cam_geodist_3000_3500))

up_cam_geodist_avg_bins <- c(up_cam_geodist_0_50_avg,
up_cam_geodist_50_100_avg,
up_cam_geodist_100_200_avg,
up_cam_geodist_200_400_avg,
up_cam_geodist_400_600_avg,
up_cam_geodist_600_1000_avg,
up_cam_geodist_1000_1500_avg,
up_cam_geodist_1500_2000_avg,
up_cam_geodist_2000_2500_avg,
up_cam_geodist_2500_3000_avg,
up_cam_geodist_3000_3500_avg)

up_cam_geodist_med_bins <- c(up_cam_geodist_0_50_med,
up_cam_geodist_50_100_med,
up_cam_geodist_100_200_med,
up_cam_geodist_200_400_med,
up_cam_geodist_400_600_med,
up_cam_geodist_600_1000_med,
up_cam_geodist_1000_1500_med,
up_cam_geodist_1500_2000_med,
up_cam_geodist_2000_2500_med,
up_cam_geodist_2500_3000_med,
up_cam_geodist_3000_3500_med)

up_cam_geodist_bins_table <-
rbind(up_cam_geodist_avg_bins,up_cam_geodist_med_bins,up_cam_freq_geodist_bins)

colnames(up_cam_geodist_bins_table) <- c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km","600-1000 km",
"1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km")
rownames(up_cam_geodist_bins_table) <- c("Mean Jaccard Weight","Median Jaccard Weight","Numb. Sites in Bin")

up_cam_geodist_bins_table

write.csv(up_cam_geodist_bins_table, file = 'WeightAnalysis/up_cam_geodist_bins_table.csv')

### Plot the lines #####

plot(low_cam_geodist_bins_table[1,],col="blue",pch=16,axes=FALSE, ann=FALSE)
points(mid_cam_geodist_bins_table[1,],col="green",pch=17)
points(up_cam_geodist_bins_table[1,],col="gold",pch=18)
axis(1, at=1:11, lab=c("0-50 ","50-100","100-200","200-400","400-600","600-1000",
"1000-1500","1500-2000","2000-2500","2500-3000","3000-3500"))
axis(2, at=c(0.7,0.75,0.80,0.85,0.90,0.95,1))
box()
title(xlab="Distance Bins (km)")
title(ylab="Bray–Curtis Dissimilarity Index Weight ")
title(main="Average Weight by Geographic Distance")

##### Lower Maastrichtian ######

# read in the lat/long file for the substage
maa_low_loc <- read.csv('R_Config_for_EN_OUTPUT/low maa/maa_low_gen_great3_avgloc.csv')

# get just the lat long locations from the table
maa_low_simple_loc <- maa_low_loc[,c(3,2)]

# NOTE: Remember that these are average locations based on the fossil occ found in each grid cell, and therefore already approximations
# which may be biased by not converting the fossil locations to paleocoordinates originally...

# Get the map information for this substage (age chosen is very approximate for this substage because limited options)
maa_low_maps <- reconstruct("plates", age=70) # get paleo plate info
maa_low_coast <- reconstruct("coastlines", age=70) # get coastline info

# Reconstruct paleocoord using age of rough age of substage (approximate)
maa_low_ploc <- reconstruct(maa_low_simple_loc, age=c(70))

# get min and max paleolong
maa_low_minlong <- min(maa_low_ploc[,1]) - 10
maa_low_maxlong <- max(maa_low_ploc[,1]) + 10

# get min and max paleolat
maa_low_minlat <- min(maa_low_ploc[,2]) - 15
maa_low_maxlat <- max(maa_low_ploc[,2]) + 15

# plot the maps and points in paleo positions to get an idea about their layout
par(mfrow = c(1,1))
# get number of unique locations to use for matrix dimensions
low_maa_count <- nrow(maa_low_loc)

# create empty dataframe to put the geographic distances into (use the names of the different grid cells as row and column names)
low_maa_geodist_matrix <- data.frame(matrix(ncol = low_maa_count, nrow = low_maa_count))

rownames(low_maa_geodist_matrix) <- maa_low_loc$node_label # give the matrix row names based on original grid names

colnames(low_maa_geodist_matrix) <- maa_low_loc$node_label # give the matrix column names based on original grid names

# run for loop to get distances using geodesic distance calc
for (i in 1:nrow(maa_low_ploc)){
  low_maa_geodist_matrix[i,] <- distGeo(maa_low_simple_loc[i,], maa_low_simple_loc, a = 6378137, f = 1/298.257223563)
}

# results are in meters of distance!!!

# check the dimensions and summary info for the matrix
ncol(low_maa_geodist_matrix)
head(low_maa_geodist_matrix)

max(low_maa_geodist_matrix)
min(low_maa_geodist_matrix)
mean(as.matrix(low_maa_geodist_matrix), na.rm = TRUE)
median(as.matrix(low_maa_geodist_matrix), na.rm = TRUE)

# make the matrix into a lower triangle (to make following calc less cumbersome and remove duplicates)
low_maa_geodist_matrix[lower.tri(low_maa_geodist_matrix, diag = TRUE)] <- NA

low_maa_geodist_matrix

# create a sequence of numbers to use for making bins (these bins were chosen to try to capture variation in values while minimizing bin number)
seq <- c(0, 50000, 100000, 200000, 400000, 600000, 1000000, 1500000, 2000000, 2500000, 3000000, 3500000)

# read in the Jaccard dist matrix created using EDENetworks
low_maa_dist <- read.table("EDENetwork_outputs/Low MAA/Low_Maa_distance.txt", header = FALSE) # jaccard values matrix

low_maa_names <- read.table("EDENetwork_outputs/Low MAA/Low_Maa_distance_names.txt", header = FALSE) # node names vector
colnames(low_maa_names) <- "Node"
# give the rows and columns names based on the node names file
colnames(low_maa_dist) <- c(low_maa_names$Node)
rownames(low_maa_dist) <- c(low_maa_names$Node)

# check that the two matrices have an identical arrangement
a <- colnames(low_maa_geodist_matrix)
b <- colnames(low_maa_dist)
a == b # They do, yay!

#### Calculate the average and mean jaccard weights for the different bins

## 0-50 km ##

# get index positions of all values in geographic distance matrix within 0-50 km
low_maa_geodist_0_50 = which(low_maa_geodist_matrix < 50000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_0_50_jaccard <-
low_maa_dist[c(low_maa_geodist_0_50[,1]),c(low_maa_geodist_0_50[,2])]

# get the average and median jaccard weight values for the 0-50 km bin
low_maa_geodist_0_50_avg <- mean(as.matrix(low_maa_geodist_0_50_jaccard), na.rm = TRUE)
low_maa_geodist_0_50_med <- median(as.matrix(low_maa_geodist_0_50_jaccard), na.rm = TRUE)

## 50-100 km ##

# get index positions of all values in geographic distance matrix within 50-100 km
low_maa_geodist_50_100 = which(low_maa_geodist_matrix > 50000 & low_maa_geodist_matrix < 100000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_50_100_jaccard <-
low_maa_dist[c(low_maa_geodist_50_100[,1]),c(low_maa_geodist_50_100[,2])]

# get the average and median jaccard weight values for the 50-100 km bin
low_maa_geodist_50_100_avg <- mean(as.matrix(low_maa_geodist_50_100_jaccard), na.rm = TRUE)
low_maa_geodist_50_100_med <- median(as.matrix(low_maa_geodist_50_100_jaccard), na.rm = TRUE)

## 100-200 km ##

# get index positions of all values in geographic distance matrix within 100-200 km
low_maa_geodist_100_200 = which(low_maa_geodist_matrix > 100000 & low_maa_geodist_matrix < 200000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_100_200_jaccard <-
low_maa_dist[c(low_maa_geodist_100_200[,1]),c(low_maa_geodist_100_200[,2])]

# get the average and median jaccard weight values for the 100-200 km bin
low_maa_geodist_100_200_avg <- mean(as.matrix(low_maa_geodist_100_200_jaccard), na.rm = TRUE)
low_maa_geodist_100_200_med <- median(as.matrix(low_maa_geodist_100_200_jaccard), na.rm = TRUE)

### 200-400 km ###

# get index positions of all values in geographic distance matrix within 200-400 km
low_maa_geodist_200_400 = which(low_maa_geodist_matrix > 200000 & low_maa_geodist_matrix < 400000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_200_400_jaccard <-
low_maa_dist[c(low_maa_geodist_200_400[,1]),c(low_maa_geodist_200_400[,2])]
# get the average and median jaccard weight values for the 200-400 km bin
low_maa_geodist_200_400_avg <- mean(as.matrix(low_maa_geodist_200_400_jaccard), na.rm = TRUE)
low_maa_geodist_200_400_med <- median(as.matrix(low_maa_geodist_200_400_jaccard), na.rm = TRUE)

### 400-600 km ###

# get index positions of all values in geographic distance matrix within 400-600 km
low_maa_geodist_400_600 = which(low_maa_geodist_matrix > 400000 & low_maa_geodist_matrix < 600000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_400_600_jaccard <-
low_maa_dist[c(low_maa_geodist_400_600[,1]),c(low_maa_geodist_400_600[,2])]
# get the average and median jaccard weight values for the 400-600 km bin
low_maa_geodist_400_600_avg <- mean(as.matrix(low_maa_geodist_400_600_jaccard), na.rm = TRUE)
low_maa_geodist_400_600_med <- median(as.matrix(low_maa_geodist_400_600_jaccard), na.rm = TRUE)

### 600-1000 km ###

# get index positions of all values in geographic distance matrix within 600-1000 km
low_maa_geodist_600_1000 = which(low_maa_geodist_matrix > 600000 & low_maa_geodist_matrix < 1000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_600_1000_jaccard <-
low_maa_dist[c(low_maa_geodist_600_1000[,1]),c(low_maa_geodist_600_1000[,2])]
# get the average and median jaccard weight values for the 600-1000 km bin
low_maa_geodist_600_1000_avg <- mean(as.matrix(low_maa_geodist_600_1000_jaccard), na.rm = TRUE)
low_maa_geodist_600_1000_med <- median(as.matrix(low_maa_geodist_600_1000_jaccard), na.rm = TRUE)

### 1000-1500 km ###
# get index positions of all values in geographic distance matrix within 1000-1500 km
low_maa_geodist_1000_1500 = which(low_maa_geodist_matrix >1000000 & low_maa_geodist_matrix < 1500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_1000_1500_jaccard <- low_maa_dist[c(low_maa_geodist_1000_1500[,1]),c(low_maa_geodist_1000_1500[,2])]

# get the average and median jaccard weight values for the 1000-1500 km bin
low_maa_geodist_1000_1500_avg <- mean(as.matrix(low_maa_geodist_1000_1500_jaccard), na.rm = TRUE)
low_maa_geodist_1000_1500_med <- median(as.matrix(low_maa_geodist_1000_1500_jaccard), na.rm = TRUE)

## 1500-2000 km ##

# get index positions of all values in geographic distance matrix within 1500-2000 km
low_maa_geodist_1500_2000 = which(low_maa_geodist_matrix >1500000 & low_maa_geodist_matrix < 2000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_1500_2000_jaccard <- low_maa_dist[c(low_maa_geodist_1500_2000[,1]),c(low_maa_geodist_1500_2000[,2])]

# get the average and median jaccard weight values for the 1500-2000 km bin
low_maa_geodist_1500_2000_avg <- mean(as.matrix(low_maa_geodist_1500_2000_jaccard), na.rm = TRUE)
low_maa_geodist_1500_2000_med <- median(as.matrix(low_maa_geodist_1500_2000_jaccard), na.rm = TRUE)

## 2000-2500 km ##

# get index positions of all values in geographic distance matrix within 2000-2500 km
low_maa_geodist_2000_2500 = which(low_maa_geodist_matrix >2000000 & low_maa_geodist_matrix < 2500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_2000_2500_jaccard <- low_maa_dist[c(low_maa_geodist_2000_2500[,1]),c(low_maa_geodist_2000_2500[,2])]

# get the average and median jaccard weight values for the 2000-2500 km bin
low_maa_geodist_2000_2500_avg <- mean(as.matrix(low_maa_geodist_2000_2500_jaccard), na.rm = TRUE)
low_maa_geodist_2000_2500_med <- median(as.matrix(low_maa_geodist_2000_2500_jaccard), na.rm = TRUE)

## 2500-3000 km ##

# get index positions of all values in geographic distance matrix within 2500-2500 km
low_maa_geodist_2500_3000 = which(low_maa_geodist_matrix >2500000 & low_maa_geodist_matrix < 3000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix

```r
low_maa_geodist_2500_3000_jaccard <-
low_maa_dist[c(low_maa_geodist_2500_3000[,1]),c(low_maa_geodist_2500_3000[,2])]

# get the average and median jaccard weight values for the 2500-2500 km bin
low_maa_geodist_2500_3000_avg <- mean(as.matrix(low_maa_geodist_2500_3000_jaccard), na.rm = TRUE)
low_maa_geodist_2500_3000_med <- median(as.matrix(low_maa_geodist_2500_3000_jaccard), na.rm = TRUE)

## 3000-3500 km ##

# get index positions of all values in geographic distance matrix within 3000-3500 km
low_maa_geodist_3000_3500 = which(low_maa_geodist_matrix >3000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
low_maa_geodist_3000_3500_jaccard <-
low_maa_dist[c(low_maa_geodist_3000_3500[,1]),c(low_maa_geodist_3000_3500[,2])]

# get the average and median jaccard weight values for the 3000-3500 km bin
low_maa_geodist_3000_3500_avg <- mean(as.matrix(low_maa_geodist_3000_3500_jaccard), na.rm = TRUE)
low_maa_geodist_3000_3500_med <- median(as.matrix(low_maa_geodist_3000_3500_jaccard), na.rm = TRUE)

low_maa_freq_geodist_bins <- c(nrow(low_maa_geodist_0_50),
nrow(low_maa_geodist_50_100),
nrow(low_maa_geodist_100_200),
nrow(low_maa_geodist_200_400),
nrow(low_maa_geodist_400_600),
nrow(low_maa_geodist_600_1000),
nrow(low_maa_geodist_1000_1500),
nrow(low_maa_geodist_1500_2000),
nrow(low_maa_geodist_2000_2500),
nrow(low_maa_geodist_2500_3000),
nrow(low_maa_geodist_3000_3500))

low_maa_geodist_avg_bins <- c(low_maa_geodist_0_50_avg,
   low_maa_geodist_50_100_avg,
   low_maa_geodist_100_200_avg,
   low_maa_geodist_200_400_avg,
   low_maa_geodist_400_600_avg,
   low_maa_geodist_600_1000_avg,
   low_maa_geodist_1000_1500_avg,
   low_maa_geodist_1500_2000_avg,
   low_maa_geodist_2000_2500_avg,
   low_maa_geodist_2500_3000_avg,
   low_maa_geodist_3000_3500_avg)

low_maa_geodist_med_bins <- c(low_maa_geodist_0_50_med,
   low_maa_geodist_50_100_med,
   low_maa_geodist_100_200_med,
   low_maa_geodist_200_400_med,
   low_maa_geodist_400_600_med,
   low_maa_geodist_600_1000_med,
   low_maa_geodist_1000_1500_med,
   low_maa_geodist_1500_2000_med,
   low_maa_geodist_2000_2500_med,
   low_maa_geodist_2500_3000_med,
   low_maa_geodist_3000_3500_med)
```
low_maa_geodist_bins_table <-
  rbind(low_maa_geodist_avg_bins,low_maa_geodist_med_bins,low_maa_freq_geodist_bins)

colnames(low_maa_geodist_bins_table) <- c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km","600-1000 km",
  "1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km")
rownames(low_maa_geodist_bins_table) <- c("Mean Jaccard Weight","Median Jaccard Weight","Numb. Sites in Bin")

low_maa_geodist_bins_table

write.csv(low_maa_geodist_bins_table, file = 'WeightAnalysis/low_maa_geodist_bins_table.csv')

### Plot the lines ######

plot(low_cam_geodist_bins_table[1,],col="blue",pch=16,axes=FALSE, ann=FALSE)
points(mid_cam_geodist_bins_table[1,],col="green",pch=17)
points(up_cam_geodist_bins_table[1,],col="gold",pch=18)
points(low_maa_geodist_bins_table[1,],col="orange",pch=19)
axis(1, at=1:11, lab=c("0-50 ","50-100 ","100-200 ","200-400 ","400-600 ","600-1000 ",
  "1000-1500 ","1500-2000 ","2000-2500 ","2500-3000 ","3000-3500 "))
axis(2, at=c(0.7,0.75,0.80,0.85,0.90,0.95,1))
box()
title(xlab="Distance Bins (km)")
title(ylab="Bray–Curtis Dissimilarity Index Weight ")
title(main="Average Weight by Geographic Distance")

##### Upper Maastrichtian #####

# read in the lat/long file for the substage
maa_up_loc <- read.csv('R_Config_for_EN_OUTPUT/up maa/maa_up_gen_great3_avgloc.csv')

# get just the lat long locations from the table
maa_up_simple_loc <- maa_up_loc[,c(3,2)]

# NOTE: Remember that these are average locations based on the fossil occurrence found in each grid cell, and therefore already approximations
# which may be biased by not converting the fossil locations to paleocoordinates originally...

# Get the map information for this substage (age chosen is very approximate for this substage because limited options)
maa_up_maps <- reconstruct("plates", age=70) # get paleo plate info
maa_up_coast <- reconstruct("coastlines", age=70) # get coastline info

# Reconstruct paleocoord using age of rough age of substage (approximate)
maa_up_ploc <- reconstruct(maa_up_simple_loc, age=c(70))

# get min and max paleolong
maa_up_minlong <- min(maa_up_ploc[,1]) - 10
maa_up_maxlong <- max(maa_up_ploc[,1]) + 10

# get min and max paleolat
maa_up_minlat <- min(maa_up_ploc[,2]) - 15
maa_up_maxlat <- max(maa_up_ploc[,2]) + 15

# plot the maps and points in paleo positions to get an idea about their layout
par(mfrow=c(1,1))
plot(maa_up_maps, col = 'grey', border = NA, xlim = c(maa_up_minlong,maa_up_maxlong), ylim = c(maa_up_minlat,maa_up_maxlat))
lines(maa_up_coast, col = "black", xlim = c(maa_up_minlong,maa_up_maxlong), ylim = c(maa_up_maxlat,maa_up_minlat))
points(maa_up_ploc, pch=3, col="red")

# get number of unique locations to use for matrix dimensions
up_maa_count <- nrow(maa_up_loc)

# create empty dataframe to put the geographic distances into (use the names of the different grid cells as row and column names)
up_maa_geodist_matrix <- data.frame(matrix(ncol = up_maa_count, nrow = up_maa_count))
colnames(up_maa_geodist_matrix) <- maa_up_loc$node_label # give the matrix column names based on original grid names
rownames(up_maa_geodist_matrix) <- maa_up_loc$node_label # give the matrix row names based on original grid names

# run for loop to get distances using geodesic distance calc
for (i in 1:nrow(maa_up_ploc)){
  up_maa_geodist_matrix[i,] <- distGeo(maa_up_simple_loc[i,],maa_up_simple_loc, a=6378137, f=1/298.257223563)
}

# results are in meters of distance!!!

# check the dimention and summary info for the matrix
ncol(up_maa_geodist_matrix)
head(up_maa_geodist_matrix)

max(up_maa_geodist_matrix)
min(up_maa_geodist_matrix)
mean(as.matrix(up_maa_geodist_matrix), na.rm = TRUE)
median(as.matrix(up_maa_geodist_matrix), na.rm = TRUE)

# make the matrix into a upper triangle (to make foluping calc less cumbersome and remove duplicates)
up_maa_geodist_matrix[lower.tri(up_maa_geodist_matrix, diag = TRUE)] <- NA
up_maa_geodist_matrix

# create a sequence of numbers to use for making bins (these bins were chosen to try to capture variation
# in values while minimizing bin number)
seq <- c(0,50000, 100000, 200000, 400000, 600000, 1000000, 1500000, 2000000, 2500000, 3000000,
3500000)

# read in the Jaccard dist matrix created using EDENetworks
up_maa_dist <- read.table("EDENetwork_outputs/Up MAA/UMaa_distance.txt", header = FALSE) #
jaccard values matrix

up_maa_names <- read.table("EDENetwork_outputs/Up MAA/UMaa_distance_names.txt",header =
FALSE) # node names vector
colnames(up_maa_names) <- "Node"

# give the rows and columns names based on the node names file
colnames(up_maa_dist) <- c(up_maa_names$Node)
rownames(up_maa_dist) <- c(up_maa_names$Node)

# check that the two matrices have an identical arrangement
a <- colnames(up_maa_geodist_matrix)
b <- colnames(up_maa_dist)
a == b # They do, yay!

#### Calculate the average and mean jaccard weights for the different bins

## 0-50 km ##
# get index positions of all values in geographic distance matrix within 0-50 km
up_maa_geodist_0_50 = which(up_maa_geodist_matrix <50000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_0_50_jaccard <-
up_maa_dist[c(up_maa_geodist_0_50[,1]),c(up_maa_geodist_0_50[,2])]

# get the average and median jaccard weight values for the 0-50 km bin
up_maa_geodist_0_50_avg <- mean(as.matrix(up_maa_geodist_0_50_jaccard), na.rm = TRUE)
up_maa_geodist_0_50_med <- median(as.matrix(up_maa_geodist_0_50_jaccard), na.rm = TRUE)

## 50-100 km ##
# get index positions of all values in geographic distance matrix within 50-100 km
up_maa_geodist_50_100 = which(up_maa_geodist_matrix >50000 & up_maa_geodist_matrix < 100000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_50_100_jaccard <-
up_maa_dist[c(up_maa_geodist_50_100[,1]),c(up_maa_geodist_50_100[,2])]

# get the average and median jaccard weight values for the 50-100 km bin
up_maa_geodist_50_100_avg <- mean(as.matrix(up_maa_geodist_50_100_jaccard), na.rm = TRUE)
up_maa_geodist_50_100_med <- median(as.matrix(up_maa_geodist_50_100_jaccard), na.rm = TRUE)

## 100-200 km ##

# get index positions of all values in geographic distance matrix within 100-200 km
up_maa_geodist_100_200 = which(up_maa_geodist_matrix >100000 & up_maa_geodist_matrix <
200000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_100_200_jaccard <-
up_maa_dist[c(up_maa_geodist_100_200[,1]),c(up_maa_geodist_100_200[,2])]

# get the average and median jaccard weight values for the 100-200 km bin
up_maa_geodist_100_200_avg <- mean(as.matrix(up_maa_geodist_100_200_jaccard), na.rm = TRUE)
up_maa_geodist_100_200_med <- median(as.matrix(up_maa_geodist_100_200_jaccard), na.rm = TRUE)

## 200-400 km ##

# get index positions of all values in geographic distance matrix within 200-400 km
up_maa_geodist_200_400 = which(up_maa_geodist_matrix >200000 & up_maa_geodist_matrix <
400000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_200_400_jaccard <-
up_maa_dist[c(up_maa_geodist_200_400[,1]),c(up_maa_geodist_200_400[,2])]

# get the average and median jaccard weight values for the 200-400 km bin
up_maa_geodist_200_400_avg <- mean(as.matrix(up_maa_geodist_200_400_jaccard), na.rm = TRUE)
up_maa_geodist_200_400_med <- median(as.matrix(up_maa_geodist_200_400_jaccard), na.rm = TRUE)

## 400-600 km ##

# get index positions of all values in geographic distance matrix within 400-600 km
up_maa_geodist_400_600 = which(up_maa_geodist_matrix >400000 & up_maa_geodist_matrix <
600000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_400_600_jaccard <-
up_maa_dist[c(up_maa_geodist_400_600[,1]),c(up_maa_geodist_400_600[,2])]

# get the average and median jaccard weight values for the 400-600 km bin
up_maa_geodist_400_600_avg <- mean(as.matrix(up_maa_geodist_400_600_jaccard), na.rm = TRUE)
up_maa_geodist_400_600_med <- median(as.matrix(up_maa_geodist_400_600_jaccard), na.rm = TRUE)

## 600-1000 km ##

# get index positions of all values in geographic distance matrix within 600-1000 km
up_maa_geodist_600_1000 = which(up_maa_geodist_matrix >600000 & up_maa_geodist_matrix <
1000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_600_1000_jaccard <- 
up_maa_dist[c(up_maa_geodist_600_1000[,1]),c(up_maa_geodist_600_1000[,2])]

# get the average and median jaccard weight values for the 600-1000 km bin
up_maa_geodist_600_1000_avg <- mean(as.matrix(up_maa_geodist_600_1000_jaccard), na.rm = TRUE)
up_maa_geodist_600_1000_med <- median(as.matrix(up_maa_geodist_600_1000_jaccard), na.rm = TRUE)

## 1000-1500 km ##

# get index positions of all values in geographic distance matrix within 1000-1500 km
up_maa_geodist_1000_1500 = which(up_maa_geodist_matrix >1000000 & up_maa_geodist_matrix <
1500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_1000_1500_jaccard <- 
up_maa_dist[c(up_maa_geodist_1000_1500[,1]),c(up_maa_geodist_1000_1500[,2])]

# get the average and median jaccard weight values for the 1000-1500 km bin
up_maa_geodist_1000_1500_avg <- mean(as.matrix(up_maa_geodist_1000_1500_jaccard), na.rm = TRUE)
up_maa_geodist_1000_1500_med <- median(as.matrix(up_maa_geodist_1000_1500_jaccard), na.rm = TRUE)

## 1500-2000 km ##

# get index positions of all values in geographic distance matrix within 1500-2000 km
up_maa_geodist_1500_2000 = which(up_maa_geodist_matrix >1500000 & up_maa_geodist_matrix <
2000000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_1500_2000_jaccard <- 
up_maa_dist[c(up_maa_geodist_1500_2000[,1]),c(up_maa_geodist_1500_2000[,2])]

# get the average and median jaccard weight values for the 1500-2000 km bin
up_maa_geodist_1500_2000_avg <- mean(as.matrix(up_maa_geodist_1500_2000_jaccard), na.rm = TRUE)
up_maa_geodist_1500_2000_med <- median(as.matrix(up_maa_geodist_1500_2000_jaccard), na.rm = TRUE)

## 2000-2500 km ##

# get index positions of all values in geographic distance matrix within 2000-2500 km
up_maa_geodist_2000_2500 = which(up_maa_geodist_matrix >2000000 & up_maa_geodist_matrix <
2500000, arr.ind = TRUE)
# extract corresponding values from the jaccard distance matrix
up_maa_geodist_2000_2500_jaccard <- 
up_maa_dist[c(up_maa_geodist_2000_2500[,1]),c(up_maa_geodist_2000_2500[,2])]

# get the average and median jaccard weight values for the 2000-2500 km bin
\[
\text{up\_maa\_geodist\_2000\_2500\_avg} \leftarrow \text{mean(as.matrix(up\_maa\_geodist\_2000\_2500\_jaccard), na.rm = TRUE)}
\]

\[
\text{up\_maa\_geodist\_2000\_2500\_med} \leftarrow \text{median(as.matrix(up\_maa\_geodist\_2000\_2500\_jaccard), na.rm = TRUE)}
\]

## 2500-3000 km ##

\# get index positions of all values in geographic distance matrix within 2500-2500 km
\[
\text{up\_maa\_geodist\_2500\_3000} = \text{which(up\_maa\_geodist\_matrix >250000 & up\_maa\_geodist\_matrix < 3000000, arr.ind = TRUE)}
\]

\# extract corresponding values from the jaccard distance matrix
\[
\text{up\_maa\_geodist\_2500\_3000\_jaccard} \leftarrow \text{up\_maa\_dist[c(up\_maa\_geodist\_2500\_3000[,1]),c(up\_maa\_geodist\_2500\_3000[,2])]}
\]

\# get the average and median jaccard weight values for the 2500-2500 km bin
\[
\text{up\_maa\_geodist\_2500\_3000\_avg} \leftarrow \text{mean(as.matrix(up\_maa\_geodist\_2500\_3000\_jaccard), na.rm = TRUE)}
\]

\[
\text{up\_maa\_geodist\_2500\_3000\_med} \leftarrow \text{median(as.matrix(up\_maa\_geodist\_2500\_3000\_jaccard), na.rm = TRUE)}
\]

## 3000-3500 km ##

\# get index positions of all values in geographic distance matrix within 3000-3500 km
\[
\text{up\_maa\_geodist\_3000\_3500} = \text{which(up\_maa\_geodist\_matrix >300000, arr.ind = TRUE)}
\]

\# extract corresponding values from the jaccard distance matrix
\[
\text{up\_maa\_geodist\_3000\_3500\_jaccard} \leftarrow \text{up\_maa\_dist[c(up\_maa\_geodist\_3000\_3500[,1]),c(up\_maa\_geodist\_3000\_3500[,2])]}\]

\# get the average and median jaccard weight values for the 3000-3500 km bin
\[
\text{up\_maa\_geodist\_3000\_3500\_avg} \leftarrow \text{mean(as.matrix(up\_maa\_geodist\_3000\_3500\_jaccard), na.rm = TRUE)}
\]

\[
\text{up\_maa\_geodist\_3000\_3500\_med} \leftarrow \text{median(as.matrix(up\_maa\_geodist\_3000\_3500\_jaccard), na.rm = TRUE)}
\]

\[
\text{up\_maa\_freq\_geodist\_bins} \leftarrow \text{c(nrow(up\_maa\_geodist\_0\_50), nrow(up\_maa\_geodist\_50\_100), nrow(up\_maa\_geodist\_100\_200), nrow(up\_maa\_geodist\_200\_400), nrow(up\_maa\_geodist\_400\_600), nrow(up\_maa\_geodist\_600\_1000), nrow(up\_maa\_geodist\_1000\_1500), nrow(up\_maa\_geodist\_1500\_2000), nrow(up\_maa\_geodist\_2000\_2500), nrow(up\_maa\_geodist\_2500\_3000), nrow(up\_maa\_geodist\_3000\_3500))}
\]

\[
\text{up\_maa\_geodist\_avg\_bins} \leftarrow \text{c(up\_maa\_geodist\_0\_50\_avg, up\_maa\_geodist\_50\_100\_avg, up\_maa\_geodist\_100\_200\_avg, up\_maa\_geodist\_200\_400\_avg, up\_maa\_geodist\_400\_600\_avg, up\_maa\_geodist\_600\_1000\_avg, up\_maa\_geodist\_1000\_1500\_avg, up\_maa\_geodist\_1500\_2000\_avg, up\_maa\_geodist\_2000\_2500\_avg, up\_maa\_geodist\_2500\_3000\_avg, up\_maa\_geodist\_3000\_3500\_avg)}
\]
up_maa_geodist_med_bins <- c(up_maa_geodist_0_50_med, up_maa_geodist_50_100_med, up_maa_geodist_100_200_med, up_maa_geodist_200_400_med, up_maa_geodist_400_600_med, up_maa_geodist_600_1000_med, up_maa_geodist_1000_1500_med, up_maa_geodist_1500_2000_med, up_maa_geodist_2000_2500_med, up_maa_geodist_2500_3000_med, up_maa_geodist_3000_3500_med)

up_maa_geodist_bins_table <- rbind(up_maa_geodist_avg_bins, up_maa_geodist_med_bins, up_maa_freq_geodist_bins)

colnames(up_maa_geodist_bins_table) <- c("0-50 km","50-100 km","100-200 km","200-400 km","400-600 km","600-1000 km", "1000-1500 km","1500-2000 km","2000-2500 km","2500-3000 km","3000-3500 km")
rownames(up_maa_geodist_bins_table) <- c("Mean Jaccard Weight","Median Jaccard Weight","Numb. Sites in Bin")

up_maa_geodist_bins_table

write.csv(up_maa_geodist_bins_table, file = 'WeightAnalysis/up_maa_geodist_bins_table.csv')

low_cam_geodist_bins_table
mid_cam_geodist_bins_table
up_cam_geodist_bins_table
low_maa_geodist_bins_table

### Plot the lines ######

setwd("C:/Users/ceara/Documents/Province Project/Proposal tables and figures")

pdf(file = "Geographic_Dist_Bins_average_weights.pdf")

# average values plot

plot(low_cam_geodist_bins_table[,1],col="blue",type="o",pch=16,axes=FALSE, ann=FALSE,ylim=c(0.65,0.99))
points(mid_cam_geodist_bins_table[1,],col="green",pch=17,type="o")
points(up_cam_geodist_bins_table[1,],col="gold",pch=18,type="o")
points(low_maa_geodist_bins_table[1,],col="orange",pch=19,type="o")
points(up_maa_geodist_bins_table[1,],col="maroon1",pch=15,type="o")
axis(1, at=1:11, lab=c("0-50 ","50-100","100-200","200-400","400-600","600-1000",
"1000-1500","1500-2000","2000-2500","2500-3000","3000-3500"))
axis(2, at=c(0.65,0.7,0.75,0.8,0.85,0.9,0.95,0.99))
box()
title(xlab="Distance Bins (km)")
title(ylab="Jaccard Distance Weight")
title(main="Average Weight by Geographic Distance")
legend(x = 8.75, y = 0.75, legend = c("Low Camp.","Mid Camp.", "Up Camp.", "Low Maastr.","Up Maastr."), col = c("blue","darkgreen","gold","orange","maroon1"),
pch = c(16,17,18,19,15))

# median values plot
plot(low_cam_geodist_bins_table[2,],col="blue",type="o",pch=16,axes=FALSE,
ann=FALSE,ylim=c(0.65,0.99))
points(mid_cam_geodist_bins_table[2,],col="green",pch=17,type="o")
points(up_cam_geodist_bins_table[2,],col="gold",pch=18,type="o")
points(low_maa_geodist_bins_table[2,],col="orange",pch=19,type="o")
points(up_maa_geodist_bins_table[2,],col="maroon1",pch=15,type="o")
axis(1, at=1:11, lab=c("0-50 ","50-100","100-200","200-400","400-600","600-1000",
"1000-1500","1500-2000","2000-2500","2500-3000","3000-3500"))
axis(2, at=c(0.65,0.7,0.75,0.8,0.85,0.9,0.95,0.99))
box()
title(xlab="Distance Bins (km)")
title(ylab="Jaccard Distance Weight")
title(main="Median Weight by Geographic Distance")
legend(x = 8.75, y = 0.75, legend = c("Low Camp.","Mid Camp.", "Up Camp.", "Low Maastr.","Up Maastr.")), col = c("blue","darkgreen","gold","orange","maroon1"),
pch = c(16,17,18,19,15))
de.v.off()
pdf(file = "PaleoLocation_Maps_Substages.pdf")

plot(cam_low_maps, col = 'grey', border = NA, xlim = c(cam_low_minlong,maa_up_maxlong), ylim =
c(cam_low_minlat,cam_low_maxlat))
lines(cam_low_coast, col = "black", xlim = c(cam_low_minlong,cam_low_maxlong), ylim =
c(cam_low_maxlat,cam_low_minlat))
points(cam_low_ploc, pch=21, bg = "blue", col = "black")

plot(cam_mid_maps, col = 'grey', border = NA, xlim = c(cam_mid_minlong,cam_mid_maxlong), ylim =
c(cam_mid_minlat,cam_mid_maxlat))
lines(cam_mid_coast, col = "black", xlim = c(cam_low_minlong,cam_low_maxlong), ylim =
c(cam_low_maxlat,cam_low_minlat))
points(cam_mid_ploc, pch=21, bg = "darkgreen", col="black")
plot(cam_up_maps, col = 'grey', border = NA, xlim = c(cam_up_minlong, cam_up_maxlong), ylim = c(cam_up_minlat, cam_up_maxlat))
lines(cam_up_coast, col = "black", xlim = c(cam_low_minlong, cam_low_maxlong), ylim = c(cam_low_maxlat, cam_low_minlat))
points(cam_up_ploc, pch=21, bg = "gold", col="black")

plot(maa_low_maps, col = 'grey', border = NA, xlim = c(maa_low_minlong, maa_up_maxlong), ylim = c(maa_low_minlat, maa_low_maxlat))
lines(maa_low_coast, col = "black", xlim = c(cam_low_minlong, cam_low_maxlong), ylim = c(cam_low_maxlat, cam_low_minlat))
points(maa_low_ploc, pch=21, bg = "orange", col="black")

plot(maa_up_maps, col = 'grey', border = NA, xlim = c(maa_up_minlong, maa_up_maxlong), ylim = c(maa_up_minlat, maa_up_maxlat))
lines(maa_up_coast, col = "black", xlim = c(cam_low_minlong, cam_low_maxlong), ylim = c(cam_low_maxlat, cam_low_minlat))
points(maa_up_ploc, pch=21, bg = "maroon1", col="black")

dev.off()
APPENDIX B. Supplementary Materials for “A network analysis approach to understanding trends in functional diversity in the Late Cretaceous of North America” by Purcell and Myers (in press)

Appendix B-1. Detailed Methods and Results

Relationship between FR and GR in Database:
A basic correlation plot of generic vs functional richness was created based on the values present in each grid cell for the raw database in each substage. The plot showed a strong positive trend with fewer points at the higher levels of GR (>100 genera) that potentially begin to plateau (Figure S1). A squared linear regression computed using the lm function in R was found to have the best fit (R² = 0.958) of the simple models tested, and visually fits the data (Figure S1). Evaluation plots of this regression (Figure S2) suggest that the data itself is non-normal based on the Q-Q plot (further confirmed with a Shapiro-Wilk test in R: p-value < 0.05), has a linear relationship based on the Residual vs. Fitted plot, contains only one significant outlier based on the Residual vs. Leverage plot, and has fairly homogeneous variance except at very low fitted values (i.e., < ~40) based on the Scale-Location plot. Overall, these results suggest that the regression is a good fit for the data, despite the fact that the data is non-normal.

The strong correlation between the two richness variables indicates that the two are closely linked and low values of generic richness will very likely correspond with low functional richness in further analysis. However, at very high generic richness levels (i.e., >100), it is likely that the two will be less strongly correlated. This may impact interpretation at different scales, given that the number of unique genera at coarser spatial aggregations in the dataset, such as within latitudinal bins or within the entire study region, will be higher than at more localize scales, such as at the level of the node. It should also be noted that all 60km nodes with >100 genera are representative of the Maastrichtian. This indicates that richness is better represented in the Maastrichtian than in the Campanian, though not all Maastrichtian nodes are as well sampled.
Figure S1. Plot of generic versus functional richness values for each node in each substage. Point color indicates substage interval (blue = Lower Camp., green = Middle Camp., yellow = Upper Camp., orange = Lower Maastr., and red = Upper Maastr.). The blue line represents a squared linear regression line fitted to the data ($R^2 = 0.958$) using the `stat_smooth` function from the ggplot2 package, with grey shading indicating the 95% confidence interval.
Figure S2. Evaluation plots of the squared linear regression fitted to the data in R using the lm function \( y = 0.095x^2 + 2.64; \ R^2 = 0.949 \), including a) fitted values plotted against residual values, b) a normal Q-Q plot, c) fitted values plotted against the square root of the standardized residuals, and d) a plot of Cook’s distance in the data, wherein leverage is plotted against the standardized residuals.
Bootstrap results for variation in FR values:

Bootstrap analysis of nodes and latitude bins found strong positive correlation between increasing FR variability and generic richness (Figure S3). Only when a node has a FR value close to the maximum for the dataset (20) does the standard error appear to potentially level off (Figure S3). This analysis tests how variable FR is when the unique genera in a node are subsampled. Given the assumption that genera are perfectly evenly distributed in FEs, FR values would not change, producing no correlation between standard error and FR. However, a positive correlation indicates that genera are not evenly distributed. A squared regression model fits this data distribution well ($R^2 = 0.96$), indicating a depreciation of the correlation at approximately more than 20 FEs.

Figure S3. Bootstrap of FR results for nodes in all substages. Blue line represents a squared linear regression fitted to the data.
SQS to determine lower sampling limits

To determine a lower sample size limit that can be assumed to capture the majority of FE in each substage, SQS subsampling (Alroy 2010) was performed on each substage using FEs as taxa and calculating their frequencies in each interval based on the number of unique genera they represent (i.e., no duplicate genera in a node; Table S1). This subsampling procedure indicates that for the Campanian substages, generic richness values greater than 16 and for the Maastrichtian substages generic richness values greater than 18 are likely to capture the majority of common FEs in each substage. This shift in the generic richness level required to adequately subsample the different stages is most likely due to a shift in the dominance of specific FEs and decreasing functional evenness, as is observed at both the paleo-latitude and substage level. When using only specific Classes of taxa, the limits are much lower, particularly for Cephalopods and Gastropods. Bivalves, which have a high functional diversity (Table 3) have SQS sample size estimates of approximately ten in the Campanian and 11 and 13 in the Lower and Upper Maastrichtian, respectively. Gastropods, also a functionally diverse group, are dominated by a few important functional entities, meaning that their SQS sample size estimate is approximately two in the Campanian and three in the Maastrichtian. The Cephalopods represent only four FEs, and their SQS sample size estimate is three for the Campanian and two for the Maastrichtian.

Table S1. Lower sample size estimates provided by SQS subsampling for each substage:

<table>
<thead>
<tr>
<th>Substage</th>
<th>Entire Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Maastr.</td>
<td>18</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>17.9</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>15.7</td>
</tr>
<tr>
<td>Middle Camp.</td>
<td>15.7</td>
</tr>
<tr>
<td>Lower Camp.</td>
<td>15.3</td>
</tr>
</tbody>
</table>
Note on Shannon Diversity Index:
Initially, the Shannon Diversity Index (SEI) was calculated alongside Simpson’s Measure of Evenness (SME), but given that the SEI values followed identical patterns to SME but with values which were not as easy to read and compare, and given that SEI is more susceptible to bias and error (Magurran 2003), it was not included in the final manuscript. However, SEI results are included here in the Supplementary Material for transparency. SEI is calculated based on the following formula:

\[
\text{Shannon Equitability Index (H1)} = \frac{-\sum(p_i \cdot \ln(p_i))}{\ln(S)}
\]  

(EQ 1)

SEI ranges from 0 to 1 with 1 indicating perfect evenness.

In all cases, SME estimates more significant proportional changes in evenness through time, as can be seen here. SME is an evenness metric which is weighted by the abundance or dominance of specific species (or in this case, FE) while SEI is a standardized form of the Shannon index, which quantifies aspects of diversity based on richness (Magurran, 2004). The SEI is therefore more prone to error when sample sizes are low, and not all elements (i.e., FE) are represented in the sample. For this reason, more emphasis is placed on SME in this analysis, though SEI is included for comparison. Though the two metrics of evenness produce such strikingly different proportional changes, they each show shifts in the same direction in all cases.
Functional Diversity Patterns:
Not all FE are present in all substages but the relative abundance of each is generally very stable through time (Figure 5, Table S4). Of the FE which have continuous presence across the study interval, only two show an increase or decrease in the relative abundance of genera greater than 0.05: facultatively mobile, infaunal suspension feeders and mobile, unattached, epifaunal carnivores, and these only once in each. These FEs are two of the three FEs with highest relative abundance in each substage. The third FE with consistently high relative abundance throughout the study interval is immobile, attached, epifaunal suspension feeders. These three FEs account for at least 8% of all genera in each substage, though which of the three contains the largest relative proportion of genera in each substage does change across the interval. In the Lower Campanian, immobile, attached, epifaunal suspension feeders have the highest relative abundance of genera (16%) and in the Middle Campanian, facultatively mobile, unattached, infaunal suspension feeders have the highest relative abundance of genera (17%). In the Upper Campanian and both Maastrichtian substages, mobile, unattached, epifaunal carnivores have the highest relative abundance (30%, 26%, and 29%, respectively).

Most FEs in each substage contain less than 1% of all genera in each substage, and some FEs are lost or gained across substages (Figure 5, Table S4). Between the Lower and Middle Campanian the total of 30 FE is maintained though there is some shift in the specific FE present. Facultatively mobile, unattached, epifaunal herbivores, which only make up 1% of the relative abundance of genera, are lost and immobile, attached, infaunal suspension feeders, which make up only 3% of the Middle Campanian relative abundance are gained. Between the Middle and Late Campanian, however, no FEs are lost and facultatively mobile, unattached, epifaunal herbivores, making up 2% of the relative abundance of genera, are regained and mobile, unattached, epifaunal, detritivores (1% relative abundance of genera) and immobile, attached, epifaunal, photosymbiotic taxa (1% relative abundance of genera) are gained. Between the Upper Campanian and Lower Maastrichtian, immobile, attached, epifaunal, photosymbiotic taxa are lost but immobile, unattached, boring, suspension feeders (2% relative abundance of genera), facultatively mobile, attached, infaunal, detritivores (1% relative abundance of genera), and immobile,
attached, semi-infaunal, photosymbiots (1% relative abundance of genera) are gained. Finally, all FE are present in the Upper Maastrichtian, meaning that facultatively mobile, unattached, boring suspension feeders (2% relative abundance of genera), and immobile, unattached, semi-infaunal, photosymbiotic taxa (1% relative abundance of genera) are gained and immobile, attached, epifaunal photosymbionts (2% relative abundance of genera) are regained.

In total, eight FE which were not originally present in the Lower Campanian are gained at some point across the study interval, but each has very low relative abundance which do not greatly change across substages (1-3%). Only two FE are “lost” at some point across the study interval, but all are recovered by the Upper Maastrichtian. Again, all FE which appear to be lost have very low relative abundances. These patterns indicate that these FE are probably rare and their loss/gain may be a product of fossilization bias or inadequate sampling. Based on previous analyses which have looked at global patterns of FE (Foster and Twitchet, 2014; Edie et al., 2018), it seems highly unlikely that these FE had gone globally extinct across the study intervals, though they may have been regionally extirpated for a time.
Table S2. Table of the functional entities present in each substage.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IA-E-SF</td>
<td>IA-E-SF</td>
<td>IA-E-SF</td>
<td>IA-E-SF</td>
<td>IA-E-SF</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>FU-I-C</td>
<td>FU-I-C</td>
<td>FU-I-C</td>
<td>FU-I-C</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>FU-I-CH</td>
<td>FU-I-CH</td>
<td>FU-I-CH</td>
<td>FU-I-CH</td>
</tr>
<tr>
<td>MU-Pi-D</td>
<td>MU-Pi-D</td>
<td>MU-Pi-D</td>
<td>MU-Pi-D</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>IU-B-SF</td>
<td>IU-B-SF</td>
<td>IU-B-SF</td>
<td>IU-B-SF</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>IU-I-SF</td>
<td>IU-I-SF</td>
<td>IU-I-SF</td>
<td>IU-I-SF</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>FA-E-SF</td>
<td>FA-E-SF</td>
<td>FA-E-SF</td>
<td>FA-E-SF</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>FU-E-SF</td>
<td>FU-E-SF</td>
<td>FU-E-SF</td>
<td>FU-E-SF</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>MU-N-SF</td>
<td>MU-N-SF</td>
<td>MU-N-SF</td>
<td>MU-N-SF</td>
</tr>
<tr>
<td>MU-Pi-C</td>
<td>MU-Pi-C</td>
<td>MU-Pi-C</td>
<td>MU-Pi-C</td>
<td>MU-Pi-C</td>
</tr>
<tr>
<td>FU-I-SF</td>
<td>FU-I-SF</td>
<td>FU-I-SF</td>
<td>FU-I-SF</td>
<td>FU-I-SF</td>
</tr>
<tr>
<td>MU-Pi-H</td>
<td>MU-Pi-H</td>
<td>MU-Pi-H</td>
<td>MU-Pi-H</td>
<td>MU-Pi-H</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>MU-I-C</td>
<td>MU-I-C</td>
<td>MU-I-C</td>
<td>MU-I-C</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>MU-I-CH</td>
<td>MU-I-CH</td>
<td>MU-I-CH</td>
<td>MU-I-CH</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>MU-E-O</td>
<td>MU-E-O</td>
<td>MU-E-O</td>
<td>MU-E-O</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>MU-E-SF</td>
<td>MU-E-SF</td>
<td>MU-E-SF</td>
<td>MU-E-SF</td>
</tr>
<tr>
<td>FA-Pi-SF</td>
<td>FA-Pi-SF</td>
<td>FA-Pi-SF</td>
<td>FA-Pi-SF</td>
<td>FA-Pi-SF</td>
</tr>
<tr>
<td>IA-Pi-SF</td>
<td>IA-Pi-SF</td>
<td>IA-Pi-SF</td>
<td>IA-Pi-SF</td>
<td>IA-Pi-SF</td>
</tr>
<tr>
<td>MU-E-C</td>
<td>MU-E-C</td>
<td>MU-E-C</td>
<td>MU-E-C</td>
<td>MU-E-C</td>
</tr>
<tr>
<td>MU-I-SF</td>
<td>MU-NB-SF</td>
<td>MU-I-SF</td>
<td>MU-I-SF</td>
<td>MU-I-SF</td>
</tr>
<tr>
<td>MU-NB-SF</td>
<td>MU-NB-SF</td>
<td>MU-NB-SF</td>
<td>MU-NB-SF</td>
<td>MU-NB-SF</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>IU-E-SF</td>
<td>IU-E-SF</td>
<td>IU-E-SF</td>
<td>IU-E-SF</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>MU-N-C</td>
<td>MU-N-C</td>
<td>MU-N-C</td>
<td>MU-N-C</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>MU-NB-C</td>
<td>FU-I-D</td>
<td>FU-I-D</td>
<td>FU-I-D</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>MU-I-D</td>
<td>MU-NB-C</td>
<td>MU-NB-C</td>
<td>MU-NB-C</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>IA-I-SF</td>
<td>MU-I-D</td>
<td>MU-I-D</td>
<td>MU-I-D</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

SI Purcell et al., 20XX 326
Summary of Regional Functional vs Taxonomic Diversity change in Substages:
Table S3. Table of proportional change in metrics across substages in the raw database.

<table>
<thead>
<tr>
<th>Proportional Change Across Substages:</th>
<th>GR</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Camp. - Mid Camp.</td>
<td>0.27</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Mid Camp. - Up Camp.</td>
<td>0.81</td>
<td>0.10</td>
<td>-0.50</td>
<td>-0.11</td>
</tr>
<tr>
<td>Up Camp. - Low Maastr.</td>
<td>-0.09</td>
<td>0.06</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Low Maastr. - Up Maastr.</td>
<td>0.08</td>
<td>0.09</td>
<td>-0.19</td>
<td>-0.03</td>
</tr>
<tr>
<td>Average:</td>
<td>0.27</td>
<td>0.06</td>
<td>-0.13</td>
<td>-0.02</td>
</tr>
<tr>
<td>Median:</td>
<td>0.17</td>
<td>0.07</td>
<td>-0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td>S.D.:</td>
<td>0.39</td>
<td>0.04</td>
<td>0.27</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Table S4. Relative abundance of each FE in each substage. Highest relative abundance values in each substage are in bold. The three most abundant FEs in all substages are denoted with an asterisk.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>*IA-E-SF</td>
<td>0.161</td>
<td>0.115</td>
<td>0.097</td>
<td>0.120</td>
<td>0.111</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>0.021</td>
<td>0.016</td>
<td>0.009</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>0.021</td>
<td>0.033</td>
<td>0.012</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-PI-D</td>
<td>0.021</td>
<td>0.027</td>
<td>0.006</td>
<td>0.007</td>
<td>0.012</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>0.021</td>
<td>0.005</td>
<td>0.018</td>
<td>0.023</td>
<td>0.022</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>0.021</td>
<td>0.016</td>
<td>0.006</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>0.021</td>
<td>0.044</td>
<td>0.033</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>0.021</td>
<td>0.016</td>
<td>0.021</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>0.021</td>
<td>0.066</td>
<td>0.033</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>0.014</td>
<td>0.016</td>
<td>0.018</td>
<td>0.020</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>0.014</td>
<td>0.011</td>
<td>0.043</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>*FU-I-SF</td>
<td>0.133</td>
<td>0.170</td>
<td>0.100</td>
<td>0.113</td>
<td>0.090</td>
</tr>
<tr>
<td>MU-PI-H</td>
<td>0.014</td>
<td>0.022</td>
<td>0.009</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>0.014</td>
<td>0.011</td>
<td>0.018</td>
<td>0.020</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>0.007</td>
<td>0.005</td>
<td>0.003</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>0.007</td>
<td>0.005</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>FU-E-H</td>
<td>0.007</td>
<td>NA</td>
<td>0.006</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>FU-PI-H</td>
<td>0.007</td>
<td>0.005</td>
<td>0.015</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>0.007</td>
<td>0.016</td>
<td>0.012</td>
<td>0.010</td>
<td>0.031</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>0.007</td>
<td>0.011</td>
<td>0.012</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>FA-PI-SF</td>
<td>0.007</td>
<td>0.005</td>
<td>0.003</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>IA-PI-SF</td>
<td>0.007</td>
<td>0.005</td>
<td>0.003</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>*MU-E-C</td>
<td>0.091</td>
<td>0.077</td>
<td>0.301</td>
<td>0.257</td>
<td>0.294</td>
</tr>
<tr>
<td>MU-I-SF</td>
<td>0.007</td>
<td>0.016</td>
<td>0.009</td>
<td>0.013</td>
<td>0.009</td>
</tr>
<tr>
<td>MU-NB-SF</td>
<td>0.084</td>
<td>0.066</td>
<td>0.049</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>0.063</td>
<td>0.038</td>
<td>0.024</td>
<td>0.013</td>
<td>0.012</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>0.049</td>
<td>0.044</td>
<td>0.024</td>
<td>0.013</td>
<td>0.006</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>0.049</td>
<td>0.038</td>
<td>0.024</td>
<td>0.033</td>
<td>0.028</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>0.042</td>
<td>0.049</td>
<td>0.043</td>
<td>0.013</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>0.042</td>
<td>0.033</td>
<td>0.030</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>IA-I-SF</td>
<td>NA</td>
<td>0.011</td>
<td>0.006</td>
<td>0.013</td>
<td>0.003</td>
</tr>
<tr>
<td>MU-E-D</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>IA-E-P</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
<td>NA</td>
<td>0.009</td>
</tr>
<tr>
<td>IA-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>FA-I-D</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>IA-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>FU-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.006</td>
</tr>
<tr>
<td>IU-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Basin-level Substage Trends:

When data representative of the WIS only are subset and analyzed, only five FEs are lost: facultatively mobile, attached, infaunal deposit feeders (bivalve); facultatively mobile, unattached, boring suspension feeders (bivalve); immobile, attached, boring suspension feeders (bivalve, sponge); immobile, attached, epifaunal photosymbionts (bivalve, coral); immobile, attached, semi-infaunal photosymbionts (bivalve); and immobile unattached, semi-infaunal photosybionts (bivalve). These results show that only one FE with a boring lifestyle (immobile, unattached, boring suspension feeders) are present, represented by bivalves, and no photosymbionts are present once GCP data are removed. Indeed, these same five FEs not present in the WIS alone among the eight FEs that are “gained” through time in the database (Figure 5). While these FEs are consistently rare in the database, accounting for less than 0.9% of all genera in a given substage, their absence from the WIS suggests that these FEs may indeed not have existed in the WIS during the Campanian and Maastrichtian.

When considering the WIS alone, generic richness rises through the Campanian to peak in the Upper Campanian, before declining through the Maastrichtian (Table S5; Figure S2). FR parallels this general trend, but proportionally changes much less across the Campanian (only 4% and 7% increase from Lower to Upper Campanian), before falling dramatically through the Maastrichtian (13% and 30% decrease; Table S6). SME decreases first by only 4% from the Lower to Middle Campanian, then by 20% from the Middle to Upper Campanian before rising by first 23% then 54% across the Upper Campanian to Lower Maastrichtian and the Lower to Upper Maastrichtian, respectively. This suggests that functional evenness fell in the later part of the Campanian in the WIS as generic richness increased, even when FR remained fairly stable, but when generic richness fell in the later substages, evenness increased. This pattern is similar to the one observed when both the WIS and GCP data are considered together (Table S3, Figure 4), indicating that as generic richness decreases, evenness increases as taxa more evenly distributed within remaining FEs. Declining FR may also contribute to this phenomenon. Indeed, the loss of FEs with low relative abundance of genera most likely contributes to rising evenness, as only those FEs with high redundancy are left.
The GCP region alone experiences a consistent increase in both generic and functional richness throughout the study interval (Table S7; Figure S3). Interestingly, the inverse pattern of rising generic richness coupled with a clear decline in SME across the Middle to Upper Campanian is easily observed in the GCP alone (Table S8). Again, this indicates that, as generic richness increases, and despite a parallel increase in FR, uneven packing within FEs increases. Similarly, to when FEs with low relative abundance of genera are lost and evenness increases, most FE that are gained through time in the GCP have very low relative abundances, increasing the uneven distribution of taxa within FEs (Table S10, Figure S5). It should also be noted, however, that shifts in relative abundance of genera in the GCP FEs is more dramatic that those observed in the WIS or overall database (Figure S4 and 5).
Table S5. Table of the generic richness, FR, SME, and SEI values in each substage in the WIS.

<table>
<thead>
<tr>
<th>WIS Data</th>
<th>Generic Richness</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>108</td>
<td>28</td>
<td>0.53</td>
<td>0.87</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>151</td>
<td>29</td>
<td>0.51</td>
<td>0.88</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>199</td>
<td>31</td>
<td>0.41</td>
<td>0.84</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>119</td>
<td>27</td>
<td>0.51</td>
<td>0.87</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>42</td>
<td>19</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td>Mean</td>
<td>123.8</td>
<td>26.8</td>
<td>0.55</td>
<td>0.87</td>
</tr>
<tr>
<td>Median</td>
<td>119</td>
<td>28</td>
<td>0.51</td>
<td>0.87</td>
</tr>
<tr>
<td>S.D.</td>
<td>57.79</td>
<td>4.60</td>
<td>0.14</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table S6. Table of proportional change in metrics across substages in the WIS.

<table>
<thead>
<tr>
<th>WIS Data</th>
<th>Proportional Change Across Substages:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>Low Camp. - Mid Camp.</td>
<td>0.40</td>
</tr>
<tr>
<td>Mid Camp. - Up Camp.</td>
<td>0.32</td>
</tr>
<tr>
<td>Up Camp. - Low Maastr.</td>
<td>-0.40</td>
</tr>
<tr>
<td>Low Maastr. - Up Maastr.</td>
<td>-0.65</td>
</tr>
<tr>
<td>Average:</td>
<td>-0.08</td>
</tr>
<tr>
<td>Median:</td>
<td>-0.04</td>
</tr>
<tr>
<td>S.D.:</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Figure S4. Line plots of changes in functional ecology metrics across substages for WIS. The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric.
Figure S5. Plot of relative abundance of each FE through time based WIS data alone. Circle size indicates relative abundance of a FE within each substage and color indicates relative abundance increase across subsequent substages greater than 0.05 (orange), decrease greater than 0.05 (blue), and increase/decrease less than 0.05 (yellow). Points are grey if they cannot be compared with a previous substage. See Table 1 for keys to each FE names along the x-axis.
Table S7. Table of the generic richness, FR, SME, and SEI values in each substage in the GCP.

<table>
<thead>
<tr>
<th>GCP Data</th>
<th>Generic Richness</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>47</td>
<td>14</td>
<td>0.61</td>
<td>0.87</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>65</td>
<td>22</td>
<td>0.67</td>
<td>0.90</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>213</td>
<td>29</td>
<td>0.21</td>
<td>0.76</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>248</td>
<td>34</td>
<td>0.26</td>
<td>0.79</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>309</td>
<td>38</td>
<td>0.22</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Mean: 176.4, 27.4, 0.39, 0.82
Median: 213, 29, 0.26, 0.79
S.D.: 115.33, 9.58, 0.23, 0.06

Table S8. Table of proportional change in metrics across substages in the GCP.

<table>
<thead>
<tr>
<th>GCP Data</th>
<th>Proportional Change Across Substages:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>Low Camp. - Mid Camp.</td>
<td>0.38</td>
</tr>
<tr>
<td>Mid Camp. - Up Camp.</td>
<td>2.28</td>
</tr>
<tr>
<td>Up Camp. - Low Maastr.</td>
<td>0.16</td>
</tr>
<tr>
<td>Low Maastr. - Up Maastr.</td>
<td>0.25</td>
</tr>
<tr>
<td>Average:</td>
<td>0.77</td>
</tr>
<tr>
<td>Median:</td>
<td>0.31</td>
</tr>
<tr>
<td>S.D.:</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Figure S6. Line plots of changes in functional ecology metrics across substages for GCP. The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric.
Figure S7. Plot of relative abundance of each FE through time based on GCP data alone. Circle size indicates relative abundance of a FE within each substage and color indicates relative abundance increase across subsequent substages greater than 0.05 (orange), decrease greater than 0.05 (blue), and increase/decrease less than 0.05 (yellow). Points are grey if they cannot be compared with a previous substage. See Table 1 for keys to each FE names along the x-axis.
Table S9. Relative abundance of each FE in each substage for the WIS.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FA-E-SF</td>
<td>0.028</td>
<td>0.046</td>
<td>0.055</td>
<td>0.042</td>
<td>NA</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>0.019</td>
<td>0.013</td>
<td>0.020</td>
<td>0.034</td>
<td>0.048</td>
</tr>
<tr>
<td>FA-PI-SF</td>
<td>0.009</td>
<td>0.007</td>
<td>0.005</td>
<td>0.008</td>
<td>0.024</td>
</tr>
<tr>
<td>FU-E-H</td>
<td>0.009</td>
<td>NA</td>
<td>0.005</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>0.028</td>
<td>0.020</td>
<td>0.020</td>
<td>0.025</td>
<td>0.024</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>0.028</td>
<td>0.013</td>
<td>0.015</td>
<td>0.017</td>
<td>NA</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>0.028</td>
<td>0.040</td>
<td>0.015</td>
<td>0.025</td>
<td>0.024</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>0.065</td>
<td>0.046</td>
<td>0.030</td>
<td>0.050</td>
<td>0.071</td>
</tr>
<tr>
<td>FU-I-SF</td>
<td>0.167</td>
<td>0.185</td>
<td>0.151</td>
<td>0.176</td>
<td>0.119</td>
</tr>
<tr>
<td>FU-PI-H</td>
<td>0.009</td>
<td>NA</td>
<td>0.005</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>0.102</td>
<td>0.079</td>
<td>0.111</td>
<td>0.126</td>
<td>0.119</td>
</tr>
<tr>
<td>IA-I-SF</td>
<td>NA</td>
<td>0.013</td>
<td>0.005</td>
<td>0.008</td>
<td>NA</td>
</tr>
<tr>
<td>IA-PI-SF</td>
<td>0.009</td>
<td>0.007</td>
<td>0.005</td>
<td>0.008</td>
<td>NA</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>0.028</td>
<td>0.007</td>
<td>0.025</td>
<td>0.025</td>
<td>0.024</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>0.074</td>
<td>0.040</td>
<td>0.035</td>
<td>0.008</td>
<td>NA</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>0.028</td>
<td>0.013</td>
<td>0.010</td>
<td>0.017</td>
<td>0.024</td>
</tr>
<tr>
<td>MU-E-C</td>
<td>0.102</td>
<td>0.073</td>
<td>0.171</td>
<td>0.118</td>
<td>0.048</td>
</tr>
<tr>
<td>MU-E-D</td>
<td>NA</td>
<td>NA</td>
<td>0.005</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>0.019</td>
<td>0.013</td>
<td>0.020</td>
<td>0.017</td>
<td>NA</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>NA</td>
<td>0.013</td>
<td>0.015</td>
<td>0.008</td>
<td>0.071</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>0.009</td>
<td>0.007</td>
<td>0.015</td>
<td>0.017</td>
<td>0.024</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>NA</td>
<td>0.007</td>
<td>0.005</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>0.009</td>
<td>0.007</td>
<td>0.005</td>
<td>0.008</td>
<td>NA</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>0.028</td>
<td>0.040</td>
<td>0.040</td>
<td>0.050</td>
<td>0.024</td>
</tr>
<tr>
<td>MU-I-SF</td>
<td>0.009</td>
<td>0.020</td>
<td>0.015</td>
<td>0.025</td>
<td>NA</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>0.028</td>
<td>0.046</td>
<td>0.050</td>
<td>0.034</td>
<td>0.024</td>
</tr>
<tr>
<td>MU-NB-SF</td>
<td>0.093</td>
<td>0.079</td>
<td>0.065</td>
<td>0.067</td>
<td>0.190</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>0.028</td>
<td>0.033</td>
<td>0.015</td>
<td>0.034</td>
<td>0.048</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>0.009</td>
<td>0.073</td>
<td>0.030</td>
<td>0.017</td>
<td>0.048</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>0.009</td>
<td>0.020</td>
<td>0.020</td>
<td>0.017</td>
<td>0.024</td>
</tr>
<tr>
<td>MU-PI-D</td>
<td>0.009</td>
<td>0.013</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MU-PI-H</td>
<td>0.019</td>
<td>0.026</td>
<td>0.015</td>
<td>0.017</td>
<td>0.024</td>
</tr>
</tbody>
</table>
Table S10. Relative abundance of each FE in each substage for the GCP.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IA-E-SF</td>
<td>0.277</td>
<td>0.185</td>
<td>0.075</td>
<td>0.117</td>
<td>0.117</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>0.021</td>
<td>0.015</td>
<td>0.019</td>
<td>0.024</td>
<td>0.016</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>0.021</td>
<td>NA</td>
<td>0.005</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>0.021</td>
<td>0.015</td>
<td>0.005</td>
<td>0.008</td>
<td>0.023</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>0.021</td>
<td>0.015</td>
<td>0.028</td>
<td>0.036</td>
<td>0.029</td>
</tr>
<tr>
<td>FU-I-SF</td>
<td>0.021</td>
<td>0.138</td>
<td>0.056</td>
<td>0.089</td>
<td>0.084</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>0.149</td>
<td>0.062</td>
<td>0.028</td>
<td>NA</td>
<td>0.003</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>0.106</td>
<td>0.062</td>
<td>0.047</td>
<td>0.008</td>
<td>0.016</td>
</tr>
<tr>
<td>MU-NB-SF</td>
<td>0.085</td>
<td>0.046</td>
<td>0.047</td>
<td>0.032</td>
<td>0.029</td>
</tr>
<tr>
<td>MU-E-C</td>
<td>0.064</td>
<td>0.062</td>
<td>0.380</td>
<td>0.286</td>
<td>0.301</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>0.064</td>
<td>0.031</td>
<td>0.033</td>
<td>0.032</td>
<td>0.029</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>0.064</td>
<td>0.062</td>
<td>0.019</td>
<td>0.016</td>
<td>0.013</td>
</tr>
<tr>
<td>MU-N-D</td>
<td>0.043</td>
<td>0.046</td>
<td>0.009</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>0.043</td>
<td>0.046</td>
<td>0.023</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>NA</td>
<td>0.031</td>
<td>0.009</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>NA</td>
<td>0.031</td>
<td>0.005</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>NA</td>
<td>0.031</td>
<td>0.014</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>NA</td>
<td>0.015</td>
<td>0.014</td>
<td>0.020</td>
<td>0.013</td>
</tr>
<tr>
<td>FU-PI-H</td>
<td>NA</td>
<td>0.015</td>
<td>0.023</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td>MU-PI-H</td>
<td>NA</td>
<td>0.015</td>
<td>0.005</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>IA-PI-SF</td>
<td>NA</td>
<td>0.015</td>
<td>NA</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>MU-I-SF</td>
<td>NA</td>
<td>0.015</td>
<td>0.014</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>NA</td>
<td>0.046</td>
<td>0.023</td>
<td>0.040</td>
<td>0.042</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>NA</td>
<td>NA</td>
<td>0.019</td>
<td>0.032</td>
<td>0.026</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>NA</td>
<td>NA</td>
<td>0.019</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>FU-E-H</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
<td>0.020</td>
<td>0.023</td>
</tr>
<tr>
<td>IA-E-P</td>
<td>NA</td>
<td>NA</td>
<td>0.005</td>
<td>NA</td>
<td>0.010</td>
</tr>
<tr>
<td>IA-I-SF</td>
<td>NA</td>
<td>NA</td>
<td>0.005</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>NA</td>
<td>NA</td>
<td>0.052</td>
<td>0.056</td>
<td>0.055</td>
</tr>
<tr>
<td>IA-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>FA-PI-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>FA-I-D</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>IA-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>MU-E-D</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>FU-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.006</td>
</tr>
<tr>
<td>IU-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Class-level Influences:

Generic richness for cephalopods increases greatly across the Campanian (by 46% and 17%) but declines from the Upper Campanian to Lower Maastrichtian (54%; Table S11; Figure S8). SME is very high within cephalopod FEs, ranging from 0.76 to 1.05, and shifts by no more than 16% across any substage (Table S12). This means that evenness values for cephalopods are consistent, and only change by more than 8% between the Lower and Middle Campanian (16% increase) and Upper Campanian to Lower Maastrichtian (16% decrease; Table S12). These changes are caused by shifts in the distribution of existing genera within the four FEs (Figure S9). Mobile, unattached, nekton-benthic suspension feeders is consistently the most common FE, but it and all but one other cephalopod FE experiences a rise or decline in relative abundance greater than 5% across these intervals (Table S17). Overall, cephalopods make up a functionally highly stable portion of the database, despite dramatic changes to generic richness.

Bivalves are the most common taxonomic group in the database, including a total of 184 genera and making up a total of 24 of the 38 FEs, meaning that they most likely have the strongest influence on overall patterns in the database. Bivalve generic and functional richness increase across the study interval and SME values are fairly stable (Table S13; Figure S10), indicating that packing within bivalve FEs is consistent through time (Figure S10). SME only changes more than 7% between the Lower and Middle Campanian, where SME decreases by 23% (Table S14). This decrease in evenness occurs at the same time as a dramatic increase in bivalve generic diversity (33%) and a slight increase in bivalve FR (6%). However, subsequent increases in both generic and functional richness, such as between the Middle and Upper Campanian, are not coupled with a decrease in SME but rather a very slight increase (7%). Across the later substages, SME declines only very slightly (2% and 3%) while generic richness increases slightly (8%) and decreases slightly (5%) and FR increases (5% and 15%). SME is decoupled from both generic and FR, remaining fairly stable in substages after the Middle Campanian, despite changes to the other two metrics.
Facultatively unattached, infaunal suspension feeders are represented by the most genera in each substage, and increase in dominance from the Lower to Middle Campanian while other FEs, including those while high relative abundance in the Lower Campanian, decline in dominance (Figure S11; Table S18). Furthermore, eight FEs are “gained” throughout the study interval among bivalves alone (Figure S11). Gaining FEs in the collective database was assumed to result in declining evenness, but in the case of bivalves, only a single FE is gained between the Lower and Middle Campanian (immobile, attached, infaunal suspension feeders) declines. Overall, these patterns suggest that declining evenness between the Early to Middle Campanian among bivalves is caused by increased packing of genera into the facultatively mobile, unattached infaunal suspension feeder group, the addition of the immobile, attached, infaunal suspension feeders, and loss of genera in other FEs. Throughout the rest of the substages, evenness is highly stable among bivalves despite changes to both generic and functional richness, as the addition of genera is coupled with the addition of FE and no significant packing into any one group.

Gastropods are the second most common class of taxa in the database, making up 184 genera and eleven FEs. They are therefore likely to greatly influence general patterns in the database. Notably, gastropod generic richness increases by 408% across the Middle to Upper Campanian, but it experiences shifts of greater than 14% across all other substages as well (rises 44% from the Lower to Middle Campanian, declines 14% from the Upper Campanian to the Lower Maastrichtian, and rises 17% through the Maastrichtian; Table S15; Figure S12). FR is stable in the Early to Middle Campanian, but rises by 57% between the Middle and Upper Campanian when generic richness greatly increases, before falling between the Upper Campanian and Lower Maastrichtian (18%), and rising in the Maastrichtian (22%; Table S16). Functional evenness in gastropod genera is much lower than that seen in bivalves, except in the Middle Campanian, where gastropod SME is 0.52 and bivalves are only 0.39, and fluctuates strongly. In particular, a sharp decrease in SME from the Middle to Upper Campanian (66%) coincides with the increase to both generic and functional richness.
Mobile, unattached, epifaunal carnivores are consistently the most dominate gastropod FE, taking up at least 50% of all genera in a given substage (Table S19; Figure S13). However, between the Middle and Upper Campanian, this FE experiences a dramatic increase from a relative abundance of 50% to 70% while, at the same time, two of the other most abundant FEs (MU-PI-C and MU-PI-H) experience a decline in relative abundance from 15% and 12% to 3% and 2%, respectively. Furthermore, four FEs are gained between these substages. These factors all contribute to the sharp decrease in evenness between the Middle and Upper Campanian. Throughout the rest of the intervals excepting the Late Maastrichtian, the relative abundance of mobile, unattached, epifaunal carnivores increases or decreases by more than 5% and as FE are lost, gained, and change in relative abundance. Clearly, gastropods are highly variable in generic richness as well as FR and functional evenness relative to bivalves. Their presence, therefore, likely contributes to the greater variability of the database as a whole relative to one including bivalves alone.
Figure S8. Line plots of changes in functional ecology metrics across substages for cephalopods only. The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric.
Figure S9. Plot of relative abundance of each FE through time based cephalopod data alone. Circle size indicates relative abundance of a FE within each substage and color indicates relative abundance increase across subsequent substages greater than 0.05 (orange), decrease greater than 0.05 (blue), and increase/decrease less than 0.05 (yellow). Points are grey if they cannot be compared with a previous substage. See Table 1 for keys to each FE names along the x-axis.
Figure S10. Line plots of changes in functional ecology metrics across substages for bivalves only. The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric.
Figure S11. Plot of relative abundance of each FE through time based bivalve data alone. Circle size indicates relative abundance of a FE within each substage and color indicates relative abundance increase across subsequent substages greater than 0.05 (orange), decrease greater than 0.05 (blue), and increase/decrease less than 0.05 (yellow). Points are grey if they cannot be compared with a previous substage. See Table 1 for keys to each FE names along the x-axis.
Figure S12. Line plots of changes in functional ecology metrics across substages for gasropods only. The left y-axis in all plots depicts generic richness values, while the right y-axis presents values for the other metric.
Figure S13. Plot of relative abundance of each FE through time based gastropod data alone. Circle size indicates relative abundance of a FE within each substage and color indicates relative abundance increase across subsequent substages greater than 0.05 (orange), decrease greater than 0.05 (blue), and increase/decrease less than 0.05 (yellow). Points are grey if they cannot be compared with a previous substage. See Table 1 for keys to each FE names along the x-axis.
Table S11. Table of the generic richness, FR, SME, and SEI values in each substage in cephalopods only.

<table>
<thead>
<tr>
<th>Substage</th>
<th>Generic Richness</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>28</td>
<td>4</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>41</td>
<td>4</td>
<td>1.05</td>
<td>0.99</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>48</td>
<td>4</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>22</td>
<td>4</td>
<td>0.83</td>
<td>0.89</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>23</td>
<td>4</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>Mean</td>
<td>32.4</td>
<td>4</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>Median</td>
<td>28</td>
<td>4</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>S.D.</td>
<td>11.55</td>
<td>0.00</td>
<td>0.11</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table S12. Table of proportional change in metrics across substages in cephalopods only.

<table>
<thead>
<tr>
<th>Proportional Change Across Substages:</th>
<th>GR</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Camp.-Mid Camp.</td>
<td>0.46</td>
<td>0.00</td>
<td>0.16</td>
<td>0.07</td>
</tr>
<tr>
<td>Mid Camp.-Up Camp.</td>
<td>0.17</td>
<td>0.00</td>
<td>-0.06</td>
<td>-0.02</td>
</tr>
<tr>
<td>Up Camp.-Low Maastr.</td>
<td>-0.54</td>
<td>0.00</td>
<td>-0.16</td>
<td>-0.08</td>
</tr>
<tr>
<td>Low Maastr.-Up Maastr.</td>
<td>0.05</td>
<td>0.00</td>
<td>-0.08</td>
<td>-0.04</td>
</tr>
<tr>
<td>Average</td>
<td>0.03</td>
<td>0.00</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>Median</td>
<td>0.11</td>
<td>0.00</td>
<td>-0.07</td>
<td>-0.03</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.42</td>
<td>0.00</td>
<td>0.14</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table S13. Table of the generic richness, FR, SME, and SEI values in each substage in bivalves only.

<table>
<thead>
<tr>
<th>Substage</th>
<th>Generic Richness</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>72</td>
<td>16</td>
<td>0.51</td>
<td>0.84</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>96</td>
<td>17</td>
<td>0.39</td>
<td>0.80</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>116</td>
<td>19</td>
<td>0.42</td>
<td>0.82</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>125</td>
<td>20</td>
<td>0.41</td>
<td>0.82</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>119</td>
<td>23</td>
<td>0.40</td>
<td>0.82</td>
</tr>
<tr>
<td>Mean</td>
<td>105.6</td>
<td>19</td>
<td>0.43</td>
<td>0.82</td>
</tr>
<tr>
<td>Median</td>
<td>116</td>
<td>19</td>
<td>0.41</td>
<td>0.82</td>
</tr>
<tr>
<td>S.D.</td>
<td>21.71</td>
<td>2.74</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table S14. Table of proportional change in metrics across substages in bivalves only.

<table>
<thead>
<tr>
<th></th>
<th>Proportional Change Across Substages:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>Low Camp.-Mid Camp.</td>
<td>0.33</td>
</tr>
<tr>
<td>Mid Camp.-Up Camp.</td>
<td>0.21</td>
</tr>
<tr>
<td>Up Camp.-Low Maastr.</td>
<td>0.08</td>
</tr>
<tr>
<td>Low Maastr.-Up Maastr.</td>
<td>-0.05</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td><strong>0.14</strong></td>
</tr>
<tr>
<td><strong>Median:</strong></td>
<td><strong>0.14</strong></td>
</tr>
<tr>
<td><strong>S.D.:</strong></td>
<td><strong>0.16</strong></td>
</tr>
</tbody>
</table>

Table S15. Table of the generic richness, FR, SME, and SEI values in each substage in gastropods only.

<table>
<thead>
<tr>
<th></th>
<th>Generic Richness</th>
<th>FR</th>
<th>SME</th>
<th>SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Camp.</td>
<td>18</td>
<td>7</td>
<td>0.47</td>
<td>0.75</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>26</td>
<td>7</td>
<td>0.52</td>
<td>0.79</td>
</tr>
<tr>
<td>Upper Camp.</td>
<td>132</td>
<td>11</td>
<td>0.18</td>
<td>0.49</td>
</tr>
<tr>
<td>Lower Maastr.</td>
<td>114</td>
<td>9</td>
<td>0.27</td>
<td>0.61</td>
</tr>
<tr>
<td>Upper Maastr.</td>
<td>133</td>
<td>11</td>
<td>0.20</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>Mean:</strong></td>
<td>84.6</td>
<td>9</td>
<td>0.33</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>Median:</strong></td>
<td>114</td>
<td>9</td>
<td>0.27</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>S.D.:</strong></td>
<td>57.71</td>
<td>2.00</td>
<td>0.16</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table S16. Table of proportional change in metrics across substages in gastropods only.

<table>
<thead>
<tr>
<th></th>
<th>Proportional Change Across Substages:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>Low Camp.-Mid Camp.</td>
<td>0.44</td>
</tr>
<tr>
<td>Mid Camp.-Up Camp.</td>
<td>4.08</td>
</tr>
<tr>
<td>Up Camp.-Low Maastr.</td>
<td>-0.14</td>
</tr>
<tr>
<td>Low Maastr.-Up Maastr.</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td><strong>1.14</strong></td>
</tr>
<tr>
<td><strong>Median:</strong></td>
<td><strong>0.31</strong></td>
</tr>
<tr>
<td><strong>S.D.:</strong></td>
<td><strong>1.97</strong></td>
</tr>
</tbody>
</table>
Table S17. Relative abundance of each FE in each substage for cephalopods only.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MU-NB-SF</td>
<td>0.429</td>
<td>0.293</td>
<td>0.333</td>
<td>0.500</td>
<td>0.522</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>0.250</td>
<td>0.195</td>
<td>0.146</td>
<td>0.182</td>
<td>0.087</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>0.214</td>
<td>0.220</td>
<td>0.292</td>
<td>0.182</td>
<td>0.217</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>0.107</td>
<td>0.293</td>
<td>0.229</td>
<td>0.136</td>
<td>0.174</td>
</tr>
</tbody>
</table>

Table S18. Relative abundance of each FE in each substage for bivalves only.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FU-I-SF</td>
<td>0.264</td>
<td>0.323</td>
<td>0.284</td>
<td>0.272</td>
<td>0.244</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>0.042</td>
<td>0.031</td>
<td>0.043</td>
<td>0.040</td>
<td>0.042</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>0.028</td>
<td>0.021</td>
<td>0.034</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>0.014</td>
<td>0.010</td>
<td>0.009</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>0.014</td>
<td>0.010</td>
<td>0.043</td>
<td>0.040</td>
<td>0.034</td>
</tr>
<tr>
<td>FA-PI-SF</td>
<td>0.014</td>
<td>0.010</td>
<td>0.009</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>IA-PI-SF</td>
<td>0.014</td>
<td>0.010</td>
<td>0.009</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>MU-I-SF</td>
<td>0.014</td>
<td>0.031</td>
<td>0.026</td>
<td>0.032</td>
<td>0.025</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>0.167</td>
<td>0.167</td>
<td>0.155</td>
<td>0.160</td>
<td>0.168</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>0.125</td>
<td>0.073</td>
<td>0.069</td>
<td>0.032</td>
<td>0.034</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>0.097</td>
<td>0.073</td>
<td>0.069</td>
<td>0.080</td>
<td>0.076</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>0.042</td>
<td>0.031</td>
<td>0.026</td>
<td>0.024</td>
<td>0.025</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>0.042</td>
<td>0.063</td>
<td>0.034</td>
<td>0.040</td>
<td>0.042</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>0.042</td>
<td>0.010</td>
<td>0.052</td>
<td>0.056</td>
<td>0.059</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>0.042</td>
<td>0.031</td>
<td>0.017</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>0.042</td>
<td>0.083</td>
<td>0.086</td>
<td>0.096</td>
<td>0.101</td>
</tr>
<tr>
<td>IA-I-SF</td>
<td>NA</td>
<td>0.021</td>
<td>0.017</td>
<td>0.032</td>
<td>0.008</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>IA-E-P</td>
<td>NA</td>
<td>NA</td>
<td>0.009</td>
<td>NA</td>
<td>0.017</td>
</tr>
<tr>
<td>FA-I-D</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>IA-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>IA-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>FU-B-SF</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.017</td>
</tr>
<tr>
<td>IU-PI-P</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Table S19. Relative abundance of each FE in each substage for gastropods only.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MU-E-C</td>
<td>0.556</td>
<td>0.500</td>
<td>0.705</td>
<td>0.623</td>
<td>0.662</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>0.111</td>
<td>0.077</td>
<td>0.106</td>
<td>0.140</td>
<td>0.113</td>
</tr>
<tr>
<td>MU-PI-H</td>
<td>0.111</td>
<td>0.154</td>
<td>0.023</td>
<td>0.026</td>
<td>0.015</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>0.056</td>
<td>0.115</td>
<td>0.045</td>
<td>0.053</td>
<td>0.038</td>
</tr>
<tr>
<td>FU-E-H</td>
<td>0.056</td>
<td>NA</td>
<td>0.015</td>
<td>0.026</td>
<td>0.023</td>
</tr>
<tr>
<td>FU-PI-H</td>
<td>0.056</td>
<td>0.038</td>
<td>0.038</td>
<td>0.044</td>
<td>0.038</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>0.056</td>
<td>0.077</td>
<td>0.030</td>
<td>0.044</td>
<td>0.038</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>NA</td>
<td>0.038</td>
<td>0.008</td>
<td>NA</td>
<td>0.030</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>NA</td>
<td>0.008</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>NA</td>
<td>NA</td>
<td>0.015</td>
<td>0.035</td>
<td>0.030</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>NA</td>
<td>NA</td>
<td>0.008</td>
<td>0.009</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Summary of Functional vs Taxonomic Diversity change in Latitudinal Bins:

*Metrics in General:*

When the data is converted to paleo-coordinates and separated into seven 5° paleo-latitudinal bins for analysis, almost all bins have continuous occupation across the study interval (Table 4). Only in the Middle Campanian for the 30-35°N bin, and in the Lower and Upper Maastrichtian for the 60-65°N bin is there missing data for comparison. However, only the 30-35 and 50-55 bins have greater than the minimum estimated sample size of unique genera for each interval (see Table S1), though most substages have sufficient generic richness to represent common FEs across all but the northern-most and southern-most paleo-latitudinal bins (Table 4). Generic richness within paleo-latitude bins has greater spread of values through time but FR values have similar spreads, except in the Middle Campanian, where the first and third quartile values are closer to the median than in the other intervals (Figure S14). However, median and mean values of both generic and functional richness show similar trends within latitudinal bins, increasing across the Campanian before decreasing across the Maastrichtian. SME values have the highest spread in the Upper Maastrichtian and mean and median values appear to decrease slightly (i.e., become less even) across the Campanian then increase slightly (become more even) across the Maastrichtian (Figure S14). SEI values follow a similar pattern overall.
Figure S14. Box and whisker plots of generic richness, FR, SME, and SEI for paleo-latitudinal bins in each substage. Median values are denoted by a horizontal line and means by triangles.
Both generic and functional richness shift across time based on paleo-latitude bins (Table S20, Figure S15). Median and average proportional change in FR between substages, taking into account all paleo-latitudinal bins, indicates that FR increases between the Lower and Middle Campanian (~16% median) and between the Middle to Upper Campanian (22% median), and fell between substages from the Upper Campanian to the Lower Maastrichtian and Lower to Upper Maastrichtian (~24% median and ~3% median, respectively). Only between the Lower and Upper Maastrichtian does the median fall particularly close to zero, suggesting little overall change, but the first and third interquartile range values indicate high variability relative to other substage comparisons.

All box plots showing proportional change values for functional richness cross the zero line, either along the hinge (first or third quartile values) or along the whisker (1.5* interquartile range). This indicates that some paleo-latitude bins experience opposing directions of change in FR, despite the prevailing trends between substages. Proportional change in generic richness follow a similar general trend, except that the spread of values is consistently larger than those observed for FR and proportional changes are more extreme. For example, all proportional changes in generic richness increase between the Middle and Upper Campanian, indicating that generic richness values increase within all paleo-latitude bins between these substages. This once again suggests that generic and functional richness, despite being tightly linked, are at times decoupled, even at this scale. When paleo-latitudinal bins with less than the estimated SQS generic richness level are removed from calculations, these patterns are generally maintained, but the spread of values is less extreme in all cases (Figure S16). Only in the Upper Campanian to Lower Maastrichtian comparison do statistic value change appreciably to a median value of -43% rather than -21%, indicating stronger loss of FR.

SME values are more variable across substages, decreasing between Campanian substages and rising between the Upper Campanian and Maastrichtian substages (Figure S15). However, only the Middle to Upper Campanian comparison shows unanimous proportional decrease in values. When paleo-latitude bins with less than the estimated SQS generic richness level are removed, the spread of
proportional change values for all substages is diminished and mean and median values for SME fall closer to zero, excluding the Middle to Upper Campanian comparison where no values were removed. This suggest that the removal of generic richness values which may not appropriately represent FEs dampens both negative and positive proportional changes in evenness. Overall, proportional changes in SME indicate that the Lower to Middle Campanian has fairly stable evenness but evenness drops between the Middle to Upper Campanian, then rises from the Upper Campanian to Upper Maastrichtian.

SEI values, unlike SME, are highly stable across all intervals, particularly when paleo-latitude bins with generic richness lower than the SQS estimated requirements are removed (Figure S16). This indicates overall evenness (as opposed to dominance) in the data at the paleo-latitude-bin level of aggregation, however error introduced by lower richness levels for this metric is assumed to be high. Slight shifts in SEI proportional change values do mimic patterns observed with SME, indicating that shifts in the distribution of genera within FEs does occur through time.
Figure S15. Box and whisker plots of proportional change in paleo-latitude bins between substages for each metric value. Triangles indicate mean values and horizontal lines indicate median values.
Table S20. Summary statistics for proportional change in generic richness, FR, SME, and SEI for paleo-latitude bins across adjacent substages with all values included, regardless of if generic richness for that paleo-latitude bin was less than estimated SQS sample size requirement.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td>Mean: 0.43</td>
<td>2.07</td>
<td>1.93</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>Median: 0.39</td>
<td>1.70</td>
<td>-0.43</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>SD: 0.94</td>
<td>2.10</td>
<td>5.92</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td>Mean: 0.31</td>
<td>0.66</td>
<td>0.65</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>Median: 0.16</td>
<td>0.22</td>
<td>-0.24</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>SD: 0.88</td>
<td>1.17</td>
<td>2.34</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td>Mean: -0.04</td>
<td>-0.37</td>
<td>0.91</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Median: -0.06</td>
<td>-0.35</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>SD: 0.14</td>
<td>0.27</td>
<td>1.97</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td>Mean: 0.01</td>
<td>-0.10</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Median: 0.00</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>SD: 0.03</td>
<td>0.14</td>
<td>0.22</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table S21. Summary statistics for proportional change in generic richness, functional richness ($\alpha_f$), and functional evenness (SME) for paleo-latitude bins across adjacent substages. Camp. = Campanian, Maastr. = Maastrichtian. Values were removed from calculations if they did not meet the minimum SQS sample size estimate.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td>Mean: 0.38</td>
<td>1.29</td>
<td>-0.48</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>Median: 0.39</td>
<td>1.18</td>
<td>-0.60</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>SD: 0.54</td>
<td>0.93</td>
<td>0.39</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td>Mean: 0.21</td>
<td>0.19</td>
<td>-0.30</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>Median: 0.16</td>
<td>0.17</td>
<td>-0.41</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>SD: 0.28</td>
<td>0.25</td>
<td>0.29</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td>Mean: 0.00</td>
<td>-0.37</td>
<td>1.23</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Median: -0.03</td>
<td>-0.35</td>
<td>0.66</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>SD: 0.12</td>
<td>0.27</td>
<td>2.02</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td>Mean: 0.01</td>
<td>-0.10</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Median: 0.01</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>SD: 0.03</td>
<td>0.14</td>
<td>0.23</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Figure S16. Box and whisker plots of proportional change between substages for each metric value with paleo-latitude bin values corresponding to generic richness less than the required estimate from SQS removed (i.e., <16 unique genera for Campanian substages, and <18 unique genera for Maastrichtian substages). Triangles indicate mean values and horizontal lines indicate median values.
**Proportional Change in Metrics Across Latitude:**

Proportional change across paleo-latitudes within paleo-latitude bins indicates that, in general, substages experience decreasing generic and functional richness across paleo-latitude from south to north, except between the 40-45°N to 45-50°N bins (Figure 4). Both generic and functional richness decrease dramatically moving north across paleo-latitude above the 40-45°N bin. The Upper Maastrichtian is one exception to this pattern, showing an increase of close to ~88% in FR across the 45-50°N and 50-55°N bins. Between the 40-45°N to 45-50°N bins, FR and generic richness increase by more than 40% in all substages (Table S22). Proportional change in generic richness across these bins, while following similar patterns, is not identical. For example, the Middle Campanian experiences a proportional increase of 235% generic richness but only a 40% increase in FR, but the Upper Campanian experiences a proportional increase of generic richness of 16% while FR for this substage only increases by 61%. Once again, the two factors are decoupled though similar.

Across the 35-40°N to 40-45°N transition, both functional and generic richness decrease in all substages while between the 30-35°N and 35-40°N paleo-latitude bins, proportional change in richness is more variable. In this southernmost transition, the Upper and Lower Campanian experience very noticeable increases in both generic and functional richness, while the Lower and Upper Maastrichtian experience a fall of only 6% and 3%, respectively (Table S22). It should be pointed out, that these southern-most bins are dominated by Maastrichtian data, and therefore this relative stability may be influenced by better sampling for the Maastrichtian than for the Campanian substages. Indeed, the more dramatic decreases in proportional richness values in the most northern latitude transitions may also be a product of poor sampling. Removing values which have lower generic richness than the SQS estimated sample size for FR in each substage produces very similar patterns, except that the northernmost and southernmost values are removed for the Campanian substages (Table S23; Figure S17). This suggests that these extraneous values may be caused by poor sampling and strengthens the conclusion that both generic and functional richness decrease in similar proportions across latitude in all substages, except between the 40-45°N to 45-50°N bins where values increase notably.
Shifts in dominance across latitude are also highly variable, though, in general, SME increases in substages across latitudes between 45°N and 60°N. This indicates that evenness increases between latitudes moving north. In the 40-45°N to 45-50°N transition, SEM values decrease in all substages, indicating that evenness drops across latitudes in this range. South of 40°N, SEM show both increase and decrease in different substages, with an extremely noticeable increase between 35-40°N and 40-45°N for the Upper Maastrichtian (Figure 4). When values with fewer than the SQS sample size estimate are removed, these patterns are maintained.
Table S22. Values of proportional change across paleo-latitude bins in each substages

<table>
<thead>
<tr>
<th></th>
<th>30-35°N to 35-40°N</th>
<th>35-40°N to 40-45°N</th>
<th>40-45°N to 45-50°N</th>
<th>45-50°N to 50-55°N</th>
<th>50-55°N to 55-60°N</th>
<th>55-60°N to 60-65°N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>2.73</td>
<td>-0.66</td>
<td>3.79</td>
<td>-0.12</td>
<td>-0.37</td>
<td>-0.86</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>-0.38</td>
<td>2.35</td>
<td>-0.48</td>
<td>-0.61</td>
<td>-0.96</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>9.14</td>
<td>-0.09</td>
<td><strong>0.16</strong></td>
<td>-0.20</td>
<td>-0.27</td>
<td>-0.92</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>-0.26</td>
<td>-0.88</td>
<td>4.74</td>
<td>-0.56</td>
<td>-0.63</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td><strong>0.14</strong></td>
<td>-0.99</td>
<td>4.00</td>
<td>0.80</td>
<td>-0.41</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>2.50</td>
<td>-0.50</td>
<td>2.14</td>
<td>0.00</td>
<td>-0.23</td>
<td>-0.71</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>-0.09</td>
<td>0.40</td>
<td>-0.18</td>
<td>-0.30</td>
<td>-0.94</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>4.60</td>
<td>-0.36</td>
<td>0.61</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.84</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>-0.06</td>
<td>-0.67</td>
<td>1.70</td>
<td>-0.41</td>
<td>-0.44</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.03</td>
<td>-0.94</td>
<td>3.00</td>
<td>0.88</td>
<td>-0.40</td>
<td>NA</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>-0.50</td>
<td>0.76</td>
<td>-0.44</td>
<td>0.25</td>
<td>0.23</td>
<td>-1.00</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>0.17</td>
<td>-0.28</td>
<td>0.18</td>
<td>0.18</td>
<td>-1.00</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>-0.49</td>
<td>-0.65</td>
<td>2.42</td>
<td>-0.15</td>
<td>0.40</td>
<td>0.45</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>0.46</td>
<td>1.26</td>
<td>-0.40</td>
<td>0.53</td>
<td>0.28</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.21</td>
<td>5.99</td>
<td>-0.49</td>
<td>0.38</td>
<td>0.14</td>
<td>NA</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>-0.07</td>
<td>0.05</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.48</td>
<td>-0.03</td>
<td>0.06</td>
<td>-0.06</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>0.06</td>
<td>0.08</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.01</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.05</td>
<td>0.21</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.00</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure S17. Beeswarm plots of proportional change between paleo-latitude bins for each metric value with paleo-latitude bin values corresponding to generic richness less than the required estimate from SQS removed (i.e., <16 unique genera for Campanian substages, and <18 unique genera for Maastrichtian substages).
Table S23. Values of proportional change across paleo-latitude bins in each substages with values below SQS sample size estimate removed.

<table>
<thead>
<tr>
<th></th>
<th>30-35°N to 35-40°N</th>
<th>35-40°N to 40-45°N</th>
<th>40-45°N to 45-50°N</th>
<th>45-50°N to 50-55°N</th>
<th>50-55°N to 55-60°N</th>
<th>55-60°N to 60-65°N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-0.12</td>
<td>-0.37</td>
<td>NA</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>-0.38</td>
<td>2.35</td>
<td>-0.48</td>
<td>-0.61</td>
<td>NA</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>NA</td>
<td>-0.09</td>
<td>0.16</td>
<td>-0.20</td>
<td>-0.27</td>
<td>NA</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>-0.26</td>
<td>-0.88</td>
<td>4.74</td>
<td>-0.56</td>
<td>-0.63</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>0.14</td>
<td>-0.99</td>
<td>4.00</td>
<td>0.80</td>
<td>-0.41</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.00</td>
<td>-0.23</td>
<td>NA</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>-0.09</td>
<td>0.40</td>
<td>-0.18</td>
<td>-0.30</td>
<td>NA</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>NA</td>
<td>-0.36</td>
<td>0.61</td>
<td>-0.07</td>
<td>-0.07</td>
<td>NA</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>-0.06</td>
<td>-0.67</td>
<td>1.70</td>
<td>-0.41</td>
<td>-0.44</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.03</td>
<td>-0.94</td>
<td>3.00</td>
<td>0.88</td>
<td>-0.40</td>
<td>NA</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.25</td>
<td>0.23</td>
<td>NA</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>0.17</td>
<td>-0.28</td>
<td>0.18</td>
<td>0.18</td>
<td>NA</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>NA</td>
<td>-0.65</td>
<td>2.42</td>
<td>-0.15</td>
<td>0.40</td>
<td>NA</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>0.46</td>
<td>1.26</td>
<td>-0.40</td>
<td>0.53</td>
<td>0.28</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.21</td>
<td>5.99</td>
<td>-0.49</td>
<td>0.38</td>
<td>0.14</td>
<td>NA</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Camp.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.04</td>
<td>0.03</td>
<td>NA</td>
</tr>
<tr>
<td>Mid Camp.</td>
<td>NA</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>NA</td>
</tr>
<tr>
<td>Up Camp.</td>
<td>NA</td>
<td>-0.32</td>
<td>0.48</td>
<td>-0.03</td>
<td>0.06</td>
<td>NA</td>
</tr>
<tr>
<td>Low Maastr.</td>
<td>0.06</td>
<td>0.08</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.01</td>
<td>NA</td>
</tr>
<tr>
<td>Up Maastr.</td>
<td>-0.05</td>
<td>0.21</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.00</td>
<td>NA</td>
</tr>
</tbody>
</table>
Nestedness verses Turnover in Functional Entities over Paleolatitude

Nestedness, or the subsetting of FE assemblages across latitude, was tested using the betapart package in R (Baselga and Orme 2012) following Schumm et al. (2019). This package allows users to calculate the pairwise Sørensen dissimilarity between FE presence-absence data, which can be interpreted as a metric of nestedness. Sørensen values of 0 indicate turnover fully accounts for the total dissimilarity and 1 indicates that nestedness accounts for the dissimilarity. Results of nestedness analysis for paleolatitude bins indicate that nestedness is low (Sørensen dissimilarity <0.5) between most paleolatitude bins, and rarely high (Sørensen dissimilarity >0.5) between spatially adjacent bins (Table. S24-S28). Thus, nestedness does not account for dissimilarity well across latitude.

Table S24. Pairwise Sørensen dissimilarity values of FEs between Lower Campanian paleolatitude bins. Values greater than 0.5 are in bold.

<table>
<thead>
<tr>
<th></th>
<th>30-35°N</th>
<th>35-40°N</th>
<th>40-45°N</th>
<th>45-50°N</th>
<th>50-55°N</th>
<th>55-60°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>35-40°N</td>
<td>0.556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-45°N</td>
<td>0.205</td>
<td>0.286</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-50°N</td>
<td>0.692</td>
<td>0.175</td>
<td>0.517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55°N</td>
<td>0.519</td>
<td>0.143</td>
<td>0.517</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-60°N</td>
<td>0.310</td>
<td>0.062</td>
<td>0.357</td>
<td>0.098</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td>60-65°N</td>
<td>0.028</td>
<td>0.284</td>
<td>0.033</td>
<td>0.630</td>
<td>0.630</td>
<td>0.436</td>
</tr>
</tbody>
</table>

Table S25. Pairwise Sørensen dissimilarity values of FEs between Middle Campanian paleolatitude bins. Values greater than 0.5 are in bold.

<table>
<thead>
<tr>
<th></th>
<th>35-40°N</th>
<th>40-45°N</th>
<th>45-50°N</th>
<th>50-55°N</th>
<th>55-60°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>40-45°N</td>
<td>0.038</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-50°N</td>
<td>0.115</td>
<td>0.158</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55°N</td>
<td>0.018</td>
<td>0.066</td>
<td>0.098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-60°N</td>
<td>0.109</td>
<td>0.097</td>
<td>0.256</td>
<td>0.168</td>
<td></td>
</tr>
<tr>
<td>60-65°N</td>
<td>0.913</td>
<td>0.000</td>
<td>0.931</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table S26. Pairwise Sørensen dissimilarity values of FEs between Upper Campanian paleolatitude bins. Values greater than 0.5 are in bold.

<table>
<thead>
<tr>
<th></th>
<th>30-35°N</th>
<th>35-40°N</th>
<th>40-45°N</th>
<th>45-50°N</th>
<th>50-55°N</th>
<th>55-60°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35°N</td>
<td>0.697</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-40°N</td>
<td>0.339</td>
<td>0.205</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-45°N</td>
<td>0.706</td>
<td>0.015</td>
<td>0.221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-50°N</td>
<td>0.550</td>
<td>0.015</td>
<td>0.189</td>
<td>0.034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55°N</td>
<td>0.533</td>
<td>0.050</td>
<td>0.154</td>
<td>0.071</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>55-60°N</td>
<td>0.028</td>
<td>0.563</td>
<td>0.477</td>
<td>0.758</td>
<td>0.742</td>
<td>0.724</td>
</tr>
<tr>
<td>60-65°N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S27. Pairwise Sørensen dissimilarity values of FEs between Lower Maastrichtian paleolatitude bins. Values greater than 0.5 are in bold.

<table>
<thead>
<tr>
<th></th>
<th>30-35°N</th>
<th>35-40°N</th>
<th>40-45°N</th>
<th>45-50°N</th>
<th>50-55°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35°N</td>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-40°N</td>
<td>0.471</td>
<td>0.450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-45°N</td>
<td>0.082</td>
<td>0.045</td>
<td>0.459</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-50°N</td>
<td>0.313</td>
<td>0.266</td>
<td>0.208</td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>50-55°N</td>
<td>0.561</td>
<td>0.479</td>
<td>0.029</td>
<td>0.500</td>
<td>0.218</td>
</tr>
<tr>
<td>55-60°N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S28. Pairwise Sørensen dissimilarity values of FEs between Upper Maastrichtian paleolatitude bins. Values greater than 0.5 are in bold.

<table>
<thead>
<tr>
<th></th>
<th>30-35°N</th>
<th>35-40°N</th>
<th>40-45°N</th>
<th>45-50°N</th>
<th>50-55°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35°N</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-40°N</td>
<td>0.897</td>
<td>0.895</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-45°N</td>
<td>0.644</td>
<td>0.636</td>
<td>0.600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-50°N</td>
<td>0.423</td>
<td>0.412</td>
<td>0.765</td>
<td>0.228</td>
<td></td>
</tr>
<tr>
<td>50-55°N</td>
<td>0.609</td>
<td>0.600</td>
<td>0.318</td>
<td>0.022</td>
<td>0.167</td>
</tr>
<tr>
<td>55-60°N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

365
Summary of Functional vs Taxonomic Diversity change in 360km2 Nodes across Substages:

At the 360km aggregation scale, generic and functional richness distributions are fairly similar in all substages (Figure S18, Table S29), though the Lower Campanian has a noticeably smaller spread of FR values in the first and third quartiles than later substages and the Upper Maastrichtian has a greater spread of values in this range than other substages. Generic richness outliers are present in most substages, but only the Lower Campanian and Lower Maastrichtian have FR outliers. SME values has small distributions, in all substages with some outliers in the Mid Campanian and Lower Maastrichtian. SEM values have outliers at zero in all cases, except the Middle Campanian, where the first interquartile hinge is at zero. Again, zeros in this metric indicate “perfect evenness” which most commonly occurs when only a single FE exists in that node. Therefore, these values are not considered important in general.

Mean and median FR values for each substage at the 360km node scale are fluctuate across the substages, ranging from 6.21-14.5 and 4-11, respectively. Generic richness mean and median values follow a similar pattern, but means are consistently much larger than median values, ranging from 12.36-56.85 while median values range from only 6-19. Mean and median values for SME are fairly consistent, ranging from 0.59-0.97 and 0.43-0.85, respectively. However, the Lower, and to a lesser degree, Upper Maastrichtian distributions have a noticeable increase in all values (Figure S16), indicating lower overall evenness in these substages compared to others.

Proportional change in 360km node metrics indicate that FR increases across the Campanian, decreases between the Upper Campanian and Lower Maastrichtian, and is fairly stable overall between the Lower to Upper Maastrichtian (Figure 5, Table S30). Median FR proportional change values range from only 2% between the Lower and Upper Maastrichtian to -39% (decrease) between the Upper Campanian and Lower Maastrichtian. The distribution of values around these means are fairly consistent, except in the Upper Campanian to Lower Maastrichtian when hinge values are noticeably closer to the median though at least one outlier point is observed in all substages. The very small spread of first and third interquartile values between the Upper Campanian and Lower Maastrichtian suggests
that the median value is highly representative of the data for this comparison. Mean FR values of proportional change are higher than median values, ranging from 98% to 208% increase in FR values between nodes. These high mean values are most likely influenced by outliers, which are present in all substages (Figure 5). Proportional change in generic richness follows a very similar pattern to FR, but mean and median values both indicate a decrease in generic richness between the Lower and Upper Maastrichtian rather than stability in the metric. Again, the two are similar but decoupled. In all cases richness distributions cross the zero line either along the interquartile range or along the whisker (1.5*IQR). This suggests that not all generic or functional richness values within 360km nodes experience only increase or decrease across substages.

SME values indicate that evenness declined through the Campanian (median: -24% and -27%) and rose between the Upper Campanian and Lower Maastrichtian (median: 15%), but was fairly stable between the Lower and Upper Maastrichtian (median: -1%) (Figure 5, Table S30). Again, mean values are frequently higher than median values for this metric, most likely influenced by outliers. All distributions for SME cross the zero line, indicating that evenness change is not unanimously increasing or decreasing in 360km nodes between substages. Overall, SME values at this scale of analysis suggest declining evenness in the Campanian, then rising evenness in the Upper Campanian to Lower Maastrichtian, followed by relative stability in the Maastrichtian.

Summary of Functional vs Taxonomic Diversity change in 60km2 Nodes across Substages:
At the 60km aggregation scale, FR distributions are fairly similar in all substages (Figure S19, Table S32), though Upper Maastrichtian has a noticeably larger interquartile range of FR values than earlier substages. Th Upper Maastrichtian generic richness distribution, on the other hand, is more similar in its interquartile range to other substages, except the Middle Campanian where the interquartile range is noticeably larger. Generic richness outliers are present in all substages, and only the Middle Campanian and Upper Maastrichtian lack FR outliers. SME values present interquartile ranges with similar distances from the median value in each substage and outliers in all substages. SEM values either have outliers at
zero in all cases or their first interquartile hinge is positioned at zero. Again, zeros in this metric indicate “perfect evenness” which most commonly occurs when only a single FE exists in that node. Therefore, these values are not considered important and may be misleading.

Median FR values for each substage at the 60km node scale are fairly similar across all the substages, ranging from 3-4, but mean values are more variable, ranging from 3.79-10.11. This is most likely a result of outliers in the distribution. Generic richness mean and median values follow a similar pattern, where medians are fairly similar (3-6) and mean values are much more diverse (5.7-33.5). Again, mean values for generic richness are probably influenced by outliers in the distributions. Median values for SME show a clear shift from higher values in the Campanian, ranging from 0.83-0.99, to lower values in the Maastrichtian (0.37 and 0.38). This suggests a decrease in evenness between the Campanian and Maastrichtian. Mean values of SME follow a similar trend but are more variable and assumed to be more influenced by outlier values.

Proportional change in 60km node metrics indicate that FR increases across the Lower to Middle Campanian, is stable across the Middle to Upper Campanian, decreases between the Upper Campanian and Lower Maastrichtian, and is fairly stable between the Lower to Upper Maastrichtian (Figure 5, Table S33). Median FR proportional change values range from 0% between the Middle and Upper Campanian to 90% (increase) between the Lower to Middle Campanian. The upper and lower interquartile range values are farther from the median in both the Lower to Middle Campanian and Lower to Upper Maastrichtian comparisons than they are in the other two comparisons, and all contain numerous outlier values. All distributions cross the zero line, suggesting that not all node values shift in the same direction (increase or decrease). Due to outliers with extremely high proportional increases, mean values are consistently higher than median values of proportional change in FR.

Proportional change in generic richness follows a very similar pattern to FR, though median values indicate an increase in generic richness between the Middle and Upper Campanian rather than stability. Again, the two are similar but decoupled. In all cases richness distributions cross the zero line.
This suggests that not all generic or functional richness values within 60km nodes experience only increase or decrease across substages.

SME values indicate that evenness was stable from the Lower to Middle Campanian (median: -0%), decreased between the Middle and Upper Campanian (median: -21%), increased between the Upper Campanian and Lower Maastrichtian (median: 13%), and decreased between the Lower and Upper Maastrichtian (median: -27%) (Figure 5, Table S33). Mean values are frequently higher (more positive) than median values for this metric, influenced by outliers which have significant increases in evenness through time. All distributions for SME cross the zero line, indicating that evenness change is not unanimously increasing or decreasing in 60km nodes between substages. **Overall, SME values at this scale of analysis suggest first stable then decreasing evenness in the Campanian, then rising evenness in the Upper Campanian to Lower Maastrichtian, followed by evenness decrease in the Maastrichtian.**
Figure S18. Box and whisker plots of generic richness, FR, SME, and SEI for 360km nodes in each substage. Median values are denoted by a horizontal line and means by triangles.
Table 29. Summary statistics of generic and functional richness, SME and SEI values from 360km nodes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>12.39</td>
<td>19.03</td>
<td>31.39</td>
<td>31.08</td>
<td>56.85</td>
</tr>
<tr>
<td>Median</td>
<td>6.00</td>
<td>6.00</td>
<td>19.00</td>
<td>12.50</td>
<td>18.50</td>
</tr>
<tr>
<td>SD</td>
<td>14.10</td>
<td>26.38</td>
<td>36.85</td>
<td>49.58</td>
<td>73.74</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.21</td>
<td>8.10</td>
<td>10.35</td>
<td>9.71</td>
<td>14.05</td>
</tr>
<tr>
<td>Median</td>
<td>4.50</td>
<td>4.00</td>
<td>9.00</td>
<td>7.50</td>
<td>11.00</td>
</tr>
<tr>
<td>SD</td>
<td>5.55</td>
<td>8.08</td>
<td>8.19</td>
<td>9.12</td>
<td>12.71</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.64</td>
<td>0.84</td>
<td>0.59</td>
<td>0.97</td>
<td>0.60</td>
</tr>
<tr>
<td>Median</td>
<td>0.61</td>
<td>0.65</td>
<td>0.65</td>
<td>0.85</td>
<td>0.43</td>
</tr>
<tr>
<td>SD</td>
<td>0.70</td>
<td>0.94</td>
<td>0.46</td>
<td>0.80</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.80</td>
<td>0.68</td>
<td>0.69</td>
<td>0.83</td>
<td>0.74</td>
</tr>
<tr>
<td>Median</td>
<td>0.92</td>
<td>0.91</td>
<td>0.88</td>
<td>0.92</td>
<td>0.82</td>
</tr>
<tr>
<td>SD</td>
<td>0.34</td>
<td>0.43</td>
<td>0.39</td>
<td>0.26</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table S30. Summary statistics of proportional change in generic and functional richness, SME and SEI values from 360km nodes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td>Mean</td>
<td>1.56</td>
<td>6.25</td>
<td>9.91</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.69</td>
<td>0.76</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.93</td>
<td>13.43</td>
<td>46.16</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td>Mean</td>
<td>1.14</td>
<td>2.08</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.32</td>
<td>0.28</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.15</td>
<td>3.90</td>
<td>7.01</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td>Mean</td>
<td>-0.05</td>
<td>-0.21</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>-0.24</td>
<td>-0.27</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.36</td>
<td>0.34</td>
<td>1.08</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td>Mean</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.05</td>
<td>0.09</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Figure S19. Box and whisker plots of generic richness, FR, SME, and SEI for 60km nodes in each substage. Median values are denoted by a horizontal line and means by triangles.
Table S31. Summary statistics of generic and functional richness, SME and SEI values from 60km nodes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>5.70</td>
<td>9.86</td>
<td>10.85</td>
<td>12.39</td>
<td>33.05</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>4.00</td>
<td>6.00</td>
<td>5.00</td>
<td>3.00</td>
<td>6.00</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>5.88</td>
<td>10.11</td>
<td>15.68</td>
<td>26.56</td>
<td>44.50</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>3.79</td>
<td>6.04</td>
<td>5.65</td>
<td>5.26</td>
<td>10.11</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>3.00</td>
<td>3.50</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>3.28</td>
<td>4.94</td>
<td>5.20</td>
<td>6.10</td>
<td>10.35</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.85</td>
<td>1.09</td>
<td>0.88</td>
<td>0.79</td>
<td>0.54</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>0.83</td>
<td>0.99</td>
<td>0.86</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.93</td>
<td>1.14</td>
<td>0.94</td>
<td>1.05</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.67</td>
<td>0.75</td>
<td>0.72</td>
<td>0.67</td>
<td>0.57</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>0.92</td>
<td>0.94</td>
<td>0.92</td>
<td>0.92</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.43</td>
<td>0.39</td>
<td>0.40</td>
<td>0.43</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Table S32. Summary statistics of proportional change in generic and functional richness, SME and SEI values from 60km nodes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>3.79</td>
<td>1.18</td>
<td>1.39</td>
<td>8.54</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>1.25</td>
<td>0.12</td>
<td>-0.38</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>6.58</td>
<td>2.49</td>
<td>10.92</td>
<td>28.09</td>
</tr>
<tr>
<td><strong>Functional Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>2.56</td>
<td>0.68</td>
<td>0.42</td>
<td>2.67</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>0.93</td>
<td>0.00</td>
<td>-0.40</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>4.10</td>
<td>1.83</td>
<td>2.87</td>
<td>7.14</td>
</tr>
<tr>
<td><strong>SME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.16</td>
<td>-0.11</td>
<td>0.37</td>
<td>-0.32</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>0.00</td>
<td>-0.21</td>
<td>0.13</td>
<td>-0.27</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.74</td>
<td>0.50</td>
<td>0.95</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>SEI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.08</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.07</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Summary of Functional vs Taxonomic Diversity change in 360km² Nodes across Distance:

Proportional changes in metrics between nodes within a single substage, plotted by the geographic distance between the nodes indicate that change in richness values are not strongly correlated with distance for the Campanian substages (Figure S20-S23). In the Maastrichtian substages, increasing geographic distance is weakly correlated with increasing proportional change in both generic and functional richness (Figure S20-S21).

SME values similarly do not show strong correlations between proportional changes to evenness and geographic distance, though the in all substages except the Lower Campanian, a very weak negative trend may exist, indicating that evenness values change less as distances increase (Figure S22). This negative trend is most likely a product of heteroscedasticity in the data, given that low values of proportional change are present across all distances but high values become increasing less common as distances increase. Given that most nodes are fairly closely spaced there will be necessity be fewer data points at very great distances to compare.

SEI values show even less of a correlation between proportional change between nodes and geographic distance (Figure S23). Negative trend lines observed in the figures are most likely a product of the highly heteroscedastic nature of the data and due to the decreasing number of data points at greater distances. The overall distribution of data points forms a fairly uniform cloud which has fewer values at greater distances, suggesting that, when baseline values of -1 are not taken into account, there is no strong correlation between proportional change in SEI and distance.
Figure S20. Proportional change in generic richness between all occupied 360km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S21. Proportional change in functional richness between all occupied 360km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S22. Proportional change in SME between all occupied 360km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S23. Proportional change in SEI between all occupied 360km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Summary of Functional vs Taxonomic Diversity change in 60km2 Nodes across Distance:

Proportional changes in metrics between 60km nodes within a single substage, plotted by the geographic distance between the nodes indicate that change in richness values are not strongly correlated with distance for the Campanian substages (Figure S24-27). In the Maastrichtian substages, increasing geographic distance is weakly correlated with increasing proportional change in both generic and functional richness (Figure S24-S25).

Proportional changes in SME values for 60km nodes are very slightly negatively correlated with the geographic between nodes at this scale of analysis (Figure S26). SEI values show no correlation between proportional change across nodes and geographic distance between those nodes when values of -1 are not considered (Figure S27). Negative trend lines observed in the figures are most likely a product of the highly heteroscedastic nature of the data and due to the decreasing number of data points at greater distances. The overall distribution of data points forms a fairly uniform cloud which has fewer values at greater distances, suggesting that, when baseline values of -1 are not taken into account, there is no strong correlation between proportional change in SEI and distance. These patterns are extremely similar to those observed at the 360km node scale, excepting that a negative trend in the proportional change in SME with distance is dampened.
Figure S24. Proportional change in generic richness between all occupied 60km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S25. Proportional change in functional richness between all occupied 60km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S26. Proportional change in SME between all occupied 60km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Figure S27. Proportional change in SEI between all occupied 60km nodes in each substage plotted against distance (m). Blue lines indicate linear regression lines calculated for each.
Summary of 360km2 Functional Entity Networks:

When all substages are modeled in a single network, strong network connections exist across the entire study interval and no clear pattern of provinciality is observed (Figure S30). Two major components are observed in the network at percolation (TH=0.36), but they do not appear to represent any consistent geographic region, instead overlapping similarly to the components observed in the Middle Campanian. Furthermore, the geographically distinct secondary component in the Upper Maastrichtian GCP region shares strong node similarities with both WIS and GCP regions in previous substages, indicating that this is not a geographically consistent set of FE assemblages.
Figure S28. Network models of Campanian substages based on 360km nodes. All networks show connections present at percolation with the threshold level denoted below the substage name. Models on the left are not plotted based on any geographic position, models on the right are plotted based on paleo-latitude coordinates. Node colors indicate paleo-latitudes (warmer = southern paleo-latitudes and cooler = northern paleo-latitudes). Color and size of links between nodes indicates the level of similarity (thicker and lighter = more similar). Middle Campanian components are denoted with arrows.
Figure S29. Network models of Maastrichtian substages based on 360km nodes. All networks show connections present at percolation with the threshold level denoted below the substage name. Models on the left are not plotted to show components and models on the right are plotted based on paleo-latitude coordinates. Node colors indicate paleo-latitudes (warmer = southern paleo-latitudes and cooler = northern paleo-latitudes). Color and size of links between nodes indicates the level of similarity (thicker and lighter = more similar). The two components present in the Upper Maastrichtian network are indicated with arrows.
Table S33. List of FE in each major network component for the Middle Campanian and Upper Maastrichtian based on 360km nodes. FEs which are shared by all nodes in the component are denoted with asterisks. FEs which are shared by all nodes in the component are denoted with asterisks. These components were separated based on manually thresholded networks created below percolation: TH=0.52 for the Middle Camp. and TH=0.75 for the Upper Maastr. FEs representative of gastropods are indicated with a gastropod symbol (•).

<table>
<thead>
<tr>
<th>Middle Campanian</th>
<th>Upper Maastrichtian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary component (WIS region)</strong></td>
<td><strong>Secondary Component (WIS region)</strong></td>
</tr>
<tr>
<td>MU-E-C</td>
<td>MU-NB-SF*</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>MU-E-C •</td>
</tr>
<tr>
<td>MU-PI-C</td>
<td>MU-NB-SF</td>
</tr>
<tr>
<td>FU-I-C*</td>
<td>MU-E-C •</td>
</tr>
<tr>
<td>MU-I-D*</td>
<td>FU-E-SF •</td>
</tr>
<tr>
<td>MU-PI-H</td>
<td>IU-B-SF</td>
</tr>
<tr>
<td>FU-I-D</td>
<td>FU-I-SF</td>
</tr>
<tr>
<td>FU-I-SF*</td>
<td>IU-E-SF</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>IA-E-SF*</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>MU-N-SF</td>
</tr>
<tr>
<td>IA-E-SF*</td>
<td>MU-NB-SF*</td>
</tr>
<tr>
<td>MU-N-SF</td>
<td>MU-I-SF</td>
</tr>
<tr>
<td>MU-NB-SF*</td>
<td>IA-I-SF</td>
</tr>
<tr>
<td>MU-I-CH</td>
<td>MU-N-C</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>MU-NB-C</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>MU-P1-C •</td>
</tr>
<tr>
<td>IU-E-SF</td>
<td>IU-I-C</td>
</tr>
<tr>
<td>MU-E-SF</td>
<td>FU-I-CH</td>
</tr>
<tr>
<td>FA-P1-SF</td>
<td>MU-I-CH</td>
</tr>
<tr>
<td>IA-P1-SF</td>
<td>FA-I-D</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>FU-E-H •</td>
</tr>
<tr>
<td>FU-P1-H</td>
<td>MU-E-H •</td>
</tr>
<tr>
<td>MU-P1-D</td>
<td>MU-P1-H •</td>
</tr>
<tr>
<td>MU-I-C</td>
<td>MU-E-O •</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>IU-P1-P</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>MU-P1-D</td>
</tr>
<tr>
<td>FU-B-SF</td>
<td>IU-I-SF</td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>IA-P1-SF</td>
</tr>
<tr>
<td>FA-P1-SF</td>
<td>IA-E-P</td>
</tr>
<tr>
<td>IA-P1-P</td>
<td>IU-I-SF</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>IA-P1-SF</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>IU-I-SF</td>
</tr>
<tr>
<td>MU-E-D</td>
<td>IU-I-SF</td>
</tr>
</tbody>
</table>
Figure S30. Network model of the entire database using 360km² nodes. Nodes are colored by substage and plotted on paleo-map reconstructions (Kocsis and Raja 2020) based on relative paleo-location (nodes may be offset to make visualization easier) Node size indicates betweenness centrality and link color and size indicates the degree of functional similarity (lighter and thicker = more similar).
Summary of 60km² Functional Entity Networks:

In the Lower Maastrichtian a small highly similar group of nodes (which only disconnect from other nodes at thresholds below percolation) geographically overlies the more extensive but less similar primary component, mostly in the WIS region (Figure 8). The larger, less similar component is dominated by functionally diverse nodes with highly variable FEs, none of which are shared by all nodes in the component, while the highly similar component is dominated by functionally depauperate nodes with a single shared FE of mobile, unattached, nekto-benthic suspension feeders (cephalopods; Table S34). The third component in the Lower Maastrichtian that is geographically located in the eastern GCP region is dominated by only mobile, unattached, nektonic, suspension feeders (cephalopods). Therefore, in this Lower Maastrichtian network, high FE assemblage similarity is predominately controlled by a disparity between FE diversity in different nodes, which can most likely be attributed to sampling bias. Since generic richness strongly correlates with both sampling and FR values, it is likely that the separation of components in this substage is a product of sampling disparity between nodes.

In the Upper Maastrichtian, where functional similarity has a clear geographical signal, sampling bias may be similarly influencing network topology. In this substage, WIS region sampling is not as strong as GCP region sampling, which may explain the functionally less diverse WIS component that does not share any FEs between all nodes and the highly diverse GCP component whose nodes all share five total FEs. The Upper Maastrichtian components are characterized by a functional depauperate WIS assemblage dominated by mobile, unattached, nekto-benthic and nektonic suspension feeders (cephalopods) and facultatively mobile, epifaunal suspension feeders (gastropods/bivalves) and a highly functionally diverse GCP assemblage.
Table S34. List of FE in each major network component for the Lower and Upper Maastrichtian based on 60km nodes. FEs which are shared by all nodes in the component are denoted with asterisks. These components were separated based on manually thresholded networks created below percolation: TH=0.45 for the Lower Maastr and TH=0.50 for the Upper Maastr. In the Upper Maastrichtian, FEs representative of gastropods are indicated with a gastropod symbol (●).

<table>
<thead>
<tr>
<th>Lower Maastrichtian</th>
<th>Upper Maastrichtian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1 (WIS region)</td>
<td>Component 2 (WIS region)</td>
</tr>
<tr>
<td>MU-I-D</td>
<td>MU-NB-SF*</td>
</tr>
<tr>
<td>FU-I-D</td>
<td></td>
</tr>
<tr>
<td>FA-I-SF</td>
<td>FU-E-SF</td>
</tr>
<tr>
<td>FU-I-SF</td>
<td>IU-I-SF</td>
</tr>
<tr>
<td>FU-E-SF</td>
<td>FU-E-SF</td>
</tr>
<tr>
<td>IU-I-SF</td>
<td>IA-E-SF</td>
</tr>
<tr>
<td>MU-NB-SF</td>
<td>MU-N-SF</td>
</tr>
<tr>
<td>MU-E-C</td>
<td>MU-NB-SF</td>
</tr>
<tr>
<td>MU-Pi-C</td>
<td>MU-I-Ch</td>
</tr>
<tr>
<td>FU-I-CH</td>
<td>MU-E-H</td>
</tr>
<tr>
<td>MU-Pi-H</td>
<td>FU-Pi-H</td>
</tr>
<tr>
<td>IU-B-SF</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>FA-E-SF</td>
<td>IA-B-SF</td>
</tr>
<tr>
<td>FA-Pi-SF</td>
<td>FA-E-SF</td>
</tr>
<tr>
<td>FU-I-C</td>
<td>FU-Pi-H</td>
</tr>
<tr>
<td>FA-I-D</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-E-H</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>MU-N-C</td>
</tr>
<tr>
<td>MU-N-C</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-NB-C</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>IA-E-SF</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-E-O</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-Pi-P</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-Pi-C</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-Pi-D</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-Pi-H</td>
<td>MU-Pi-D</td>
</tr>
<tr>
<td>MU-Pi-H</td>
<td>MU-Pi-D</td>
</tr>
</tbody>
</table>
Table S35. List of paleo-environments provided by the PBDB for select occurrences found in network components for the Lower and Upper Maastrichtian (see Table S34). Environmental descriptions which include carbonaceous attributes are in bold. In the Upper Maastrichtian components, a potential difference between siliciclastic dominated and carbonaceous-dominated environments may be present, but more data is necessary to determine this pattern than is possible given the scope of this project.

<table>
<thead>
<tr>
<th>Lower Maastrichtian</th>
<th>Upper Maastrichtian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1 (WIS region)</td>
<td>Component 2 (WIS region)</td>
</tr>
<tr>
<td>Offshore, siliciclastic sandstone</td>
<td>Siliciclastic shale</td>
</tr>
<tr>
<td>Calcareous carbonaceous SS</td>
<td>Siliciclastic sediments</td>
</tr>
<tr>
<td>Gray Sandstone</td>
<td>Offshore limestone</td>
</tr>
<tr>
<td>Grey Sandstone</td>
<td>Siltstone</td>
</tr>
<tr>
<td>Grey Sandstone</td>
<td>Chalk</td>
</tr>
<tr>
<td>Sandstone</td>
<td>Chalk</td>
</tr>
<tr>
<td>Concretionary sideritic calcareous shale</td>
<td>Chalk</td>
</tr>
<tr>
<td>Concretionary shale</td>
<td></td>
</tr>
<tr>
<td>Calcareous shale and grey limestone (offshore)</td>
<td></td>
</tr>
</tbody>
</table>


Figure S31. Bray-Curtis dissimilarity values between 360km nodes plotted by distance.
Figure S32. Bray-Curtis dissimilarity values between 60km nodes plotted by distance.
Summary of 60km² Class-level Functional Entity Networks:

Class-level analysis of 60km² networks indicates that bivalves and cephalopods, which make up the majority of FEs in the database, form strong network connections across all regions and lack a geographic or provincial signal (Figure S33-S36). However, it should be noted that Late Maastrichtian bivalves are not common and form poor network connections in general, though not necessarily based on geography (Figure S36). The strongly connected, functionally depauperate components in the Lower Maastrichtian are represented by specific, shared cephalopod FEs (Table S4). Their sampling is not strongly provincial, and therefore most likely produces the lack of provinciality in most substage networks where they form the dominate taxa. Gastropods, on the other hand, are strongly provincial (Figure S37-S38). Indeed, the three of the five FEs which are shared by all nodes in the GCP component of the Upper Maastrichtian are represented by gastropods and eleven of the 38 FEs possible in this region are represented by gastropods. Gastropods, therefore, are considered to have the prevailing influence on provinciality in FE assemblages in the Upper Maastrichtian.
Figure S33. Campanian cephalopod substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
Figure S34. Maastrichtian cephalopods substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
Figure S35. Campanian bivalve substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
Figure S36. Maastrichtian bivalve substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
Figure S37. Campanian gastropod substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
Figure S38. Maastrichtian gastropod substage networks plotted without any geographic coordinates (left) and based on paleo-coordinates (right). Node colors indicate paleo-latitude where warmer colors indicate southern paleo-latitudes and cooler colors indicate northern paleo-latitude.
References


Appendix B-2. Functional Ecology References for Table S36


Avelar, W., and Cunha, A.D., 2009, The anatomy and functional morphology of Diplodon rhombeus fontainianus (Orbigny, 1835) (Mollusca Bivalvia, Hyriidae):.


Beu, A.G., and Maxwell, P.A., 1987a, A revision of the fossil and living gastropods related to Plesiotriton Fischer, 1884 (Family Cancellariidae, Subfamily Plesiotritoninae n. subfam.). With an appendix:


Griffiths, R.J., 1981, Predation on the bivalves Choromytilus meridionalis (Kr.) by the gastropod Natica (Tectonatica) tecta Anton: Journal of Molluscan Studies, v. 47, p. 112–120.


Hayami, I., 1984, Natural history and evolution of Cryptopecten (a Cenozoic- Recent pectinid genus): The University Museum, The University of Tokyo, Bulletin, 24, p. 1–149.


Kiessling, W., 2004, Ecology opinions:


Levine, T.D., Hansen, H.B., and Gerald, G.W., 2013, Effects of shell shape, size, and sculpture in burrowing and anchoring abilities in the freshwater mussel Potamilus alatus (Unionidae):.


McLay, C.L., 2006, Retroplumidae (Crustacea, Decapoda) from the Indo-Malayan archipelago (Indonesia, Philippine) and the Melanesian arc islands (Solomon Islands, Fiji and New Caledonia), and paleogeographical comments: Tropical Deep-Sea Benthos, p. 375–391.


Perron, F.E., 1978, Seasonal Burrowing Behavior and Ecology of Aporrhais occidentalis (Gastropoda: Strombacea); https://about.jstor.org/terms.


Smith, A.B., 2009, Classification of the Echinoidea: Online database.


Stephenson, L.W., 1941, The larger invertebrate fossils of the Navarro Group of Texas: University of Texas Publication number 4101, p. 1–641.


Appendix B-3. R Code

######## Code for mapping FD data and analyzing FD based on mFD package (Magneville et al. 2021) ########

##### Load necessary packages ######
library(ggplot2) # for making plots
library(ggmap) # for plotting on maps
library(tidyverse)
library(sf) # shapefiles
library(mapview)
library(raster)
library(rgdal)
library(dismo)
library(XML)
library(maps)
library(mFD) # primary package for running FE analysis
library(reshape2) # manipulate dfs
library(vegan)
library(dplyr) # df stuff
library(tidyr)
library(gridExtra) # for plotting multiple plots on one image
library(goodpractice)
library(tidyr)

##### Download data and set directory #####
# Set working directory to get the file
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files")
# Read in database file

data <- read.csv("FE_data_compress_fe_3-7-23_basins.csv") # DATABASE HAS BASIN ID'S

# Option to subset out different basin's data

# data <- subset(data, Basin == "WIS") # SUBSET OUT WIS BASIN OCC
# data <- subset(data, Basin == "GCP") # SUBSET OUT GCP BASIN OCC

nrow(data)
colnames(data) # check file

# The data here has all columns from the "raw" file, including:
# "State", "County", "Town", "Phylum", "Class", "Order", "Family",
# "Updated_Genus", "Updated_Sp", "Updated_subtax", "Early_Int.Substag",
# "Late_Int.Substag", "Corr_Stage", "HigherTax_Class", "Formation", "Group",
# "Member", "Zone_Comments", "Substage_from_Zone.Mbr",
# "Inform_Zone...Merewethere.et.al.2015.", "X", "PaperCited", "collection_no",
# "Museum", "EcolRef", "motility", "life_habitat", "feeding"

#### Function for SUBSAMPLING WITH SQS ####

# function to calc sqs value, based on
https://strata.uga.edu/8370/rtips/shareholderQuorumSubsampling.html

# this looks like it was put together by Steve Holland.... not sure how to credit

run_sqs <- function(abundance, quota=0.8, trials=100, ignore.singletons=FALSE, exclude.dominant=FALSE) {

  # abundance is a vector of integers representing the abundance of every species

  if ((quota <= 0 || quota >= 1)) {
    stop("The SQS quota must be greater than 0.0 and less than 1.0")
  }
# compute basic statistics
specimens <- sum(abundance)
numTaxa <- length(abundance)
singletons <- sum(abundance==1)
doubletons <- sum(abundance==2)
highest <- max(abundance)
mostFrequent <- which(abundance==highest)[1]

if (exclude.dominant == FALSE) {
    highest <- 0
    mostFrequent <- 0
}

# compute Good's u
u <- 0
if (exclude.dominant == TRUE) {
    u <- 1 - singletons / (specimens - highest)
} else {
    u <- 1 - singletons / specimens
}

if (u == 0) {
    stop('Coverage is zero because all taxa are singletons')
}

# re-compute taxon frequencies for SQS
frequencyInitial <- abundance - (singletons + doubletons / 2) / numTaxa
frequency <- frequencyInitial / (specimens - highest)

# return if the quorum target is higher than estimated coverage
if ((quota > sum(frequency)) || (quota >= sum(abundance))) {
  stop('SQS quota is too large, relative to the estimated coverage')
}

# create a vector, length equal to total number of specimens,
# each value is the index of that species in the abundance array
ids <- unlist(mapply(rep, 1:numTaxa, abundance))

# subsampling trial loop
richness <- rep(0, trials) # subsampled taxon richness
for (trial in 1:trials) {
  pool <- ids # pool from which specimens will be sampled
  specimensRemaining <- length(pool) # number of specimens remaining to be sampled
  seen <- rep(0, numTaxa) # keeps track of whether taxa have been sampled
  subsampledFrequency <- rep(0, numTaxa) # subsampled frequencies of the taxa
  coverage <- 0

  while (coverage < quota) {
    # draw a specimen
    drawnSpecimen <- sample(1:specimensRemaining, size=1)
    drawnTaxon <- pool[drawnSpecimen]

    # increment frequency for this taxon
    subsampledFrequency[drawnTaxon] <- subsampledFrequency[drawnTaxon] + 1

    # if taxon has not yet been found, increment the coverage
if (seen[drawnTaxon] == 0) {
    if (drawnTaxon != mostFrequent && (ignore.singletons == 0 || abundance[drawnTaxon] > 1)) {
        coverage <- coverage + frequency[drawnTaxon]
    }
    seen[drawnTaxon] <- 1
    # increment the richness if the quota hasn't been exceeded,
    # and randomly throw back some draws that put the coverage over quota
    if (coverage < quota || runif(1) <= frequency[drawnTaxon]) {
        richness[trial] <- richness[trial] + 1
    } else {
        subsampledFrequency[drawnTaxon] <- subsampledFrequency[drawnTaxon] - 1
    }
}

# decrease pool of specimens not yet drawn
pool[drawnSpecimen] <- pool[specimensRemaining]
specimensRemaining <- specimensRemaining - 1

# compute subsampled richness
s2 <- richness[richness>0]
subsampledRichness <- exp(mean(log(s2))) * length(s2)/length(richness)
return(round(subsampledRichness, 1))

# get the full db list of taxa and then get their frequencies
genus_vector <- data[,c("Updated_Genus")]

genus_freq <- table(genus_vector) # all data frequencies of genera

# run sqs on entire db abundance vector
data_sqs_val <- run_sqs(genus_freq)
data_sqs_val

# function to get abundance list for each substage
sub_func_for_sqs <- function(data, substage) {
  dat <- as.data.frame(subset(data, Substage_from_Zone.Mbr == substage))
  dat <- dat[,c("Updated_Genus")]
  dat2 <- table(dat)
  return(dat2)
}

maa_up_for_sqs <- sub_func_for_sqs(data, "MAA (up)")
maa_low_for_sqs <- sub_func_for_sqs(data, "MAA (low)")
cam_up_for_sqs <- sub_func_for_sqs(data, "CAM (up)")
cam_mid_for_sqs <- sub_func_for_sqs(data, "CAM (mid)")
cam_low_for_sqs <- sub_func_for_sqs(data, "CAM (low)")

# run sqs on the substages
maa_up_sqs_val <- run_sqs(maa_up_for_sqs, quota=0.9)
maa_low_sqs_val <- run_sqs(maa_low_for_sqs, quota=0.9)
cam_up_sqs_val <- run_sqs(cam_up_for_sqs, quota=0.9)
cam_mid_sqs_val <- run_sqs(cam_mid_for_sqs, quota=0.9)
cam_low_sqs_val <- run_sqs(cam_low_for_sqs, quota=0.9)

maa_up_sqs_val
maa_low_sqs_val
cam_up_sqs_val
cam_mid_sqs_val
cam_low_sqs_val

#### CONFIGURE DATA ####

#### Overlay Grid and Create Smaller Data Frame (remove unnecessary columns) ####

# Set new WD to read in grid made in ArcGIS pro
setwd("C:/Users/ceara/Documents/Province Project/FE_Analysis_Code-and-Files/OccurrenceData_copiedfromProvinceProject/VettedCombinedDB")

sixtykmgrid <- st_read("SpatialGrids/grid_60km.shp")
# sixtykmgrid <- st_read("SpatialGrids/grid_360km.shp") # option to use 360km grid instead

# Transform database into shapefile
sf_data <- st_as_sf(data, coords = c("Long", "Lat"), crs = 4326)

# Function to spatially join DB with grids
spatial_join <- function(shapefile_data,grid){
  data1 <- st_join(shapefile_data,left=TRUE,grid["PageName"])
  return(data1)
}

# Spatially join Grid cell Page Names to data table
join60_data <- spatial_join(sf_data,sixtykmgrid)
join60_data <- as.data.frame(join60_data)
nrow(join60_data) # number of occurrences
# number of unique genera in the database
length(unique(join60_data$Updated_Genus))

#### Option to SUBSET OUT JUST SPECIFIC CLASS OF ORGANISMS FOR COMPARISON ####

## Subsetting out just specific Class types, using only the most abundant  
# classes of Bivalvia, Gastropoda, and Cephalopoda

# Look at the frequencies of different taxonomic classes
table(join60_data$Class)

# Make subsetted df for each of the three most common
bivalve_db <- subset(join60_data, Class == "Bivalvia")
gastropod_db <- subset(join60_data, Class == "Gastropoda")
cephalopod_db <- subset(join60_data, Class == "Cephalopoda")

# Apply the specific taxonomic classes as the base data to use in the rest of the code
# join60_data <- bivalve_db
# join60_data <- gastropod_db
# join60_data <- cephalopod_db

#### Option to SUBSET OUT GRID CELLS with greater than 3 unique genera from DB ####

# Function to remove all rows that represent grid cells with fewer than 3 unique genera
get_greater_3_database <- function(data){
    # Function to get vector of grid cell names with greater than 3 unique genera
    get_great_3_grids <- function(data,Age){
        # Get just the columns of Genus, Age, and Grid cell name
        gen_age_grid <- data[,c("Updated_Genus","Substage_from_Zone.Mbr","PageName")]
    }
}
# Remove duplicates from the list
unique_gen_age_grid <- unique(gen_age_grid)

# Get just the substage information in question
substage_unique_gen_age_grid <- unique_gen_age_grid %>% filter(Substage_from_Zone.Mbr == Age)

# Get the number of unique genera in each grid cell in the substage
substage_grid_freq <- data.frame(table(substage_unique_gen_age_grid$PageName))

# Extract just the Grid cell names that have 3 or more unique genera
substage_grid_greater_3_gen <- substage_grid_freq %>% filter(Freq > 2)
colnames(substage_grid_greater_3_gen) <- c("PageName", "freq_gen")

return(substage_grid_greater_3_gen)
}

# use above function to create list of Grid cells with >3 occ

maa_up_grid_greater_3_gen <- get_great_3_grids(data, "MAA (up)")
maa_low_grid_greater_3_gen <- get_great_3_grids(data, "MAA (low)")
cam_up_grid_greater_3_gen <- get_great_3_grids(data, "CAM (up)")
cam_mid_grid_greater_3_gen <- get_great_3_grids(data, "CAM (mid)")
cam_low_grid_greater_3_gen <- get_great_3_grids(data, "CAM (low)")

# get subset of the original db for each substage

maa_up_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr == "MAA (up)"))
maa_low_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr == "MAA (low)"))
cam_up_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr == "CAM (up)"))
cam_mid_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr == "CAM (mid)")))
cam_low_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr == "CAM (low)")

# merge the two dfs together to get extra column that has NA if less than 3 genera in the grid cell in that SS
maa_up_data1_merge <- merge(maa_up_data1, maa_up_grid_greater_3_gen, by = "PageName", all.x = TRUE)
maa_low_data1_merge <- merge(maa_low_data1, maa_low_grid_greater_3_gen, by = "PageName", all.x = TRUE)
cam_up_data1_merge <- merge(cam_up_data1, cam_up_grid_greater_3_gen, by = "PageName", all.x = TRUE)
cam_mid_data1_merge <- merge(cam_mid_data1, cam_mid_grid_greater_3_gen, by = "PageName", all.x = TRUE)
cam_low_data1_merge <- merge(cam_low_data1, cam_low_grid_greater_3_gen, by = "PageName", all.x = TRUE)

# remove rows with NA in the freq_gen column
maa_up_data1_data2 <- as.data.frame(subset(maa_up_data1_merge, !is.na(freq_gen)))
maa_low_data1_data2 <- as.data.frame(subset(maa_low_data1_merge, !is.na(freq_gen)))
cam_up_data1_data2 <- as.data.frame(subset(cam_up_data1_merge, !is.na(freq_gen)))
cam_mid_data1_data2 <- as.data.frame(subset(cam_mid_data1_merge, !is.na(freq_gen)))
cam_low_data1_data2 <- as.data.frame(subset(cam_low_data1_merge, !is.na(freq_gen)))

# bind the SS dfs back together
all_substage_greater_3_data <- rbind(maa_up_data1_data2, maa_low_data1_data2, cam_up_data1_data2, cam_mid_data1_data2, cam_low_data1_data2)

# remove the column of freq_gen
all_substage_greater_3_data <- all_substage_greater_3_data[, c(1:ncol(all_substage_greater_3_data)-1)]
return(all_substage_greater_3_data)
}

# Use above function to remove all rows from the DB that represent grid cells with
# fewer than 3 unique genera
join60_great_3_data <- get_greater_3_database(join60_data)

##### IF GOING TO SUBSAMPLE BY WAY OF USING ONLY GRID CELLS WITH >3 UNIQUE GENERA, 
## USE THE ABOVE CODE AND THEN THE NEXT LINE TO REPLACE THE DB USED IN THE REST OF
## THE CODE WITH ONE WITH >3 UNIQUE GENERA GRID CELLS

# Apply the truncated df (with only >3 occ) as the base data to use in the rest of the code
# join60_data <- join60_great_3_data

##### Option to SUBSET OUT GRID CELLS with greater than SQS limit of unique genera from DB ######

# Function to remove all rows that represent grid cells with fewer than 3 unique genera
get_greater_sqs_database <- function(data){
    # Function to get vector of grid cell names with greater than CAM for SQS limit of 16 unique genera
    get_great_cam_sqs_grids <- function(data,Age){
        # Get just the columns of Genus, Age, and Grid cell name
        gen_age_grid <- data[,c("Updated_Genus","Substage_from_Zone.Mbr","PageName")]
        # Remove duplicates from the list
        unique_gen_age_grid <- unique(gen_age_grid)
        # Get just the substage information in question
        substage_unique_gen_age_grid <- unique_gen_age_grid %>% filter(Substage_from_Zone.Mbr == Age)
    }
    # Get just the columns of Genus, Age, and Grid cell name
    gen_age_grid <- data[,c("Updated_Genus","Substage_from_Zone.Mbr","PageName")]
    # Remove duplicates from the list
    unique_gen_age_grid <- unique(gen_age_grid)
    # Get just the substage information in question
    substage_unique_gen_age_grid <- unique_gen_age_grid %>% filter(Substage_from_Zone.Mbr == Age)
}
# Get the number of unique genera in each grid cell in the substage
substage_grid_freq <- data.frame(table(substage_unique_gen_age_grid$PageName))

# Extract just the Grid cell names that have 3 or more unique genera
substage_grid_greater_cam_sqs_gen <- substage_grid_freq %>% filter(Freq >15)
colnames(substage_grid_greater_cam_sqs_gen) <- c("PageName","freq_gen")

return(substage_grid_greater_cam_sqs_gen)
}

# Function to get vector of grid cell names with greater than MAA for SQS limit of 18 unique genera
great_maa_sqs_grids <- function(data,Age){
  # Get just the columns of Genus, Age, and Grid cell name
  gen_age_grid <- data[,c("Updated_Genus","Substage_from_Zone.Mbr","PageName")]
  
  # Remove duplicates from the list
  unique_gen_age_grid <- unique(gen_age_grid)

  # Get just the substage information in question
  substage_unique_gen_age_grid <- unique_gen_age_grid %>% filter(Substage_from_Zone.Mbr == Age)

  # Get the number of unique genera in each grid cell in the substage
  substage_grid_freq <- data.frame(table(substage_unique_gen_age_grid$PageName))

  # Extract just the Grid cell names that have 3 or more unique genera
  substage_grid_greater_maa_sqs_gen <- substage_grid_freq %>% filter(Freq >17)
colnames(substage_grid_greater_maa_sqs_gen) <- c("PageName","freq_gen")

  return(substage_grid_greater_maa_sqs_gen)
}
# use above function to create list of Grid cells with >3 occ
maa_up_grid_greater_maa_sqs_gen <- get_great_maa_sqs_grids(data,"MAA (up)")
maa_low_grid_greater_maa_sqs_gen <- get_great_maa_sqs_grids(data,"MAA (low)")
cam_up_grid_greater_cam_sqs_gen <- get_great_cam_sqs_grids(data,"CAM (up)")
cam_mid_grid_greater_cam_sqs_gen <- get_great_cam_sqs_grids(data,"CAM (mid)")
cam_low_grid_greater_cam_sqs_gen <- get_great_cam_sqs_grids(data,"CAM (low)")

# get subset of the original db for each substage
maa_up_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr =="MAA (up)"))
maa_low_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr =="MAA (low)"))
cam_up_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr =="CAM (up)"))
cam_mid_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr =="CAM (mid)"))
cam_low_data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr =="CAM (low)"))

# merge the two dfs together to get extra column that has NA if less than 3 genera in the grid cell in that SS
maa_up_data1_merge <-
merge(maa_up_data1,maa_up_grid_greater_maa_sqs_gen,by="PageName",all.x = TRUE)
maa_low_data1_merge <-
merge(maa_low_data1,maa_low_grid_greater_maa_sqs_gen,by="PageName",all.x = TRUE)
cam_up_data1_merge <-
merge(cam_up_data1,cam_up_grid_greater_cam_sqs_gen,by="PageName",all.x = TRUE)
cam_mid_data1_merge <-
merge(cam_mid_data1,cam_mid_grid_greater_cam_sqs_gen,by="PageName",all.x = TRUE)
cam_low_data1_merge <-
merge(cam_low_data1,cam_low_grid_greater_cam_sqs_gen,by="PageName",all.x = TRUE)

# remove rows with NA in the freq_gen column
maa_up_data1_data2 <- as.data.frame(subset(maa_up_data1_merge, !is.na(freq_gen)))
maa_low_data1_data2 <- as.data.frame(subset(maa_low_data1_merge, !is.na(freq_gen)))
cam_up_data1_data2 <- as.data.frame(subset(cam_up_data1_merge, !is.na(freq_gen)))
cam_mid_data1_data2 <- as.data.frame(subset(cam_mid_data1_merge, !is.na(freq_gen)))
cam_low_data1_data2 <- as.data.frame(subset(cam_low_data1_merge, !is.na(freq_gen)))

# bind the SS dfs back together
all_substage_greater_sqs_data <- rbind(maa_up_data1_data2, maa_low_data1_data2,
cam_up_data1_data2, cam_mid_data1_data2,
cam_low_data1_data2)

# remove the column of freq_gen
all_substage_greater_sqs_data <- all_substage_greater_sqs_data[,c(1:ncol(all_substage_greater_sqs_data)-1)]

return(all_substage_greater_sqs_data)

# Use above function to remove all rows from the DB that represent grid cells with
# fewer than sqs number of unique genera
join60_great_sqs_data <- get_greater_sqs_database(join60_data)

### IF GOING TO SUBSAMPLE BY WAY OF USING ONLY GRID CELLS WITH >SQS estimate of UNIQUE
# GENERA,
## USE THE ABOVE CODE AND THEN THE NEXT LINE TO REPLACE THE DB USED IN THE REST OF
## THE CODE WITH ONE WITH >3 UNIQUE GENERA GRID CELLS

# join60_data <- join60_great_sqs_data
Get the number of nodes removed from each substage.

```r
x = subset(join60_data, Substage_from_Zone.Mbr=="MAA (up)"
             )
a = length(unique(x$PageName))
y = subset(join60_great_sqs_data, Substage_from_Zone.Mbr=="MAA (up)"
             )
b = length(unique(y$PageName))
```

```r
x1 = subset(join60_data, Substage_from_Zone.Mbr=="MAA (low)"
             )
a1 = length(unique(x1$PageName))
y1 = subset(join60_great_sqs_data, Substage_from_Zone.Mbr=="MAA (low)"
             )
b1 = length(unique(y1$PageName))
```

```r
x2 = subset(join60_data, Substage_from_Zone.Mbr=="CAM (up)"
             )
a2 = length(unique(x2$PageName))
y2 = subset(join60_great_sqs_data, Substage_from_Zone.Mbr=="CAM (up)"
             )
b2 = length(unique(y2$PageName))
```

```r
x3 = subset(join60_data, Substage_from_Zone.Mbr=="CAM (mid)"
             )
a3 = length(unique(x3$PageName))
y3 = subset(join60_great_sqs_data, Substage_from_Zone.Mbr=="CAM (mid)"
             )
b3 = length(unique(y3$PageName))
```

```r
x4 = subset(join60_data, Substage_from_Zone.Mbr=="CAM (low)"
             )
a4 = length(unique(x4$PageName))
y4 = subset(join60_great_sqs_data, Substage_from_Zone.Mbr=="CAM (low)"
             )
b4 = length(unique(y4$PageName))
```
# Number removed
\[(a-b)+(a_1-b_1)+(a_2-b_2)+(a_3-b_3)+(a_4-b_4)\]

# Original number of nodes
\[a+a_1+a_2+a_3+a_4\]

#### CHANGE DIRECTORY FOR RESULTS TO BE SAVED IN! (This should be changed based on subsampling/subsetting of the original df) ####

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/base_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/greater_3_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_greater_3_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_data_360_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_greater_3_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_wis_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_gcp_data_all_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_bivalve_data_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_gastropod_data_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_cephalopod_data_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_wis_360_data_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_sqs_all_data_outputs")
```

```r
setwd("C:/Users/Ceara/Documents/Province Project/FE_Analysis_Code-and-Files/FE_base_R_Analysis_OutputFiles/compress_sqs_360_data_outputs")
```
#### Create DF of trait information (motility, life_habitat, and feeding) and simplify DB columns ####

```r
# Create data frame of functional traits for imputing into mFD analysis
trait_df <- data.frame(c("motility", "life_habitat", "feeding"), c("N", "N", "N"))
names(trait_df) <- c("trait_name", "trait_type")
head(trait_df)
```

```r
# Convert data to simple dfs that can be used in mFD analyses (subset columns)
simplify_data <- function(data){
data1 <- data.frame(data) # make into a df
data2 <- subset(data1, select = c("Updated_Genus", "Substage_from_Zone.Mbr", "motility", "life_habitat", "feeding", "PageName", "geometry"))
return(data2)
}
simple60_data <- simplify_data(join60_data)
ncol(simple60_data)
colnames(simple60_data)

#### Separate Data into Substages ####

```r
# Function for splitting into substages, remove duplicate Genus-FE rows, make
genus name the index
substg_split <- function(data,x){
data1 <- as.data.frame(subset(data, Substage_from_Zone.Mbr ==x))
```

```r
```
data2 <- distinct(data1, Updated_Genus, PageName, keep_all = TRUE) # delete duplicate genus-Grid Cell values
return(data2)
}

# Create individual substage trait data frames for different genera using substg_split function
maa_up_60_sbstg <- substg_split(simple60_data, "MAA (up)")
maa_low_60_sbstg <- substg_split(simple60_data, "MAA (low)")
cam_up_60_sbstg <- substg_split(simple60_data, "CAM (up)")
cam_mid_60_sbstg <- substg_split(simple60_data, "CAM (mid)")
cam_low_60_sbstg <- substg_split(simple60_data, "CAM (low)")

# Write CSV for each subsampled substage dataset with grid cell values
write.csv(maa_up_60_sbstg, file = "MaaUP60_substage_FE.csv")
write.csv(maa_low_60_sbstg, file = "MaaLOW60_substage_FE.csv")
write.csv(cam_up_60_sbstg, file = "CamUP60_substage_FE.csv")
write.csv(cam_mid_60_sbstg, file = "CamMID60_substage_FE.csv")
write.csv(cam_low_60_sbstg, file = "CamLOW60_substage_FE.csv")

##### Create df and pres-abs matrix of Substages to use in mFD analysis #####
# Function to remove extra columns from the substage data frames (keep only FE # traits and grid loc "PageName" with genus as index still)
substg_trait_df <- function(data){
data1 <- as.data.frame(subset(data, select = c("Updated_Genus", "motility", "life_habitat", "feeding", "PageName")))
data2 <- data1 %>% distinct(Updated_Genus, .keep_all = TRUE)
data2 <- data.frame(data2[,-1], row.names = data2[,1]) # make the first column with genus names the index
return(data2)
}

# Remove unnecessary columns from the data frames to get just the trait
# information for each gen
maa_up_60_trait_data <- substg_trait_df(maa_up_60_sbstg)
maa_low_60_trait_data <- substg_trait_df(maa_low_60_sbstg)
cam_up_60_trait_data <- substg_trait_df(cam_up_60_sbstg)
cam_mid_60_trait_data <- substg_trait_df(cam_mid_60_sbstg)
cam_low_60_trait_data <- substg_trait_df(cam_low_60_sbstg)

# Check number of rows (# genera, also generic richness because no duplicates)
nrow(maa_up_60_trait_data)
nrow(maa_low_60_trait_data)
nrow(cam_up_60_trait_data)
nrow(cam_mid_60_trait_data)
nrow(cam_low_60_trait_data)

# Function to create a simple df of trait values with all variables as factors
# need this later in the code, when calculating sp.to.fe,
df_factor_simple <- function(data){
data2 <- as.data.frame(unclass(data[,c(1:4)]), stringsAsFactors = TRUE)  # make the columns into factors
row.names(data2) <- row.names(data)  # make the first column with genus names the index
data3 <- data2[,c("motility","life_habitat","feeding")]
return(data3)
}

# create simple factor variable df for all substages
maa_up_60_trait_data_factor <- df_factor_simple(maa_up_60_trait_data)
maa_low_60_trait_data_factor <- df_factor_simple(maa_low_60_trait_data)
cam_up_60_trait_data_factor <- df_factor_simple(cam_up_60_trait_data)
cam_mid_60_trait_data_factor <- df_factor_simple(cam_mid_60_trait_data)
cam_low_60_trait_data_factor <- df_factor_simple(cam_low_60_trait_data)

class(cam_low_60_trait_data_factor$motility) # check class of column

# Get a presence-absence matrix for the substages based on grid cells
### NOTE: I redo this later by creating Latitudinal Bins that will act as "collections", instead of grid cells
# see below section on latitudinal bins.

## Function for data substages removing all columns but Genus
# names and Grid Cell names, and for creating presence-absence matrixes for
# each grid cell as collection locations
substg_genus_grid <- function(data){
  data1 <- as.data.frame(subset(data),[c("Updated_Genus","PageName")]) # subset out just the genus name and site
  data2 <- dcast(data1, PageName~Updated_Genus, length) # transform into a pres-abs matrix
  data3 <- data.frame(data2[-1], row.names = data2[,1]) # make the site names the row names
  return(data3)
}

# Create data frames for pres-abs for each substage based on grid cells
# use the substage df that still has Grid Names and geometry
maa_up_60_pres_ab <- substg_genus_grid(maa_up_60_sbstg)
maa_low_60_pres_ab <- substg_genus_grid(maa_low_60_sbstg)
cam_up_60_pres_ab <- substg_genus_grid(cam_up_60_sbstg)
cam_mid_60_pres_ab <- substg_genus_grid(cam_mid_60_sbstg)
cam_low_60_pres_ab <- substg_genus_grid(cam_low_60_sbstg)

# check the number of columns (# genera) against the number of rows in original
# df (again, # genera)
ncol(maa_up_60_pres_ab)
ncol(maa_low_60_pres_ab)
ncol(cam_up_60_pres_ab)
ncol(cam_mid_60_pres_ab)
ncol(cam_low_60_pres_ab)

#### RUN mFD ANALYSIS FOR DATABASE ####
#### Summarize Substage data using sp.tr.summary function ####

# Summarize data using function from mFD package:
maa_up_60_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
                            sp_tr = maa_up_60_trait_data_factor,
                            stop_if_NA = TRUE)

maa_low_60_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
                            sp_tr = maa_low_60_trait_data_factor,
                            stop_if_NA = TRUE)

cam_up_60_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
                            sp_tr = cam_up_60_trait_data_factor,
                            stop_if_NA = TRUE)
cam_mid_60_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
    sp_tr = cam_mid_60_trait_data_factor,
    stop_if_NA = TRUE)

cam_low_60_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
    sp_tr = cam_low_60_trait_data_factor,
    stop_if_NA = TRUE)

# Check trait types
maa_up_60_traits_summ$"tr_types"

# Check trait type details on levels
maa_up_60_traits_summ$"mod_list"

# Create df of Trait summary list for each substage:
maa_up_60_traits_list <- maa_up_60_traits_summ$"tr_summary_list"
maa_low_60_traits_list <- maa_low_60_traits_summ$"tr_summary_list"
cam_up_60_traits_list <- cam_up_60_traits_summ$"tr_summary_list"
cam_mid_60_traits_list <- cam_mid_60_traits_summ$"tr_summary_list"
cam_low_60_traits_list <- cam_low_60_traits_summ$"tr_summary_list"

# write the summary lists of traits to csv files:
write.csv(maa_up_60_traits_list,file="substages_outputs/maa_up_60_traits_list.csv")
write.csv(maa_low_60_traits_list,file="substages_outputs/maa_low_60_traits_list.csv")
write.csv(cam_up_60_traits_list,file="substages_outputs/cam_up_60_traits_list.csv")
write.csv(cam_mid_60_traits_list,file="substages_outputs/cam_mid_60_traits_list.csv")
write.csv(cam_low_60_traits_list,file="substages_outputs/cam_low_60_traits_list.csv")
# NOTE: This just gives summary information of the trait data, not actual 
# counts of FE, etc. Can run on more SS if necessary using above lines.

#### Compute sp.to.fe mFD function to get functional entity information for each Substage ####

## Function for summarizing genera into functional entities

# This function will take the genera from each substage, classify them into FE, then give information about 
# the distribution of genera in FE, list of FE, etc.

```
sum_fe <- function(data){
  data1 <- mFD::sp.to.fe(
    sp_tr       = data, # df of trait values for each genus, with genera names as row index IDs
    tr_cat      = trait_df, # df of traits, includes just "motility", "feeding", and "life_habitat"
    fe_nm_type  = "fe_rank", # tell to name FE by ranked # genera in each (can't use trait names, too many)
    check_input = TRUE) # tell function to give error description if issues in dfs
  return(data1)
}
```

### Possible outputs from above function: ###

# $fe_nms > gives a vector of the FE names (will be ranked list based on # genera in each, not based on traits)

# $sp_fe > gives a vector of genera names and their FE in the substage; order based on decreasing # genera

# $fe_tr > gives a df of the FE and their traits (motility, life habitat, feeding) in columns with FE as rows

# $fe_nb_sp > gives a vector with species number per FE. If all FE have only one species, a warning message

# is returned. FE are ordered according to the decreasing number of species they gather.

# $details_fe > gives a list containing:
# $fe_codes > a vector containing character referring to traits values (like a barcode) with names as
# in fe_nm_type and sorted according to fe_nb_sp

# tr_uval > a list containing for each trait a vector of its unique values or a data frame for fuzzy-coded traits

# fuzzy_E > a list with for each fuzzy-coded trait a data frame with names of entities (E) and names of species (sp)

# tr_nb_uval > a vector with number of unique values per trait (or combinations for fuzzy-coded traits)

# max_nb_fe > the maximum number of FE possible given number of unique values per trait.

# Classify each substage into FE using above function
maa_up_60_traits <- sum_fe(maa_up_60_trait_data_factor)
maa_low_60_traits <- sum_fe(maa_low_60_trait_data_factor)
cam_up_60_traits <- sum_fe(cam_up_60_trait_data_factor)
cam_mid_60_traits <- sum_fe(cam_mid_60_trait_data_factor)
cam_low_60_traits <- sum_fe(cam_low_60_trait_data_factor)

## NOTE: the above trait outputs will be used later to calc FR and evenness for each SS overall using the $fe_nb_sp

# option; see later section.

# Look at the FE names
maa_up_60_traits$"fe_nm"
maa_low_60_traits$"fe_nm"
cam_up_60_traits$"fe_nm"
cam_mid_60_traits$"fe_nm"
cam_low_60_traits$"fe_nm" # different numbers of FE found for each substage, but roughly the same amount
# Look at which FE each genus belongs to

```r
sp_fe_maa_up_60_traits <- maa_up_60_traits$"sp_fe"
sp_fe_maa_low_60_traits <- maa_low_60_traits$"sp_fe"
sp_fe_cam_up_60_traits <- cam_up_60_traits$"sp_fe"
sp_fe_cam_mid_60_traits <- cam_mid_60_traits$"sp_fe"
sp_fe_cam_low_60_traits <- cam_low_60_traits$"sp_fe"
```

# Look at number of genera per FE

```r
fe_nb_sp_maa_up_60_traits <- maa_up_60_traits$"fe_nb_sp" # yes, genus names kept!! hurray!
fe_nb_sp_maa_up_60_traits
```

##### Use SQS to check subsampling size for adequate representation of FEs #######

## This is intended to give an idea about how low sampling can go before FE are
# not well represented in each SS. Important for nodes, where there are very
# few total genera, and even fewer unique ones...
## Use the above table lists of FE for each genera.

# Create Frequency lists for each FE in each SS:

```r
maa_up_60_fe_freq <- table(sp_fe_maa_up_60_traits) # all data frequencies of genera
maa_low_60_fe_freq <- table(sp_fe_maa_low_60_traits) # all data frequencies of genera
cam_up_60_fe_freq <- table(sp_fe_cam_up_60_traits) # all data frequencies of genera
cam_mid_60_fe_freq <- table(sp_fe_cam_up_60_traits) # all data frequencies of genera
cam_low_60_fe_freq <- table(sp_fe_cam_up_60_traits) # all data frequencies of genera
```

# Use SQS subsampling function (see above) to get estimate for sampling level:

```r
maa_up_sqs_sample_for_fe <- run_sqs(maa_up_60_fe_freq)
maa_low_sqs_sample_for_fe <- run_sqs(maa_low_60_fe_freq)
```
cam_up_sqs_sample_for_fe <- run_sqs(cam_up_60_fe_freq)
cam_mid_sqs_sample_for_fe <- run_sqs(cam_mid_60_fe_freq)
cam_low_sqs_sample_for_fe <- run_sqs(cam_low_60_fe_freq)

# CHANGE IF USING WIS ONLY: Bind into a df and give row names for each SS:
sqs_sample_size_fe <- data.frame(c(maa_up_sqs_sample_for_fe, maa_low_sqs_sample_for_fe,
                                  cam_up_sqs_sample_for_fe, cam_mid_sqs_sample_for_fe,
                                  cam_low_sqs_sample_for_fe))

# sqs_sample_size_fe <- data.frame(c(maa_low_sqs_sample_for_fe,
                                  # cam_up_sqs_sample_for_fe, cam_mid_sqs_sample_for_fe,
                                  # cam_low_sqs_sample_for_fe))

rownames(sqs_sample_size_fe) <- c("Maa UP","Maa LOW","Cam UP","Cam MID","Cam LOW")

# rownames(sqs_sample_size_fe) <- c("Maa LOW","Cam UP","Cam MID","Cam LOW")

colnames(sqs_sample_size_fe) <- "SQS Size Estimate"

# Write to csv
write.csv(sqs_sample_size_fe,file="grid_collections_outputs/sqs_sample_size_fe.csv")

#### ANALYSIS FOR GRID CELLS AS COLLECTIONS ####

#### Compute alpha.fd.fe mFD function to get alpha and beta diversity information for each Substage ####

## Function to calc functional trait values
calc_fe <- function(trait_data,pres_ab_data){

...
alpha_fd_fe_MAAup <- mFD::alpha.fd.fe(
  asb_sp_occ = pres_ab_data,
  sp_to_fe = trait_data,
  ind_nm = c("fred", "fored", "fvlun"),
  check_input = TRUE,
  details_returned = TRUE)
}

# Calculate functional richness, redundancy, over redundancy, and vulnerability for each substage in grid cells
maa_up_60_fe <- calc_fe(maa_up_60_traits, maa_up_60_pres_ab)
maa_low_60_fe <- calc_fe(maa_low_60_traits, maa_low_60_pres_ab)
cam_up_60_fe <- calc_fe(cam_up_60_traits, cam_up_60_pres_ab)
cam_mid_60_fe <- calc_fe(cam_mid_60_traits, cam_mid_60_pres_ab)
cam_low_60_fe <- calc_fe(cam_low_60_traits, cam_low_60_pres_ab)

# Get a matrix of fe values for each grid cell in a time bin
maa_up_60_fdfe <- (maa_up_60_fe$"asb_fdfe")
maa_low_60_fdfe <- (maa_low_60_fe$"asb_fdfe")
cam_up_60_fdfe <- (cam_up_60_fe$"asb_fdfe")
cam_mid_60_fdfe <- (cam_mid_60_fe$"asb_fdfe")
cam_low_60_fdfe <- (cam_low_60_fe$"asb_fdfe")

# Look at the df of the alpha-beta diversity outputs
maa_up_60_fdfe
dmaa_low_60_fdfe
cam_up_60_fdfe
cam_mid_60_fdfe
cam_low_60_fdfe
# Export results tables summarizing the number of spp, fe, and functional richness and evenness values for each substage

write.csv(maa_up_60_fdfe,file="grid_collections_outputs/MaaUP60_fe_alpha_beta_results.csv")
write.csv(maa_low_60_fdfe,file="grid_collections_outputs/MaaLOW60_fe_alpha_beta_results.csv")
write.csv(cam_up_60_fdfe,file="grid_collections_outputs/CamUP60_fe_alpha_beta_results.csv")
write.csv(cam_mid_60_fdfe,file="grid_collections_outputs/CamMID60_fe_alpha_beta_results.csv")
write.csv(cam_low_60_fdfe,file="grid_collections_outputs/CamLOW60_fe_alpha_beta_results.csv")

# Function to check for diff between two data frames genus name lists, only use
# if error in cal FR above

cHECK_DIFF_SP <- function(trait_data,pres_ab_data){
  df <- as.data.frame(trait_data$"sp_fe")
  df <- cbind(newColName = rownames(df), df)
  rownames(df) <- NULL
  df <- df[,1]

  dd <- as.data.frame(colnames(pres_ab_data))
  dd <- dd[,1]
  head(dd)

  df <- sort(df)
  dd <- sort(dd)
  diff <- setdiff(df,dd)
  return(diff)
}

#### Look at relationship between Generic and Functional Richness ####

#### Look at relationship between Generic and Functional Richness ####
# Use iNext to create rarifaction curve of the number genera per number of FEs
library(iNEXT)

data(spider)
str(spider)

out <- iNEXT(spider, q=0, datatype="abundance", endpoint=500)

aa <- fe_data1[,2:3]

bb <- iNEXT(aa, q=0, datatype="abundance", size=NULL, endpoint=NULL, knots=40, se=TRUE, conf=0.95, nboot=50)

out <- iNEXT(spider, q=c(0, 1, 2), datatype="abundance", endpoint=500)
# Sample-size-based R/E curves, separating plots by "site"
ggiNEXT(out, type=1, facet.var="site")

# function to get the Age data as first column
df_func <- function(data){
  ### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###
data1 <- cbind(Bin = rownames(data), data) # make the grid name the 1st column again
rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
data1 <- as.data.frame(data1)
i <- c(2:6)
data1[, i] <- apply(data1[, i], 2, function(x) as.numeric(as.character(x))) # Specify own function within apply to make numeric
return(as.data.frame(data1))
}

fe_data1 <- df_func(maa_up_60_fdfe)
fe_data2 <- df_func(maa_low_60_fdfe)
fe_data3 <- df_func(cam_up_60_fdfe)
fe_data4 <- df_func(cam_mid_60_fdfe)
fe_data5 <- df_func(cam_low_60_fdfe)

data_fe_all <- rbind(fe_data1, fe_data2, fe_data3, fe_data4, fe_data5)

# perform Shapiro test for normality on the GR and FR values from grid cells
shapiro.test(data_fe_all[, "nb_sp"])
shapiro.test(data_fe_all[, "nb_fe"])

# make plot of all the GR vs FR values for each grid cell in all substages
p <- ggplot(fe_data1, aes(x = nb_sp, y = nb_fe)) +
  geom_point(color = "darkred") +
  geom_point(data = fe_data2, aes(x = nb_sp, y = nb_fe), color = "#D55E00") +
  geom_point(data = fe_data3, aes(x = nb_sp, y = nb_fe), color = "goldenrod3") +
  geom_point(data = fe_data4, aes(x = nb_sp, y = nb_fe), color = "darkolivegreen4") +
  geom_point(data = fe_data5, aes(x = nb_sp, y = nb_fe), color = "steelblue4")

# add different models to the data to see which fits best
p + stat_smooth(method = "lm", formula = y ~ x, size = 1, se = FALSE, colour = "black") +
  stat_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1, se = FALSE, colour = "blue") +
  stat_smooth(method = "loess", formula = y ~ x, size = 1, se = FALSE, colour = "red") +
  stat_smooth(method = "gam", formula = y ~ s(x), size = 1, se = FALSE, colour = "green") +
  stat_smooth(method = "gam", formula = y ~ s(x, k = 3), size = 1, se = FALSE, colour = "violet") +
stat_smooth(method = "lm", formula = y ~ sqrt(x), size = 1, se = FALSE, colour = "goldenrod") +
stat_smooth(method = "lm", formula = log(y) ~ log(x), size = 1, se = FALSE, colour = "brown4")

# look at most likely models of those above
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1, se = FALSE, colour = "blue") +
stat_smooth(method = "loess", formula = y ~ x, size = 1, se = FALSE, colour = "red") +
stat_smooth(method = "gam", formula = y ~ s(x, k = 3), size = 1, se = FALSE, colour = "violet") +
stat_smooth(method = "lm", formula = y ~ sqrt(x), size = 1, se = FALSE, colour = "goldenrod")

# look at model that appears to fit best of those above
p + stat_smooth(method = "lm", formula = y ~ sqrt(x), size = 1, colour = "goldenrod") +

p + stat_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1, colour = "blue")

# It appears, based on the above plots, that other than the loess and gam models,
# the square root and squared models fit the data best, so tested below:

# create square root model
model_sr<-lm(fe_data1$nb_sp ~ I(sqrt(fe_data1$nb_fe)), data = fe_data1)
summary(model_sr)

# create squared model (this one appears to have a higher R2 so probably slightly better fit)
model_sq<-lm(fe_data1$nb_sp ~ I((fe_data1$nb_fe)^2), data = fe_data1)
summary(model_sq)

names(summary(model_sq))

summary(model_sq)$r.squared # extremely high R2 value, indicates good fit
summary(model_sq)$coefficients
summary(model_sq)$adj.r.squared

# plot the information related to the regression model
par(mfrow = c(2, 2))
plot(model_sq)

# export plots as pdf
pdf("grid_collections_outputs/squared_regression_model_GRvsFR_plots.pdf", width = 11, height = 11) # Open a new pdf file
par(mfrow = c(2, 2))
plot(model_sq)
dev.off() # Close the file

# export plot of GR vs FR within grid cells for each SS and regression line:
pdf("grid_collections_outputs/GRvsFR_regression_plots.pdf", width = 6, height = 6) # Open a new pdf file
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1, colour = "blue")
dev.off() # Close the file

# Based on the above plots, only one point counts as an outlier which may influence
# the regression (>1 cook's distance), but the residuals indicate a linear
# relationship in the data (Residual vs Fitted). The Scale-location plot suggests
# that the data have fairly homogeneous variance values, except at low fitted
# values where the variance is quite low. The Q-Q plots show non-normality,
# which is expected given the Shapiro-Wilk tests performed above.
Results show that there is a high correlation between GR and FR, which appears to potentially plateau at the very highest levels of GR (>150), and potentially follow a squared model function (i.e., peaking at ~150 genera before the correlation becomes negative).

Either way, this should be considered when examining the results of the analysis. When using lower sampling (correlated strongly with GR), GR is low, and therefore FR will probably be low as well. Might be the case that you can't really examine FD at local scales in the fossil record, given the limitations on the data. This issue may also become even more difficult when using subsampling like sqs, since it will necessarily decrease sampling overall and impacting sampling especially in localized regions.

Use alpha.fd.fe.plot to Plot the results for each grid cell in each substage
(par(mfrow) wont work for ggplots so this is my work around)

Note: can't keep title of each subplot when putting them together. very annoying. Not sure how to fix this since it seems like it's just part of the base code for plotting these. Will have to manually add titles if necessary.

Run a for loop function to create the plots and add them to the empty list

plot_alpha_beta_fe <- function(fe_data,sbstg_data) {
  fe_plots = list() # create an empty list to store the grid plots in
  for(i in 1:length(unique(sbstg_data$PageName))) {
    grid <- unique(sbstg_data$PageName)
    fe_plots[[i]] = mFD::alpha.fd.fe.plot(
      alpha_fd_fe = fe_data,
      plot_asb_nm = c(grid[i]),
      plot_ind_nm = c("fred", "fored", "fvuln"),
    )
  }
  return(fe_plots)
}

Run this function with your data to create the plots and store them in a list called fe_plots.
name_file = NULL,
color_fill_fored = "darkolivegreen2",
color_line_fred = "darkolivegreen4",
color_fill_bar = "grey80",
color_fill_fvuln = "lightcoral",
color_arrow_fvuln = "indianred4",
size_line_fred = 0.5,
size_arrow_fvuln = 0.5,
check_input = TRUE)
}
return(fe_plots)
}

maa_up_60_fe_plots <- plot_alpha_beta_fe(maa_up_60_fe, maa_up_60_sbstg)
maa_low_60_fe_plots <- plot_alpha_beta_fe(maa_low_60_fe, maa_low_60_sbstg)
cam_up_60_fe_plots <- plot_alpha_beta_fe(cam_up_60_fe, cam_up_60_sbstg)
cam_mid_60_fe_plots <- plot_alpha_beta_fe(cam_mid_60_fe, cam_mid_60_sbstg)
cam_low_60_fe_plots <- plot_alpha_beta_fe(cam_low_60_fe, cam_low_60_sbstg)

# export plots as single pdf
pdf("grid_collections_outputs/alpha_beta_allgrids_MAA_up_plots.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_up_60_fe_plots,
              top = "Upper Maastr. Alpha/Beta Diversity per Grid Cell") # Write the grid.arrange in the file
dev.off() # Close the file

# export plots as single pdf
grid.arrange(grobs = maa_low_60_fe_plots,
              top = "Lower Maastr. Alpha/Beta Diversity per Grid Cell") # Write the grid.arrange in the file
dev.off() # Close the file

# export plots as single pdf
pdf("grid_collections_outputs/alpha_beta_allgrids_CAM_up_plots.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_up_60_fe_plots,
              top = "Upper Camp. Alpha/Beta Diversity per Grid Cell") # Write the grid.arrange in the file
dev.off() # Close the file

# export plots as single pdf
grid.arrange(grobs = cam_mid_60_fe_plots,
              top = "Middle Camp. Alpha/Beta Diversity per Grid Cell") # Write the grid.arrange in the file
dev.off() # Close the file

# export plots as single pdf
grid.arrange(grobs = cam_low_60_fe_plots,
              top = "Lower Camp. Alpha/Beta Diversity per Grid Cell") # Write the grid.arrange in the file
dev.off() # Close the file

# Clearly, many grid cells contain very few functional entities AND genera
# think about removing cells with fewer than 3 genera...
#### Create df of FR for each Grid Cell through time####

# This section is firstly trying to make matrices which will show richness values in each
# grid cell across the five substages. Many grid cells do not have consistent occupation,
# however. Next, will be adding the general location of the grid cells (lat/lon) to the FR df
# so that these tables can be exported and plotted in on maps.

# Function to configure to columns with site names as 1st and FR values as second for FUNCTIONAL
# RICHNESS
config_for_loc_agg <- function(data, Age){
  # Create df that just take the functional richness values in each grid cell
  data_rich_summ <- (data"asb_fdfe")[,2]

  data_rich_summ <- data.frame(data_rich_summ)

  data_rich_summ <- cbind(newColName = rownames(data_rich_summ), data_rich_summ) # make grid
cell first column

  rownames(data_rich_summ) <- 1:nrow(data_rich_summ)

  colnames(data_rich_summ) <- c("Site","Numb_FE")

  data_rich_summ$Age <- Age

  return(data_rich_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_func_rich_summ <- config_for_loc_agg(maa_up_60_fe, "MAA up")
maa_low_60_func_rich_summ <- config_for_loc_agg(maa_low_60_fe, "MAA low")
cam_up_60_func_rich_summ <- config_for_loc_agg(cam_up_60_fe, "CAM up")
cam_mid_60_func_rich_summ <- config_for_loc_agg(cam_mid_60_fe, "CAM mid")
cam_low_60_func_rich_summ <- config_for_loc_agg(cam_low_60_fe, "CAM low")
# Bind the data into a single df with 2 columns (stacked the different dfs)

```r
func_rich_stack <- rbind(maa_up_60_func_rich_summ, maa_low_60_func_rich_summ,
                         cam_up_60_func_rich_summ, cam_mid_60_func_rich_summ,
                         cam_low_60_func_rich_summ)
```

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values

```r
func_rich_table_all <- dcast(func_rich_stack, Site ~ Age, value.var = "Numb_FE")
```

# This is a table showing the number of FE in each grid cell for each substage

```r
head(func_rich_table_all)
```

```r
write.csv(func_rich_table_all, file="grid_collections_outputs/Grid_FR_through_time.csv")
```

#### Create df of GR for each Grid Cell through time####

# Function to configure to columns with site names as 1st and FR values as second for GENERIC RICHNESS

```r
config_for_loc_agg_gen_rich <- function(data, Age){
    # Create df that just take the functional richness values in each grid cell
    data_rich_summ <- (data$"asb_fdfe")[,1]
    
    data_rich_summ <- data.frame(data_rich_summ)
    data_rich_summ <- cbind(newColName = rownames(data_rich_summ), data_rich_summ) # make grid cell first column
    rownames(data_rich_summ) <- 1:nrow(data_rich_summ)
    colnames(data_rich_summ) <- c("Site","Numb_G")
}
```

```r
config_for_loc_agg_gen_rich(data, Age)
```
data_rich_summ$Age <- Age
return(data_rich_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_generic_rich_summ <- config_for_loc_agg_gen_rich(maa_up_60_fe, "MAA up")
maa_low_60_generic_rich_summ <- config_for_loc_agg_gen_rich(maa_low_60_fe, "MAA low")
cam_up_60_generic_rich_summ <- config_for_loc_agg_gen_rich(cam_up_60_fe, "CAM up")
cam_mid_60_generic_rich_summ <- config_for_loc_agg_gen_rich(cam_mid_60_fe, "CAM mid")
cam_low_60_generic_rich_summ <- config_for_loc_agg_gen_rich(cam_low_60_fe, "CAM low")

# Bind the data into a single df with 2 columns (stacked the different dfs)
generic_rich_stack <- rbind(maa_up_60_generic_rich_summ, maa_low_60_generic_rich_summ,
                                 cam_up_60_generic_rich_summ, cam_mid_60_generic_rich_summ,
                                 cam_low_60_generic_rich_summ)

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
generic_rich_table_all <- dcast(generic_rich_stack, Site ~ Age, value.var = "Numb_G")

# This is a table showing the number of FE in each grid cell for each substage
head(generic_rich_table_all)

write.csv(generic_rich_table_all, file="grid_collections_outputs/Grid_GR_through_time.csv")

##### Get the average location for each grid cell per substage (use to group by lat and compare across lat) #####
# Function to separate geometry back out into lat-long information:

sep_lat_substage <- function(data, substage) {

  # separate the geometry values back into lat/lon
  data <- data %>%
    mutate(geometry = gsub('[POINT ()]', '', geometry)) %>%
    separate(col = geometry, into = c('Longitude', 'Latitude'), sep = '\,')

  # remove the first character in lat string, don't know why but the above code adds a "c"
  data$Longitude <- sub('.', '', data$Longitude)

  # make the lat/lon columns into numeric values
  data$Longitude <- as.numeric(data$Longitude)
  data$Latitude <- as.numeric(data$Latitude)

  # separate the df by substages so can take average locations for each grid cell at each substage
  substg_data <- as.data.frame(subset(data, Substage_from_Zone.Mbr == substage))
  substg_data %>% distinct(Updated_Genus, PageName, .keep_all = TRUE) # delete duplicate value

  return(substg_data)
}

# Create new df of all data with the lat/lon separated back out

maa_up_60_loc_data <- sep_lat_substage(join60_data, "MAA (up)")
maa_low_60_loc_data <- sep_lat_substage(join60_data, "MAA (low)")
cam_up_60_loc_data <- sep_lat_substage(join60_data, "CAM (up)")
cam_mid_60_loc_data <- sep_lat_substage(join60_data, "CAM (mid)")
cam_low_60_loc_data <- sep_lat_substage(join60_data, "CAM (low)")
# Function to separate geometry back out into lat-long information for entire db:
sep_lat <- function(data){

  # separate the geometry values back into lat/lon
  data <- data %>%
    mutate(geometry = gsub('[POINT (]','", geometry)) %>%
    separate(col = geometry, into = c('Longitude', 'Latitude'), sep = '\\,')

  # remove the first character in lat string, don't know why but the above code adds a "c"
  data$Longitude <- sub('.', '', data$Longitude)

  # make the lat/lon columns into numeric values
  data$Longitude <- as.numeric(data$Longitude)
  data$Latitude <- as.numeric(data$Latitude)

  return(data)
}

join60_data_lat_sep <- sep_lat(join60_data)
# this above format not used in separating by lat bins (below section), instead
# use the db table made by combining substages that have been given paleo coords,
# see next few lines.

# load the chronosphere package to run this code. This should allow for reconstructions of plat and plong:
# https://www.evolv-ed.net/post/chronosphere-paleomap/chronosphere-paleomap/
library(chronosphere)
# Function to get the plat and plon information for each occ:

```r
estimate_ploc <- function(data, age){
  # get just the lat long locations from the table
  data_loc_2col <- data[,c("Longitude","Latitude")]

  # Reconstruct paleocoord using age of rough age of substage (approximate, will use the same age for M Cam because limited options...)
  data_ploc <- as.data.frame(reconstruct(data_loc_2col, age=c(age)))

  colnames(data_ploc) <- c("plong","plat")

  ### Add the paleocoord back into the original location df
  data$plong <- data_ploc$plong
  data$plat <- data_ploc$plat
  return(data)
}
```

# Use above function to get the plat and plon information for each substage site

```r
maa_up_60_paleo_loc_data <- estimate_ploc(maa_up_60_loc_data, 80)
maa_low_60_paleo_loc_data <- estimate_ploc(maa_low_60_loc_data, 80)
cam_up_60_paleo_loc_data <- estimate_ploc(cam_up_60_loc_data, 75)
cam_mid_60_paleo_loc_data <- estimate_ploc(cam_mid_60_loc_data, 70)
cam_low_60_paleo_loc_data <- estimate_ploc(cam_low_60_loc_data, 70)
```

# Write paleo location information to a new file set (MIGHT WANT THESE LATER IF CHRONOSPHERE STOPS WORKING AGAIN...)
write.csv(maa_up_60_paleo_loc_data, file="maa_up_60_paleo_loc_data.csv")
write.csv(maa_low_60_paleo_loc_data, file="maa_low_60_paleo_loc_data.csv")
write.csv(cam_up_60_paleo_loc_data, file="cam_up_60_paleo_loc_data.csv")
write.csv(cam_mid_60_paleo_loc_data, file="cam_mid_60_paleo_loc_data.csv")
write.csv(cam_low_60_paleo_loc_data, file="cam_low_60_paleo_loc_data.csv")

# read in the csv data that was produced before for paleocoor information, to save comp time
maa_up_60_paleo_loc_data <- read.csv("maa_up_60_paleo_loc_data.csv")
maa_low_60_paleo_loc_data <- read.csv("maa_low_60_paleo_loc_data.csv")
cam_up_60_paleo_loc_data <- read.csv("cam_up_60_paleo_loc_data.csv")
cam_mid_60_paleo_loc_data <- read.csv("cam_mid_60_paleo_loc_data.csv")
cam_low_60_paleo_loc_data <- read.csv("cam_low_60_paleo_loc_data.csv")

# join the above dfs together to create a complete df of all occurrence information
join60_data_paleo_loc <- rbind(maa_up_60_paleo_loc_data,
                               maa_low_60_paleo_loc_data,
                               cam_up_60_paleo_loc_data,
                               cam_mid_60_paleo_loc_data,
                               cam_low_60_paleo_loc_data)

# write the entire db as a new file with paleo locations:
write.csv(join60_data_paleo_loc, file="join60_data_60_paleo_loc_data.csv")

## NOTE: the above df used later when binning by latitude, used to create "bins" that can be used instead of

# grid cells when calculating alpha-beta functional diversity

# Create function to aggregate location information based on grid cell and merge with grid IDs
agg_paleo_loc <- function(loc_data, rich_data) {
  agg_lat <- aggregate(x = loc_data$plat, by = list(loc_data$PageName), FUN = mean)
  agg_lon <- aggregate(x = loc_data$plong, by = list(loc_data$PageName), FUN = mean)
  grid_loc <- merge(agg_lat, agg_lon, "Group.1")
  colnames(grid_loc) <- c("Site", "pLat", "pLon")
  rich_latlong <- merge(rich_data, grid_loc, by = "Site", all.x = TRUE)
  return(rich_latlong)
}

# Aggregate and merge location information for each substage (above function)
# these locations will be used to plot the locations of values on maps based on average grid cell locations
maa_up_60_grid_paleo_local <- agg_paleo_loc(maa_up_60_paleo_loc_data, maa_up_60_func_rich_summ)
maa_low_60_grid_paleo_local <- agg_paleo_loc(maa_low_60_paleo_loc_data, maa_low_60_func_rich_summ)
cam_up_60_grid_paleo_local <- agg_paleo_loc(cam_up_60_paleo_loc_data, cam_up_60_func_rich_summ)
cam_mid_60_grid_paleo_local <- agg_paleo_loc(cam_mid_60_paleo_loc_data, cam_mid_60_func_rich_summ)
cam_low_60_grid_paleo_local <- agg_paleo_loc(cam_low_60_paleo_loc_data, cam_low_60_func_rich_summ)

# Aggregate and merge location information for entire database (above function)
# these locations will be used to group lat bins in plots based on average grid cell locations
grid_local <- agg_paleo_loc(join60_data_paleo_loc, func_rich_table_all)
#### Create histograms of grid metric values for each substage ####

# function to create histograms of all metrics calculated by mFD for alpha and beta diversity

hist_plots <- function(data){
  sp_hist <- ggplot() +
  geom_histogram(as.data.frame(data), mapping = aes(x = nb_sp),
                 binwidth=5, fill="skyblue4", color="#e9ecef", alpha=0.9) +
  labs(title="Grid Cell Generic Richness (GR)",
       x="# Genera", y = "Frequency of Grid Cells") +
  theme(plot.title = element_text(hjust = 0.5))

  fe_hist <- ggplot() +
  geom_histogram(as.data.frame(data), mapping = aes(x = nb_fe),
                 binwidth=3, fill="skyblue4", color="#e9ecef", alpha=0.9) +
  labs(title="Grid Cell Functional Richness (FR)",
       x="# Functional Entities", y = "Frequency of Grid Cells") +
  theme(plot.title = element_text(hjust = 0.5))

  fred_hist <- ggplot() +
  geom_histogram(as.data.frame(data), mapping = aes(x = fred),
                 binwidth=0.3, fill="skyblue4", color="#e9ecef", alpha=0.9) +
  labs(title="Grid Cell Functional Redundancy (FRed)",
       x="GR/FR", y = "Frequency of Grid Cells") +
  theme(plot.title = element_text(hjust = 0.5))

  fored_hist <- ggplot() +
  geom_histogram(as.data.frame(data), mapping = aes(x = fored),
                 binwidth=0.1, fill="skyblue4", color="#e9ecef", alpha=0.9) +
  labs(title="Grid Cell Functional Over-Redundancy (FORed)",
       x="FRed Values", y = "Frequency of Grid Cells") +
  theme(plot.title = element_text(hjust = 0.5))

  return(list(sp_hist = sp_hist, fe_hist = fe_hist, fred_hist = fred_hist, fored_hist = fored_hist))
}

hist_plots(data)
fvuln_hist <- ggplot() +
    geom_histogram(as.data.frame(data), mapping = aes(x = fvuln),
        binwidth=0.15, fill="skyblue4", color="#e9ecef", alpha=0.9) +
    labs(title="Grid Cell Functional Vulnerability (FVln)", x="FVln Values", y = "Frequency of Grid Cells") +
    theme(plot.title = element_text(hjust = 0.5))

dist_list <- list(sp_hist,fe_hist,fred_hist,fored_hist,fvuln_hist)

# Create series of histograms for metric values in each substage
maa_up_60_grid_hist <- hist_plots(maa_up_60_fdfe)
maa_low_60_grid_hist <- hist_plots(maa_low_60_fdfe)
cam_up_60_grid_hist <- hist_plots(cam_up_60_fdfe)
cam_mid_60_grid_hist <- hist_plots(cam_mid_60_fdfe)
cam_low_60_grid_hist <- hist_plots(cam_low_60_fdfe)

df("grid_collections_outputs/hist_MAA_up_allgrids.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_up_60_grid_hist,
    top = "Upper Maastr. Histograms of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() #close the file

df("grid_collections_outputs/hist_MAA_low_allgrids.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_low_60_grid_hist,
    top = "Lower Maastr. Histograms of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() #close the file

df("grid_collections_outputs/hist_CAM_up_allgrids.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_up_60_grid_hist, 
    top = "Upper Camp. Histograms of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

pdf("grid_collections_outputs/hist_CAM_mid_allgrids.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_mid_60_grid_hist, 
    top = "Middle Camp. Histograms of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

grid.arrange(grobs = cam_low_60_grid_hist, 
    top = "Lower Camp. Histograms of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

##### Create bar plots of FE metrics for each grid cell in each Substage #####
# Function to create series of bar plots for each substage's grid cells
bar_plot <- function(fe_data, loc_data, substage){

    ### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###
data1 <- cbind(newColName = rownames(fe_data), fe_data) # make the genus name the 1st column again
rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
data1 <- as.data.frame(data1)
i <- c(2:6)
data1[, i] <- apply(data1[, i], 2, function(x) as.numeric(as.character(x)))
colnames(data1) <- c("Grid", "nb_sp", "nb_fe", "fred", "fores", "fvuln") # set colnames

### Extract Location Information: ###
loc_data1 <- loc_data[cbind("Site", "substage", "pLat", "pLon")]
# Extract columns for just one substage
colnames(loc_data1) <- c("Grid", "substage", "pLat", "pLon")
# Rename columns
loc_data1$lat_1 <- loc_data1$pLat # Add another column to use for binning lat
# Next, use mutate and cut to create binned categories of lat
loc_data2 <- loc_data1 %>% mutate(lat_1 = cut(loc_data1$lat_1,
breaks=c(-Inf, seq(round(min(loc_data1$pLat)/5)*5,
round(max(loc_data1$pLat)/5)*5, by = 5), Inf))

### Merge the new bins with the fe df: ###
data3 <- merge(data1, loc_data2, by = "Grid", all.x = TRUE)
# get the unique lat bins available for this substage and change colname
bins <- data.frame(unique(data3$lat_1))
colnames(bins) <- "lat_1"
# Create column with the sorts of labels actually want to use based on ifelse
bins$lat_bins <- with(bins, ifelse(lat_1 == "(-Inf,30]", '23-30°N',
ifelse(lat_1 == "[30,35]", '30-35°N',
ifelse(lat_1 == "[35,40]", '35-40°N',
ifelse(lat_1 == "[40,45]", '40-45°N',
ifelse(lat_1 == "[45,50]", '45-50°N',
ifelse(lat_1 == "[50,55]", '50-55°N',
ifelse(lat_1 == "[55,60]", '55-60°N',
'60-65°N'))))))})))
# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)

### Arrange the df based on lat bins and then create a object with the correct order: ###
data4 <- arrange(data3,Lat_1)
grids <- data4$Grid

### Plot the data using geom_col: ###
sp_bar <- ggplot(data3, aes(x=Grid, y=nb_sp, fill=Lat_1)) +
  geom_col(color="#e9ecef") +
  geom_text(aes(label=nb_sp), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +
  labs(title="Generic Richness",x="Grid Cell", y = "Number of Genera", fill="Paleo-Latitude") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_manual(values=c("red4",
                        "tomato3",
                        "darkorange3",
                        "goldenrod2",
                        "darkolivegreen4",
                        "aquamarine4",
                        "steelblue"),
                        labels = bin_labels)+
  scale_x_discrete(limits=grids) +
  ylim(0, 1 + max(data3$nb_sp))

fe_bar <- ggplot(data3, aes(x=Grid, y=nb_fe, fill=Lat_1)) +
  geom_col(color="#e9ecef") +
geom_text(aes(label=signif(nb_fe,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +
labs(title="Functional Richness", x="Grid Cell", y = "Number of Functional Entities", fill="Paleo-Latitude") +
theme(plot.title = element_text(hjust = 0.5),
axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
scale_fill_manual(values=c("red4",
"tomato3",
"darkorange3",
"goldenrod2",
"darkolivegreen4",
"aquamarine4",
"steelblue"),
labels = bin_labels)+
scale_x_discrete(limits=grids) +
ylim(0, 1 + max(data3$nb_fe))

fred_bar <- ggplot(data3, aes(x=Grid, y=fred, fill=Lat_1)) +
geom_col(color="#e9ecef") +
geom_text(aes(label=signif(fred,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +
labs(title="Functional Redundancy (FRed)", x="Grid Cell", y = "FRed Value", fill="Paleo-Latitude") +
theme(plot.title = element_text(hjust = 0.5),
axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
scale_fill_manual(values=c("red4",
"tomato3",
"darkorange3",
"goldenrod2",
"darkolivegreen4",
"aquamarine4",
"steelblue"),
labels = bin_labels)+

scale_x_discrete(limits=grids) +
ylim(0, 1 + max(data3$fored))

fores_bar <- ggplot(data3, aes(x=Grid, y=fored, fill=Lat_1)) +
  geom_col(color="#e9ecef") +
  geom_text(aes(label=signif(fored,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +
  labs(title="Functional Over-Redundancy", x="Grid Cell", y = "FORed Value", fill="Paleo-Latitude") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_manual(values=c("red4",
                       "tomato3",
                       "darkorange3",
                       "goldenrod2",
                       "darkolivegreen4",
                       "aquamarine4",
                       "steelblue"),
                      labels = bin_labels)+
  scale_x_discrete(limits=grids) +
  ylim(0, 1 + max(data3$fored))

fvuln_bar <- ggplot(data3, aes(x=Grid, y=fvuln, fill=Lat_1)) +
  geom_col(color="#e9ecef") +
  geom_text(aes(label=signif(fvuln,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +
  labs(title="Functional Vulnerability (FVuln)", x="Grid Cell", y = "FVuln Value", fill="Paleo-Latitude") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_manual(values=c("red4",
                           "tomato3","
"darkorange3",
"goldenrod2",
"darkolivegreen4",
"aquamarine4",
"steelblue"),
labels = bin_labels)+
scale_x_discrete(limits=grids) +
ylim(0, 1 + max(data3$fvuln))

bar_list <- list(sp_bar,fe_bar,fred_bar,fored_bar,fvuln_bar)
}

# Create series of bar plots for metric values in each substage
maa_up_60_grid_bar <- bar_plot(maa_up_60_fdfe,grid_local,"MAA up")
maa_low_60_grid_bar <- bar_plot(maa_low_60_fdfe,grid_local,"MAA low")
cam_up_60_grid_bar <- bar_plot(cam_up_60_fdfe,grid_local,"CAM up")
cam_mid_60_grid_bar <- bar_plot(cam_mid_60_fdfe,grid_local,"CAM mid")
cam_low_60_grid_bar <- bar_plot(cam_low_60_fdfe,grid_local,"CAM low")

# Extract
tpdf("grid_collections_outputs/bar_MAA_up_allgrids.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_up_60_grid_bar,
              top = "Upper Maastr. Bar Plots of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() #close the file

tpdf("grid_collections_outputs/bar_MAA_low_allgrids.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_low_60_grid_bar,
           top = "Lower Maastr. Bar Plots of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

pdf("grid_collections_outputs/bar_CAM_up_allgrids.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_up_60_grid_bar,
           top = "Upper Camp. Bar Plots of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

pdf("grid_collections_outputs/bar_CAM_mid_allgrids.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_mid_60_grid_bar,
           top = "Middle Camp. Bar Plots of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

pdf("grid_collections_outputs/bar_CAM_low_allgrids.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_low_60_grid_bar,
           top = "Lower Camp. Bar Plots of Diversity Metrics per Grid Cell") # Write the grid.arrange in the file
dev.off() # close the file

#### Simpson's Measure of Evenness per Grid Cell ####

## function for calculating Simpson's measure of evenness that can be used with the apply function
# n = total # genera in a FE in a site/time; N = total # genera in all FE at a site/time;
# S = total # FE at the site/time; D = sum(n*(n-1))/N(N-1); inversD = inverse of D value;
# final = inverseD/S

### Simpson's Measure of Evenness per Grid Cell ###
simpsons_calc <- function(data){
  data1 <- data.frame(data$"details_fdfe")

  # simpsons measure of evenness calculation:
  simp_calc <- function(data1){
    data1 <- data.frame(data1)

    n <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)
    if (length(n)>1 && max(n)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE
      N <- sum(n)
      S <- length(n)
      D = sum(n*(n-1))/(N*(N-1))
      inversD <- 1/D
      sD <- inversD/S
    } else {
      0
    }
  }

  final <- apply(data1,1,simp_calc) # 1 indicates that it applies function to rows

  data3 <- data.frame(final)
  data4 <- cbind(newColName = rownames(data3), data3)
  rownames(data4) <- 1:nrow(data4)
  colnames(data4) <- c("Site","Simp_Even_Index")

  return(data4)
}


# Calculate Simpson's measure of evenness for each grid cell in substages using above function

maa_up_60_simp_even <- simpsons_calc(maa_up_60_fe)

maa_low_60_simp_even <- simpsons_calc(maa_low_60_fe)

cam_up_60_simp_even <- simpsons_calc(cam_up_60_fe)

cam_mid_60_simp_even <- simpsons_calc(cam_mid_60_fe)

cam_low_60_simp_even <- simpsons_calc(cam_low_60_fe)

#### Create df of Simpsons for each Grid Cell through time ####

# Function to configure to columns with site names as 1st and evenness values as second, age as 3rd
config_for_loc_even_agg <- function(data, Age){
    # Create df that just take the functional richness values in each grid cell
    data_even_summ <- data

    data_even_summ <- data.frame(data_even_summ)

    colnames(data_even_summ) <- c("Site","Evenness")

    data_even_summ$Age <- Age

    return(data_even_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_simp_summ <- config_for_loc_even_agg(maa_up_60_simp_even, "MAA up")

maa_low_60_simp_summ <- config_for_loc_even_agg(maa_low_60_simp_even, "MAA low")

cam_up_60_simp_summ <- config_for_loc_even_agg(cam_up_60_simp_even, "CAM up")

cam_mid_60_simp_summ <- config_for_loc_even_agg(cam_mid_60_simp_even, "CAM mid")

cam_low_60_simp_summ <- config_for_loc_even_agg(cam_low_60_simp_even, "CAM low")
# Bind the data into a single df with 2 columns (stacked the different dfs)

```
simp_stack <-
  rbind(maa_up_60_simp_summ, maa_low_60_simp_summ, cam_up_60_simp_summ, cam_mid_60_simp_summ, cam_low_60_simp_summ)
```

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values

```
simp_table_all <- dcast(simp_stack, Site ~ Age, value.var = "Evenness")
```

# This is a table showing the number of FE in each grid cell for each substage

```
head(simp_table_all)
```

# Export the df as a csv file

```
write.csv(simp_table_all, file="grid_collections_outputs/Grid_Simpsons_through_time.csv")
```

#### Create Bar plots of Simpsons Measure of Evenness ####

# Function to create plot bar graphs of simpson's evenness for each grid cell per substage

```
bar_plot_even <- function(even_data, loc_data, substage_column, substage_name){
  colnames(even_data) <- c("Grid","Simp_Even_Index")

  ### Extract Location Information: ###
  loc_data1 <- loc_data[c("Site", substage_column, "pLat", "pLon")]
  colnames(loc_data1) <- c("Grid", substage_column, "pLat", "pLon")
  loc_data1$Lat_1 <- loc_data1$pLat

  # Next, use mutate and cut to create binned categories of lat
  loc_data2 <- loc_data1 %>% mutate(Lat_1 = cut(loc_data1$Lat_1, breaks=c(-Inf, seq(round(min(loc_data1$pLat)/5)*5, round(max(loc_data1$pLat)/5)*5, by = 5), Inf)))
}
```
### Merge the new bins with the fe df: ###

data3 <- merge(even_data, loc_data2, by = "Grid", all.x = TRUE)

# get the unique lat bins available for this substage and change colname
bins <- data.frame(unique(data3$Lat_1))
colnames(bins) <- "Lat_1"

# Create column with the sorts of labels actually want to use based on ifelse
bins$lat_bins <- with(bins, ifelse(Lat_1 == "(-Inf,30]", '23-30°N',
                              ifelse(Lat_1 == (30,35], '30-35°N',
                              ifelse(Lat_1 == (35,40], '35-40°N',
                              ifelse(Lat_1 == (40,45], '40-45°N',
                              ifelse(Lat_1 == (45,50], '45-50°N',
                              ifelse(Lat_1 == (50,55], '50-55°N',
                              ifelse(Lat_1 == (55,60], '55-60°N',
                              '60-65°N'))) )))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)

### Arrange the df based on lat bins and then create a object with the correct order: ###
data4 <- arrange(data3, Lat_1)

grids <- data4$Grid

### Plot the data using geom_col: ###
bar <- ggplot(data3, aes(x=Grid, y=Simp_Even_Index, fill=Lat_1)) +
  geom_col(color="#e9ecef") +
geom_text(aes(label=signif(Simp_Even_Index,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +

labs(title=paste(substage_name,"Simpson's Diversity Index"),x="Grid Cell", y = "Evenness") +

theme(plot.title = element_text(hjust = 0.5),

axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +

labs(fill="Paleo-Latitude") +

scale_fill_manual(values=c("red4",
"tomato3",
"darkorange3",
"goldenrod2",
"darkolivegreen4",
"aquamarine4",
"steelblue"),

labels = bin_labels)+

scale_x_discrete(limits=grids) +

ylim(0, 0.25 + max(data3$Simp_Even_Index))

return(bar)

}

# Use above function to create bar plots of Simpson's Evenness for each SS

maa_up_simp_even_barplots <- bar_plot_even(maa_up_60_simp_even,grid_local,"MAA up", "Upper Maastr.")

maa_low_simp_even_barplots <- bar_plot_even(maa_low_60_simp_even,grid_local,"MAA up", "Lower Maastr.")

cam_up_simp_even_barplots <- bar_plot_even(cam_up_60_simp_even,grid_local,"MAA up", "Upper Camp.")

cam_mid_simp_even_barplots <- bar_plot_even(cam_mid_60_simp_even,grid_local,"MAA up", "Middle Camp.")

cam_low_simp_even_barplots <- bar_plot_even(cam_low_60_simp_even,grid_local,"MAA up", "Lower Camp.")
# Make list of the barplots of simpson's evenness

simp_even_barplots <- list(maa_up_simp_even_barplots, maa_low_simp_even_barplots, cam_up_simp_even_barplots, cam_mid_simp_even_barplots, cam_low_simp_even_barplots)

# Export bar plots of the Simpson's diversity index per grid cell for each substage

pdf("grid_collections_outputs/bar_plots_SimpEven_grids.pdf", width = 50, height = 25) # Open a new pdf file

grid.arrange(grobs = simp_even_barplots, top = "Bar Plots of Simpson's Evenness per Grid Cell") # Write the grid.arrange in the file

dev.off() #close file

##### Shannon Equability Index for each grid cell####

# function to get overall shannons evenness for each substage

shannons_calc <- function(data){

  # equation for shannons calc

  shannon_equit <- function(data){

    data1 <- data.frame(data)

    data1 <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)

    if (length(data1)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE

      S <- length(which(data1!=0)) # get the numb of FE

      Si <- sum(data1) # find out how many genera total present in all FE

    }

  }

  shannon_equit(data)

}


Pvals <- data1/Si # get proportions of each FE based on dividing # genera in each by total # genera
nlPvals <- log(Pvals) # take natural log of the proportions
H <- -sum(nlPvals*Pvals) # calculate Shannon diversity index (H) by multiplying nl of p by p and summing and multiply by neg 1
LNS <- log(S) # get natural log of number of FE present
Hi <- H/LNS # calculate Shannon Equitability Index (Hi) by dividing H by nat log of S
}} else {
  0
}
}

data1 <- data.frame(data$"details_fdfe")

final <- apply(data1,1,shannon_equit) # 1 indicates that it applies function to rows

data3 <- data.frame(final)
data4 <- cbind(newColName = rownames(data3), data3) # make row name (i.e., grid cell) first column again
rownames(data4) <- 1:nrow(data4)
colnames(data4) <- c("Site","Shan_Equit_Index")

return(data4)

#
# Calculate Shannons Equability Index for each grid cell in substages using above function
maa_up_60_shan_even <- shannons_calc(maa_up_60_fe)
maa_low_60_shan_even <- shannons_calc(maa_low_60_fe)
cam_up_60_shan_even <- shannons_calc(cam_up_60_fe)
cam_mid_60_shan_even <- shannons_calc(cam_mid_60_fe)
cam_low_60_shan_even <- shannons_calc(cam_low_60_fe)

#### Create df of Shannon for each Grid Cell through time ####

# Function to configure to columns with site names as 1st and evenness values as second, age as 3rd
config_for_loc_even_agg <- function(data, Age){
  # Create df that just take the functional richness values in each grid cell
  data_even_summ <- data

  data_even_summ <- data.frame(data_even_summ)
colnames(data_even_summ) <- c("Site","Evenness")

data_even_summ$Age <- Age
return(data_even_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_shan_summ <- config_for_loc_even_agg(maa_up_60_shan_even, "MAA up")
maa_low_60_shan_summ <- config_for_loc_even_agg(maa_low_60_shan_even, "MAA low")
cam_up_60_shan_summ <- config_for_loc_even_agg(cam_up_60_shan_even, "CAM up")
cam_mid_60_shan_summ <- config_for_loc_even_agg(cam_mid_60_shan_even, "CAM mid")
cam_low_60_shan_summ <- config_for_loc_even_agg(cam_low_60_shan_even, "CAM low")

# Bind the data into a single df with 2 columns (stacked the different dfs)
shan_stack <-
  rbind(maa_up_60_shan_summ, maa_low_60_shan_summ, cam_up_60_shan_summ, cam_mid_60_shan_summ, cam_low_60_shan_summ)
# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
shan_table_all <- dcast(shan_stack, Site ~ Age, value.var = "Evenness")

# This is a table showing the number of FE in each grid cell for each substage
head(shan_table_all)

# Export the df as a csv file
write.csv(shan_table_all, file="grid_collections_outputs/Grid_Shannon_through_time.csv")

##### Create Bar plots of Shannon Equability Index #####

# Plot bar graphs of simpson's evenness for each grid cell per substage
bar_plot_shan_even <- function(even_data, loc_data, substage_column, substage_name){
  colnames(even_data) <- c("Grid","Shan_Even_Index")

  ### Extract Location Information: ###
  loc_data1 <- loc_data[,c("Site",substage_column,"pLat","pLon")]
  colnames(loc_data1) <- c("Grid",substage_column,"pLat","pLon")
  loc_data1$Lat_1 <- loc_data1$pLat                 # Add another column to use for binning lat
  loc_data2 <- loc_data1%>%mutate(Lat_1 = cut(loc_data1$Lat_1,
                                         breaks=c(-Inf, seq(round(min(loc_data1$Lat_1)/5)*5,
                                                            round(max(loc_data1$Lat_1)/5)*5, by = 5), Inf)))

  ### Merge the new bins with the fe df: ###

}
data3 <- merge(even_data, loc_data2, by = "Grid", all.x = TRUE)

# get the unique lat bins available for this substage and change colname
bins <- data.frame(unique(data3$Lat_1))
colnames(bins) <- "Lat_1"

# Create column with the sorts of labels actually want to use based on ifelse
bins$lat_bins <- with(bins, ifelse(Lat_1 == "(-Inf,30]", '23-30°N',
                                         ifelse(Lat_1 == "(30,35]", '30-35°N',
                                             ifelse(Lat_1 == "(35,40]", '35-40°N',
                                                 ifelse(Lat_1 == "(40,45]", '40-45°N',
                                                     ifelse(Lat_1 == "(45,50]", '45-50°N',
                                                         ifelse(Lat_1 == "(50,55]", '50-55°N',
                                                             ifelse(Lat_1 == "(55,60]", '55-60°N',
                                                                 '60-65°N' )))))))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)

### Arrange the df based on lat bins and then create a object with the correct order: ###
data4 <- arrange(data3, Lat_1)
grids <- data4$Grid

### Plot the data using geom_col: ###
bar <- ggplot(data3, aes(x=Grid, y=Shan_Even_Index, fill=Lat_1)) +
    geom_col(color="#e9ecef") +
geom_text(aes(label=signif(Shan_Even_Index,2)), position=position_dodge(width=0.9), vjust=-0.25, size = 2.5) +

labs(title=paste(substage_name, "Shannon Equitability Index"), x="Grid Cell", y="Evenness", fill = "Paleo-Latitude") +

theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +

scale_fill_manual(values=c("red4",
                           "tomato3",
                           "darkorange3",
                           "goldenrod2",
                           "darkolivegreen4",
                           "aquamarine4",
                           "steelblue"),
    labels = bin_labels)+

scale_x_discrete(limits=grids) +

ylim(0, 0.1 + max(data3$Shan_Even_Index))

return(bar)

#
#
#
# Plot bar plots for each substage
#

maa_up_shan_even_barplots <- bar_plot_shan_even(maa_up_60_shan_even,grid_local,"MAA up","Upper Maastr.")

maa_low_shan_even_barplots <- bar_plot_shan_even(maa_low_60_shan_even,grid_local,"MAA low","Lower Maastr.")

cam_up_shan_even_barplots <- bar_plot_shan_even(cam_up_60_shan_even,grid_local,"CAM up","Upper Camp.")

cam_mid_shan_even_barplots <- bar_plot_shan_even(cam_mid_60_shan_even,grid_local,"CAM mid","Middle Camp.")
cam_low_shan_even_barplots <- bar_plot_shan_even(cam_low_60_shan_even,grid_local,"CAM low","Lower Camp.")

# Create list of shannon bar plots
bar_list_shan <- list(maa_up_shan_even_barplots, maa_low_shan_even_barplots,
cam_up_shan_even_barplots, cam_mid_shan_even_barplots,
cam_low_shan_even_barplots)

# plot bar plots of the Shannon Equitability index per grid cell for each substage
pdf("grid_collections_outputs/bar_plots_ShanEven_grid.pdf", width = 50, height = 25) # Open a new pdf file
grid.arrange(grobs = bar_list_shan,
top = "Bar Plots of Shannon Equitability Index per Grid Cell") # Write the grid.arrange in the file
dev.off() #close file

####### ANALYSIS LATITUDE BINS AS COLLECTIONS ########

####### Create pres-absence matrices to use in mFD analysis ######

# Function for taking the substage df that have latitudinal bins (see above section on location information)
# and creating a presence-absence matrix based on bins as "collection" locations
substg_gen_latbins_pres_abs <- function(data){
  # select out the relevant columns from the df
data1 <-
  subset(data,select=c("Updated_Genus","Substage_from_Zone.Mbr","motility","life_habitat","feeding",
                      "plong","plat"))
  # create new column with lat information to mess with
data1$Lat_1 <- data1$plat
# create breaks in the df based on latitude

```r
data2 <- data1 %>% mutate(Lat_1 = cut(data1$Lat_1,
      breaks=c(-Inf, seq(round(min(data1$plat)/5)*5,
                         round(max(data1$plat)/5)*5, by = 5), Inf))
```

# subset out just the genus name and lat bin

```r
data3 <- as.data.frame(subset(data2, select=c("Updated_Genus","Lat_1")))
```

# remove duplicates

```r
data3 %>% distinct(Updated_Genus,Lat_1, .keep_all = TRUE)
```

# transform into a pres-abs matrix

```r
data4 <- dcast(data3, Lat_1~Updated_Genus, length)
```

# make the lat bin names the row names

```r
data5 <- data.frame(data4[-1], row.names = data4[,1])
```

# make all values over 1 into 1 and keep all else 0 (presence-absence)

```r
data6 <- ifelse(data5 > 0, 1, 0)
```

```r
return(data6)
```

# Create data frames for pres-abs for each substage based on latitudinal bins (use paleo-lat dfs)

```r
maa_up_60_lat_pres_ab <- substg_gen_latbins_pres_abs(maa_up_60_paleo_loc_data)
maa_low_60_lat_pres_ab <- substg_gen_latbins_pres_abs(maa_low_60_paleo_loc_data)
cam_up_60_lat_pres_ab <- substg_gen_latbins_pres_abs(cam_up_60_paleo_loc_data)
cam_mid_60_lat_pres_ab <- substg_gen_latbins_pres_abs(cam_mid_60_paleo_loc_data)
```
cam_low_60_lat_pres_ab <- substg_gen_latbins_pres_abs(cam_low_60_paleo_loc_data)

## Function to calc functional trait values

calc_fe <- function(trait_data, pres_ab_data){
  alpha_fd_fe_MAAup <- mFD::alpha.fd.fe(
    asb_sp_occ = pres_ab_data,
    sp_to_fe = trait_data,
    ind_nm = c("fred", "fored", "fviu"),
    check_input = TRUE,
    details_returned = TRUE)
}

#### Compute alpha.fd.fe from mFD for Latitudinal Bins ####

## NOTE: use the same df of trait values for entire SS overall calculated using sp.to.fe"

# Calculate functional richness, redundancy, over redundancy, and vulnerability for each substage in lat bins

# see above section for the function used here (same function used for getting alpha-beta diversity for grid cells)

maa_up_60_fe_lat <- calc_fe(maa_up_60_traits, maa_up_60_lat_pres_ab)
maa_low_60_fe_lat <- calc_fe(maa_low_60_traits, maa_low_60_lat_pres_ab)
cam_up_60_fe_lat <- calc_fe(cam_up_60_traits, cam_up_60_lat_pres_ab)
cam_mid_60_fe_lat <- calc_fe(cam_mid_60_traits, cam_mid_60_lat_pres_ab)
cam_low_60_fe_lat <- calc_fe(cam_low_60_traits, cam_low_60_lat_pres_ab)

# get a matrix of fe values for each latitudinal bin in a substage
maa_up_60_fdfe_lat <- (maa_up_60_fe_lat$"asb_fdfe")
maa_low_60_fdfe_lat <- (maa_low_60_fe_lat$"asb_fdfe")
cam_up_60_fdfe_lat <- (cam_up_60_fe_lat$"asb_fdfe")
cam_mid_60_fdfe_lat <- (cam_mid_60_fe_lat$"asb_fdfe")
cam_low_60_fdfe_lat <- (cam_low_60_fe_lat$"asb_fdfe")

# Function to fix row names (so all the same)
make_lat_names <- function(data){
  data <- data.frame(data)
  data$bins <- rownames(data)
  data$bins <- with(data, ifelse(bins == "(-Inf,30\[)", '(-Inf,30\[, '
    ifelse(bins == "(-Inf,35\[)", '(30,35\[, '
      ifelse(bins == "(30,35\[)", '(30,35\[, '
        ifelse(bins == "(35,40\[)", '(35,40\[, '
          ifelse(bins == "(-Inf,45\[)", '(40,45\[, '
            ifelse(bins == "(40,45\[)", '(40,45\[, '
              ifelse(bins == "(45,50\[)", '(45,50\[, '
                ifelse(bins == "(50,55\[)", '(50,55\[, '
                  ifelse(bins == "(55,60\[)", '(55,60\[, '
                    ifelse(bins == "(55, Inf\[)", '(55,60\[, '
                      ifelse(bins == '(60,65\[)", '(60, Inf\[, '
                        '(60, Inf\[) ))))))))))))
}
rownames(data) <- data$bins
return(data[,c("nb_sp","nb_fe","fred","fored","fvuln")])

# Use above function to make sure lat bin names are all the same when they mean the same thing
maa_up_60_fdfe_lat <- make_lat_names(maa_up_60_fdfe_lat)
maa_low_60_fdfe_lat <- make_lat_names(maa_low_60_fdfe_lat)
cam_up_60_fdfe_lat <- make_lat_names(cam_up_60_fdfe_lat)
cam_mid_60_fdfe_lat <- make_lat_names(cam_mid_60_fdfe_lat)
cam_low_60_fdfe_lat <- make_lat_names(cam_low_60_fdfe_lat)

# export results tables summarizing the number of spp, fe, and functional richness and evenness values for each substage
write.csv(maa_up_60_fdfe_lat,file="lat_bins_outputs/LatBins_MaaUP60_fe_alpha_beta_results.csv")
write.csv(maa_low_60_fdfe_lat,file="lat_bins_outputs/LatBins_MaaLOW60_fe_alpha_beta_results.csv")
write.csv(cam_up_60_fdfe_lat,file="lat_bins_outputs/LatBins_CamUP60_fe_alpha_beta_results.csv")
write.csv(cam_mid_60_fdfe_lat,file="lat_bins_outputs/LatBins_CamMID60_fe_alpha_beta_results.csv")
write.csv(cam_low_60_fdfe_lat,file="lat_bins_outputs/LatBins_CamLOW60_fe_alpha_beta_results.csv")

##### Create Bar Plots of Metrics across lat bins for all substages: ######

# Function to create series of bar plots for each substage's latitudinal bins
bar_plot_lat <- function(fe_data){
### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###

data1 <- cbind(Bin = rownames(fe_data), fe_data) # make the genus name the 1st column again
rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
data1 <- as.data.frame(data1)

i <- c(2:6)
data1[, i] <- apply(data1[, i], 2, # Specify own function within apply to make numeric
                   function(x) as.numeric(as.character(x)))

# get the unique lat bins available for this substage and change colname
bins <- data.frame(unique(data1$Bin))
colnames(bins) <- "Lat_1"

# Create column with the sorts of labels actually want to use based on ifelse
bins$lat_bins <- with(bins, ifelse(Lat_1 == "[(-Inf,30"]", '23-30°N,
                    ifelse(Lat_1 == "[(-Inf,35"]", '30-35°N,
                    ifelse(Lat_1 == "(30,35]", '30-35°N,
                    ifelse(Lat_1 == "(35,40]", '35-40°N,
                    ifelse(Lat_1 == "(40,45]", '40-45°N,
                    ifelse(Lat_1 == "(45,50]", '45-50°N,
                    ifelse(Lat_1 == "(50,55]", '50-55°N,
                    ifelse(Lat_1 == "(55,60]", '55-60°N,
                    ifelse(Lat_1 == "(55, Inf]", '55-60°N,
                    '60-65°N'))))))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)

### Plot the data using geom_col: ###

sp_bar <- ggplot(data1, aes(x=Bin, y=nb_sp, fill=Bin) +
geom_col(color="#e9ecef") +
labs(title="Generic Richness", x="Paleo-Latitude Bin", y = "Number of Genera", fill="Paleo-Latitude") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
  scale_fill_manual(values=c("red4",
                          "tomato3",
                          "darkorange3",
                          "goldenrod2",
                          "darkolivegreen4",
                          "aquamarine4",
                          "steelblue"),
                          labels = bin_labels)+
scale_x_discrete(labels= bin_labels) +
ylim(0, 5 + max(data1$nb_sp))+
geom_text(aes(label=signif(nb_sp,3)), position=position_dodge(width=0.9), vjust=-0.25)

fe_bar <- ggplot(data1, aes(x=Bin, y=nb_fe, fill=Bin)) +
  geom_col(color="#e9ecef") +
  labs(title="Functional Richness", x="Paleo-Latitude Bin", y = "Number of Functional Entities", fill="Paleo-Latitude") +
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
  scale_fill_manual(values=c("red4", # set scale of colors manually, need 5 colors since 5 lat bins
                          "tomato3",
                          "darkorange3",
                          "goldenrod2",
                          "darkolivegreen4","
"aquamarine4",
"steelblue"),
labels = bin_labels)+
scale_x_discrete(labels= bin_labels) +
ylim(0, 1 + max(data1$nb_fe)) +
geom_text(aes(label=signif(nb_fe,3)), position=position_dodge(width=0.9), vjust=-0.25)

fred_bar <- ggplot(data1, aes(x=Bin, y=fred, fill=Bin)) +
  geom_col(color="#e9ecef") +
  labs(title="Functional Redundancy (FRed)", x="Paleo-Latitude Bin", y = "FRed Value", fill="Paleo-Latitude") +
  theme(plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
  guides(fill = "none") +
  scale_fill_manual(values=c("red4",
    "tomato3",
    "darkorange3",
    "goldenrod2",
    "darkolivegreen4",
    "aquamarine4",
    "steelblue"),
    labels = bin_labels)+
scale_x_discrete(labels= bin_labels) +
  ylim(0, 0.25 + max(data1$fred)) +
  geom_text(aes(label=signif(fred,3)), position=position_dodge(width=0.9), vjust=-0.25)

fored_bar <- ggplot(data1, aes(x=Bin, y=fored, fill=Bin)) +
  geom_col(color="#e9ecef") +
labs(title="Functional Over-Redundancy",x="Paleo-Latitude Bin", y = "FORed Value", fill="Paleo-Latitude") +

theme(plot.title = element_text(hjust = 0.5),
       axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
scale_fill_manual(values=c("red4",
                        "tomato3",
                        "darkorange3",
                        "goldenrod2",
                        "darkolivegreen4",
                        "aquamarine4",
                        "steelblue"),
                        labels = bin_labels)+
scale_x_discrete(labels= bin_labels) +
ylim(0, 0.25 + max(data1$fored)) +
geom_text(aes(label=signif(fored,3)), position=position_dodge(width=0.9), vjust=-0.25)

fvuln_bar <- ggplot(data1, aes(x=Bin, y=fvuln, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Functional Vulnerability (FVuln)","Paleo-Latitude Bin", y = "FVuln Value", fill="Paleo-Latitude") +
theme(plot.title = element_text(hjust = 0.5),
       axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
scale_fill_manual(values=c("red4",
                        "tomato3",
                        "darkorange3",
                        "goldenrod2",
                        "darkolivegreen4",
                        "aquamarine4",
                        "steelblue"),
                        labels = bin_labels)+

488
"aquamarine4",
"steelblue"),
labels = bin_labels)+

scale_x_discrete(labels= bin_labels) +

ylim(0, 0.25 + max(data1$fvuln))+

geom_text(aes(label=signif(fvuln,3)), position=position_dodge(width=0.9), vjust=-0.25)

bar_list <- list(sp_bar,fe_bar,fred_bar,fored_bar,fvuln_bar)
}

# Create series of bar plots for metric values in each substage using above function

maa_up_60_lat_bar <- bar_plot_lat(maa_up_60_fdfe_lat)
maa_low_60_lat_bar <- bar_plot_lat(maa_low_60_fdfe_lat)
cam_up_60_lat_bar <- bar_plot_lat(cam_up_60_fdfe_lat)
cam_mid_60_lat_bar <- bar_plot_lat(cam_mid_60_fdfe_lat)
cam_low_60_lat_bar <- bar_plot_lat(cam_low_60_fdfe_lat)

pdf("lat_bins_outputs/LatBins_MAA_up_bars.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_up_60_lat_bar,

top = "Upper Maastr. Bar Plots of Diversity Metrics per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() #close the file

pdf("lat_bins_outputs/LatBins_MAA_low_bars.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = maa_low_60_lat_bar,

top = "Lower Maastr. Bar Plots of Diversity Metrics per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() #close the file
pdf("lat_bins_outputs/LatBins_CAM_up_bars.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_up_60_lat_bar,
              top = "Upper Camp. Bar Plots of Diversity Metrics per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() # close the file

pdf("lat_bins_outputs/LatBins_CAM_mid_bars.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_mid_60_lat_bar,
              top = "Middle Camp. Bar Plots of Diversity Metrics per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() # close the file

pdf("lat_bins_outputs/LatBins_CAM_low_bars.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = cam_low_60_lat_bar,
              top = "Lower Camp. Bar Plots of Diversity Metrics per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() # close the file

##### Create Line Plots of richness for each Latitudinal Bin in each Substage #####

# Function to create line plots of GR vs FR across Lat bins
plot_line_lat <- function(fe_data1,fe_data2,fe_data3,fe_data4,fe_data5,yaxis,legendlabel){
    # function to get the Age data as first column
df_func <- function(data){
        ### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###
data1 <- cbind(Bin = rownames(data), data) # make the genus name the 1st column again
        rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
data1 <- as.data.frame(data1)
i <- c(2:6)
data1[, i] <- apply(data1[, i], 2,            # Specify own function within apply to make numeric
                   function(x) as.numeric(as.character(x)))

return(as.data.frame(data1))
}

fe_data1 <- df_func(fe_data1)
fe_data2 <- df_func(fe_data2)
fe_data3 <- df_func(fe_data3)
fe_data4 <- df_func(fe_data4)
fe_data5 <- df_func(fe_data5)

# function to get the bin names (latitude bins) for each SS
get_bin_names <- function(data){
  # get the unique lat bins available for this substage and change colname
  bins <- data.frame(unique(data$Bin))
  colnames(bins) <- "Lat_1"

  # Create column with the sorts of labels actually want to use based on ifelse
  bins$lat_bins <-with(bins, ifelse(Lat_1 == "(-Inf,30]", '23-30°N',
                                ifelse(Lat_1 == "(-Inf,35]", '30-35°N',
                                    ifelse(Lat_1 == "(30,35]", '30-35°N',
                                        ifelse(Lat_1 == "(35,40]", '35-40°N',
                                            ifelse(Lat_1 == "(40,45]", '40-45°N',
                                                ifelse(Lat_1 == "(45,50]", '45-50°N',
                                                    ifelse(Lat_1 == "(50,55]", '50-55°N',
                                                      ifelse(Lat_1 == "(55,60]", '55-60°N',
                                                        ifelse(Lat_1 == "(60,Inf)", '60°N+', TRUE))))))))))})
ifelse(Lat_1 == "[55, Inf]", '55-60°N',
  '60-65°N'))))))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)
return(bin_labels)
}

bin_labels1 <- get_bin_names(fe_data1)
bin_labels2 <- get_bin_names(fe_data2)
bin_labels3 <- get_bin_names(fe_data3)
bin_labels4 <- get_bin_names(fe_data4)
bin_labels5 <- get_bin_names(fe_data5)

bin1_line <- ggplot(fe_data1, aes(x=Bin)) +
  geom_line(aes(y = nb_sp, color = "#D55E00",group=1)) +
  geom_line(aes(y = 5*nb_fe, color = "steelblue",group=2),linetype = "dashed") +
  geom_point(aes(y = nb_sp), color = "#D55E00") +
  geom_point(aes(y = 5*nb_fe), color="steelblue") +
  scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.2,name=yaxis)
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Upper Maastr.")) +
xlab(label="Latitude") +
```
scale_x_discrete(labels= bin_labels1) +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))

bin2_line <- ggplot(fe_data2, aes(x=Bin)) +
  geom_line(aes(y = nb_sp, color = "#D55E00",group=1)) +
  geom_line(aes(y = 5*nb_fe, color="steelblue",group=2),linetype = "dashed") +
  geom_point(aes(y = nb_sp), color = "#D55E00") +
  geom_point(aes(y = 5*nb_fe), color="steelblue") +
  scale_y_continuous(
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.2,name=yaxis)
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Lower Maastr.")) +
  xlab("Latitude") +
  scale_x_discrete(labels= bin_labels2) +
  theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))

bin3_line <- ggplot(fe_data3, aes(x=Bin)) +
  geom_line(aes(y = nb_sp, color = "#D55E00",group=1)) +
  geom_line(aes(y = 5*nb_fe, color="steelblue",group=2),linetype = "dashed") +
  geom_point(aes(y = nb_sp), color = "#D55E00") +
  geom_point(aes(y = 5*nb_fe), color="steelblue") +
  scale_y_continuous(
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.2,name=yaxis)
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Lower Maastr.")) +
  xlab("Latitude") +
  scale_x_discrete(labels= bin_labels2) +
  theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))
```
sec.axis = sec_axis(~ . * 0.2,name=yaxis)

) +
scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Upper Camp.")) +
xlab(label="Latitude") +
scale_x_discrete(labels= bin_labels3) +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))

bin4_line <- ggplot(fe_data4, aes(x=Bin)) +
geom_line(aes(y = nb_sp, color = "#D55E00",group=1)) +
geom_line(aes(y = 5*nb_fe, color="steelblue",group=2),linetype = "dashed") +
geom_point(aes(y = nb_sp), color = "#D55E00") +
geom_point(aes(y = 5*nb_fe), color="steelblue") +
scale_y_continuous(
  # Features of the first axis
  name = "# Genera",
  # Add a second axis and specify its features
  sec.axis = sec_axis(~ . * 0.2,name=yaxis)
)

+ scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Middle Camp.")) +
xlab(label="Latitude") +
scale_x_discrete(labels= bin_labels4) +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))

bin5_line <- ggplot(fe_data5, aes(x=Bin)) +
geom_line(aes(y = nb_sp, color = "#D55E00",group=1)) +
geom_line(aes(y = 5*nb_fe, color="steelblue",group=2),linetype = "dashed") +
geom_point(aes(y = nb_sp), color = "#D55E00") +
geom_point(aes(y = 5*nb_fe), color="steelblue") + 

scale_y_continuous(
    # Features of the first axis
    name = "Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.2,name=yaxis)
) +

scale_color_discrete(labels=c("Generic Richness", legendlabel)) +

labs(title=paste("Lower Camp.")) +

xlab(label="Latitude") +

scale_x_discrete(labels= bin_labels5) +

theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5))

bin_list <- list(bin1_line,bin2_line,bin3_line,bin4_line,bin5_line)
return(bin_list)

# create list of plots for changes in fr across lat in diff substages
plot_bins_gr_fr <- plot_line_lat(maa_up_60_fdfe_lat,maa_low_60_fdfe_lat,cam_up_60_fdfe_lat,
    cam_mid_60_fdfe_lat,cam_low_60_fdfe_lat,
    "# Functional Entities","Functional Richness")

pdf("lat_bins_outputs/avg_GR_FR_lat_bin_allsubstages.pdf", width = 20, height = 25) # Open a new pdf file

grid.arrange(grobs = plot_bins_gr_fr,
    top = "Average Generic and Functional Richness across Latitude") # Write the grid.arrange in the file

devo.off() #close the file
# Export table of FR and GR through time across Latitude Bins

# Function to configure to columns with site names as 1st and evenness values as second, age as 3rd
config_for_loc_rich_agg_latbins <- function(data, Age)
{
  # Create df that just take the functional richness values in each grid cell
  data_rich_summ <- data.frame(data[,c("nb_sp","nb_fe")])
  data_rich_summ$bins <- row.names(data_rich_summ)
  row.names(data_rich_summ) <- 1:nrow(data_rich_summ)
  data_rich_summ <- data_rich_summ[,c("bins","nb_sp","nb_fe")]
  data_rich_summ$Age <- Age
  return(data_rich_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_rich_summ_lat_bins <- config_for_loc_rich_agg_latbins(maa_up_60_fdfe_lat, "MAA up")
maa_low_60_rich_summ_lat_bins <- config_for_loc_rich_agg_latbins(maa_low_60_fdfe_lat, "MAA low")
cam_up_60_rich_summ_lat_bins <- config_for_loc_rich_agg_latbins(cam_up_60_fdfe_lat, "CAM up")
cam_mid_60_rich_summ_lat_bins <- config_for_loc_rich_agg_latbins(cam_mid_60_fdfe_lat, "CAM mid")
cam_low_60_rich_summ_lat_bins <- config_for_loc_rich_agg_latbins(cam_low_60_fdfe_lat, "CAM low")

# Bind the GR data into a single df with 2 columns (stacked the different dfs)
GR_stack <-
rbind(maa_up_60_rich_summ_lat_bins[,c("bins","nb_sp","Age")],maa_low_60_rich_summ_lat_bins[,c("bins","nb_sp","Age")],
cam_up_60_rich_summ_lat_bins[,c("bins","nb_sp","Age")], cam_mid_60_rich_summ_lat_bins[,c("bins","nb_sp","Age")], cam_low_60_rich_summ_lat_bins[,c("bins","nb_sp","Age")]}}

# Bind the FR data into a single df with 2 columns (stacked the different dfs)
FR_stack <- rbind(maa_up_60_rich_summ_lat_bins[,c("bins","nb_fe","Age")], maa_low_60_rich_summ_lat_bins[,c("bins","nb_fe","Age")], cam_up_60_rich_summ_lat_bins[,c("bins","nb_fe","Age")], cam_mid_60_rich_summ_lat_bins[,c("bins","nb_fe","Age")], cam_low_60_rich_summ_lat_bins[,c("bins","nb_fe","Age")])

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
GR_lat_bins_table_all <- dcast(GR_stack, bins ~ Age, value.var = "nb_sp")

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
FR_lat_bins_table_all <- dcast(FR_stack, bins ~ Age, value.var = "nb_fe")

# This are tables showing the number of genera and FE in each grid cell for each substage
head(GR_lat_bins_table_all)
head(FR_lat_bins_table_all)

# Export the dfS as a csv file
write.csv(GR_lat_bins_table_all, file="lat_bins_outputs/Lat_bins_GR_through_time.csv")
write.csv(FR_lat_bins_table_all, file="lat_bins_outputs/Lat_bins_FR_through_time.csv")

#### Calculate and create Bar Plot of Simpsons Evenness for Latitudinal Bins ####
# simpsons measure of evenness calculation function

simpsons_calc <- function(data){
  data1 <- data.frame(data$"details_fdfe")

  # simpsons measure of evenness calculation:
  simp_calc <- function(data1){
    data1 <- data.frame(data1)
    n <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)
    if (length(n)>1 && max(n)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE
      N <- sum(n)
      S <- length(n)
      D = sum(n*(n-1))/(N*(N-1))
      inversD <- 1/D
      sD <- inversD/S
    } else {
      0
    }
  }
  final <- apply(data1,1,simp_calc) # 1 indicates that it applies function to rows
  data3 <- data.frame(final)
  data4 <- cbind(newColName = rownames(data3), data3)
  rownames(data4) <- 1:nrow(data4)
  colnames(data4) <- c("Site","Simp_Even_Index")
  return(data4)
}

final <- apply(data1,1,simp_calc) # 1 indicates that it applies function to rows

data3 <- data.frame(final)
data4 <- cbind(newColName = rownames(data3), data3)
rownames(data4) <- 1:nrow(data4)
colnames(data4) <- c("Site","Simp_Even_Index")

return(data4)
}
# Calculate Simpson's measure of evenness for each lat bin in substages using above function
maa_up_60_simp_even_lat <- simpsons_calc(maa_up_60_fe_lat)
maa_low_60_simp_even_lat <- simpsons_calc(maa_low_60_fe_lat)
cam_up_60_simp_even_lat <- simpsons_calc(cam_up_60_fe_lat)
cam_mid_60_simp_even_lat <- simpsons_calc(cam_mid_60_fe_lat)
cam_low_60_simp_even_lat <- simpsons_calc(cam_low_60_fe_lat)

# plot bar graphs of simpson’s evenness for each grid cell per substage
bar_plot_even_lat <- function(maa_up_even_data, maa_low_even_data, 
cam_up_even_data,cam_mid_even_data,cam_low_even_data){
  colnames(maa_up_even_data) <- c("Bin","Simp_Even_Index")
  colnames(maa_low_even_data) <- c("Bin","Simp_Even_Index")
  colnames(cam_up_even_data) <- c("Bin","Simp_Even_Index")
  colnames(cam_mid_even_data) <- c("Bin","Simp_Even_Index")
  colnames(cam_low_even_data) <- c("Bin","Simp_Even_Index")

  # function to get the bin names (latitude bins) for each SS
  get_bin_names <- function(data){
    # get the unique lat bins available for this substage and change colname
    bins <- data.frame(unique(data$Bin))
    colnames(bins) <- "Lat_1"

    # Create column with the sorts of labels actually want to use based on ifelse
    bins$lat_bins <-with(bins, ifelse(Lat_1 == "(-Inf,30\]", '23-30°N', 
                                  ifelse(Lat_1 == "(-Inf,35\]", '30-35°N', 
                                  ifelse(Lat_1 == "(-Inf,35\]", '30-35°N',))
  
  return(bins)
}

get_bin_names(data)

ifelse(Lat_1 == "[30,35]", '30-35°N',
 ifelse(Lat_1 == "[35,40]", '35-40°N',
 ifelse(Lat_1 == "[40,45]", '40-45°N',
 ifelse(Lat_1 == "[45,50]", '45-50°N',
 ifelse(Lat_1 == "[50,55]", '50-55°N',
 ifelse(Lat_1 == "[55,60]", '55-60°N',
 ifelse(Lat_1 == "(55, Inf]", '55-60°N',
 '60-65°N')))))))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)
return(bin_labels)
}

bin_labels1 <- get_bin_names(maa_up_even_data)
bin_labels2 <- get_bin_names(maa_low_even_data)
bin_labels3 <- get_bin_names(cam_up_even_data)
bin_labels4 <- get_bin_names(cam_mid_even_data)
bin_labels5 <- get_bin_names(cam_low_even_data)

### Plot the data using geom_col: ###
maa_up_bar <- ggplot(maa_up_even_data, aes(x=Bin, y=Simp_Even_Index, fill=Bin)) +
  geom_col(color="#e9ecef") +
  labs(title="Upper Maastr. Simpson's Diversity Index",x="Paleo-Latitude Bin", y = "SME") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
  guides(fill= "none") +
  scale_fill_manual(values=c("red4",

500
maa_low_bar <- ggplot(maa_low_even_data, aes(x=Bin, y=Simp_Even_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Lower Maastr. Simpson's Diversity Index", x="Paleo-Latitude Bin", y = "SME") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill= "none") +
scale_fill_manual(values=c("red4",
                        "tomato3",
                        "darkorange3",
                        "goldenrod2",
                        "darkolivegreen4",
                        "aquamarine4",
                        "steelblue")) +
  scale_x_discrete(labels= bin_labels2) +
  ylim(0, 0.25 + max(maa_low_even_data$Simp_Even_Index)) +
  geom_text(aes(label=signif(Simp_Even_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

cam_up_bar <- ggplot(cam_up_even_data, aes(x=Bin, y=Simp_Even_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
Supplementary Materials for Purcell and Myers, 20XX

```r
labs(title="Upper Camp. Simpson's Diversity Index", x="Paleo-Latitude Bin", y = "SME") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
scale_fill_manual(values=c("red4",
                        "tomato3",
                        "darkorange3",
                        "goldenrod2",
                        "darkolivegreen4",
                        "aquamarine4",
                        "steelblue"))+

scale_x_discrete(labels= bin_labels3) +
 ylim(0, 0.25 + max(cam_up_even_data$Simp_Even_Index)) +
 geom_text(aes(label=signif(Simp_Even_Index,3)), position=position_dodge(width=0.9, vjust=-0.25))

cam_mid_bar <- ggplot(cam_mid_even_data, aes(x=Bin, y=Simp_Even_Index, fill=Bin)) +
  geom_col(color="#e9ecef") +
  labs(title="Middle Camp. Simpson's Diversity Index", x="Paleo-Latitude Bin", y = "SME") +
  theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill = "none") +
  scale_fill_manual(values=c("red4",
                            "tomato3",
                            "darkorange3",
                            "goldenrod2",
                            "darkolivegreen4",
                            "aquamarine4",
                            "steelblue"))+
  scale_x_discrete(labels= bin_labels4) +
```

502
ylim(0, 0.25 + max(cam_mid_even_data$Simp_Even_Index)) +
geom_text(aes(label=signif(Simp_Even_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

cam_low_bar <- ggplot(cam_low_even_data, aes(x=Bin, y=Simp_Even_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Lower Camp. Simpson's Diversity Index", x="Paleo-Latitude Bin", y = "SME") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill= "none") +
scale_fill_manual(values=c("red4",
                       "tomato3",
                       "darkorange3",
                       "goldenrod2",
                       "darkolivegreen4",
                       "aquamarine4",
                       "steelblue"))+
scale_x_discrete(labels= bin_labels5) +
ylim(0, 0.25 + max(cam_low_even_data$Simp_Even_Index)) +
geom_text(aes(label=signif(Simp_Even_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

bar_list <- list(maa_up_bar, maa_low_bar, cam_up_bar, cam_mid_bar, cam_low_bar)
}

simp_even_barplots_lat <-
bar_plot_even_lat(maa_up_60_simp_even_lat, maa_low_60_simp_even_lat, cam_up_60_simp_even_lat,
  cam_mid_60_simp_even_lat, cam_low_60_simp_even_lat)

# plot bar plots of the Simpson's diversity index per grid cell for each substage
pdf("lat_bins_outputs/bar_plots_SimpEven_lat_bins.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = simp_even_barplots_lat,
    top = "Bar Plots of Simpson's Evenness per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() # close file

#### Create line plots of Simpsons for Lat Bins ####

# function to compare GR and simpsons evenness in line plots for each substage
plot_line_lat_simp <- function(fe_data1, fe_data2, fe_data3, fe_data4, fe_data5,
    data1, data2, data3, data4, data5, yaxis, legendlabel){

    # function to get the Age data as first column and make relevant columns numeric
    df_func <- function(data){
        ### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###
data1 <- cbind(Bin = rownames(data), data) # make the lat bin name the 1st column again
        rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
        data1 <- as.data.frame(data1)
        i <- c(2:6)
data1[ , i] <- apply(data1[ , i], 2, function(x) as.numeric(as.character(x)))
        return(as.data.frame(data1))
    }

    # use above function on the fe df output to get lat bins as first column and make nb_fe and nb_sp numeric
    fe_data1 <- df_func(fe_data1$asb_fdfe)
    fe_data2 <- df_func(fe_data2$asb_fdfe)
    fe_data3 <- df_func(fe_data3$asb_fdfe)
    fe_data4 <- df_func(fe_data4$asb_fdfe)
fe_data5 <- df_func(fe_data5$asb_fdfe)

bind_data1 <- cbind(data1[,1],data1[,2])
bind_data2 <- cbind(data2[,1],data2[,2])
bind_data3 <- cbind(data3[,1],data3[,2])
bind_data4 <- cbind(data4[,1],data4[,2])
bind_data5 <- cbind(data5[,1],data5[,2])

# change the column names
colnames(bind_data1) <- c("Bin","GR","other")
colnames(bind_data2) <- c("Bin","GR","other")
colnames(bind_data3) <- c("Bin","GR","other")
colnames(bind_data4) <- c("Bin","GR","other")
colnames(bind_data5) <- c("Bin","GR","other")

# function to get the bin names (latitude bins) for each SS
get_bin_names <- function(data){
  # get the unique lat bins available for this substage and change colname
  bins <- data.frame(unique(data$Bin))
  colnames(bins) <- "Lat_1"

  # Create column with the sorts of labels actually want to use based on ifelse
  bins$lat_bins <- with(bins, ifelse(Lat_1 == "(-Inf,30]", '23-30°N',
                      ifelse(Lat_1 == "(-Inf,35]", '30-35°N',
                      ifelse(Lat_1 == "(30,35]", '30-35°N',
                      ifelse(Lat_1 == "(35,40]", '35-40°N',
                      ifelse(Lat_1 == "[35,40]", '35-40°N',
                      ifelse(Lat_1 == "[40,Inf)", '40°N+",
                      "lat_bins", "lat_bins", "lat_bins", "lat_bins", "lat_bins", "lat_bins")
  )
)
}
ifelse(Lat_1 == "[40,45]", '40-45°N',
    ifelse(Lat_1 == "(45,50]", '45-50°N',
        ifelse(Lat_1 == "(50,55]", '50-55°N',
            ifelse(Lat_1 == "(55,60]", '55-60°N',
                ifelse(Lat_1 == "(55, Inf]", '55-60°N',
                    '60-65°N')))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)
return(bin_labels)

bin_labels1 <- get_bin_names(bind_data1)
bin_labels2 <- get_bin_names(bind_data2)
bin_labels3 <- get_bin_names(bind_data3)
bin_labels4 <- get_bin_names(bind_data4)
bin_labels5 <- get_bin_names(bind_data5)

bin1_line <- ggplot(bind_data1, aes(x=Bin)) +
    geom_line(aes(y = GR, color = "#D55E00",group=1)) +
    geom_line(aes(y = 50*other, color="steelblue",group=2),linetype = "dashed") + # have multiplied the other variable by 50
    geom_point(aes(y = GR), color = "#D55E00") +
    geom_point(aes(y = 50*other), color="steelblue") +
    scale_y_continuous(
        # Features of the first axis
        name = "# Genera",
        # Add a second axis and specify its features
        Guides(y = guide_axis(title="# Genera", label="...")))
sec.axis = sec_axis(~ . * 0.02,name=yaxis) # scale the other y axis to match multiplication of other variable

) +

scale_color_discrete(labels=c("Generic Richness", legendlabel)) +

labs(title=paste("Upper Maastr.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +

scale_x_discrete(labels= bin_labels1)

bin2_line <- ggplot(bind_data2, aes(x=Bin)) +

geom_line(aes(y = GR, color = "#D55E00",group=1)) +
geom_line(aes(y = 50*other, color="steelblue",group=2),linetype = "dashed") + # have multiplied the other variable by 50
geom_point(aes(y = GR), color = "#D55E00") +
geom_point(aes(y = 50*other), color="steelblue") +

scale_y_continuous(

# Features of the first axis

name = "# Genera",

# Add a second axis and specify its features

sec.axis = sec_axis(~ . * 0.02,name=yaxis)# scale the other y axis to match multiplication of other variable

) +

scale_color_discrete(labels=c("Generic Richness", legendlabel)) +

labs(title=paste("Lower Maastr.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +

scale_x_discrete(labels= bin_labels2)

bin3_line <- ggplot(bind_data3, aes(x=Bin)) +

geom_line(aes(y = GR, color = "#D55E00",group=1)) +

# Features of the first axis

name = "# Genera",

# Add a second axis and specify its features

sec.axis = sec_axis(~ . * 0.02,name=yaxis)# scale the other y axis to match multiplication of other variable

) +

scale_color_discrete(labels=c("Generic Richness", legendlabel)) +

labs(title=paste("Lower Maastr.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +

scale_x_discrete(labels= bin_labels2)
geom_line(aes(y = 50*other, color="steelblue", group=2), linetype = "dashed") + # have multiplied the other variable by 50
geom_point(aes(y = GR), color = "#D55E00") +
geom_point(aes(y = 50*other), color="steelblue") +
scale_y_continuous(
  # Features of the first axis
  name = "# Genera",
  # Add a second axis and specify its features
  sec.axis = sec_axis(~ . * 0.02, name=yaxis) # scale the other y axis to match multiplication of other variable
)
+
scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Upper Camp.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(), plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels3)

bin4_line <- ggplot(bind_data4, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00", group=1)) +
  geom_line(aes(y = 50*other, color="steelblue", group=2), linetype = "dashed") + # have multiplied the other variable by 50
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 50*other), color="steelblue") +
scale_y_continuous(
  # Features of the first axis
  name = "# Genera",
  # Add a second axis and specify its features
  sec.axis = sec_axis(~ . * 0.02, name=yaxis) # scale the other y axis to match multiplication of other variable
)
+
scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Middle Camp.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels4)

bin5_line <- ggplot(bind_data5, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00",group=1)) +
  geom_line(aes(y = 50*other, color="steelblue",group=2),linetype = "dashed") # have multiplied the other variable by 50
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 50*other), color="steelblue") +
  scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.02,name=yaxis) # scale the other y axis to match multiplication of other variable
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Lower Camp.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels5)

bin_list <- list(bin1_line,bin2_line,bin3_line,bin4_line,bin5_line)
return(bin_list)
# create list of plots for changes in simpsons and gr across lat in diff substages

plot_bins_gr_simp <-
plot_line_lat_simp(maa_up_60_fe_lat, maa_low_60_fe_lat, cam_up_60_fe_lat, cam_mid_60_fe_lat, cam_low_60_fe_lat,
                   maa_up_60_simp_even_lat, maa_low_60_simp_even_lat, cam_up_60_simp_even_lat, cam_mid_60_simp_even_lat, cam_low_60_simp_even_lat, "D/S","Simpson's Measure of Evenness")

#### NOTE: I have rescaled the evenness data by multiplying by 100, ggplot not able to plot 2 scales
# but I also rescaled and the second y axis to match (by 0.01).

# Save plots of changes in simpsons and gr across latitude

pdf("lat_bins_outputs/lat_bin_GR_Simp_allsubstages.pdf", width = 20, height = 25) # Open a new pdf file
grid.arrange(grobs = plot_bins_gr_simp, top = "Average Generic Richness and Simpson's Measure of Evenness across Latitude") # Write the grid.arrange in the file
dev.off() #close the file

#### Export table of Simpsons through time across Latitude Bins ####

# Function to configure to columns with site names as 1st and evenness values as second, age as 3rd
cfg_for_loc_even_agg_latbins <- function(data, Age){
  # Create df that just take the functional evenness values in each lat bin
  data_even_summ <- data.frame(data)
  colnames(data_even_summ) <- c("bins","even")
```r
data_even_summ$Age <- Age
return(data_even_summ)
```

```r
# Use config function (above) to alter then added new column with Age

maa_up_60_simp_summ_lat_bins <- config_for_loc_even_agg_latbins(maa_up_60_simp_even_lat, "MAA up")

maa_low_60_simp_summ_lat_bins <- config_for_loc_even_agg_latbins(maa_low_60_simp_even_lat, "MAA low")

cam_up_60_simp_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_up_60_simp_even_lat, "CAM up")

cam_mid_60_simp_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_mid_60_simp_even_lat, "CAM mid")

cam_low_60_simp_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_low_60_simp_even_lat, "CAM low")

# Function to fix row names (so all the same)
make_lat_names_even <- function(data){
  data <- data.frame(data)
  data$bins <-with(data, ifelse(bins == "(-Inf,30\]", '(30,35\]',
                         ifelse(bins == "(-Inf,35\]", '(30,35\]',
                             ifelse(bins == "(30,35\]", '(30,35\]',
                               ifelse(bins == "'(35,40\]', '(35,40\]',
                                 ifelse(bins == "'(40,45\]', '(40,45\]',
                                   ifelse(bins == "'(45,50\]', '(45,50\]',
                                     ifelse(bins == "'(50,55\]', '(50,55\]',
                                       ifelse(bins == "'(55,60\]', '(55,60\]',
                                         ifelse(bins == "'(55, Inf\]", '(55,60\]',
                                           data$bins)))))))))
}
ifelse(bins == '[60,65]', '[60, Inf],
     '[60, Inf] )))))))))

return(data)

# Use above function to make sure lat bin names are all the same when they mean the same thing
maa_up_60_simp_summ_lat_bins <- make_lat_names_even(maa_up_60_simp_summ_lat_bins)
maa_low_60_simp_summ_lat_bins <- make_lat_names_even(maa_low_60_simp_summ_lat_bins)
cam_up_60_simp_summ_lat_bins <- make_lat_names_even(cam_up_60_simp_summ_lat_bins)
cam_mid_60_simp_summ_lat_bins <- make_lat_names_even(cam_mid_60_simp_summ_lat_bins)
cam_low_60_simp_summ_lat_bins <- make_lat_names_even(cam_low_60_simp_summ_lat_bins)

# Bind the GR data into a single df with 2 columns (stacked the different dfs)
simp_stack_lat_bins <- rbind(maa_up_60_simp_summ_lat_bins, maa_low_60_simp_summ_lat_bins,
                            cam_up_60_simp_summ_lat_bins, cam_mid_60_simp_summ_lat_bins,
                            cam_low_60_simp_summ_lat_bins)

# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
simp_stack_lat_bins_table_all <- dcast(simp_stack_lat_bins, bins ~ Age, value.var = "even")

# This are tables showing the Simpsons values in each lat bin for each substage
head(simp_stack_lat_bins_table_all)

# Export the df as a csv file
write.csv(simp_stack_lat_bins_table_all, file="lat_bins_outputs/Lat_bins_Simp_through_time.csv")

##### Calculate and create Bar Plot of Shannon Equitability for Latitudinal Bins #####

# Shannon equitability calculation function
shannons_calc <- function(data){
  # equation for shannons calc
  shannon_equit <- function(data){
    data1 <- data.frame(data)
    data1 <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)

    if (length(data1)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE
      S <- length(which(data1!=0)) # get the numb of FE
      Si <- sum(data1) # find out how many genera total present in all FE
      Pvals <- data1/Si # get proportions of each FE based on dividing # genera in each by total # genera
      nlPvals <- log(Pvals) # take natural log of the proportions
      H <- -sum(nlPvals*Pvals) # calculate Shannon diversity index (H) by multiplying nl of p by p and summing and multiply by neg 1
      LNS <- log(S) # get natural log of number of FE present
      Hi <- H/LNS # calculate Shannon Equitability Index (Hi) by dividing H by nat log of S
    } else {
      0
    }
  }

  data1 <- data.frame(data$"details_fdfe")

  final <- apply(data1,1,shannon_equit) # 1 indicates that it applies function to rows

  data3 <- data.frame(final)
  data4 <- cbind(newColName = rownames(data3), data3) # make row name (i.e., grid cell) first column again

rownames(data4) <- 1:nrow(data4)
colnames(data4) <- c("Site","Shan_Equit_Index")

return(data4)
}

# Calculate Shannon Equitability Index for each lat bin in substages using above function
maa_up_60_shan_even_lat <- shannons_calc(maa_up_60_fe_lat)
maa_low_60_shan_even_lat <- shannons_calc(maa_low_60_fe_lat)
cam_up_60_shan_even_lat <- shannons_calc(cam_up_60_fe_lat)
cam_mid_60_shan_even_lat <- shannons_calc(cam_mid_60_fe_lat)
cam_low_60_shan_even_lat <- shannons_calc(cam_low_60_fe_lat)

# plot bar graphs of Shannon Equitability Index for each grid cell per substage
bar_plot_equit_lat <- function(maa_up_even_data, maa_low_even_data, cam_up_even_data, cam_mid_even_data, cam_low_even_data){

colnames(maa_up_even_data) <- c("Bin","Shan_Equit_Index")
colnames(maa_low_even_data) <- c("Bin","Shan_Equit_Index")
colnames(cam_up_even_data) <- c("Bin","Shan_Equit_Index")
colnames(cam_mid_even_data) <- c("Bin","Shan_Equit_Index")
colnames(cam_low_even_data) <- c("Bin","Shan_Equit_Index")

# function to get the bin names (latitude bins) for each SS
get_bin_names <- function(data){
    # get the unique lat bins available for this substage and change colname
    bins <- data.frame(unique(data$Bin))
colnames(bins) <- "Lat_1"
# Create column with the sorts of labels actually want to use based on ifelse
bins$lat_bins <- with(bins, ifelse(Lat_1 == "(-Inf,30]\", '23-30°N',
    ifelse(Lat_1 == "(-Inf,35]\", '30-35°N',
    ifelse(Lat_1 == "[30,35]\", '30-35°N',
    ifelse(Lat_1 == "(35,40]\", '35-40°N',
    ifelse(Lat_1 == "(40,45]\", '40-45°N',
    ifelse(Lat_1 == "(45,50]\", '45-50°N',
    ifelse(Lat_1 == "(50,55]\", '50-55°N',
    ifelse(Lat_1 == "(55,60]\", '55-60°N',
    ifelse(Lat_1 == "(55, Inf]\", '55-60°N',
    '60-65°N'))))))))

# sort the vector so they're in order from lowest to highest
bin_labels <- sort(bins$lat_bins)
return(bin_labels)

bin_labels1 <- get_bin_names(maa_up_even_data)
bin_labels2 <- get_bin_names(maa_low_even_data)
bin_labels3 <- get_bin_names(cam_up_even_data)
bin_labels4 <- get_bin_names(cam_mid_even_data)
bin_labels5 <- get_bin_names(cam_low_even_data)

### Plot the data using geom_col: ###
maa_up_bar <- ggplot(maa_up_even_data, aes(x=Bin, y=Shan_Equit_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Upper Maastr. Shannon Equitability Index", x="Paleo-Latitude Bin", y = "SEI") +
theme(plot.title = element_text(hjust = 0.5),
axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill="none") +
scale_fill_manual(values=c("red4",
  "tomato3",
  "darkorange3",
  "goldenrod2",
  "darkolivegreen4",
  "aquamarine4",
  "steelblue")) +
scale_x_discrete(labels=bin_labels1) +
ylim(0, 0.25 + max(maa_up_even_data$Shan_Equit_Index)) +
gem_text(aes(label=signif(Shan_Equit_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

maa_low_bar <- ggplot(maa_low_even_data, aes(x=Bin, y=Shan_Equit_Index, fill=Bin)) +
  geom_col(color="#e9ecef") +
  labs(title="Lower Maastr. Shannon Equitability Index", x="Paleo-Latitude Bin", y = "SEI") +
  theme(plot.title = element_text(hjust = 0.5),
  axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill="none") +
scale_fill_manual(values=c("red4",
  "tomato3",
  "darkorange3",
  "goldenrod2",
  "darkolivegreen4",
  "aquamarine4",
  "steelblue")) +
scale_x_discrete(labels=bin_labels2) +
ylim(0, 0.25 + max(maa_low_even_data$Shan_Equit_Index)) +
geom_text(aes(label=signif(Shan_Equit_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

cam_up_bar <- ggplot(cam_up_even_data, aes(x=Bin, y=Shan_Equit_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Upper Camp. Shannon Equitability Index", x="Paleo-Latitude Bin", y = "SEI") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill="none") +
scale_fill_manual(values=c("red4",
                         "tomato3",
                         "darkorange3",
                         "goldenrod2",
                         "darkolivegreen4",
                         "aquamarine4",
                         "steelblue"))+
scale_x_discrete(labels= bin_labels3) +
ylim(0, 0.25 + max(cam_up_even_data$Shan_Equit_Index)) +
geom_text(aes(label=signif(Shan_Equit_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

cam_mid_bar <- ggplot(cam_mid_even_data, aes(x=Bin, y=Shan_Equit_Index, fill=Bin)) +
geom_col(color="#e9ecef") +
labs(title="Middle Camp. Shannon Equitability Index", x="Paleo-Latitude Bin", y = "SEI") +
theme(plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
guides(fill="none") +
scale_fill_manual(values=c("red4",
                         "tomato3",
                         "darkorange3",
                         "steelblue"))
```r

"goldenrod2",
"darkolivegreen4",
"aquamarine4",
"steelblue") +

scale_x_discrete(labels = bin_labels4) +

ylim(0, 0.25 + max(cam_mid_even_data$Shan_Equit_Index)) +

geom_text(aes(label=signif(Shan_Equit_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

cam_low_bar <- ggplot(cam_low_even_data, aes(x=Bin, y=Shan_Equit_Index, fill=Bin)) +

geom_col(color="#e9ecef") +

labs(title="Lower Camp. Shannon Equitability Index",x="Paleo-Latitude Bin", y = "SEI") +

theme(plot.title = element_text(hjust = 0.5),

axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +

guides(fill="none") +

scale_fill_manual(values=c("red4",

"tomato3",

"darkorange3",

"goldenrod2",

"darkolivegreen4",

"aquamarine4",

"steelblue") +

scale_x_discrete(labels = bin_labels5) +

ylim(0, 0.25 + max(cam_low_even_data$Shan_Equit_Index)) +

geom_text(aes(label=signif(Shan_Equit_Index,3)), position=position_dodge(width=0.9), vjust=-0.25)

bar_list <- list(maa_up_bar, maa_low_bar, cam_up_bar, cam_mid_bar, cam_low_bar)
}
```
shan_even_barplots_lat <- bar_plot_equit_lat(maa_up_60_shan_even_lat, maa_low_60_shan_even_lat, cam_up_60_shan_even_lat, cam_mid_60_shan_even_lat, cam_low_60_shan_even_lat)

# plot bar plots of the Simpson's diversity index per grid cell for each substage
pdf("lat_bins_outputs/bar_plots_ShanEquit_lat_bins.pdf", width = 15, height = 25) # Open a new pdf file
grid.arrange(grobs = shan_even_barplots_lat, top = "Bar Plots of Shannon Equitability per Latitudinal Bin") # Write the grid.arrange in the file
dev.off() #close file

#### Create line plots of Shannon for Lat Bins ####

# function to compare GR and shannon equitability index in line plots for each substage
plot_line_lat_shan <- function(fe_data1,fe_data2,fe_data3,fe_data4,fe_data5, data1,data2,data3,data4,data5,yaxis,legendlabel){

# function to get the Age data as first column and make relevant columns numeric
df_func <- function(data){

### Get the functional ecology information for the substage (this is a dataframe) and reconfig: ###
data1 <- cbind(Bin = rownames(data), data) # make the lat bin name the 1st column again
rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
data1 <- as.data.frame(data1)

i <- c(2:6)
data1[, i] <- apply(data1[, i], 2, # Specify own function within apply to make numeric
function(x) as.numeric(as.character(x))
return(as.data.frame(data1))
}

# use above function on the fe df output to get lat bins as first column and make nb_fe and nb_sp numeric
fe_data1 <- df_func(fe_data1$asb_fdfe)
fe_data2 <- df_func(fe_data2$asb_fdfe)
fe_data3 <- df_func(fe_data3$asb_fdfe)
fe_data4 <- df_func(fe_data4$asb_fdfe)
fe_data5 <- df_func(fe_data5$asb_fdfe)

# bind the lat bin and gr (nb_sp column) with the fr for each
bind_data1 <- cbind(fe_data1[,c(1,2)],data1[,2])
bind_data2 <- cbind(fe_data2[,c(1,2)],data2[,2])
bind_data3 <- cbind(fe_data3[,c(1,2)],data3[,2])
bind_data4 <- cbind(fe_data4[,c(1,2)],data4[,2])
bind_data5 <- cbind(fe_data5[,c(1,2)],data5[,2])

# change the column names
colnames(bind_data1) <- c("Bin","GR","other")
colnames(bind_data2) <- c("Bin","GR","other")
colnames(bind_data3) <- c("Bin","GR","other")
colnames(bind_data4) <- c("Bin","GR","other")
colnames(bind_data5) <- c("Bin","GR","other")
# function to get the bin names (latitude bins) for each SS
get_bin_names <- function(data){
  # get the unique lat bins available for this substage and change colname
  bins <- data.frame(unique(data$Bin))
  colnames(bins) <- "Lat_1"

  # Create column with the sorts of labels actually want to use based on ifelse
  bins$lat_bins <- with(bins, ifelse(Lat_1 == "(-Inf,30]\", '23-30°N',
                                ifelse(Lat_1 == "(-Inf,35]\", '30-35°N',
                                ifelse(Lat_1 == "(30,35]\", '30-35°N',
                                ifelse(Lat_1 == "(35,40]\", '35-40°N',
                                ifelse(Lat_1 == "(40,45]\", '40-45°N',
                                ifelse(Lat_1 == "(45,50]\", '45-50°N',
                                ifelse(Lat_1 == "(50,55]\", '50-55°N',
                                ifelse(Lat_1 == "(55, Inf]\", '55-60°N',
                                         '60-65°N'))) ))) ))))

  # sort the vector so they're in order from lowest to highest
  bin_labels <- sort(bins$lat_bins)
  return(bin_labels)
}

bin_labels1 <- get_bin_names(bind_data1)
bin_labels2 <- get_bin_names(bind_data2)
bin_labels3 <- get_bin_names(bind_data3)
bin_labels4 <- get_bin_names(bind_data4)
bin_labels5 <- get_bin_names(bind_data5)
bin1_line <- ggplot(bind_data1, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00", group=1)) +
  geom_line(aes(y = 50*other, color="steelblue"),group=2,linetype = "dashed") + # have multiplied the other variable by 50
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 50*other), color="steelblue") +
  scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.02,name=yaxis) # scale the other y axis to match multiplication of other variable
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Upper Maastr.")) +
  xlab(label="Latitude") +
  theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
  scale_x_discrete(labels= bin_labels1)

bin2_line <- ggplot(bind_data2, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00", group=1)) +
  geom_line(aes(y = 50*other, color="steelblue"),group=2,linetype = "dashed") + # have multiplied the other variable by 50
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 50*other), color="steelblue") +
  scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
  ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Upper Maastr.")) +
  xlab(label="Latitude") +
  theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
  scale_x_discrete(labels= bin_labels1)
sec.axis = sec_axis(~ * 0.02, name=yaxis)# scale the other y axis to match multiplication of other variable
}

) +

scale_color_discrete(labels=c("Generic Richness", legendlabel)) + labs(title=paste("Lower Maastr.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels2)

bin3_line <- ggplot(bind_data3, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00", group=1)) +
  geom_line(aes(y = 50*other, color = "steelblue"), group=2, linetype = "dashed") + # have multiplied the other variable by 50
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 50*other), color = "steelblue") +
  scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ * 0.02, name=yaxis)# scale the other y axis to match multiplication of other variable
    ) +
  scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
  labs(title=paste("Upper Camp.")) +
  xlab(label="Latitude") +
  theme(legend.title=element_blank(), plot.title = element_text(hjust = 0.5)) +
  scale_x_discrete(labels= bin_labels3)

bin4_line <- ggplot(bind_data4, aes(x=Bin)) +
  geom_line(aes(y = GR, color = "#D55E00", group=1)) +
geom_line(aes(y = 50*other, color="steelblue"),group=2,linetype = "dashed") + # have multiplied the other variable by 50
geom_point(aes(y = GR), color = "#D55E00") +
geom_point(aes(y = 50*other), color="steelblue") +
scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.02,name=yaxis) # scale the other y axis to match multiplication of other variable
)

+ scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Middle Camp.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels4)

bin5_line <- ggplot(bind_data5, aes(x=Bin)) +
    geom_line(aes(y = GR, color = "#D55E00",group=1)) +
    geom_line(aes(y = 50*other, color="steelblue"),group=2,linetype = "dashed") + # have multiplied the other variable by 50
    geom_point(aes(y = GR), color = "#D55E00") +
    geom_point(aes(y = 50*other), color="steelblue") +
    scale_y_continuous(
        # Features of the first axis
        name = "# Genera",
        # Add a second axis and specify its features
        sec.axis = sec_axis(~ . * 0.02,name=yaxis) # scale the other y axis to match multiplication of other variable
    ) +
scale_color_discrete(labels=c("Generic Richness", legendlabel)) +
labs(title=paste("Lower Camp.")) +
xlab(label="Latitude") +
theme(legend.title=element_blank(),plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels= bin_labels5)

bin_list <- list(bin1_line,bin2_line,bin3_line,bin4_line,bin5_line)
return(bin_list)

# create list of plots for changes in simpsons and gr across lat in diff substages
plot_bins_gr_shan <- plot_line_lat_shan(maa_up_60_fe_lat,maa_low_60_fe_lat,
cam_up_60_fe_lat,cam_mid_60_fe_lat,
cam_low_60_fe_lat,
maa_up_60_shan_even_lat,
maa_low_60_shan_even_lat,
cam_up_60_shan_even_lat,
cam_mid_60_shan_even_lat,
cam_low_60_shan_even_lat, "Hi","Shannon Equitability Index")

#### NOTE: I have rescaled the evenness data by multiplying by 50, ggplot not able to plot 2 scales
# but I also resealed and the second y axis to match (by 0.02).

# Save plots of changes in simpsons and gr across latitude
pdf("lat_bins_outputs/lat_bin_GR_Shan_allsubstages.pdf", width = 20, height = 25) # Open a new pdf file
grid.arrange(grobs = plot_bins_gr_shan,
Supplementary Materials for Purcell and Myers, 20XX

```r

# Write the grid.arrange in the file
dev.off() # close the file


#### Export table of Shannon through time across Latitude Bins ####

# Function to configure to columns with site names as 1st and evenness values as second, age as 3rd
cconfig_for_loc_even_agg_latbins <- function(data, Age){
  # Create df that just take the functional evenness values in each lat bin
  data_even_summ <- data.frame(data)
  colnames(data_even_summ) <- c("bins", "even")
  data_even_summ$Age <- Age
  return(data_even_summ)
}

# Use config function (above) to alter then added new column with Age
maa_up_60_shan_summ_lat_bins <- config_for_loc_even_agg_latbins(maa_up_60_shan_even_lat, "MAA up")
maa_low_60_shan_summ_lat_bins <- config_for_loc_even_agg_latbins(maa_low_60_shan_even_lat, "MAA low")
cam_up_60_shan_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_up_60_shan_even_lat, "CAM up")
cam_mid_60_shan_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_mid_60_shan_even_lat, "CAM mid")
cam_low_60_shan_summ_lat_bins <- config_for_loc_even_agg_latbins(cam_low_60_shan_even_lat, "CAM low")
```

526
# Function to fix row names (so all the same)
make_lat_names_even <- function(data) {
  data <- data.frame(data)
  data$bins <- with(data, ifelse(bins == "(-Inf,30\],\ '(Inf,30]\],\ 
    ifelse(bins == "(-Inf,35\],\ '(30,35]\],\ 
      ifelse(bins == "(30,35]\],\ '(30,35]\],\ 
        ifelse(bins == "(35,40]\],\ '(35,40]\],\ 
          ifelse(bins == "(40,45]\],\ '(40,45]\],\ 
            ifelse(bins == "(45,50]\],\ '(45,50]\],\ 
              ifelse(bins == "(50,55]\],\ '(50,55]\],\ 
                ifelse(bins == "(55,60]\],\ '(55,60]\],\ 
                  ifelse(bins == "(60, Inf]\],\ '(60, Inf]\],\ 
                    ifelse(bins == '(60, Inf]\],\ '(60, Inf]\])))))))))
  return(data)
}

# Use above function to make sure lat bin names are all the same when they mean the same thing
maa_up_60_shan_summ_lat_bins <- make_lat_names_even(maa_up_60_shan_summ_lat_bins)
maa_low_60_shan_summ_lat_bins <- make_lat_names_even(maa_low_60_shan_summ_lat_bins)
cam_up_60_shan_summ_lat_bins <- make_lat_names_even(cam_up_60_shan_summ_lat_bins)
cam_mid_60_shan_summ_lat_bins <- make_lat_names_even(cam_mid_60_shan_summ_lat_bins)
cam_low_60_shan_summ_lat_bins <- make_lat_names_even(cam_low_60_shan_summ_lat_bins)

# Bind the GR data into a single df with 2 columns (stacked the different dfs)
shan_stack_lat_bins <- rbind(maa_up_60_shan_summ_lat_bins, maa_low_60_shan_summ_lat_bins,
cam_up_60_shan_summ_lat_bins, cam_mid_60_shan_summ_lat_bins,
cam_low_60_shan_summ_lat_bins)
# Used dcast to reconfigure the data so that Age and Site would be identifiers with unique values
shan_stack_lat_bins_table_all <- dcast(shan_stack_lat_bins, bins ~ Age, value.var = "even")

# This are tables showing the Simpsons values in each lat bin for each substage
head(shan_stack_lat_bins_table_all)

# Export the df as a csv file
write.csv(shan_stack_lat_bins_table_all, file="lat_bins_outputs/Lat_bins_Shan_through_time.csv")

####### ANALYSIS USING ALL DATA TO GET VALUES FOR EACH SUBSTAGE OVERALL (NO GRIDS OR LAT BINS) #######

#Function for removing duplicate Genus-FE-Age rows, make genera row index, and randomly subsample the data (maybe)
substg_clean <- function(data){
  data1 <- as.data.frame(data)
  data2 <-distinct(data1,Updated_Genus,motility,life_habitat,feeding,.keep_all=TRUE) # delete duplicate value
  data3 <- data.frame(data2[-1], row.names = data2[,1]) # make the first column with genus names the index
  #data4 <- data3[sample(nrow(data3), 84), ] # randomly subsample to get 84 occ
  return(data3)
}

# Use above function to simplify the df to just the age, traits, grid name, and geometry
all_substgs<- substg_clean(simple60_data)
## function for grouping data by substage and removing all columns but Genus names and Grid Cell names as a pres-abs matrix

```
substg_gen_grid_all <- function(data){
  data1 <- cbind(newColName = rownames(data), data) # make the genus name the 1st column again
  rownames(data1) <- 1:nrow(data1) # make a new index for the rownames
  data2 <- as.data.frame(subset(data1)[,c(1,2)]) # subset out just the genus name and age
  colnames(data2) <- c("Updated_Genus","Age") # Give the columns names
  data3 <- dcast(data2, Age~Updated_Genus, length) # transform into a pres-abs matrix
  data4 <- data.frame(data3[-1], row.names = data3[,1]) # make the site names the row names
  return(data4)
}
```

# create df for pres-abs for all data

```
substg_pres_ab_all <- substg_gen_grid_all(all_sbstgs)
```

# function to create a simple df of trait values with all variables as factors
# need this later in the code, when calculating sp.to.fe otherwise will lose genera names

```
df_factor_simple_all <- function(data){
  data1 <- cbind(newColName = rownames(data), data) # make the genus name the 1st column again
  data2 <- as.data.frame(unclass(data1[c(1,3:5)]), stringsAsFactors = TRUE)
  data3 <- data.frame(data2[-1], row.names = data2[,1])
  return(data3)
}
```

# create simple factor variable df for all substages

```
all_data_simp <- df_factor_simple_all(all_sbstgs)
```

# summarize data using function from mFD pakage:
all_traits_summ <- mFD::sp.tr.summary(tr_cat = trait_df,
          sp_tr = all_data_simp,
          stop_if_NA = TRUE)

# check trait types
all_traits_summ$"tr_types"

# check trait type details on levels
all_traits_summ$"mod_list"

### This section uses the trait data created for each substage prior to alpha-beta FE analysis
# (which looks at "collections" of grid cells). It therefore is based on FE summary information
# for the entire SS as a whole, not divided by location.

#### Find Overall values for each SS without identifying collection locations ####

# Function to get GR, FR, Simpsons, and Shannons values for each SS overall:
rich_even_summ <- function(data_traits, Age){

  # extract vector of the number of genera in each FE for each SS overall:
  data_fe_nb_sp <- data_traits$fe_nb_sp

  # get the total # of FE in each SS, this is the Functional Richness (FR) for the SS overall
  data_nb_fe <- length(data_fe_nb_sp)

  # Extract vector of unique genera and their FE for each SS:
  data_sp_fe <- data_traits$sp_fe

  # get the total # of genera in each SS, this is the Generic Richness (GR) for the SS overall
  data_nb_g <- length(data_sp_fe)
# simpsons measure of evenness calculation:
simp_calc <- function(data1){
data1 <- data.frame(data1)

n <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)

if (length(n)>1 && max(n)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE

N <- sum(n)
S <- length(n)
D = sum(n*(n-1))/(N*(N-1))
inversD <- 1/D
sD <- inversD/S
}
else {
0
}

# get the Simpson's Measure of Evenness Value:
data_simp <- simp_calc(data_fe_nb_sp)

# equation for shannons calc
shannon_equit <- function(data){
data1 <- data.frame(data)

data1 <- data1[data1!=0] # get vector of only values greater than 0 (no empty FE)

if (length(data1)>1){ # use if statement to only run the calc on nodes with >1 FE and more than 1 genera in at least one FE

S <- length(which(data1!=0)) # get the numb of FE
Si <- sum(data1) # find out how many genera total present in all FE

S <- length(which(data1!=0)) # get the numb of FE
Si <- sum(data1) # find out how many genera total present in all FE
Pvals <- data1/Si # get proportions of each FE based on dividing # genera in each by total # genera

nlPvals <- log(Pvals) # take natural log of the proportions

H <- -sum(nlPvals*Pvals) # calculate Shannon diversity index (H) by multiplying nl of p by p and summing and multiply by neg 1

LNS <- log(S) # get natural log of number of FE present

Hi <- H/LNS # calculate Shannon Equitability Index (Hi) by dividing H by nat log of S

} else {

0

}

}

data_shan <- shannon_equit(data_fe_nb_sp)

final_df <- data.frame(firstcolumn = data_nb_g, secondcolumn= data_nb_fe, thirdcolumn= data_simp, fourthcolumn= data_shan)

rownames(final_df) <- Age

colnames(final_df) <- c("GR","FR","SME","SEI")

return(final_df)

#

# Use above function to get the overall values of GR, FR, Simpsons, and Shannons for each SS

maa_up_60_rich_even <- rich_even_summ(maa_up_60_traits,"Maa UP")

maa_low_60_rich_even <- rich_even_summ(maa_low_60_traits,"Maa LOW")

cam_up_60_rich_even <- rich_even_summ(cam_up_60_traits,"Cam UP")

cam_mid_60_rich_even <- rich_even_summ(cam_mid_60_traits,"Cam MID")

cam_low_60_rich_even <- rich_even_summ(cam_low_60_traits,"Cam LOW")
# bind the df of each SS richness and evenness from above into single DF

```r
all_rich_even <- rbind(cam_low_60_rich_even,cam_mid_60_rich_even,cam_up_60_rich_even,maa_low_60_rich_even,maa_up_60_rich_even)

write.csv(all_rich_even, file="substages_outputs/all_richness_evenness_through_time.csv")
```

##### Plot of functional and generic richness through 5 substages ######

```r
all_rich_even_age <- cbind(Age = rownames(all_rich_even), all_rich_even) # make the genus name the 1st column again
rownames(all_rich_even_age) <- 1:nrow(all_rich_even_age) # make a new index for the rownames

# plot the richness values (GR and FR) for each substage to compare
gr_fr_plot <- ggplot(all_rich_even_age, aes(x=Age)) +
  geom_line(aes(y = GR, color = "#D55E00",group=1)) +
  geom_line(aes(y = 10*FR, color="steelblue",group=2),linetype="dashed") + # multiply FR by 10 so more visible next to GR
  geom_point(aes(y = GR), color = "#D55E00") +
  geom_point(aes(y = 10*FR), color="steelblue") + # multiply FR by 10 so more visible next to GR
  scale_y_continuous(
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.1,name="# FE") # make sure scale divides y axis proportional to above multiplication
  ) +
  scale_color_discrete(labels=c("Generic Richness", "Functional Richness")) +
  labs(title=paste("Generic and Functional Richness through Time")) +
  theme(text = element_text(size = 16),legend.position="none",plot.title = element_text(hjust = 0.5)) +
```
scale_x_discrete(labels=c("Lower Camp.", "Middle Camp.","Upper Camp.","Lower Maastr.","Upper Maastr.")))}

# plot the GR vs Simp for each substage to compare
gr_simp_plot <- ggplot(all_rich_even_age, aes(x=Age)) +
geom_line(aes(y = GR, color = "#D55E00",group=1)) +
geom_line(aes(y = 500*SME, color="steelblue",group=2),linetype="dashed") + # multiply SME by 500 so more visible next to GR
geom_point(aes(y = GR), color = "#D55E00") +
geom_point(aes(y = 500*SME), color="steelblue") + # multiply SME by 500 so more visible next to GR
scale_y_continuous(
    # Features of the first axis
    name = "# Genera",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ . * 0.02,name="SME") # make sure scale divides y axis proportional to above mutiplication
) +
scale_color_discrete(labels=c("Generic Richness", "Simpson's Measure of Evenness")) +
labs(title=paste("Generic Richness and Simpson's Measure of Evenness through Time'")) +
theme(text = element_text(size = 16),legend.position="none",plot.title = element_text(hjust = 0.5)) +
scale_x_discrete(labels=c("Lower Camp.", "Middle Camp.","Upper Camp.","Lower Maastr.","Upper Maastr.")))}

# plot the GR vs Shan for each substage to compare
gr_shan_plot <- ggplot(all_rich_even_age, aes(x=Age)) +
geom_line(aes(y = GR, color = "#D55E00",group=1)) +
geom_line(aes(y = 500*SEI, color="steelblue",group=2),linetype="dashed") + # multiply SEI by 500 so more visible next to GR
geom_point(aes(y = GR), color = "#D55E00") +
geom_point(aes(y = 500*SEI), color="steelblue") + # multiply SEI by 500 so more visible next to GR
```r
scale_y_continuous(
  # Features of the first axis
  name = "Genera",

  # Add a second axis and specify its features
  sec.axis = sec_axis(~ . * 0.02,name="SEI") # make sure scale divides y axis proportional to above multiplication
)

+ scale_color_discrete(labels=c("Generic Richness", "Shannon Equitability Index")) +
  labs(title=paste("Generic Richness and Shannon Equitability Index through Time")) +
  theme(text = element_text(size = 16),legend.position="none",plot.title = element_text(hjust = 0.5)) +
  scale_x_discrete(labels=c("Lower Camp.", "Middle Camp.","Upper Camp.","Lower Maastr.","Upper Maastr."))

rich_even_plots <- list(gr_fr_plot,gr_simp_plot,gr_shan_plot)

grid.arrange(grobs = rich_even_plots,
  top = "Change in Richness and Evenness Values Through Time") # Write the grid.arrange in the file

# Save plots of changes in metrics across substages
pdf("substages_outputs/Substage_metrics_change_lineplots.pdf", width = 13, height = 16) # Open a new pdf file
grid.arrange(grobs = rich_even_plots,
  top = "Change in Richness and Evenness Values Through Time") # Write the grid.arrange in the file
dev.off() #close the file
```
### Make "bubble plot" of functional entity change in # genera (as relative abundance) per substage ###

# Function to get # genera present in each FE for a substage

fe_counts_names <- function(data){
  counts <- as.data.frame(data$"fe_nb_sp")
  names <- as.data.frame(data$details_fe$fe_codes)
  counts1 <- cbind(newColName = rownames(counts), counts) # make grid cell first column
  rownames(counts1) <- 1:nrow(counts1)
  colnames(counts1) <- c("FE","genera")

  names1 <- cbind(newColName = rownames(names), names) # make grid cell first column
  rownames(names1) <- 1:nrow(names1)
  colnames(names1) <- c("FE","names")

  count_names <- merge(counts1, names1, by ="FE", x.all=TRUE) # merge the FE genera counts with names
  count_names <- count_names[,c(3,2)] # rearrange columns
}

# Create df of # genera per FE in each substage using above function

maa_up_60_fe_counts <- fe_counts_names(maa_up_60_traits)
maa_low_60_fe_counts <- fe_counts_names(maa_low_60_traits)
cam_up_60_fe_counts <- fe_counts_names(cam_up_60_traits)
cam_mid_60_fe_counts <- fe_counts_names(cam_mid_60_traits)
cam_low_60_fe_counts <- fe_counts_names(cam_low_60_traits)
# Function to get relative abundance of genera for each FE

```r
find_rel_abund <- function(data){
  data1 <- data.frame(data[,1])
  for(i in 1:length(data[,2])){
    data1$rel_ab[i] <- data[i,2]/sum(data[,2])
  }
  colnames(data1) <- c("names","rel_ab")
  return(data1)
}
```

# Get relative abundance of genera in each FE in each substage using above function

```r
maa_up_60_fe_rel_abd <- find_rel_abund(maa_up_60_fe_counts)
maa_low_60_fe_rel_abd <- find_rel_abund(maa_low_60_fe_counts)
cam_up_60_fe_rel_abd <- find_rel_abund(cam_up_60_fe_counts)
cam_mid_60_fe_rel_abd <- find_rel_abund(cam_mid_60_fe_counts)
cam_low_60_fe_rel_abd <- find_rel_abund(cam_low_60_fe_counts)
```

# Compile above df into list to merge

```r
rel_abd_list <- list(cam_low_60_fe_rel_abd,cam_mid_60_fe_rel_abd,
                      cam_up_60_fe_rel_abd,maa_low_60_fe_rel_abd,
                      maa_up_60_fe_rel_abd)
```

# merge all the df into one (this makes sure no FE are omitted in any SS)

```r
rel_abd_list_complete <- rel_abd_list %>% reduce(full_join, by='names')
```
colnames(rel_abd_list_complete) <- c("names","CamLOW","CamMID","CamUP","MaaLOW","MaaUP")

# Function to create trait names for the plot
make_IDs <- function(data){
  # split the names column by "."
  names_split <- str_split_fixed(data$names, ".", 4)

  # split the subsequent columns by str values to get just trait IDs
  motility_str <- str_split_fixed(names_split[ ,1], "Y", 2)
  habitat_str <- str_split_fixed(names_split[ ,3], "AT", 2)
  feeding_str <- str_split_fixed(names_split[ ,4], "G", 2)

  # concatenate and capitalize the ID values into a single vector
  trait_names <- paste(toupper(motility_str[ ,2]), toupper(habitat_str[ ,2]), toupper(feeding_str[ ,2]), sep="-")

  return(trait_names)
}

# Get the trait name IDs using the above function
trait_names <- make_IDs(rel_abd_list_complete)

# Add trait names IDs to the abundance df
rel_abd_list_complete$names <- trait_names

write.csv(rel_abd_list_complete,file="substages_outputs/trait_names_rel_abundance.csv")

# Function to create a bubble plot like Foster and Twitchet 2014 Fig 3
bubble_gen_per_fe <- function(data){
    # make copy of the compiled list of # genera per FE in each substage to manipulate
    data2 <- data

    # Change all values to either 1-5 based on substage (exclude NAs)
    data2$MaaUP[data2$MaaUP >0 ] <- 5
    data2$MaaLOW[data2$MaaLOW >0 ] <- 4
    data2$CamUP[data2$CamUP >0 ] <- 3
    data2$CamMID[data2$CamMID >0 ] <- 2
    data2$CamLOW[data2$CamLOW >0 ] <- 1

    # merge the df with # genera in each substage with the df with 1-5 code numbers for all non NA cells
    data3 <- merge(data,data2,by="names",x.all=TRUE)

    # create new rows of the diff in relative abundance between two successive intervals
    data3$cl_cm <- data3$CamLOW.x-data3$CamMID.x
    data3$cm_cu <- data3$CamMID.x-data3$CamUP.x
    data3$cu_ml <- data3$CamUP.x-data3$MaaLOW.x
    data3$ml_mu <- data3$MaaLOW.x-data3$MaaUP.x

    # create threshold breaks in the columns of diff in relative abundances
    data3$TH_cl_cm<-cut(data3$cl_cm,breaks=c(-Inf,-0.05,0.05,Inf),labels=c("<=-0.05","no",">0.05"))
    data3$TH_cm_cu<-cut(data3$cm_cu,breaks=c(-Inf,-0.05,0.05,Inf),labels=c("<=-0.05","no",">0.05"))
    data3$TH_cu_ml<-cut(data3$cu_ml,breaks=c(-Inf,-0.05,0.05,Inf),labels=c("<=-0.05","no",">0.05"))
    data3$TH_ml_mu<-cut(data3$ml_mu,breaks=c(-Inf,-0.05,0.05,Inf),labels=c("<=-0.05","no",">0.05"))

    # create bubble plot of the df to show # genera (based on relative size) in each FE across substages
    r <- ggplot(data3, aes(x=names)) +
geom_point(aes(y = MaaUP.y, size=MaaUP.x, fill = TH_ml_mu),shape=21) + # add points with select color, size by # genera
geom_point(aes(y = MaaLOW.y, size=MaaLOW.x, fill = TH_cu_ml),shape=21) +
geom_point(aes(y = CamUP.y, size=CamUP.x, fill = TH_cm_cu),shape=21) +
geom_point(aes(y = CamMID.y, size=CamMID.x, fill = TH_cl_cm),shape=21) +
geom_point(aes(y = CamLOW.y, size=CamLOW.x), fill = "darkgrey",shape=21) + # put this color atrib outside aes because it will set absolutely and not follow any variable
labs(title="Change in # Genera in Functional Entities through Time") +
labs(size="Relative Abundances") + # add a legend title, but only for size (will remove color from legend)
guides(fill="none") +
xlab("Functional Entity") + ylab("Age") + # add the x and y axis labels
theme(text = element_text(size = 16),plot.title = element_text(hjust = 0.5), # make the title centered
axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), # change x axis ticks to be vert
axis.text.y = element_text(angle = 90, vjust = 0.5, hjust=0.5),
legend.position = "top",
legend.direction="horizontal") +
scale_y_discrete(limits=c("Lower Camp.", "Middle Camp.","Upper Camp.","Lower Maastr.","Upper Maastr.")) + # change the name of y axis ticks
scale_fill_manual(values = c("<=-0.05" = "#D55E00",
"no"="#E69F00",
">0.05"="#0072B2")) +
scale_size_continuous(range = c(1, 12), breaks = c(-Inf,0.001,0.02,0.25,0.5,1))

return(r)
}

# Use above function to create bubble plot of the # genera per FE in each substage
gen_fe_bubble_plot <- bubble_gen_per_fe(rel_abd_list_complete)

gen_fe_bubble_plot
# Export the bubble plot

df("substages_outputs/FE_Bubble_Plot_raw.pdf", width = 14, height = 9) # Open a new pdf file
gen_fe_bubble_plot
dev.off() # close the file
APPENDIX C. Supplementary Materials for Chapter 4

The following outlines detailed methods and results, including environmental data collection, aggregation, and interpolation, as well as niche analysis methods using the ecospat package. Detailed results tables and figures are included here.

Appendix C-1. Detailed Methods

Sedimentary data aggregation

All environmental proxy data was collected using best practices outlined in Myers et al. (2015) and as applied in Purcell et al. (2023). Lithologic information, including grain size and type, stratigraphic structures, bedding style and thickness, bioturbation level, and other notable features were recorded for each locality included. Grain size and type were converted to percentages while stratigraphic structures, thickness, and all other variables were converted into categorical code values for each unit described (Table S1). Units were aggregated based on three levels of temporal binning: (1) within ammonite biozone intervals (hereafter referred to as biozone intervals), (2) substages, and (3) stage levels (see Table S2 for explanation of variable calculations when aggregated). Not all sedimentary variables recorded and aggregated in the dataset were included in the final analysis, either because the variable is not considered to be well represented in the dataset (e.g., limestone bedding style) or if the variable was recorded to examine potential bias (e.g., bioturbation confidence). Confidence values indicate that very few sedimentary locations were based on sites with literature descriptions that could be interpreted with high confidence (Table S2), however the ratio of covered to uncovered portions of an aggregated section indicated that almost all sedimentary locations were described almost completely and did not include more than 25% covered units, which would create “gaps in the data aggregation (Table S2). Ammonite zones were selected based on the number of stratigraphic localities described, with a minimum of 15 total locations necessary to represent a stratigraphic unit. Ammonite biozone time bins are based primarily on the biostratigraphic level described in the source material itself.
Table S1. Explanation of code values used in sedimentary variables.

<table>
<thead>
<tr>
<th>Bedding Thickness</th>
<th>Avg Size Value (m)</th>
<th>Limestone Bedding Style</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laminated (&lt;1cm)</td>
<td>0.005</td>
<td>Planar</td>
<td>1</td>
</tr>
<tr>
<td>Very Thin (1-3cm)</td>
<td>0.02</td>
<td>Wavy</td>
<td>2</td>
</tr>
<tr>
<td>Thin (3-10cm)</td>
<td>0.065</td>
<td>Irregular/Rubbly</td>
<td>3</td>
</tr>
<tr>
<td>Medium (10-30cm)</td>
<td>0.2</td>
<td>Nodular</td>
<td>4</td>
</tr>
<tr>
<td>Thick (30-100cm)</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Thick (&gt;100cm)</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Siliciclastic Sedimentary Structures</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planar/indistinct/unstated</td>
<td>1</td>
</tr>
<tr>
<td>Cross Lam/Trough Crossbedding</td>
<td>2</td>
</tr>
<tr>
<td>Ripples/Wavy/Graded</td>
<td>3</td>
</tr>
<tr>
<td>Hummocky/Nodular/Scored/Chnl structures</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bioturbation</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>No mentioned explicitly</td>
<td>0.5</td>
</tr>
<tr>
<td>At least 1 burrow type mentioned</td>
<td>1</td>
</tr>
<tr>
<td>Multiple burrows/Burrows common</td>
<td>2</td>
</tr>
<tr>
<td>Burrows abundant</td>
<td>3</td>
</tr>
<tr>
<td>Bioturbated and above</td>
<td>4</td>
</tr>
<tr>
<td>Basin</td>
<td>6</td>
</tr>
</tbody>
</table>
Table S2. Explanation of each variable calculated in the final data table of sedimentary information collected from literature sources.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation of Table Computations/Rational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sedimentary Grain Percentages</td>
<td>Sum of lengths (m) of grain type/size within all units divided by total aggregation length (m). Interpretation of sediment type amounts based on stratigraphic descriptions and columns can be seen in Tables 1 and 2 in Myers et al. (2015).</td>
</tr>
<tr>
<td>Average Bed Thickness</td>
<td>Thickness separated by siliciclastic and limestone beds. Sum of all percentages of silic or limestone beds within a unit times the average bed thickness recorded. This value was then multiplied by the unit percent within the aggregated section. Finally, normalized average thickness values were summed for the total aggregated section. If no thickness value recorded in stratigraphic descriptions or displayed in stratigraphic columns from the source materials, assumed average bed thickness of 0.15m.</td>
</tr>
<tr>
<td>Siliciclastic Bed Style</td>
<td>Siliciclastic bed styles were given a value from 1 to 4 based on assumed energy level during deposition (Table S1). These code values were then multiplied by the unit percentage within the total aggregated section and summed. If siliciclastic beds were present but style was not described, the unit was given a siliciclastic bed style of 1 (planar).</td>
</tr>
<tr>
<td>Average bioturbation</td>
<td>Bioturbation within a unit was given a value from 1 to 4 based on degree of mixing, number and types of burrows, and burrow distinction (Table S1). These code values were then multiplied by the unit percentage within the total aggregated section and summed. If bioturbation was not mentioned, a value of 0.5 was given, and only if a unit was described as lacking bioturbation explicitly was the unit give a value of 0 (no bioturbation).</td>
</tr>
<tr>
<td>Limestone Bedding Style</td>
<td>Limestone bed styles were given a value from 1 to 4 based on assumed energy level during deposition (Table S1). These code values were then multiplied by the unit percentage within the total aggregated section and summed. If limestone beds were present but style was not described, the unit was given a limestone bed style of 1 (planar).</td>
</tr>
<tr>
<td>Ratio of Covered vs. Uncovered Length</td>
<td>To measure some aspect of uncertainty in unit descriptions, the ratio of aggregated unit lengths that were undescribed (usually covered units) to described unit lengths was calculated.</td>
</tr>
<tr>
<td>Limestone Bedding style Confidence</td>
<td>The relative confidence of limestone bed style measurements was quantified by applying a binary value of either confidence (0) or no confidence (1) to limestone bed style codes. These values were then multiplied by the percentage that unit represents within the aggregated section, and the resulting values were summed. Lower values closer to zero are therefore considered to have greater confidence than higher values. Confidence was only applied to units if a limestone bed style was described in the original source material directly.</td>
</tr>
<tr>
<td>Siliciclastic Bedding Style Confidence</td>
<td>The relative confidence of siliciclastic bed style measurements was quantified by applying a binary value of either confidence (0) or no confidence (1) to siliciclastic bed style codes. These values were then multiplied by the percentage that unit represents within the aggregated section, and the resulting values were summed. Lower values closer to zero are therefore considered to have greater confidence than higher values. Confidence was only applied to units if a siliciclastic bed style was described in the original source material directly.</td>
</tr>
<tr>
<td>Bioturbation Confidence</td>
<td>The relative confidence of bioturbation level measurements was quantified by applying a binary value of either confidence (0) or no confidence (1) to bioturbation level codes. These values were then multiplied by the percentage that unit represents within the aggregated section, and the resulting values were summed. Lower values closer to zero are therefore considered to have greater confidence than higher values. Confidence was only applied to units if a bioturbation level was described in the original source material directly.</td>
</tr>
<tr>
<td>Thickness Confidence</td>
<td>The relative confidence of bed thickness measurements was quantified by applying a binary value of either confidence (0) or no confidence (1) to both siliciclastic and limestone bed thicknesses. These values were then averaged and multiplied by the percentage that unit represents within the aggregated section, and the resulting values were summed. Lower values closer to zero are therefore considered to have greater confidence than higher values. Confidence was only applied to units if bed thickness was described in the original source material directly.</td>
</tr>
</tbody>
</table>
If biostratigraphic information was not included in the source material, a secondary source was used. All stratigraphic information was classified as having either high, medium, or low stratigraphic confidence. High confidence indicates that the biostratigraphic information was well-constrained (i.e., the upper and lower bounds of the zones were observed by the original authors and/or was well supported based on nearby stratigraphy); medium confidence indicates that the biostratigraphic information was partially well-constrained (either part of the section was well-constrained or the biostratigraphic age of the units was well-described in a nearby location by a secondary source); low confidence indicates that the biostratigraphic information was only roughly inferred by a secondary source (Table S3).

Table S3. Number of stratigraphic localities and their assigned stratigraphic confidence level for each ammonite biozone interval.

<table>
<thead>
<tr>
<th>Interval</th>
<th>High Confidence</th>
<th>Medium Confidence</th>
<th>Low Confidence</th>
<th>Total Localities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoploscaphites birkelundae-</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Hoploscaphites nebrascensis</td>
<td>5</td>
<td>19</td>
<td>15</td>
<td>39</td>
</tr>
<tr>
<td>Baculites clinoabatus</td>
<td>10</td>
<td>18</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>Baculites baculus-Baculites grandis</td>
<td>28</td>
<td>6</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>Baculites reesidei-Baculites eliasi</td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Baculites compressus-Baculites cuneatus</td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Didymoceras cheyennense</td>
<td>12</td>
<td>21</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>Didymoceras nebrascense-Exiteloceras jenneyi</td>
<td>13</td>
<td>7</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>Baculites reducusc-Baculites scotti</td>
<td>38</td>
<td>10</td>
<td>19</td>
<td>67</td>
</tr>
<tr>
<td>Baculites perplexus-Baculites gregoryensis</td>
<td>24</td>
<td>17</td>
<td>30</td>
<td>71</td>
</tr>
<tr>
<td>Baculites maclearni-Baculites sp. (smooth)</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>59</td>
</tr>
<tr>
<td>Baculites obtusus</td>
<td>25</td>
<td>9</td>
<td>29</td>
<td>63</td>
</tr>
<tr>
<td>Scaphites lee-Baculites sp. (weak flank ribs)</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>59</td>
</tr>
</tbody>
</table>
Figure S1. Maps of sedimentary data locations. Points are colored based on biostratigraphic confidence (red = low, yellow = medium, green = high) and labeled with numeric values. These values can be looked up in Table S4 and in the supplementary excel file (Table S27). See Table S3 for full biozone names.
Table S4. Sedimentary data references table, including numeric ID values for map reference. See Table S3 for full interval names.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Confidence</th>
<th>State</th>
<th>Reference</th>
<th>ID Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>112</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>110</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>MT</td>
<td>Robinson et al. 1964</td>
<td>70</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>68</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>70</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>118</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>129</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>166</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>3</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>4</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>5</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>11</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>64</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>208</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>210</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>216</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>217</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>218</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>33</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>221</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>51</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>131</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>132</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>133</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>77</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>75</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>38</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>MT</td>
<td>Bishop 1973</td>
<td>62</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>195</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>194</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>196</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>58</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>192</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>206</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>205</td>
</tr>
<tr>
<td>B. baculus - B. grandis</td>
<td>Low</td>
<td>MT</td>
<td>Heald 1926</td>
<td>191</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>129</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-------</td>
<td>-----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>3</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>4</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>5</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>6</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>6</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>7</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>11</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>209</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>211</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>212</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>MT</td>
<td>Daly 1984</td>
<td>213</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>214</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>215</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>216</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>217</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>218</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>219</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>ND</td>
<td>Daly 1984</td>
<td>220</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>33</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>224</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>221</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>50</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>51</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>131</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>132</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>133</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>77</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>75</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>49</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>38</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>High</td>
<td>CO</td>
<td>Nwangwu 1977</td>
<td>116</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>SD</td>
<td>Pettyjohn 1967</td>
<td>134</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Medium</td>
<td>MT</td>
<td>Bishop 1973</td>
<td>62</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>195</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>194</td>
</tr>
<tr>
<td>B. clinobatus</td>
<td>Low</td>
<td>CO</td>
<td>Mather 1928</td>
<td>196</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>89</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>221</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>77</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>49</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>38</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------</td>
<td>-------</td>
<td>----------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>NM</td>
<td>Baltz 1967</td>
<td>80</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>190</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>199</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>200</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>207</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>67</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>105</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>CO</td>
<td>Mather 1928</td>
<td>201</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>NM</td>
<td>Fitter 1958</td>
<td>157</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>58</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>222</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Medium</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>57</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>61</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>79</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>45</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>37</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>125</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>197</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>MT</td>
<td>Fuentes et al 2011</td>
<td>150</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>MT</td>
<td>Braun 1983</td>
<td>223</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>MT</td>
<td>Braun 1983</td>
<td>107</td>
</tr>
<tr>
<td>B. compressus - B. cuneatus</td>
<td>Low</td>
<td>MT</td>
<td>Braun 1983</td>
<td>56</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>109</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>226</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>130</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>119</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>117</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>162</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>172</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>166</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>8</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>12</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>19</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>22</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>22</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>25</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>28</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>60</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>64</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>240</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>----------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>242</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>241</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>243</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>40</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>99</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>99</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>100</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>101</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>102</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>103</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>90</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>90</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>91</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>92</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>104</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>93</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>94</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>95</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>96</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>97</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>1</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>10</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>14</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>17</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>21</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>23</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>26</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>29</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>190</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>198</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>207</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>High</td>
<td>SD</td>
<td>Martin et al. 2007 in Martin &amp; Parris</td>
<td>35</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>SD</td>
<td>Martin et al. 2007 in Martin &amp; Parris</td>
<td>36</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>52</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>KS</td>
<td>Bertog et al. 2007</td>
<td>84</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>248</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>249</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>47</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>151</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>158</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>188</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>139</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>SAS</td>
<td>Crockford 1949</td>
<td>149</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>SAS</td>
<td>Crockford 1949</td>
<td>142</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>164</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>165</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>168</td>
</tr>
<tr>
<td>B. maclearni - B. sp. (smooth)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>193</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>109</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>226</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>130</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>119</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>117</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>71</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>CO</td>
<td>Scott 1969</td>
<td>85</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>162</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>172</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>166</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>15</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>19</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>16</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al. 2005</td>
<td>64</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>234</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>235</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>236</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>237</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>238</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>WY</td>
<td>Minor et al. 2022</td>
<td>239</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>42</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>99</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>100</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>101</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>102</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>103</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>90</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>91</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>92</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>104</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>93</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al. 1960</td>
<td>94</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>UT</td>
<td>Maberry 1971</td>
<td>121</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>10</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>14</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>17</td>
</tr>
<tr>
<td>Interval</td>
<td>Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------</td>
<td>-------</td>
<td>----------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>21</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>23</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>190</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>207</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SD</td>
<td>Martin et al. 2007 in Martin &amp; Parris</td>
<td>34</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SD</td>
<td>Martin et al. 2007 in Martin &amp; Parris</td>
<td>35</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>52</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>KS</td>
<td>Bertog et al. 2007</td>
<td>83</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>248</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>249</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>47</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>54</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>204</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>193</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SAS</td>
<td>McNeil and Caldwell 1981</td>
<td>154</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SAS</td>
<td>McNeil and Caldwell 1981</td>
<td>155</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SAS</td>
<td>McNeil and Caldwell 1981</td>
<td>156</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>173</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>174</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>175</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>176</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>187</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>135</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>136</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>137</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>High</td>
<td>SD</td>
<td>McNeil and Caldwell 1981</td>
<td>138</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>115</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>114</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>109</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>226</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>130</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>119</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>225</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>117</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>CO</td>
<td>Scott 1969</td>
<td>85</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>162</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>172</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>166</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>8</td>
</tr>
<tr>
<td>Interval</td>
<td>Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>-------</td>
<td>--------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>12</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>15</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>22</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>25</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>28</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>31</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>WY</td>
<td>Gill 1974</td>
<td>32</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>65</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>74</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>60</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>73</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>64</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>244</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>247</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>246</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>245</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>46</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>44</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>44</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>123</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>66</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>66</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>66</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>59</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>59</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>89</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>86</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>63</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>90</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>91</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>92</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>104</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>93</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>94</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>95</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>96</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>97</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>52</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Medium</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>67</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>NM</td>
<td>Fitter 1958</td>
<td>157</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>SAS</td>
<td>Crockford 1949</td>
<td>149</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>SAS</td>
<td>Crockford 1949</td>
<td>142</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>SAS</td>
<td>McNeil and Caldwell 1981</td>
<td>153</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>173</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>174</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>Low</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>180</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>180</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>135</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>136</td>
</tr>
<tr>
<td>B. perplexus - B. gregoryensis</td>
<td>High</td>
<td>SD</td>
<td>McNeil and Caldwell 1981</td>
<td>138</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Scott 1969</td>
<td>85</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>227</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>46</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>44</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>66</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>59</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>86</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>63</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>53</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>90</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>91</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>92</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>104</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>94</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>95</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>96</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>98</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>2</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Warner 1964</td>
<td>9</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Warner 1964</td>
<td>13</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>High</td>
<td>CO</td>
<td>Warner 1964</td>
<td>18</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>CO</td>
<td>Warner 1964</td>
<td>20</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>24</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>27</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Low</td>
<td>CO</td>
<td>Warner 1964</td>
<td>30</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>67</td>
</tr>
<tr>
<td>B. reduncus - B. scotti</td>
<td>Medium</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>105</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>108</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>112</td>
</tr>
<tr>
<td>Interval</td>
<td>Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>110</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>127</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>128</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>129</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>11</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>60</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>73</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>64</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>230</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>UT</td>
<td>Minor et al 2022</td>
<td>231</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>44</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>CO</td>
<td>Porter 1940</td>
<td>124</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>33</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>221</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>50</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>51</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>131</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>132</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>133</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>77</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>38</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>58</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>58</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>222</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>222</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>79</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>45</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>37</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>High</td>
<td>ALB</td>
<td>Tsujita 1955</td>
<td>125</td>
</tr>
<tr>
<td>B. reesidei - B. eliasi</td>
<td>Low</td>
<td>MT</td>
<td>Heald 1926</td>
<td>191</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>108</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>221</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>77</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Low</td>
<td>NM</td>
<td>Caldwell 1953</td>
<td>78</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Low</td>
<td>NM</td>
<td>Caldwell 1953</td>
<td>76</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Low</td>
<td>NM</td>
<td>Caldwell 1953</td>
<td>55</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Low</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>151</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>192</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>179</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>181</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>--------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>182</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>183</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>184</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>185</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>186</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>137</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>108</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>225</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>117</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>111</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>CO</td>
<td>Scott 1969</td>
<td>85</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>CO</td>
<td>Roehler 1990</td>
<td>161</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>CO</td>
<td>Minor et al 2022</td>
<td>228</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>229</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>88</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>33</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>53</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>NM</td>
<td>Sealey and Lucas 2022</td>
<td>224</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>2</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>9</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>13</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>18</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>20</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>24</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>27</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>CO</td>
<td>Warner 1964</td>
<td>30</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>SD</td>
<td>Hanczaryk and Gallagher 2007</td>
<td>105</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>NM</td>
<td>Hutchinson 1974</td>
<td>170</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Low</td>
<td>NM</td>
<td>Fitter 1958</td>
<td>157</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>197</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>192</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>Medium</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>185</td>
</tr>
<tr>
<td>D. nebrascense - E. jenneyi</td>
<td>High</td>
<td>MAN</td>
<td>McNeil and Caldwell 1981</td>
<td>137</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>High</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>69</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------</td>
<td>-------</td>
<td>----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>High</td>
<td>CO</td>
<td>Nwangwu 1977</td>
<td>116</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Pettyjohn 1967</td>
<td>134</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Low</td>
<td>SD</td>
<td>Pettyjohn 1967</td>
<td>152</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>High</td>
<td>SD</td>
<td>Waage 1968</td>
<td>120</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>High</td>
<td>SD</td>
<td>Waage 1968</td>
<td>159</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>160</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>163</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>169</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>171</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>177</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>178</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>189</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>140</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>141</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Medium</td>
<td>SD</td>
<td>Waage 1968</td>
<td>143</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Low</td>
<td>SD</td>
<td>Waage 1968</td>
<td>144</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Low</td>
<td>SD</td>
<td>Waage 1968</td>
<td>145</td>
</tr>
<tr>
<td>H. birkelundae - H. nebrascensis</td>
<td>Low</td>
<td>SD</td>
<td>Waage 1968</td>
<td>146</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>SD</td>
<td>Waage 1968</td>
<td>147</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>WY</td>
<td>Gill and Cobban 1966</td>
<td>113</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>115</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>WY</td>
<td>Robinson et al. 1964</td>
<td>114</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>226</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Gill et al. 1970</td>
<td>130</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>WY</td>
<td>Kiteley 1979</td>
<td>71</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>CO</td>
<td>Scott 1969</td>
<td>85</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>162</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Roehler 1990</td>
<td>148</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Gill 1974</td>
<td>12</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>12</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>19</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Gill 1974</td>
<td>16</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>106</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>65</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>74</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>72</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>73</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>WY</td>
<td>Johnson et al 2005</td>
<td>64</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>232</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Minor et al 2022</td>
<td>233</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>82</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>87</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>43</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval Confidence</td>
<td>State</td>
<td>Reference</td>
<td>ID Number</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>--------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Seymour 2012</td>
<td>41</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>99</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>99</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>100</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>101</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>102</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>102</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>90</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>UT</td>
<td>Fisher et al 1960</td>
<td>91</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>CO</td>
<td>Leckie et al 1997</td>
<td>81</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Maberry 1971</td>
<td>122</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>10</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>14</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>17</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>UT</td>
<td>Chan and Newman 1991</td>
<td>21</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>190</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>NM</td>
<td>Reeside, 1924</td>
<td>207</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>KS</td>
<td>Bertog et al. 2007</td>
<td>83</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>KS</td>
<td>Bertog et al. 2007</td>
<td>48</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>248</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>High</td>
<td>WY</td>
<td>Bertog et al. 2007</td>
<td>126</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>SD</td>
<td>Bertog et al. 2007</td>
<td>249</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>164</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>165</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>165</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Medium</td>
<td>ALB</td>
<td>Stott 1967</td>
<td>167</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>203</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>202</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>54</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>54</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Johnson and Smith 1964</td>
<td>204</td>
</tr>
<tr>
<td>S. leei - B. sp. (weak flank ribs)</td>
<td>Low</td>
<td>MT</td>
<td>Fuentes et al 2011</td>
<td>150</td>
</tr>
</tbody>
</table>
Sedimentary Variable Interpolation

Interpolations were performed using ArcGIS Pro (ArcGIS Pro 3.2.0). Cross validation comparisons of root mean squared error (RMSE) values were conducted on two example variable layers to determine the most appropriate method for the highly clustered, low sample-size datasets. Only two variable layers were compared in this way to save time on computation: percent sand in the *Hoploscaphites birkelundae-Hoploscaphites nebrasensis* and *Baculites obtusus* intervals (Table S5). Kriging was not utilized given the scarcity and non-normal distribution of values within the datasets (Webster and Oliver 2007; Wang et al. 2017; Zhao et al. 2018). The *H. birkelundae-H. nebrasensis* and *B. obtusus* percent sand variables were selected for this comparison because they contain one of the lowest and highest numbers of sedimentary localities (20 and 63, respectively) and both are highly clustered, therefore potentially representing the extremes in potential sources of bias (Figure S2). Inverse Distance Weighting (IDW) interpolation of raw values was used to create raster surfaces of these test variables, using a cell size of 30km, a power of 2, and the Standard search neighborhood and all other default values from the ArcGIS Pro IDW tool.

Table S5. RMSE results for raw data and averaged/extrapolated data.

<table>
<thead>
<tr>
<th>Interval</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Data</td>
</tr>
<tr>
<td><em>H. birkelundae-H. nebrasensis</em></td>
<td>0.268</td>
</tr>
<tr>
<td><em>S. leei-B. sp. (weak flank ribs)</em></td>
<td>0.260</td>
</tr>
<tr>
<td><em>B. obtusus</em></td>
<td>0.211</td>
</tr>
</tbody>
</table>

Given that many sedimentary localities are well within 30km of each other and contain potentially variable values even within these distances, average values within a 30km range were also calculated using the Point Statistic tool in ArcGIS Pro (calculating the mean value of 30km cells within a square neighborhood of two cells). This tool was used to calculate the mean of point values within a specified neighborhood of two 30km cells and convert those mean values to a 30km raster. Cells in the raster with no point values are given NA values. This raster was then converted to vector point values.
using the Raster to Point tool, and those values were then interpolated using the IDW method as described above. Surfaces interpolated using this spatial averaging technique had much lower RMSE values than those produced using the raw data, though this result is considered highly biased given that values are smoothed prior to interpolation. However, the raster surfaces created using these mean values were considered more reasonable based on visual inspection and contain fewer interpolation artifacts caused by the raw data, such as bulls-eye patterns and sudden sharp changes in value (Figure S2). Both distance from seeps and distance from nearest shoreline were interpolated without smoothing the data first.

Figure S2. Example interpolation results for percent sand in the *H. birkelundae-H. nebrascensis* interval based on raw data values and averaged/extrapolated data values. Blue points represent stratigraphic data locations, red points represent averaged values calculated using the Point Statistics tool, and dark grey lines represent shorelines. Note the lack of artifacts in the averaged/extrapolated interpolation compared to the raw data interpolation. Red colors indicate higher percent sand, and cool colors indicate lower percent sand.
Distance from Seeps Interpolation Methods:

Distance from seep locations was included as a variable in the analysis. This variable was calculated by first documenting general seep locations based on literature sources (see reference section). These locations did not attempt to include individual seeps, but instead mark the general location of seeps within a specific interval based on literature descriptions and map locations (Figure S3). Since seeps are often highly clustered in space, within a 10 km from each other (see Metz 2010 for discussion of seep distributions), and the raster cell size selected for this analysis is 30km based on taxonomic occurrence error allowance, it was not considered necessary to include higher resolution locality information. All seep locations were given distance from seep values of zero.

Distance from seeps was determined by plotting general seep occurrence localities and creating buffer distances around these points at 60km intervals from 60 to 480km (and an additional 2000km distance) using the Buffer tool in ArcGIS Pro (using geodesic distance and dissolving features). Points were then created along that polygon’s boundary using the Generate Points Along Line tool (using the percentage option to create points every 10% of the polygon’s border). These points were given the value of the initial buffer distance used to create them in meters using the Calculate Fields tool. Once all buffer distance polygon boundaries were converted into point distributions, these points were collected into a single feature class using the Merge tool. Merged points were then used to interpolate a new surface of distance from seeps. This analysis and all others performed in ArcGIS Pro were performed using the USA Contiguous Albers Equal Area Conic projected coordinate system which preserves area values.
Figure S3. Map of hydrothermal seep locations across the WIS, colored by age. Strandlines are from the S. leei-B. sp. (weak flank ribs), which provides the largest seaway extent.
Distance from Strandlines Interpolation

Strandlines for each biostratigraphic interval were produced based on relevant publications (Gill and Cobban, 1973; Slattery et al., 2015; Cobban et al. 1994; Roberts and Krischbaum 1995) and adjusted based on the taxonomic occurrence data as needed. Vector line feature classes were produced in ArcGIS Pro and converted to point features using the Generate Points along Line tool every 120km along the shoreline, including endpoints. The Near tool was used to calculate the angle and distance from each point on the western shoreline to the nearest point along the eastern shoreline. These values were then used to calculate latitude and longitude values along the same angle at 5%, 25%, 50%, 75% and 95% of the total distance from a point on the western shoreline to the nearest point along the eastern shoreline. Vector points were created at these locations and given their distance from the closest shoreline as an integer attribute, i.e., a point created at 75% of the distance from a western shoreline point to an eastern shoreline point were given a value of 25% of the total distance between the two points (see Appendix 1-2 for code). These points were used to interpolate raster surfaces of distance from the nearest shoreline using the IDW method as described above.

Masking by Shoreline/Outcrop Polygons

All variable raster surfaces were clipped based on relevant Upper Cretaceous outcrop polygons and strandlines boundaries (see above for discussion strandline production). USA State geologic outcrop polygons with relevant formation polygons were extracted based on age (CDC 1996; Green 1992; Green et al. 1994; Green et al. 1997; Hintze et al. 2000; North Dakota Geological Survey 2001; Ross and Jorgina 1992; Tomhave and Schulz 2004; Vuke et al 2007). Outcrop polygons were given a buffer of 30km distance using the Buffer tool, to account for taxonomic locality error and provide slightly more area for analysis on the assumption that some outcrop features will not be accounted for by the outcrop polygons, especially when geographically small. The buffered outcrop polygon was merged with biozone strandline boundary polygons to create a mask layer for each interval. Interpolation raster layers for each variable in each interval were then clipped to the mask extents of their respective interval using the Extract by Mask tool.
Environmental Variables Correlation

All interpolated layers were exported as TIF files and input into R to run analysis (Figures S4). These layers were analyzed for autocorrelation using the `pairs` function in R (Figures S5). Each interval was evaluated for correlation with variables removed that showed a Pearson correlation coefficient of 0.8 or greater. All intervals included a different combination of variables following removal of variables to reduce correlation (Table S6 and S7). Shale, though present in all intervals, strongly correlated with other variables in all intervals, and was therefore removed from all analysis. All other variables were present in at least five intervals.

Comparisons with Substage and Stage

Aggregations were also performed at the substage and stage level, when reasonable, creating five total substage time bins (lower, middle, and upper Campanian, and lower and upper Maastrichtian) and a single stage-level aggregation for the Maastrichtian. A Campanian stage bin was not created, given that the Campanian spans ~ 11.5 Ma, which was not considered to represent a meaningful level of environmental aggregation.

Aggregation and interpolations for Substages and the Maastrichtian were created using the same process used for biozone intervals. When possible, the same initial interpolation layers used for biozone intervals were recycled for use in the larger temporal bins, including the *B. baculus-B. grandis* distance to seep and distance to shore raster layers for the lower Maastrichtian and Maastrichtian and all *S. leei-B. sp.* (weak flank ribs) raster layers for the lower Campanian. To make shorelines for lower temporal resolution bins, shoreline polygons were merged and boundaries were dissolved to create a single, new polygon. In the case of the upper Campanian, where this new merged polygon showed highly irregular boundaries including right angles, shorelines were smoothed manually to reduce harsh angle artifacts (Figure S4). Raster layers for each lower resolution temporal bin were then clipped to these new shoreline boundaries using the smoothed polygons.
Table S6. Table of variables that had Pearson’s correlation coefficient values less than 0.80 in each biozone. Y=Yes, N=No, and NA=not present

<table>
<thead>
<tr>
<th>Interval</th>
<th>Dist. to Shore</th>
<th>Dist. to Seeps</th>
<th>Percent Pebble</th>
<th>Percent Sand</th>
<th>Percent Shale</th>
<th>Percent Silt</th>
<th>Percent Mud</th>
<th>Percent Limestone</th>
<th>Bedding Style</th>
<th>Bioturbation</th>
<th>Bed Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. birkelundae-</em> <em>H. nebrascensis</em></td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>B. clinolobatus</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>B. baculus-B. grandis</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B. reesidei-B. eliasi</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B. compressus-B. cuneatus</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>D. cheyennense</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>D. nebrascense-<em>E. jenneyi</em></td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>B. reduncus-B. scotti</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>B. perplexus-B. gregoryensis</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B. maclearni-B. sp. (smooth)</td>
<td>Y</td>
<td>NA</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B. obtusus</td>
<td>Y</td>
<td>NA</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>S. leei-B. sp. (weak flank ribs)</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Table S7. Table of variables used in temporal comparison of niche overlap, indicated with an asterisk. Y=less than 0.80 Spearman’s correlation with another variable, N=greater than or equal to 0.80 Spearman’s correlation with another variable, and NA=not present in interval.

<table>
<thead>
<tr>
<th></th>
<th>Dist. to Shore</th>
<th>Dist. to Seeps</th>
<th>Percent Pebble</th>
<th>Percent Sand</th>
<th>Percent Shale</th>
<th>Percent Silt</th>
<th>Percent Mud</th>
<th>Percent Limestone</th>
<th>Bedding Style</th>
<th>Bioturbation</th>
<th>Bed Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upper Campanian</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reesidei-B. eliasi</td>
<td>Y*</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
</tr>
<tr>
<td>B. compressus-B. cuneatus</td>
<td>Y*</td>
<td>N</td>
<td>Y</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
<td>Y</td>
</tr>
<tr>
<td>D. nebrascense-E. jenneyi</td>
<td>Y*</td>
<td>Y</td>
<td>NA</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
<td>N</td>
</tr>
<tr>
<td><strong>Middle Campanian</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. reduncus-B. scotti</td>
<td>Y*</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>N</td>
</tr>
<tr>
<td>B. maclearni-B. sp. (smooth)</td>
<td>Y*</td>
<td>Y</td>
<td>NA</td>
<td>Y*</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
<td>Y</td>
<td>N</td>
<td>Y*</td>
<td>Y*</td>
</tr>
</tbody>
</table>
Stratigraphic Database:

Raw stratigraphic data, collected using the methods described above, can be found in the separate excel table (Table S28).
Results

Figure S4. Raster map layers for each interval.

- **Dist_to_Shore**
- **Percent_Pebble**
- **Percent_Sand**
- **Percent_Shale**
- **Percent_Silt**
- **Percent_Mud**
- **Percent_Limestone**
- **Bedding_STYLE**
- **Bioturbation**
- **Avg_Bed_Thickness**

H. birkelundae—
H. nebrascensis
Environmental Layers
Supplementary Materials for Purcell and Myers, 20XX

Environmental Layers:

- Dist_to_Shore
- Percent_Pebble
- Percent_Sand
- Percent_Shale
- Percent_Silt
- Percent_Mud
- Percent_Limestone
- Bedding_Style
- Bioturbation
- Avg_Bed_Thickness

B. maclearni-
B. sp. (smooth)
Supplementary Materials for Purcell and Myers, 20XX

District to Shore

Percent_Pebble

Percent_Sand

Percent_Shale

Percent_Silt

Percent_Mud

Percent_Limestone

Bedding Style

Bioturbation

Avg_Bed_Thickness

Maastrichtian Environmental Layers
Figure S5. Correlation charts for environmental raster layers with variables removed to ensure all Pearson’s Correlation Coefficients were less than 0.80.

H. birkelundae–H. nebrascensis Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
B. clinobatus Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
### B. baculus - B. grandis Correlation Comparisons (>0.80 Pearson's Corr Coef removed)

<table>
<thead>
<tr>
<th></th>
<th>0e+00</th>
<th>4e+05</th>
<th>0.0</th>
<th>0.6</th>
<th>0.00</th>
<th>0.15</th>
<th>0.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out_to_Shore</td>
<td>0.59</td>
<td>0.10</td>
<td>0.55</td>
<td>0.47</td>
<td>0.10</td>
<td>0.31</td>
<td>0.42</td>
<td>0.31</td>
</tr>
<tr>
<td>Out_to_Seed</td>
<td>0.74</td>
<td>0.78</td>
<td>0.14</td>
<td>0.39</td>
<td>0.72</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent_Pebble</td>
<td>0.35</td>
<td>0.10</td>
<td>0.74</td>
<td>0.31</td>
<td>0.34</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent_Sand</td>
<td>0.32</td>
<td>0.31</td>
<td>0.00</td>
<td>0.53</td>
<td>0.41</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent_Silt</td>
<td>0.16</td>
<td>0.65</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent_Mud</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent_Limestone</td>
<td>0.35</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioluminescence</td>
<td>0.44</td>
<td></td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedrock</td>
<td></td>
<td></td>
<td></td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>
B. reesidei–B. eliasi Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
B. compressus–B. cuneatus Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
D. cheyennense Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
D. nebrascense–E. jenneyi Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
B. perplexus–B. gregoryensis Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
B. maclearni–B. sp. (smooth) Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
S. leei–B. sp. (weak flank ribs) Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Maastrichtian Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Upper Maastrichtian Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Lower Maastrichtian Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Supplementary Materials for Purcell and Myers, 20XX

Upper Campanian Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Lower Campanian Correlation Comparisons (>0.80 Pearson's Corr Coef removed)
Figure S6. Overlap comparison plots of species’ niches in different temporal bins. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the longer interval niche has expanded beyond the biozone.)
interval niche), and green represents regions of niche unfilling (i.e., where the longer interval niche has moved away from the biozone interval niche). Red lines represent the extent of environmental space realized within the longer interval, green lines represent the extent of environmental space realized within the biozone interval, and the red arrow indicates the shift of the niche centroid between the two species. Non-analogue conditions are indicated by regions where the extent of environmental parameters in the different intervals do not overlap.

Figure S7. Overlap comparison plots of ingroup *Inoceramus* species' niches. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the second listed species has expanded beyond the niche of the first listed species), and green represents regions of niche unfilling (i.e., where the second listed species niche has moved away from space occupied by the first listed species). Red lines represent the extent of environmental space realized within the interval and the red arrow indicates the shift of the niche centroid between the two species.
Figure S8a. Overlap comparison plots of comparison species’ niches. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the second listed species has expanded beyond the niche of the first listed
species), and green represents regions of niche unfilling (i.e., where the second listed species niche has moved away from space occupied by the first listed species). Red lines represent the extent of environmental space realized within the interval and the red arrow indicates the shift of the niche centroid between the two species.

Figure S8b. Overlap comparison plots of comparison species’ niches. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the second listed species has expanded beyond the niche of the first listed species), and green represents regions of niche unfilling (i.e., where the second listed species niche has moved away from space occupied by the first listed species). Red lines represent the extent of environmental space realized within the interval and the red arrow indicates the shift of the niche centroid between the two species.
Figure S9. Overlap comparison plots of *Inoceramus* species’ niches and the *Inoceramus* genus niche. Blue represents regions of niche stability, red regions of niche expansion (i.e., where the genus has expanded beyond the species), and green represents regions of niche unfilling (i.e., where the genus niche has moved away from space occupied by the species). Red lines represent the extent of environmental space realized within the interval and the red arrow indicates the shift of the niche centroid between the two species.
Table S8. Niche overlap values for temporal bin comparisons including I and D overlap values, and niche expansion, stability, and unfilling.

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion</th>
<th>Stability</th>
<th>Unfilling</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.31</td>
<td>0.44</td>
<td>0.09</td>
<td>0.91</td>
<td>0.24</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.68</td>
<td>0.84</td>
<td>0.12</td>
<td>0.88</td>
<td>0.15</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.41</td>
<td>0.58</td>
<td>0.42</td>
<td>0.58</td>
<td>0.05</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.43</td>
<td>0.59</td>
<td>0.18</td>
<td>0.82</td>
<td>0.16</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.43</td>
<td>0.63</td>
<td>0.37</td>
<td>0.63</td>
<td>0.01</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. compressus-B. cuneatus to Upper Campanian</td>
<td>0.43</td>
<td>0.62</td>
<td>0.29</td>
<td>0.71</td>
<td>0.00</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.41</td>
<td>0.62</td>
<td>0.01</td>
<td>0.99</td>
<td>0.14</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.49</td>
<td>0.62</td>
<td>0.38</td>
<td>0.62</td>
<td>0.19</td>
</tr>
<tr>
<td>I. azerbaydjanensis</td>
<td>B. maclearni-B. sp. (smooth) to Mid Campanian</td>
<td>0.09</td>
<td>0.27</td>
<td>0.69</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.55</td>
<td>0.70</td>
<td>0.19</td>
<td>0.81</td>
<td>0.16</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.46</td>
<td>0.63</td>
<td>0.00</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. perplexus-B. gregoryensis to Mid Campanian</td>
<td>0.018*</td>
<td>0.154</td>
<td>0.545</td>
<td>0.545</td>
<td>0.543</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.224</td>
<td>0.253</td>
<td>0.756</td>
<td>0.756</td>
<td>0.547</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.297</td>
<td>0.307</td>
<td>0.589</td>
<td>0.589</td>
<td>0.527</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.273</td>
<td>0.535</td>
<td>0.381</td>
<td>0.381</td>
<td>0.932</td>
</tr>
<tr>
<td>L. subundata</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.267</td>
<td>0.267</td>
<td>0.339</td>
<td>0.339</td>
<td>0.489</td>
</tr>
<tr>
<td>L. subundata</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.896</td>
<td>0.854</td>
<td>0.870</td>
<td>0.870</td>
<td>0.154</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.511</td>
<td>0.581</td>
<td>0.697</td>
<td>0.697</td>
<td>0.685</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.321</td>
<td>0.367</td>
<td>0.158</td>
<td>0.158</td>
<td>0.790</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Lower Maastrichtian</td>
<td><strong>0.014</strong>*</td>
<td><strong>0.012</strong>*</td>
<td>0.212</td>
<td>0.212</td>
<td><strong>0.006</strong>*</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>0.275</td>
<td>0.579</td>
<td>0.798</td>
<td>0.798</td>
<td>0.429</td>
</tr>
</tbody>
</table>
Table S9. Results of greater niche equivalency test for temporal bin comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.413</td>
<td>0.463</td>
<td>0.190</td>
<td>0.190</td>
<td>0.812</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenneyi vs. Upper Campanian</td>
<td>0.008*</td>
<td>0.002*</td>
<td>0.208</td>
<td>0.208</td>
<td>0.275</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.495</td>
<td>0.549</td>
<td>0.729</td>
<td>0.729</td>
<td>0.220</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenneyi vs Upper Campanian</td>
<td>0.140</td>
<td>0.255</td>
<td>0.194</td>
<td>0.194</td>
<td>0.709</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.493</td>
<td>0.495</td>
<td>0.739</td>
<td>0.739</td>
<td>0.166</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. compressus-B. cuneatus vs. Upper Campanian</td>
<td>0.820</td>
<td>0.846</td>
<td>0.904</td>
<td>0.904</td>
<td>0.066</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.064</td>
<td>0.114</td>
<td>0.040*</td>
<td>0.040*</td>
<td>0.709</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenneyi vs Upper Campanian</td>
<td>0.054</td>
<td>0.140</td>
<td>0.415</td>
<td>0.415</td>
<td>0.397</td>
</tr>
<tr>
<td>I. azerbaydjanensis</td>
<td>B. maclearni-B. sp. (smooth) vs. Mid Campanian</td>
<td>0.966</td>
<td>0.970</td>
<td>0.984</td>
<td>0.984</td>
<td>0.774</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.142</td>
<td>0.178</td>
<td>0.293</td>
<td>0.293</td>
<td>0.527</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.154</td>
<td>0.311</td>
<td>0.068</td>
<td>0.068</td>
<td>0.415</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. perplexus-B. gregoryensis to Mid Campanian</td>
<td>0.018*</td>
<td>0.154</td>
<td>0.545</td>
<td>0.545</td>
<td>0.543</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.224</td>
<td>0.253</td>
<td>0.756</td>
<td>0.756</td>
<td>0.547</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.297</td>
<td>0.307</td>
<td>0.589</td>
<td>0.589</td>
<td>0.527</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.273</td>
<td>0.535</td>
<td>0.381</td>
<td>0.381</td>
<td>0.932</td>
</tr>
<tr>
<td>L. subundata</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.267</td>
<td>0.267</td>
<td>0.339</td>
<td>0.339</td>
<td>0.489</td>
</tr>
<tr>
<td>L. subundata</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.896</td>
<td>0.854</td>
<td>0.870</td>
<td>0.870</td>
<td>0.154</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.511</td>
<td>0.581</td>
<td>0.697</td>
<td>0.697</td>
<td>0.685</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.321</td>
<td>0.367</td>
<td>0.158</td>
<td>0.158</td>
<td>0.790</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Lower Maastrichtian</td>
<td>0.014*</td>
<td>0.012*</td>
<td>0.212</td>
<td>0.212</td>
<td>0.006*</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>0.275</td>
<td>0.579</td>
<td>0.798</td>
<td>0.798</td>
<td>0.429</td>
</tr>
</tbody>
</table>
Table S10. Results of lower niche equivalency test for temporal bin comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.621</td>
<td>0.575</td>
<td>0.818</td>
<td>0.818</td>
<td>0.190</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.988</td>
<td>0.996</td>
<td>0.800</td>
<td>0.800</td>
<td>0.715</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.495</td>
<td>0.429</td>
<td>0.273</td>
<td>0.273</td>
<td>0.739</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.892</td>
<td>0.766</td>
<td>0.810</td>
<td>0.810</td>
<td>0.313</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.499</td>
<td>0.497</td>
<td>0.261</td>
<td>0.261</td>
<td>0.870</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. compressus-B. cuneatus to Upper Campanian</td>
<td>0.174</td>
<td>0.170</td>
<td>0.098</td>
<td>0.098</td>
<td>0.930</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.940</td>
<td>0.886</td>
<td>0.968</td>
<td>0.968</td>
<td>0.295</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.944</td>
<td>0.832</td>
<td>0.565</td>
<td>0.565</td>
<td>0.607</td>
</tr>
<tr>
<td>I. azerbaydjanensis</td>
<td>B. maclearni-B. sp. (smooth) to Mid Campanian</td>
<td><strong>0.034</strong></td>
<td><strong>0.034</strong></td>
<td><strong>0.020</strong></td>
<td><strong>0.020</strong></td>
<td>0.244</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.842</td>
<td>0.822</td>
<td>0.671</td>
<td>0.671</td>
<td>0.489</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.808</td>
<td>0.675</td>
<td>0.930</td>
<td>0.930</td>
<td>0.587</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. perplexus-B. gregoryensis to Mid Campanian</td>
<td>0.990</td>
<td>0.850</td>
<td>0.479</td>
<td>0.479</td>
<td>0.431</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.754</td>
<td>0.709</td>
<td>0.196</td>
<td>0.196</td>
<td>0.421</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.701</td>
<td>0.697</td>
<td>0.435</td>
<td>0.435</td>
<td>0.481</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.750</td>
<td>0.461</td>
<td>0.647</td>
<td>0.647</td>
<td>0.058</td>
</tr>
<tr>
<td>L. subundata</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.693</td>
<td>0.691</td>
<td>0.639</td>
<td>0.639</td>
<td>0.411</td>
</tr>
<tr>
<td>L. subundata</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.104</td>
<td>0.148</td>
<td>0.134</td>
<td>0.134</td>
<td>1.000</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.501</td>
<td>0.429</td>
<td>0.250</td>
<td>0.250</td>
<td>0.343</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.749</td>
<td>0.689</td>
<td>0.832</td>
<td>0.832</td>
<td>0.232</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Lower Maastrichtian</td>
<td>0.982</td>
<td>0.990</td>
<td>0.806</td>
<td>0.806</td>
<td>0.994</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>0.719</td>
<td>0.423</td>
<td>0.204</td>
<td>0.204</td>
<td>0.581</td>
</tr>
</tbody>
</table>
Table S11. Results of greater niche similarity test for temporal bin comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.243</td>
<td>0.241</td>
<td>0.171</td>
<td>0.171</td>
<td>0.147</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenneyi vs. Upper Campanian</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.022*</td>
<td>0.022*</td>
<td>0.069</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.034*</td>
<td>0.021*</td>
<td>0.080</td>
<td>0.080</td>
<td>0.067</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenneyi vs Upper Campanian</td>
<td>0.059</td>
<td>0.045*</td>
<td>0.133</td>
<td>0.133</td>
<td>0.078</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.113</td>
<td>0.063</td>
<td>0.205</td>
<td>0.205</td>
<td>0.013*</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. compressus-B. cuneatus vs. Upper Campanian</td>
<td>0.133</td>
<td>0.113</td>
<td>0.166</td>
<td>0.166</td>
<td>0.008*</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.105</td>
<td>0.023*</td>
<td>0.003*</td>
<td>0.003*</td>
<td>0.190</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenneyi vs Upper Campanian</td>
<td>0.003*</td>
<td>0.001*</td>
<td>0.126</td>
<td>0.126</td>
<td>0.177</td>
</tr>
<tr>
<td>I. azerbaydjanensis</td>
<td>B. maclearni-B. sp. (smooth) vs. Mid Campanian</td>
<td>0.165</td>
<td>0.122</td>
<td>0.161</td>
<td>0.161</td>
<td>0.190</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi vs. Upper Campanian</td>
<td>0.014</td>
<td>0.016</td>
<td>0.191</td>
<td>0.191</td>
<td>0.016*</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti vs. Mid Campanian</td>
<td>0.028</td>
<td>0.017</td>
<td>0.024</td>
<td>0.024</td>
<td>0.136</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. perplexus-B. gregoryensis to Mid Campanian</td>
<td>0.002*</td>
<td>0.006*</td>
<td>0.191</td>
<td>0.191</td>
<td>0.204</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.102</td>
<td>0.065</td>
<td>0.239</td>
<td>0.239</td>
<td>0.146</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.093</td>
<td>0.081</td>
<td>0.153</td>
<td>0.153</td>
<td>0.094</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.085</td>
<td>0.127</td>
<td>0.168</td>
<td>0.168</td>
<td>0.114</td>
</tr>
<tr>
<td>L. subundata</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.076</td>
<td>0.065</td>
<td>0.133</td>
<td>0.133</td>
<td>0.076</td>
</tr>
<tr>
<td>L. subundata</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.337</td>
<td>0.241</td>
<td>0.139</td>
<td>0.139</td>
<td>0.058</td>
</tr>
<tr>
<td>C. nebrascensiss</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.003*</td>
<td>0.004*</td>
<td>0.016*</td>
<td>0.016*</td>
<td>0.282</td>
</tr>
<tr>
<td>C. nebrascensiss</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.040*</td>
<td>0.027*</td>
<td>0.021*</td>
<td>0.021*</td>
<td>0.082</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Lower Maastrichtian</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.005*</td>
<td>0.005*</td>
<td>0.003*</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>0.105</td>
<td>0.127</td>
<td>0.118</td>
<td>0.118</td>
<td>0.153</td>
</tr>
</tbody>
</table>
Table S12. Results of lower niche similarity test of temporal bin sizes including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. convexus</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.765</td>
<td>0.764</td>
<td>0.838</td>
<td>0.838</td>
<td>0.847</td>
</tr>
<tr>
<td>I. convexus</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>1.000</td>
<td>1.000</td>
<td>0.984</td>
<td>0.984</td>
<td>0.936</td>
</tr>
<tr>
<td>I. barabini</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.961</td>
<td>0.979</td>
<td>0.910</td>
<td>0.910</td>
<td>0.929</td>
</tr>
<tr>
<td>I. barabini</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.958</td>
<td>0.971</td>
<td>0.899</td>
<td>0.899</td>
<td>0.934</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.889</td>
<td>0.940</td>
<td>0.792</td>
<td>0.792</td>
<td>0.984</td>
</tr>
<tr>
<td>I. sagensis</td>
<td>B. compressus-B. cuneatus to Upper Campanian</td>
<td>0.844</td>
<td>0.857</td>
<td>0.813</td>
<td>0.813</td>
<td>0.986</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.876</td>
<td>0.976</td>
<td>0.999</td>
<td>0.999</td>
<td>0.819</td>
</tr>
<tr>
<td>I. sublaevis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.993</td>
<td>0.996</td>
<td>0.870</td>
<td>0.870</td>
<td>0.830</td>
</tr>
<tr>
<td>I. azerbaydjanensis</td>
<td>B. maclearni-B. sp. (smooth) to Mid Campanian</td>
<td>0.871</td>
<td>0.908</td>
<td>0.867</td>
<td>0.867</td>
<td>0.828</td>
</tr>
<tr>
<td>I. oblongus</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.994</td>
<td>0.990</td>
<td>0.827</td>
<td>0.827</td>
<td>0.992</td>
</tr>
<tr>
<td>I. saskatchewanensis</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.976</td>
<td>0.982</td>
<td>0.977</td>
<td>0.977</td>
<td>0.862</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. perplexus-B. gregoryensis to Mid Campanian</td>
<td>1.000</td>
<td>1.000</td>
<td>0.834</td>
<td>0.834</td>
<td>0.812</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reduncus-B. scotti to Mid Campanian</td>
<td>0.886</td>
<td>0.907</td>
<td>0.747</td>
<td>0.747</td>
<td>0.826</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.916</td>
<td>0.930</td>
<td>0.878</td>
<td>0.878</td>
<td>0.924</td>
</tr>
<tr>
<td>C. imbricatula</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.930</td>
<td>0.898</td>
<td>0.866</td>
<td>0.866</td>
<td>0.909</td>
</tr>
<tr>
<td>L. subundata</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.910</td>
<td>0.933</td>
<td>0.877</td>
<td>0.877</td>
<td>0.926</td>
</tr>
<tr>
<td>L. subundata</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.630</td>
<td>0.762</td>
<td>0.867</td>
<td>0.867</td>
<td>1.000</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>D. nebrascense-E. jenneyi to Upper Campanian</td>
<td>0.995</td>
<td>0.994</td>
<td>0.991</td>
<td>0.991</td>
<td>0.705</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. reesidei-B. eliasi to Upper Campanian</td>
<td>0.957</td>
<td>0.973</td>
<td>0.981</td>
<td>0.981</td>
<td>0.933</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Lower Maastrichtian</td>
<td>1.000</td>
<td>1.000</td>
<td>0.989</td>
<td>0.989</td>
<td>0.997</td>
</tr>
<tr>
<td>C. nebrascensis</td>
<td>B. baculus-B. grandis to Maastrichtian</td>
<td>0.915</td>
<td>0.881</td>
<td>0.899</td>
<td>0.899</td>
<td>0.867</td>
</tr>
</tbody>
</table>
Correlation between Duration Ratio and Niche Overlap

R Squared = 0.251

Figure S10. Correlation plot of duration ratio to niche overlap (D value). This plot represents non-averaged correlation data.
Ingroup Comparison:

Table S13. Niche overlap values for *Inoceramus* species comparisons including I and D overlap values, and niche expansion, stability, and unfilling.

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion</th>
<th>Stability</th>
<th>Unfilling</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.721</td>
<td>0.779</td>
<td>0.674</td>
<td>0.326</td>
<td>0.505</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. saskatchewanensis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.518</td>
<td>0.633</td>
<td>0.217</td>
<td>0.783</td>
<td>0.540</td>
</tr>
<tr>
<td><em>I. sublaevis</em> vs. <em>I. saskatchewanensis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.491</td>
<td>0.585</td>
<td>0.231</td>
<td>0.769</td>
<td>0.383</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. barabini</em></td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.106</td>
<td>0.289</td>
<td>0.589</td>
<td>0.411</td>
<td>0.101</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.088</td>
<td>0.254</td>
<td>0.674</td>
<td>0.326</td>
<td>0.505</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sublaevis</em></td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.796</td>
<td>0.851</td>
<td>0.186</td>
<td>0.814</td>
<td>0.495</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sagensis</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.675</td>
<td>0.719</td>
<td>0.344</td>
<td>0.656</td>
<td>0.181</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. oblongus</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.545</td>
<td>0.621</td>
<td>0.406</td>
<td>0.594</td>
<td>0.295</td>
</tr>
<tr>
<td><em>I. oblongus</em> vs. <em>I. sagensis</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.777</td>
<td>0.845</td>
<td>0.125</td>
<td>0.875</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Table S14. Results of greater niche equivalency test for *Inoceramus* species comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.912</td>
<td>0.882</td>
<td>0.926</td>
<td>0.926</td>
<td>0.810</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. saskatchewanensis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.695</td>
<td>0.663</td>
<td>0.381</td>
<td>0.381</td>
<td>0.884</td>
</tr>
<tr>
<td><em>I. sublaevis</em> vs. <em>I. saskatchewanensis</em></td>
<td>B. reduncus-B. scotti</td>
<td>0.677</td>
<td>0.741</td>
<td>0.617</td>
<td>0.617</td>
<td>0.749</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. barabini</em></td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.866</td>
<td>0.884</td>
<td>0.934</td>
<td>0.934</td>
<td>0.303</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.912</td>
<td>0.882</td>
<td>0.926</td>
<td>0.926</td>
<td>0.810</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sublaevis</em></td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.066</td>
<td>0.092</td>
<td>0.202</td>
<td>0.202</td>
<td>0.591</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sagensis</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.052</td>
<td>0.072</td>
<td>0.511</td>
<td>0.511</td>
<td>0.152</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. oblongus</em></td>
<td>B. reesidei-B. eliasi</td>
<td>0.130</td>
<td>0.186</td>
<td>0.589</td>
<td>0.589</td>
<td>0.357</td>
</tr>
<tr>
<td><em>I. oblongus</em> vs. <em>I. sagensis</em></td>
<td>B. reesidei-B. eliasi</td>
<td><strong>0.012</strong>*</td>
<td><strong>0.018</strong>*</td>
<td>0.158</td>
<td>0.158</td>
<td>0.132</td>
</tr>
</tbody>
</table>
Table S15. Results of lower niche equivalency test for *Inoceramus* species comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species Intervals Compared</th>
<th>D value</th>
<th>I Value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>0.068</td>
<td>0.112</td>
<td>0.060</td>
<td>0.060</td>
<td>0.162</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. saskatchewanensis</em></td>
<td>0.317</td>
<td>0.319</td>
<td>0.689</td>
<td>0.689</td>
<td>0.140</td>
</tr>
<tr>
<td><em>I. sublaevis</em> vs. <em>I. saskatchewanensis</em></td>
<td>0.281</td>
<td>0.253</td>
<td>0.347</td>
<td>0.347</td>
<td>0.240</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. barabini</em></td>
<td>0.122</td>
<td>0.110</td>
<td>0.076</td>
<td>0.076</td>
<td>0.669</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>0.068</td>
<td>0.112</td>
<td>0.060</td>
<td>0.060</td>
<td>0.162</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sublaevis</em></td>
<td>0.930</td>
<td>0.910</td>
<td>0.798</td>
<td>0.798</td>
<td>0.443</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sagensis</em></td>
<td>0.966</td>
<td>0.934</td>
<td>0.543</td>
<td>0.543</td>
<td>0.850</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. oblongus</em></td>
<td>0.862</td>
<td>0.828</td>
<td>0.371</td>
<td>0.371</td>
<td>0.629</td>
</tr>
<tr>
<td><em>I. oblongus</em> vs. <em>I. sagensis</em></td>
<td>0.986</td>
<td>0.980</td>
<td>0.832</td>
<td>0.832</td>
<td>0.866</td>
</tr>
</tbody>
</table>

Table S16. Results of greater niche similarity test for *Inoceramus* species comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>0.431</td>
<td>0.276</td>
<td>0.392</td>
<td>0.392</td>
<td>0.247</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. saskatchewanensis</em></td>
<td>0.112</td>
<td>0.193</td>
<td>0.381</td>
<td>0.381</td>
<td>0.246</td>
</tr>
<tr>
<td><em>I. sublaevis</em> vs. <em>I. saskatchewanensis</em></td>
<td>0.221</td>
<td>0.239</td>
<td>0.353</td>
<td>0.353</td>
<td>0.179</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. barabini</em></td>
<td>0.429</td>
<td>0.361</td>
<td>0.209</td>
<td>0.209</td>
<td>0.232</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td>0.431</td>
<td>0.276</td>
<td>0.392</td>
<td>0.392</td>
<td>0.247</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sublaevis</em></td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.251</td>
<td>0.251</td>
<td>0.045*</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sagensis</em></td>
<td>0.015*</td>
<td>0.028*</td>
<td>0.026*</td>
<td>0.026*</td>
<td>0.164</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. oblongus</em></td>
<td>0.038*</td>
<td>0.068</td>
<td>0.072</td>
<td>0.072</td>
<td>0.206</td>
</tr>
<tr>
<td><em>I. oblongus</em> vs. <em>I. sagensis</em></td>
<td>0.048*</td>
<td>0.043*</td>
<td>0.058</td>
<td>0.058</td>
<td>0.037*</td>
</tr>
</tbody>
</table>
Table S17. Results of lower niche similarity test for *Inoceramus* species comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species (lower) vs. Species (higher)</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.554</td>
<td>0.704</td>
<td>0.618</td>
<td>0.618</td>
<td>0.777</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. saskatchewanensis</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.890</td>
<td>0.817</td>
<td>0.616</td>
<td>0.616</td>
<td>0.764</td>
</tr>
<tr>
<td><em>I. sublaevis</em> vs. <em>I. saskatchewanensis</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.780</td>
<td>0.772</td>
<td>0.661</td>
<td>0.661</td>
<td>0.822</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.592</td>
<td>0.672</td>
<td>0.806</td>
<td>0.806</td>
<td>0.787</td>
</tr>
<tr>
<td><em>I. convexus</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.554</td>
<td>0.704</td>
<td>0.618</td>
<td>0.618</td>
<td>0.777</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>1.000</td>
<td>0.999</td>
<td>0.769</td>
<td>0.769</td>
<td>0.947</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.985</td>
<td>0.977</td>
<td>0.975</td>
<td>0.975</td>
<td>0.864</td>
</tr>
<tr>
<td><em>I. barabini</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.959</td>
<td>0.924</td>
<td>0.916</td>
<td>0.916</td>
<td>0.785</td>
</tr>
<tr>
<td><em>I. oblongus</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.945</td>
<td>0.952</td>
<td>0.937</td>
<td>0.937</td>
<td>0.955</td>
</tr>
</tbody>
</table>
Comparison taxa Comparisons:

Table S18. Niche overlap values for *Inoceramus* species to the genus comparisons including I and D overlap values, and niche expansion, stability, and unfilling.

<table>
<thead>
<tr>
<th>Species Compared</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion</th>
<th>Stability</th>
<th>Unfilling</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>B. reduncus-B. scotti</td>
<td>0.247</td>
<td>0.425</td>
<td>0.665</td>
<td>0.335</td>
<td>0.164</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>B. reduncus-B. scotti</td>
<td>0.383</td>
<td>0.525</td>
<td>0.473</td>
<td>0.527</td>
<td>0.076</td>
</tr>
<tr>
<td>C. imbricatula vs. I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.443</td>
<td>0.654</td>
<td>0.368</td>
<td>0.632</td>
<td>0.085</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.832</td>
<td>0.918</td>
<td>0.069</td>
<td>0.931</td>
<td>0.092</td>
</tr>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.147</td>
<td>0.267</td>
<td>0.280</td>
<td>0.720</td>
<td>0.582</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.207</td>
<td>0.290</td>
<td>0.784</td>
<td>0.216</td>
<td>0.834</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.832</td>
<td>0.918</td>
<td>0.069</td>
<td>0.931</td>
<td>0.092</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.072</td>
<td>0.248</td>
<td>0.088</td>
<td>0.912</td>
<td>0.668</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sagensis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.854</td>
<td>0.925</td>
<td>0.116</td>
<td>0.884</td>
<td>0.431</td>
</tr>
<tr>
<td>L. subundata vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.700</td>
<td>0.791</td>
<td>0.167</td>
<td>0.833</td>
<td>0.093</td>
</tr>
<tr>
<td>L. subundata vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.104</td>
<td>0.260</td>
<td>0.187</td>
<td>0.813</td>
<td>0.604</td>
</tr>
<tr>
<td>L. subundata vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.571</td>
<td>0.661</td>
<td>0.426</td>
<td>0.574</td>
<td>0.658</td>
</tr>
<tr>
<td>L. subundata vs. I. sagensis</td>
<td>B. compressus-B. cuneatus</td>
<td>0.294</td>
<td>0.467</td>
<td>0.362</td>
<td>0.638</td>
<td>0.114</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td>0.611</td>
<td>0.709</td>
<td>0.062</td>
<td>0.938</td>
<td>0.371</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.655</td>
<td>0.771</td>
<td>0.084</td>
<td>0.916</td>
<td>0.248</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.547</td>
<td>0.700</td>
<td>0.126</td>
<td>0.874</td>
<td>0.315</td>
</tr>
<tr>
<td>C. imbricatula vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td>0.599</td>
<td>0.659</td>
<td>0.137</td>
<td>0.863</td>
<td>0.396</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.786</td>
<td>0.852</td>
<td>0.034</td>
<td>0.966</td>
<td>0.133</td>
</tr>
<tr>
<td>C. imbricatula vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.710</td>
<td>0.834</td>
<td>0.021</td>
<td>0.979</td>
<td>0.160</td>
</tr>
</tbody>
</table>
Table S19. Results of greater niche equivalency test for *Inoceramus* species to comparison species comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

### Greater Equivalency Test (p-values)

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.800</td>
<td>0.743</td>
<td>0.884</td>
<td>0.884</td>
<td>0.345</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.549</td>
<td>0.593</td>
<td>0.667</td>
<td>0.667</td>
<td>0.263</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. saskatchewanensis</em></td>
<td><em>B. reduncus</em>-<em>B. scotti</em></td>
<td>0.669</td>
<td>0.607</td>
<td>0.495</td>
<td>0.495</td>
<td>0.226</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.012*</td>
<td>0.006*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.788</td>
<td>0.888</td>
<td>0.717</td>
<td>0.717</td>
<td>0.908</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.782</td>
<td>0.800</td>
<td>0.946</td>
<td>0.946</td>
<td>0.958</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.012*</td>
<td>0.006*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.900</td>
<td>0.814</td>
<td>0.070</td>
<td>0.070</td>
<td>0.944</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sagensis</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.068</td>
<td>0.054</td>
<td></td>
<td>0.042*</td>
<td>0.042*</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.026*</td>
<td>0.040*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.850</td>
<td>0.886</td>
<td>0.451</td>
<td>0.451</td>
<td>0.906</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense</em>-<em>E. jenneyi</em></td>
<td>0.311</td>
<td>0.359</td>
<td>0.561</td>
<td>0.561</td>
<td>0.794</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sagensis</em></td>
<td><em>B. compressus</em>-<em>B. cuneatus</em></td>
<td>0.599</td>
<td>0.693</td>
<td>0.597</td>
<td>0.597</td>
<td>0.591</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.070</td>
<td>0.096</td>
<td>0.112</td>
<td>0.112</td>
<td>0.539</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.060</td>
<td>0.042*</td>
<td></td>
<td>0.178</td>
<td>0.178</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.259</td>
<td>0.214</td>
<td>0.303</td>
<td>0.303</td>
<td>0.463</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.112</td>
<td>0.164</td>
<td>0.148</td>
<td>0.148</td>
<td>0.593</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.014*</td>
<td>0.010*</td>
<td></td>
<td>0.030*</td>
<td>0.030*</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei</em>-<em>B. eliasi</em></td>
<td>0.046*</td>
<td>0.020*</td>
<td></td>
<td>0.026*</td>
<td>0.026*</td>
</tr>
</tbody>
</table>
Table S20. Results of lower niche equivalency test for *Inoceramus* species to comparison species comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.226</td>
<td>0.277</td>
<td>0.138</td>
<td>0.138</td>
<td>0.679</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.503</td>
<td>0.489</td>
<td>0.345</td>
<td>0.345</td>
<td>0.715</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.323</td>
<td>0.425</td>
<td>0.525</td>
<td>0.525</td>
<td>0.764</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.994</td>
<td>1.000</td>
<td>0.880</td>
<td>0.880</td>
<td>0.854</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.192</td>
<td>0.098</td>
<td>0.303</td>
<td>0.303</td>
<td>0.082</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.214</td>
<td>0.176</td>
<td>0.054</td>
<td>0.054</td>
<td>0.034*</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.994</td>
<td>1.000</td>
<td>0.880</td>
<td>0.880</td>
<td>0.854</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.068</td>
<td>0.148</td>
<td>0.932</td>
<td>0.932</td>
<td>0.044*</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sagensis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.932</td>
<td>0.942</td>
<td>0.964</td>
<td>0.964</td>
<td>0.511</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.982</td>
<td>0.952</td>
<td>0.689</td>
<td>0.689</td>
<td>0.802</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.184</td>
<td>0.122</td>
<td>0.567</td>
<td>0.567</td>
<td>0.098</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.641</td>
<td>0.601</td>
<td>0.443</td>
<td>0.443</td>
<td>0.196</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sagensis</em></td>
<td><em>B. compressus-B. cuneatus</em></td>
<td>0.421</td>
<td>0.357</td>
<td>0.417</td>
<td>0.417</td>
<td>0.429</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.932</td>
<td>0.908</td>
<td>0.866</td>
<td>0.866</td>
<td>0.421</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.932</td>
<td>0.940</td>
<td>0.848</td>
<td>0.848</td>
<td>0.581</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.735</td>
<td>0.810</td>
<td>0.737</td>
<td>0.737</td>
<td>0.549</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.900</td>
<td>0.858</td>
<td>0.848</td>
<td>0.848</td>
<td>0.423</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.988</td>
<td>0.990</td>
<td>0.974</td>
<td>0.974</td>
<td>0.794</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.946</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
<td>0.794</td>
</tr>
</tbody>
</table>
Table S21. Results of greater niche similarity test for *Inoceramus* species to comparison species comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>B. reduncus-B. scotti</td>
<td>0.411</td>
<td>0.399</td>
<td>0.295</td>
<td>0.295</td>
<td>0.325</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>B. reduncus-B. scotti</td>
<td>0.306</td>
<td>0.362</td>
<td>0.147</td>
<td>0.147</td>
<td>0.294</td>
</tr>
<tr>
<td>C. imbricatula vs. I. saskatchewanensis</td>
<td>B. reduncus-B. scotti</td>
<td>0.112</td>
<td>0.054</td>
<td>0.074</td>
<td>0.074</td>
<td>0.038*</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.199</td>
</tr>
<tr>
<td>C. imbricatula vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.398</td>
<td>0.349</td>
<td>0.305</td>
<td>0.305</td>
<td>0.304</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.356</td>
<td>0.332</td>
<td>0.626</td>
<td>0.626</td>
<td>0.536</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.001*</td>
<td>0.199</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.449</td>
<td>0.300</td>
<td>0.015*</td>
<td>0.015*</td>
<td>0.350</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.015*</td>
<td>0.006*</td>
<td>0.045*</td>
<td>0.045*</td>
<td>0.267</td>
</tr>
<tr>
<td>L. subundata vs. I. barabini</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.101</td>
<td>0.099</td>
<td>0.138</td>
<td>0.138</td>
<td>0.115</td>
</tr>
<tr>
<td>L. subundata vs. I. convexus</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.501</td>
<td>0.422</td>
<td>0.356</td>
<td>0.356</td>
<td>0.305</td>
</tr>
<tr>
<td>L. subundata vs. I. sublaevis</td>
<td>D. nebrascense-E. jenneyi</td>
<td>0.028*</td>
<td>0.050*</td>
<td>0.448</td>
<td>0.448</td>
<td>0.232</td>
</tr>
<tr>
<td>L. subundata vs. I. sagensis</td>
<td>B. compressus-B. cuneatus</td>
<td>0.265</td>
<td>0.234</td>
<td>0.158</td>
<td>0.158</td>
<td>0.167</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td>0.006*</td>
<td>0.008*</td>
<td>0.098</td>
<td>0.098</td>
<td>0.029*</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.067</td>
<td>0.047*</td>
<td>0.045*</td>
<td>0.045*</td>
<td>0.122</td>
</tr>
<tr>
<td>C. nebrascensis vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.101</td>
<td>0.062</td>
<td>0.083</td>
<td>0.083</td>
<td>0.149</td>
</tr>
<tr>
<td>C. imbricatula vs. I. barabini</td>
<td>B. reesidei-B. eliasi</td>
<td>0.017*</td>
<td>0.039*</td>
<td>0.172</td>
<td>0.172</td>
<td>0.028*</td>
</tr>
<tr>
<td>C. imbricatula vs. I. sagensis</td>
<td>B. reesidei-B. eliasi</td>
<td>0.044*</td>
<td>0.049*</td>
<td>0.035*</td>
<td>0.035*</td>
<td>0.038*</td>
</tr>
<tr>
<td>C. imbricatula vs. I. oblongus</td>
<td>B. reesidei-B. eliasi</td>
<td>0.076</td>
<td>0.053</td>
<td>0.018*</td>
<td>0.018*</td>
<td>0.054</td>
</tr>
</tbody>
</table>
Table S22. Results of lower niche similarity test for *Inoceramus* species to comparison species comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.601</td>
<td>0.608</td>
<td>0.727</td>
<td>0.727</td>
<td>0.667</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.716</td>
<td>0.683</td>
<td>0.844</td>
<td>0.844</td>
<td>0.742</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.901</td>
<td>0.953</td>
<td>0.937</td>
<td>0.937</td>
<td>0.954</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.771</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.615</td>
<td>0.678</td>
<td>0.725</td>
<td>0.725</td>
<td>0.707</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.628</td>
<td>0.637</td>
<td>0.358</td>
<td>0.358</td>
<td>0.435</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.771</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.536</td>
<td>0.725</td>
<td>0.985</td>
<td>0.985</td>
<td>0.623</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.993</td>
<td>0.998</td>
<td>0.952</td>
<td>0.952</td>
<td>0.719</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.910</td>
<td>0.907</td>
<td>0.863</td>
<td>0.863</td>
<td>0.907</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.517</td>
<td>0.566</td>
<td>0.650</td>
<td>0.650</td>
<td>0.698</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.980</td>
<td>0.942</td>
<td>0.535</td>
<td>0.535</td>
<td>0.762</td>
</tr>
<tr>
<td><em>L. subundata</em> vs. <em>I. sagensis</em></td>
<td><em>B. compressus-B. cuneatus</em></td>
<td>0.707</td>
<td>0.727</td>
<td>0.822</td>
<td>0.822</td>
<td>0.824</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.988</td>
<td>0.989</td>
<td>0.875</td>
<td>0.875</td>
<td>0.976</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.939</td>
<td>0.955</td>
<td>0.948</td>
<td>0.948</td>
<td>0.907</td>
</tr>
<tr>
<td><em>C. nebrascensis</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.875</td>
<td>0.923</td>
<td>0.923</td>
<td>0.923</td>
<td>0.846</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.972</td>
<td>0.956</td>
<td>0.827</td>
<td>0.827</td>
<td>0.969</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.927</td>
<td>0.924</td>
<td>0.939</td>
<td>0.939</td>
<td>0.950</td>
</tr>
<tr>
<td><em>C. imbricatula</em> vs. <em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.933</td>
<td>0.949</td>
<td>0.967</td>
<td>0.967</td>
<td>0.956</td>
</tr>
</tbody>
</table>
Species to Genus (all occurrences) Comparison

Table S23. Niche overlap values for *Inoceramus* species to the genus comparisons including I and D overlap values, and niche expansion, stability, and unfilling.

<table>
<thead>
<tr>
<th>Species</th>
<th>Interval</th>
<th>D value</th>
<th>I value</th>
<th>Expansion</th>
<th>Stability</th>
<th>Unfilling</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.721</td>
<td>0.826</td>
<td>0.112</td>
<td>0.888</td>
<td>0.078</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.740</td>
<td>0.832</td>
<td>0.087</td>
<td>0.913</td>
<td>0.020</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.679</td>
<td>0.844</td>
<td>0.348</td>
<td>0.652</td>
<td>0.018</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.221</td>
<td>0.466</td>
<td>0.512</td>
<td>0.488</td>
<td>0.000</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.427</td>
<td>0.549</td>
<td>0.680</td>
<td>0.320</td>
<td>0.237</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.509</td>
<td>0.643</td>
<td>0.224</td>
<td>0.776</td>
<td>0.119</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.465</td>
<td>0.676</td>
<td>0.481</td>
<td>0.519</td>
<td>0.000</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.477</td>
<td>0.687</td>
<td>0.302</td>
<td>0.698</td>
<td>0.000</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.408</td>
<td>0.636</td>
<td>0.334</td>
<td>0.666</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table S24. Results of greater niche equivalency test for *Inoceramus* species to genus (all possible occurrences) comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.132</td>
<td>0.152</td>
<td>0.160</td>
<td>0.160</td>
<td>0.766</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.126</td>
<td>0.190</td>
<td>0.150</td>
<td>0.150</td>
<td>0.473</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.359</td>
<td>0.263</td>
<td>0.725</td>
<td>0.725</td>
<td>0.497</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.697</td>
<td>0.655</td>
<td>0.735</td>
<td>0.735</td>
<td>0.399</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.441</td>
<td>0.565</td>
<td>0.930</td>
<td>0.930</td>
<td>0.990</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.347</td>
<td>0.375</td>
<td>0.415</td>
<td>0.415</td>
<td>0.874</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.511</td>
<td>0.435</td>
<td>0.816</td>
<td>0.816</td>
<td>0.307</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.527</td>
<td>0.467</td>
<td>0.535</td>
<td>0.535</td>
<td>0.295</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.607</td>
<td>0.519</td>
<td>0.485</td>
<td>0.485</td>
<td>0.377</td>
</tr>
</tbody>
</table>
Table S25. Results of lower niche equivalency test for *Inoceramus* species to genus (all possible occurrences) comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.886</td>
<td>0.860</td>
<td>0.854</td>
<td>0.854</td>
<td>0.246</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.886</td>
<td>0.776</td>
<td>0.852</td>
<td>0.852</td>
<td>0.531</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.631</td>
<td>0.707</td>
<td>0.279</td>
<td>0.279</td>
<td>0.489</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.269</td>
<td>0.307</td>
<td>0.246</td>
<td>0.246</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.595</td>
<td>0.485</td>
<td>0.056</td>
<td>0.056</td>
<td>0.006*</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.687</td>
<td>0.673</td>
<td>0.611</td>
<td>0.611</td>
<td>0.134</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.469</td>
<td>0.561</td>
<td>0.144</td>
<td>0.144</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.491</td>
<td>0.567</td>
<td>0.443</td>
<td>0.443</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.447</td>
<td>0.533</td>
<td>0.533</td>
<td>0.533</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table S26. Results of greater niche similarity test for *Inoceramus* species to genus (all possible occurrences) comparisons including p-values for I and D overlap (greater than random), and niche expansion (less than random), stability (greater than random), and unfilling (less than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (lower)</th>
<th>Stability (higher)</th>
<th>Unfilling (lower)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.114</td>
<td>0.141</td>
<td>0.225</td>
<td>0.225</td>
<td>0.100</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.165</td>
<td>0.194</td>
<td>0.144</td>
<td>0.144</td>
<td>0.125</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.060</td>
<td>0.047*</td>
<td>0.145</td>
<td>0.145</td>
<td>0.148</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.292</td>
<td>0.262</td>
<td>0.202</td>
<td>0.202</td>
<td>0.238</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.137</td>
<td>0.215</td>
<td>0.347</td>
<td>0.347</td>
<td>0.442</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.152</td>
<td>0.154</td>
<td>0.170</td>
<td>0.170</td>
<td>0.047*</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.033*</td>
<td>0.022*</td>
<td>0.041*</td>
<td>0.041*</td>
<td>0.157</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.206</td>
<td>0.164</td>
<td>0.089</td>
<td>0.089</td>
<td>0.097</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.212</td>
<td>0.174</td>
<td>0.076</td>
<td>0.076</td>
<td>0.068</td>
</tr>
</tbody>
</table>
Table S27. Results of lower niche similarity test for *Inoceramus* species to genus (all possible occurrences) comparisons including p-values for I and D overlap (less than random), and niche expansion (greater than random), stability (less than random), and unfilling (greater than random).

<table>
<thead>
<tr>
<th>Species</th>
<th>Intervals Compared</th>
<th>D value</th>
<th>I value</th>
<th>Expansion (higher)</th>
<th>Stability (lower)</th>
<th>Unfilling (higher)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I. convexus</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.907</td>
<td>0.880</td>
<td>0.764</td>
<td>0.764</td>
<td>0.923</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.858</td>
<td>0.829</td>
<td>0.869</td>
<td>0.869</td>
<td>0.877</td>
</tr>
<tr>
<td><em>I. saskatchewanensis</em></td>
<td><em>B. reduncus-B. scotti</em></td>
<td>0.936</td>
<td>0.958</td>
<td>0.831</td>
<td>0.831</td>
<td>0.825</td>
</tr>
<tr>
<td><em>I. convexus</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.729</td>
<td>0.763</td>
<td>0.836</td>
<td>0.836</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. sublaevis</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.883</td>
<td>0.804</td>
<td>0.672</td>
<td>0.672</td>
<td>0.558</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>D. nebrascense-E. jenneyi</em></td>
<td>0.840</td>
<td>0.849</td>
<td>0.831</td>
<td>0.831</td>
<td>0.952</td>
</tr>
<tr>
<td><em>I. barabini</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.972</td>
<td>0.981</td>
<td>0.965</td>
<td>0.965</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. sagensis</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.827</td>
<td>0.860</td>
<td>0.906</td>
<td>0.906</td>
<td>1.000</td>
</tr>
<tr>
<td><em>I. oblongus</em></td>
<td><em>B. reesidei-B. eliasi</em></td>
<td>0.779</td>
<td>0.827</td>
<td>0.900</td>
<td>0.900</td>
<td>1.000</td>
</tr>
</tbody>
</table>
References


Appendix C-3. Detailed References for Environmental Raster Layers

Sedimentological Literature References:


Seep Localities Literature Sources:


Outcrop Polygon References:

Conservation and Survey Division, University of Nebraska-Lincoln (CSD), 1996, bedrock.e00: Nebraska Geological Survey, Conservation and Survey Division, University of Nebraska-Lincoln, Lincoln, Nebraska, scale 1:1,000,000.


Appendix C-4. R Code

library(ecospat)
library(raster)
library(rgbif)
library(terra)
library(geodata)

library(ade4)
library(rlist)

library(raster) # for raster analysis
library(sf) # for spatial data analysis
library(dplyr) #
library(scales)
library(dismo) # a collection of ENM/SDM tools
library(tidyrr)
library(colorRamps)
library(ENMeval) # for a few new tools in ENM/SDM
library(ggplot2)
library(GGally) # Use for ggpairs fuction (not used...)
library(spThin)
library(sdm)
library(rasterVis)

$$$$$$$$$$ READ IN ENVIRONMENTAL DATA $$$$$$$$$$

setwd("C:/Users/Ceara/Documents/Province Project/ENM_Analysis_Code-and-Files")

# Function to get stacked raster layers of environmental proxies (with proper names for plotting)
get_rasters<- function(interval_abrev_strg,seep_presence){

if (seep_presence== TRUE){
    sr <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sr_av_idw_clip.tif")) # Dist to Shore
    se <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_se_av_idw_clip.tif")) # Dist to Seeps
    p <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_p_av_idw_clip.tif")) # % Pebble
    s <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_s_av_idw_clip.tif")) # % Sand
    sh <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sh_av_idw_clip.tif")) # % Shale
    sl <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sl_av_idw_clip.tif")) # % Silt
    m <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_m_av_idw_clip.tif")) # % Mud
    l <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_l_av_idw_clip.tif")) # % LS
    bs <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_bs_av_idw_clip.tif")) # Bed Style
    b <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_b_av_idw_clip.tif")) # Bioturbation
    t <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_t_av_idw_clip.tif")) # Bed Thickness

    ### Stack env parameters
    stacked_env = stack(sr,se,p,s,sl,sh,m,l,bs,b,t)

    names(stacked_env) <- c("Dist_to_Shore","Dist_to_Seeps",
                           "Percent_Pebble","Percent_Sand","Percent_Shale","Percent_Silt","Percent_Mud",
                           "Percent_Limestone","Bedding_Style","Bioturbation","Avg_Bed_Thickness")
}

} else {
    sr <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sr_av_idw_clip.tif")) # Dist to Shore
    p <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_p_av_idw_clip.tif")) # % Pebble
    s <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_s_av_idw_clip.tif")) # % Sand
    sh <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sh_av_idw_clip.tif")) # % Shale
    sl <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_sl_av_idw_clip.tif")) # % Silt
    m <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_m_av_idw_clip.tif")) # % Mud
    l <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_l_av_idw_clip.tif")) # % LS
    bs <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_bs_av_idw_clip.tif")) # Bed Style
    b <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_b_av_idw_clip.tif")) # Bioturbation
    t <- raster::raster(paste0("Env Raster Files/",interval_abrev_strg,"_t_av_idw_clip.tif")) # Bed Thickness
### Stack env parameters

```r
stacked_env = stack(sr,p,s1,sh,m,l,bs,b,t)

names(stacked_env) <- c("Dist_to_Shore","Percent_Pebble","Percent_Sand","Percent_Shale","Percent_Silt","Percent_Mud", "Percent_Limestone","Bedding_Style","Bioturbation","Avg_Bed_Thickness")
```

return(stacked_env)

# Use above function to get stacked rasters for each time bin

```r
maastr_env <- get_rasters("maastr",FALSE)
camplow_env <- get_rasters("camplow",FALSE)
campmid_env <- get_rasters("campmid",TRUE)
campup_env <- get_rasters("campup",TRUE)
maastrlow_env <- get_rasters("maastrlow",TRUE)
maastrup_env <- get_rasters("maastrup",FALSE)
```

```r
birk_env <- get_rasters("birk",FALSE)
clin_env <- get_rasters("clin",FALSE)
bacu_env <- get_rasters("bacu",TRUE)
rees_env <- get_rasters("rees",TRUE)
comp_env <- get_rasters("comp",TRUE)
chey_env <- get_rasters("chey",TRUE)
nebr_env <- get_rasters("nebr",TRUE)
redu_env <- get_rasters("redu",TRUE)
```
perp_env <- get_rasters("perp",TRUE)
macl_env <- get_rasters("macl",FALSE)
obtu_env <- get_rasters("obtu",FALSE)
leei_env <- get_rasters("leei",FALSE)

#### Remove the variables that are NA or that correlate more than 0.8 Pearson's Corr Coef from raster stacks ####

# H. birkelundae- H. nebrascensis
birk_env_clean <- birk_env[[c(1:2,5,7,10)]]
# Removed all but Dist to shore, Pebble, Silt, Limestone, and Bed thickness

# B. clinobatus
clin_env_clean <- clin_env[[c(1:3,6,10)]]
# Removed all but Dist to shore, Pebble, Sand, Mud, and Bed thickness

# B. baculus-B. grandis
bacu_env_clean <- bacu_env[[c(1:4,6:8,10:11)]]
# Removed all but Dist to shore, Dist to seeps, Pebble, Sand, Silt, Mud, LS, Bioturb, and Bed thickness

# B. reesidei-B. eliasi
rees_env_clean <- rees_env[[c(1,3:4,6:11)]]
# Removed all but Dist to shore, Pebble, Sand, Silt, Mud, LS, Bed Style, Bioturb, and Bed thickness

# B. compressus-B. cuneatus
comp_env_clean <- comp_env[[c(1:2,4,6:7,10)]]
# Removed all but Dist to shore, Dist to Seeps, Sand, Silt, Mud, and Bioturbation

# D. cheyennense
chey_env_clean <- chey_env[[c(1,4,6:8)]]
# Removed all but Dist to shore, Sand, Silt, Mud, and LS
# Suppl. M. Purcell and Myers, 20XX

# D. nebrascense-E. jenneyi

\[
\text{nebr\_env\_clean <- nebr\_env[[c(1:2,4,6:11)]]}
\]

# Removed all but Dist to shore, dist to seeps, Sand, Silt, Mud, LS, Bed style, Bioturb, and Bed thickness

# B. reduncus-B. scotti

\[
\text{redu\_env\_clean <- redu\_env[[c(1:2,4,6:10)]]}
\]

# Removed all but Dist to shore, dist to seeps, Sand, Silt, Mud, LS, Bed style, Bioturb

# B. perplexus-B. gregoryensis

\[
\text{perp\_env\_clean <- perp\_env[[c(1:4,6:11)]]}
\]

# Removed all but Dist to shore, dist to seeps, Sand, Silt, Mud, LS, Bed style, Bioturb

# B. maclearni-B. sp. (smooth)

\[
\text{macl\_env\_clean <- macl\_env[[c(1,3,5:7,9:10)]]}
\]

# Removed all but Dist to shore, dist to seeps, Sand, Silt, Mud, LS, Bed style, Bioturb

# B. obtusus

\[
\text{obtu\_env\_clean <- obtu\_env[[c(1,3,5:8,10)]]}
\]

# Removed all but Dist to shore, Sand, Silt, Mud, LS, Bed style, and Bed Thickness

# S. leei-B. sp. (weak flank ribs)

\[
\text{leei\_env\_clean <- leei\_env[[c(1:3,5:8)]]}
\]

# Removed all but Dist to shore, Pebble, Sand, Silt, Mud, LS, Bed style

pairs(obtu_env_clean)

#### Remove the variables that are not shared between intervals (substage-biozone) ####

```r
pairs(obtu_env_clean)
```
# B. baculus-B. grandis
bacu_env_shared <- bacu_env[[c(1,3,4,6:8,11)]]

# B. reesidei-B. eliasi
rees_env_shared <- rees_env[[c(1,4,6:7,10)]]
# Removed all but Dist to shore, Pebble, Sand, Silt, Mud, LS, Bed Style, Bioturb, and Bed thickness

# B. compressus-B. cuneatus
comp_env_shared <- comp_env[[c(1,4,6:7,10)]]
# Removed all but Dist to shore, Dist to Seeps, Sand, Silt, Mud, and Bioturbation

# D. nebrascense-E. jenneyi
nebr_env_shared <- nebr_env[[c(1,4,6:7,10)]]
# Removed all but Dist to shore, dist to seeps, Sand, Silt, Mud, LS, Bed style, Bioturb, and Bed thickness

# B. reduncus-B. scotti
redu_env_shared <- redu_env[[c(1,4,6:8,10)]]
# Removed all but Dist to shore, Sand, Silt, Mud, LS, Bed style, Bioturb

# B. maclearni-B. sp. (smooth)
macl_env_shared <- macl_env[[c(1,3,5,6:7,9)]]

# B. perplexus-B. gregoryensis
perp_env_shared <- perp_env[[c(1,4,6:8,10)]]

# Upper Campanian
campup_env_shared <- campup_env[[c(1,4,6:7,10)]]
# Removed all but Dist to shore, Dist to seeps, Pebble, Sand, Silt, Mud, LS, Bed style, Bioturb, Bed thickness

# Middle Campanian
campmid_env_shared <- campmid_env[[c(1,4,6:8,10)]]

# Lower Maastrichtian
maastrlow_env_shared <- maastrlow_env[[c(1,3,4,6:8,11)]]

# Maastrichtian
maastr_env_shared <- maastr_env[[c(1,2,3,5:7,10)]]

#### Read in species occurrence data for the taxa at each relevant interval of comparison ####

# set name to use in calling/naming files
bara_nebr_name = "Ibarabini_nebr"

bara_rees_name = "Ibarabini_rees"

bara_campup_name = "Ibarabini_campup"

conv_redu_name = "Iconvexus_redu"

conv_nebr_name = "Iconvexus_nebr"

conv_campmid_name = "Iconvexus_campmid"

conv_campup_name = "Iconvexus_campup"

sage_comp_name = "Isagensis_comp"

sage_rees_name = "Isagensis_rees"
sage_campup_name = "Isagensis_campup"

subl_redu_name = "Isublaevis_redu"

subl_nebr_name = "Isublaevis_nebr"

subl_campmid_name = "Isublaevis_campmid"

subl_campup_name = "Isublaevis_campup"

azer_macl_name <- "Iazerbaydjanensis_macl"

azer_campmid_name <- "Iazerbaydjanensis_campmid"

oblo_rees_name <- "Ioblongus_rees"

oblo_campup_name <- "Ioblongus_campup"

sask_redu_name <- "Isaskatchewanensis_redu"

sask_campmid_name <- "Isaskatchewanensis_campmid"

# Outgroup Taxa:

chla_nebr_name <- "Cnebrascensis_nebr"
chla_rees_name <- "Cnebrascensis_rees"
chla_bacu_name <- "Cnebrascensis_bacu"
chla_campup_name <- "Cnebrascensis_campup"
chla_maastrlow_name <- "Cnebrascensis_maastrlow"
chla_maastr_name <- "Cnebrascensis_maa"
luci_nebr_name <- "Lsubundata_nebr"
luci_comp_name <- "Lsubundata_comp"
luci_campup_name <- "Lsubundata_campup"

cten_perp_name <- "Cimbricatula_perp"
cten_redu_name <- "Cimbricatula_redu"
cten_nebr_name <- "Cimbricatula_nebr"
cten_rees_name <- "Cimbricatula_rees"
cten_campup_name <- "Cimbricatula_campup"
cten_campmid_name <- "Cimbricatula_campmid"

# Additional taxa for analysis

cmon_macl_name <- "Cmontanensis_macl"
cmon_perp_name <- "Cmontanensis_perp"
cmon_campmid_name <- "Cmontanensis_campmid"
cmon_redu_name <- "Cmontanensis_redu"
cmon_comp_name <- "Cmontanensis_comp"
cmon_campup_name <- "Cmontanensis_campup"

hnod_comp_name <- "Hnodosus_comp"
hnod_rees_name <- "Hnodosus_rees"
hnod_campup_name <- "Hnodosus_campup"

ifib_clin_name <- "Ifibrosus_clin"
ifib_maastrlow_name <- "Ifibrosus_maastrlow"
ifib_maastr_name <- "Ifibrosus_maastr"

pmee_nebr_name <- "Pmeeki_nebr"
pmee_chey_name <- "Pmeeki_chey"
pmee_comp_name <- "Pmeeki_comp"
pmee_campup_name <- "Pmeeki_campup"

psyr_leei_name <- "Psyrtale_leei"
psyr_camplow_name <- "Psyrtale_camplow"

# Read in the occurrence data for the relevant intervals
conv_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/",conv_redu_name,"_thin1.csv"))
conv_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",conv_nebr_name,"_thin1.csv"))
conv_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/",conv_campmid_name,"_thin1.csv"))
conv_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",conv_campup_name,"_thin1.csv")))

bara_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/",bara_rees_name,"_thin1.csv"))
bara_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",bara_nebr_name,"_thin1.csv"))
bara_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",bara_campup_name,"_thin1.csv")))

sage_comp <- read.csv(as.character(paste0("Thinned_taxa_csvs/",sage_comp_name,"_thin1.csv"))
sage_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/",sage_rees_name,"_thin1.csv"))
sage_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",sage_campup_name,"_thin1.csv")))

subl_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/",subl_redu_name,"_thin1.csv"))
subl_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",subl_nebr_name,"_thin1.csv"))

653
subl_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/",subl_campmid_name,"_thin1.csv")))
subl_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",subl_campup_name,"_thin1.csv")))

azer_macl <- read.csv(as.character(paste0("Thinned_taxa_csvs/",azer_macl_name,"_thin1.csv")))
azer_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/",azer_campmid_name,"_thin1.csv")))

oblo_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/",oblo_rees_name,"_thin1.csv")))
oblo_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",oblo_campup_name,"_thin1.csv")))

sask_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/",sask_redu_name,"_thin1.csv")))
sask_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/",sask_campmid_name,"_thin1.csv")))

# Outgroup Taxa:

chla_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_nebr_name,"_thin1.csv")))
chla_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_rees_name,"_thin1.csv")))

chla_bacu <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_bacu_name,"_thin1.csv")))
chla_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_campup_name,"_thin1.csv")))
chla_maastrlow <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_maastrlow_name,"_thin1.csv")))
chla_maastr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",chla_maastr_name,"_thin1.csv")))

luci_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/",luci_nebr_name,"_thin1.csv")))
luci_comp <- read.csv(as.character(paste0("Thinned_taxa_csvs/",luci_comp_name,"_thin1.csv")))
luci_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/", luci_campup_name, "_thin1.csv")))

ten_perp <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_perp_name, "_thin1.csv")))
ten_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_redu_name, "_thin1.csv")))
ten_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_nebr_name, "_thin1.csv")))
ten_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_rees_name, "_thin1.csv")))

ten_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_campup_name, "_thin1.csv")))
ten_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cten_campmid_name, "_thin1.csv")))

# Additional Taxa for an analysis

cmon_macl <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_macl_name, "_thin1.csv")))
cmon_perp <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_perp_name, "_thin1.csv")))
cmon_campmid <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_campmid_name, "_thin1.csv")))
cmon_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_redu_name, "_thin1.csv")))
cmon_comp <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_comp_name, "_thin1.csv")))
cmon_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/", cmon_campup_name, "_thin1.csv")))

hnod_comp <- read.csv(as.character(paste0("Thinned_taxa_csvs/", hnod_comp_name, "_thin1.csv")))
hnod_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/", hnod_rees_name, "_thin1.csv")))
hnod_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/", hnod_campup_name, "_thin1.csv")))

ifib_clin <- read.csv(as.character(paste0("Thinned_taxa_csvs/", ifib_clin_name, "_thin1.csv")))
ifib_maastrlow <- read.csv(as.character(paste0("Thinned_taxa_csvs/", ifib_maastrlow_name, "_thin1.csv")))
ifib_maastr <- read.csv(as.character(paste0("Thinned_taxa_csvs/", "Ifibrosus_maa", "_thin1.csv"))))
pmee_mael <- read.csv(as.character(paste0("Thinned_taxa_csvs/","pmee_nebr_name","_thin1.csv")))
pmee_perp <- read.csv(as.character(paste0("Thinned_taxa_csvs/","pmee_chey_name","_thin1.csv")))
pmee_comp <- read.csv(as.character(paste0("Thinned_taxa_csvs/","pmee_comp_name","_thin1.csv")))
pmee_campup <- read.csv(as.character(paste0("Thinned_taxa_csvs/","pmee_campup_name","_thin1.csv"))))

psyr_leei <- read.csv(as.character(paste0("Thinned_taxa_csvs/","psyr_leei_name","_thin1.csv"))))
psyr_camplow <- read.csv(as.character(paste0("Thinned_taxa_csvs/","psyr_camplow_name","_thin1.csv"))))

#### Thin Genus level information (by selecting all relevant spp occ from database) ####

# Function to thin spp occ by 30km in intervals if >6 initial occurrences, creates csves and outputs summary table
thin_taxa_by_interv<- function(taxa_names,occ_data,tax_abrev){
  # Subset out species from occ table
  spp_dat <- subset(occ_data,Taxa_name %in% taxa_names, select=c(Taxa_name,Lat,Long,interval_biozone))
  # remove duplicate locations
  spp_dat <- unique(spp_dat)
  # Get list of number of occ in each interval
  unique_interv<- as.data.frame(table(spp_dat$interval_biozone))
  unique_interv_7great <- subset(unique_interv, Freq>6)
  unique_interv_7great <- unique_interv_7great$Var1
  # Get list of df of occ for each interval with >6 occ
  inter_dat_list <- list()  # create empty list to hold df for each relevant interval
  for (i in 1:length(unique_interv_7great)){
    spp_dat_inter <- subset(spp_dat,interval_biozone==unique_interv_7great[i])  # subset out by interval name
    inter_dat_list[[i]] <-spp_dat_inter[,c("Taxa_name","Lat","Long")]) # get just the lat/lon and taxa name
  }

  # Run thinning on each df in list:
  for (i in 1:nrow(as.data.frame(unique_interv_7great))){
    # Thinning function:
}
thin_tax_in_loop <- function(tax, file_name) {
  tax <- as.data.frame(tax)
  thin(tax, long.col = "Long", lat.col = "Lat", spec.col = "Taxa_name", thin.par = 30, reps = 1,
       out.dir = "C:/Users/ceara/Documents/Province Project/ENM_Analysis_Code-and-Files/Thinned_taxa_csvs", out.base = file_name, write.log.file = FALSE, verbose = FALSE)
}

  taxa_interval_name = as.character(paste(tax_abrev, unique(unique_interv_7great[[i]]), sep = "_")) # Set the name of output file
  thin_tax_in_loop(inter_dat_list[[i]], taxa_interval_name) # run thinning function to output files
}

output_list <- list()
for (i in 1:length(unique_interv_7great)) {
  taxa_interval_name = as.character(paste(tax_abrev, unique(unique_interv_7great[[i]]), sep = "_")) # Set the name of output file
  output_list[[i]] <- read.csv(as.character(paste0("Thinned_taxa_csvs/", taxa_interval_name, "_thin1.csv")))
  output_list[[i]]$interval <- taxa_interval_name
}

# Summarize the output CSV files into a vector of occ counts (frequency)
summarize <- function(df) {as.data.frame(table(df$interval))} # function to get frequency counts of each interval
output_list <- sapply(output_list, summarize) # apply the above function to each df
rownames(output_list) <- c("Taxa-Interval", "Freq") # renamed the summary table row names

return(output_list)
}

# Make lists of names to use for genus-level analysis in relevant intervals
redu_tax_names <- c("Inoceramus convexus","Inoceramus sublaevis", "Inoceramus samsungensis")
nebr_tax_names <- c("Inoceramus barabini","Inoceramus convexus","Inoceramus sublaevis")
rees_tax_names <- c("Inoceramus barabini","Inoceramus sagensis","Inoceramus oblongus")

#### Thin Genus level information (using all occ, including Inoceramus w/out spp ID)
# Read in occ database (with spp-genus combine in column Taxa_name)
tax_sp_inter <- read.csv('FE_data_compress_fe_3-7-23_basins_Fm_update_gen-sp_name_interval_update.csv')

# Extract all relevant spp names to use for collecting Inoceramus genus data
ino_spp_df <- unique(subset(tax_sp_inter, Updated_Genus == "Inoceramus", select=c(Taxa_name)))
ino_spp_vec <- as.character(ino_spp_df[,1]) # make into a character vector

# Make lists of names to use for genus-level analysis in relevant intervals
redu_tax_names <- ino_spp_vec
ebr_tax_names <- ino_spp_vec
rees_tax_names <- ino_spp_vec

# Use above function to run thinning on genus-level data in relevant intervals
# Inoceramus_redu<- thin_taxa_by_interv(redu_tax_names,tax_sp_inter,"Inoceramus_all_redu")
#
# Inoceramus_nebr<- thin_taxa_by_interv(nebr_tax_names,tax_sp_inter,"Inoceramus_all_nebr")
#
# Inoceramus_rees<- thin_taxa_by_interv(rees_tax_names,tax_sp_inter,"Inoceramus_all_rees")

#### Read in Genus Occurrence data for the taxa at each relevant interval of comparison ####

# Read in the genus occurrence data for the relevant intervals
Inoceramus_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_redu_B. reduncus-B. scotti","_thin1.csv"))
Inoceramus_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_nebr_D. nebrascense-E. jenneyi","_thin1.csv")))
Inoceramus_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_rees_B. reesidei-B. eliasi","_thin1.csv")))

Inoceramus_all_redu <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_all_redu_B. reduncus-B. scotti","_thin1.csv")))
Inoceramus_all_nebr <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_all_nebr_D. nebrascense-E. jenneyi","_thin1.csv")))
Inoceramus_all_rees <- read.csv(as.character(paste0("Thinned_taxa_csvs/","Inoceramus_all_rees_B. reesidei-B. eliasi","_thin1.csv")))

#### MAKE MAPS? Using the directions from this site: https://plantarum.ca/2021/07/29/ecospat/ ####

# Tell what columns are lat/long
coordinates(conv_redu) <- c("Long", "Lat")

# Download the maps of North American countries to plot
NorthAmerica <- gadm(country = country_codes("North America")$ISO3, 
                      level = 0, resolution = 2, 
                      path = "maps_basemaps")
us <- gadm(country = "USA", level = 1, resolution = 2, 
          path = "maps_basemaps")
canada <- gadm(country = "CAN", level = 1, resolution = 2, 
               path = "maps_basemaps")

# bind the maps into a single feature to map
CanUS <- rbind(NorthAmerica,us, canada)

# Plot the maps
plot(CanUS, xlim = c(-115, -95), ylim = c(55, 23), 
     border = "gray")
points(conv_redu,pch=16,col = "red")

#### Run function to prep data for use in ecospat (clean NAs, make into spatial object, extract values, etc) ####

run_data_prep <- function(taxa_inter_name, interval_raster_stack){

  # Call the thinned data file
tax_test <- read.csv(as.character(paste0("Thinned_taxa_csvs/",taxa_inter_name,"_thin1.csv")))
# Convert occ pts for taxa to spatial points
coordinates(tax_test) = ~ Long + Lat

# make sure to set spatial coor system (WGS 84)
myCRS1 <- CRS("+init=epsg:4326")
crs(tax_test) <- myCRS1

# project raster stack back to WGS 84 coor system
interval_raster_proj <- projectRaster(interval_raster_stack, crs=myCRS1)

# look at conditions at train occ locations (find nulls, or pts that are not loc on variable values)
conditions_occ <- raster::extract(interval_raster_proj,tax_test) # Use the thinned, cleaned taxa data and the env stacked rasters to extract variable vals
bad_rec <- is.na(conditions_occ[,1]) # Get the locations of those extracted pts with NA (no vals)
table(bad_rec) # Look at the list of NA locations
conditions_occ[bad_rec,.] # Another way to look at it

# remove null values from occ
occ <- tax_test[!bad_rec,.] # take out the pts that are over null vals
occ # Check result

# extract env values for occurrences
occ <- cbind(occ, extract(interval_raster_proj, occ))

#Plot final data:
par(mfrow=c(1,1))
plot(interval_raster_proj[[1]]) # plot env layer
plot(occ, add=T, col="red") # Plot occ pts

output_list <- list(occ, interval_raster_proj)
# Use function to prep the data for further analysis: output is list of occ then env data
conv_redu_prep <- run_data_prep(conv_redu_name, redu_env_shared)
conv_campmid_prep <- run_data_prep(conv_campmid_name, campmid_env_shared)
conv_nebr_prep <- run_data_prep(conv_nebr_name, nebr_env_shared)
conv_campup_prep <- run_data_prep(conv_campup_name, campup_env_shared)

bara_rees_prep <- run_data_prep(bara_rees_name, rees_env_shared)
bara_nebr_prep <- run_data_prep(bara_nebr_name, nebr_env_shared)
bara_campup_prep <- run_data_prep(bara_campup_name, campup_env_shared)

sage_comp_prep <- run_data_prep(sage_comp_name, comp_env_shared)
sage_rees_prep <- run_data_prep(sage_rees_name, rees_env_shared)
sage_campup_prep <- run_data_prep(sage_campup_name, campup_env_shared)

subl_redu_prep <- run_data_prep(subl_redu_name, redu_env_shared)
subl_campmid_prep <- run_data_prep(subl_campmid_name, campmid_env_shared)
subl_nebr_prep <- run_data_prep(subl_nebr_name, nebr_env_shared)
subl_campup_prep <- run_data_prep(subl_campup_name, campup_env_shared)

azer_macl_prep <- run_data_prep(azer_macl_name, macl_env_shared)
azer_campmid_prep <- run_data_prep(azer_campmid_name, campmid_env_shared)

oblo_rees_prep <- run_data_prep(oblo_rees_name, rees_env_shared)
oblo_campup_prep <- run_data_prep(oblo_campup_name, campup_env_shared)
sask_redu_prep <- run_data_prep(sask_redu_name,redu_env_shared)
sask_campmid_prep <- run_data_prep(sask_campmid_name,campmid_env_shared)

# Outgroup Taxa:

chla_nebr_prep <- run_data_prep(chla_nebr_name,nebr_env_shared)
chla_rees_prep <- run_data_prep(chla_rees_name,rees_env_shared)
chla_bacu_prep <- run_data_prep(chla_bacu_name,bacu_env_shared)

chla_campup_prep <- run_data_prep(chla_campup_name,campup_env_shared)
chla_maastrlow_prep <- run_data_prep(chla_maastrlow_name,maastrlow_env_shared)
chla_maastr_prep <- run_data_prep(chla_maastr_name,maastr_env_shared)

luci_nebr_prep <- run_data_prep(luci_nebr_name,nebr_env_shared)
luci_comp_prep <- run_data_prep(luci_comp_name,comp_env_shared)

luci_campup_prep <- run_data_prep(luci_campup_name,campup_env_shared)

cten_campmid_prep <- run_data_prep(cten_campmid_name,campmid_env_shared)
cten_campup_prep <- run_data_prep(cten_campup_name,campup_env_shared)

cten_perp_prep <- run_data_prep(cten_perp_name,perp_env_shared)
cten_redu_prep <- run_data_prep(cten_redu_name,redu_env_shared)
cten_nebr_prep <- run_data_prep(cten_nebr_name,nebr_env_shared)
cten_rees_prep <- run_data_prep(cten_rees_name,rees_env_shared)
# Additional taxa for analysis

cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)
cmon_rees_prep <- run_data_prep(cmon_rees_name, rees_env_shared)

# Function to prep genus level data for analysis

prep_genus <- function(gen_data, interval_raster_stack) {
  # Convert occ pts for taxa to spatial points
  coordinates(gen_data) = ~ Long + Lat

  # make sure to set spatial coor system (WGS 84)
  myCRS1 <- CRS("+init=epsg:4326")
  crs(gen_data) <- myCRS1
}
# project raster stack back to WGS 84 coor system
interval_raster_proj <- projectRaster(interval_raster_stack, crs=myCRS1)

# look at conditions at train occ locations (find nulls, or pts that are not loc on variable values)
conditions_occ <- raster::extract(interval_raster_proj,gen_data) # Use the thinned, cleaned taxa data and the
env stacked rasters to extract variable vals
bad_rec <- is.na(conditions_occ[,1]) # Get the locations of those extracted pts with NA (no vals)
table(bad_rec) # Look at the list of NA locations
conditions_occ[bad_rec,] # Another way to look at it

# remove null values from occ
occ <- gen_data[!bad_rec,] # take out the pts that are over null vals
occ # Check result

# extract env values for occurrences
occ <- cbind(occ, extract(interval_raster_proj, occ))

#Plot final data:
par(mfrow=c(1,1))
plot(interval_raster_proj[[1]]) # plot env layer
plot(occ, add=T, col="red") # Plot occ pts
output_list <- list(occ, interval_raster_proj)

return(output_list)

}
# Use above function to prep genus including no spp ID data
Inoceramus_all_redu_prep <- prep_genus(Inoceramus_all_redu,redu_env_shared)
Inoceramus_all_nebr_prep <- prep_genus(Inoceramus_all_nebr,nebr_env_shared)
Inoceramus_all_rees_prep <- prep_genus(Inoceramus_all_rees,rees_env_shared)

##### Run Ecospat analysis to compare overlap (make plots and tables) ######

# Function to run ecospat overlap comparison, calc overlap and other vals, and create PCA summary and overlap plots
run_ecospat_dyn <- function(first_occ_env_list,second_occ_env_list,spp_comp_title,spp_comp_pdf_name){
  ## Extract values to matrix for each raster stack of env proxy variables:
  first_bg<- getValues(first_occ_env_list[[2]])
  second_bg <- getValues(second_occ_env_list[[2]])

  ## Clean out missing values:
  first_bg <- first_bg[complete.cases(first_bg), ]
  second_bg <- second_bg[complete.cases(second_bg), ]

  ## Combined global environment:
  joined_bg <- rbind(first_bg, second_bg)

  # Run PCA analysis on combine BG pts (all env data)
  pca_joined_bg <- dudi.pca(joined_bg, center = TRUE,
    scale = TRUE, scannf = FALSE, nf = 2)
  joined_bg_scores <- pca_joined_bg$li # get the pca scores

  # Start saving figures to pdf file
  pdf(paste("Figures/",spp_comp_pdf_name, "ecospat_overlap_comparison.pdf"))

  # look at variable contribution
ecospat.plot.contrib(contrib=pca_joined_bg$co, eigen=pca_joined_bg$eig)

# Map occurrence data into the 2d ordination (have to coerce spatialPointsDataFrame into data.frame)
# Explicitly match the colnames so only using the right ones
# Only selected the li element (doesn't include others)
first_occ_scores <-
suprow(pca_joined_bg,
       data.frame(first_occ_env_list[[1]])[, colnames(joined_bg)])$li
second_occ_scores <-
suprow(pca_joined_bg,
       data.frame(second_occ_env_list[[1]])[, colnames(joined_bg)])$li

# Map BG data to the 2d ordination
first_bg_scores <- suprow(pca_joined_bg, first_bg)$li
second_bg_scores <- suprow(pca_joined_bg, second_bg)$li

# density distribution for first interval
grid_first <- ecospat.grid.clim.dyn(
    glob = joined_bg_scores,
    glob1 = first_bg_scores,
    sp = first_occ_scores,
    R = 100,
    th.sp = 0
)

# density distribution for second interval
grid_second <- ecospat.grid.clim.dyn(
    glob = joined_bg_scores,
    glob1 = second_bg_scores,
Supplementary Materials for Purcell and Myers, 20XX

sp = second_occ_scores,
R = 100,
th.sp = 0
)

D_overlap <- ecospat.niche.overlap (grid_first, grid_second, cor=T)
# Schoener's D metric and I metric

# Make summary table of D and I metrics
d_i_overlap <- matrix(nrow=1,ncol=2)
d_i_overlap[1,1] <- D_overlap[[1]]
d_i_overlap[1,2] <- D_overlap[[2]]
colnames(d_i_overlap) <- c("D_val","I_val")

# Finally we're ready to do the Niche Quantification/Comparisons. We'll use the
# PCA scores for the global environment, the native and invasive environments,
# and the native and invasive occurrence records

first_grid <- ecospat.grid.clim.dyn(joined_bg_scores,
    first_bg_scores,
    first_occ_scores)

second_grid <- ecospat.grid.clim.dyn(joined_bg_scores,
    second_bg_scores,
    second_occ_scores)

ecospat.plot.niche.dyn(first_grid,title=paste(spp_comp_title,"nEcospat Overlap Comparison"), second_grid,
    quant = 0.05, name.axis1="PC1",name.axis2="PC2")
ecospat.shift.centroids(first_occ_scores, second_occ_scores, first_bg_scores, second_bg_scores)
## Niche Equivalency Test

## Obs: observed overlaps, sim: simulated overlaps, p.D: pvalue of the test on D
## p.I: pvalue of the test on I

## Test for greater or lower equivalency

```r
eq_testgr <- ecospat.niche.equivalency.test(first_grid, second_grid,
rep=500, overlap.alternative = "higher",
expansion.alternative = "lower",
stability.alternative = "higher",
unfilling.alternative = "lower",
ncores=4) ##rep = 1000 recommended for operational runs

eq_testlw <- ecospat.niche.equivalency.test(first_grid, second_grid,
rep=500, overlap.alternative = "lower",
expansion.alternative = "higher",
stability.alternative = "lower",
unfilling.alternative = "higher",
ncores=4) ##rep = 1000 recommended for operational runs
```

# write D and I values of equivalency test

```r
EQ_DI = cbind("Obs_D"=eq_testgr$obs$D,"Obs_I"=eq_testgr$obs$I)
```

# write p values of equivalency test

```r
```

# write p values of expansion, stability, and unfilling

```r
```

## Test for greater (niche conservatism) or lower (niche divergence) similarity

```r
sim_testgr <- ecospat.niche.similarity.test(first_grid, second_grid,
```
rep=1000, overlap.alternative = "higher",
expansion.alternative = "lower",
stability.alternative = "higher",
unfilling.alternative = "lower",
rnd.type=1,ncore=4)
sim_testlw <- ecospat.niche.similarity.test(first_grid, second_grid,
rep=1000, overlap.alternative = "lower",
expansion.alternative = "higher",
stability.alternative = "lower",
unfilling.alternative = "higher",
rnd.type=1,ncore=4)

# write D and I values of equivalency test
SIM_DI = cbind("Obs_D"=sim_testgr$obs$D,"Obs_I"=sim_testgr$obs$I)

# write p values of similarity test
p_SIM_DI =

# write p values of expansion, stability, and unfilling
p_SIM_dynam =

# Plot test distributions
ecospat.plot.overlap.test(eq_testgr, "D", "Greater Equivalency")
ecospat.plot.overlap.test(eq_testlw, "D", "Lower Equivalency")
ecospat.plot.overlap.test(sim_testgr, "D", "Greater Similarity")
ecospat.plot.overlap.test(sim_testlw, "D", "Lower Similarity")
dev.off()  # close the pdf so will stop adding plots

## Plot niche space for first (grid.clim1) and second (grid.clim2) intervals (NOT WORKING, THEY MIGHT BE FIXING IT)
# ecospat.plot.niche(grid_first, title=spp_first_inter_name, name.axis1='PC1', name.axis2='PC2', cor=FALSE)
# ecospat.plot.niche(grid_second, title=spp_second_inter_name, name.axis1='PC1', name.axis2='PC2')

### Look at niche expansion, stability, and unfilling
# NA=analysis on entire area, 0=analysis on only overlapping, 0.05=analysis on 5th quantile intersection
dynam_allAreas = ecospat.niche.dyn.index (grid_first, grid_second, intersection=NA)
dynam_overlap = ecospat.niche.dyn.index (grid_first, grid_second, intersection=0)

# write csv of niche dynamic percentages
Niche_Ex_St = cbind("AllAreas"=dynam_allAreas$dynamic.index.w,
                     "Overlapping"=dynam_overlap$dynamic.index.w)

list_outputs <- list(d_i_overlap,eq_testgr,eq_testlw,sim_testgr,sim_testlw,Niche_Ex_St)
names(list_outputs) <-
                  c("D_I_Overlap","Eq_Test_Great","Eq_Test_Low","Sim_Test_Great","Sim_Test_Low","Niche_dynamics")
return(list_outputs)

### TEMPORAL RESOLUTION COMPARISON ####
# Use above function to make plots of overlap comparison and produce list of overlap metric results
# (output is in the form of a list including 1-D/I values of overlap, 2-Greater Equiv results (also list),
# 3-Lower Equiv results, 4-Greater Similarity results, 5-Lower Similarity results, and 6-niche dynamics)
conv_redu_campmid_overlap <- run_ecospat_dyn(conv_redu_prep,conv_campmid_prep, 
                                              "I. convextus B. reduncus-B. scotti vs. Middle Campanian","Iconv_redu_to_midcamp")
conv_nebr_campup_overlap <- run_ecospat_dyn(conv_nebr_prep,conv_campup_prep,
                                           "I. convextus D. nebrascense-E. jenneyi\nvs. Upper Campanian","Iconv_nebr_to_upcamp")

bara_rees_campup_overlap <- run_ecospat_dyn(bara_rees_prep,bara_campup_prep,
                                          "I. barabini\nB. reesidei-B. eliasi\nvs. Upper Campanian","Ibara_rees_to_upcamp")

bara_nebr_campup_overlap <- run_ecospat_dyn(bara_nebr_prep,bara_campup_prep,
                                           "I. barabini\nD. nebrascense-E. jenneyi\nvs. Upper Campanian","Ibara_nebr_to_upcamp")

sage_rees_campup_overlap <- run_ecospat_dyn(sage_rees_prep,sage_campup_prep,
                                          "I. sagensis\nB. reesidei-B. eliasi\nvs. Upper Campanian","Isage_rees_to_upcamp")

sage_comp_campup_overlap <- run_ecospat_dyn(sage_comp_prep,sage_campup_prep,
                                          "I. sagensis\nB. compressus-B. cuneatus\nvs. Upper Campanian","Isage_comp_to_upcamp")

subl_redu_campmid_overlap <- run_ecospat_dyn(subl_redu_prep,subl_campmid_prep,
                                           "I. sublaevis\nB. reduncus-B. scotti\nvs. Middle Campanian","Isabl_redu_to_midcamp")

subl_nebr_campup_overlap <- run_ecospat_dyn(subl_nebr_prep,subl_campup_prep,
                                           "I. sublaevis\nD. nebrascense-E. jenneyi\nvs. Upper Campanian","Isabl_nebr_to_upcamp")

azer_macl_campmid_overlap <- run_ecospat_dyn(azer_macl_prep,azer_campmid_prep,
                                           "I. azerbaydjanensis\nB. maclearni-B. sp. (smooth)\nvs. Middle Campanian","Iazer_macl_to_midcamp")

sask_redu_campmid_overlap <- run_ecospat_dyn(sask_redu_prep,sask_campmid_prep,
                                          "I. saskatchewanensis\nB. reduncus-B. scotti\nvs. Middle Campanian","Isask_redu_to_midcamp")
oblo_rees_campup_overlap <- run_ecospat_dyn(oblo_rees_prep,oblo_campup_prep,
   "I. oblongus\nB. reesidei-B. eliasi\nvvs. Upper Campanian","loblo_rees_to_upcamp")

cten_perp_campmid_overlap <- run_ecospat_dyn(cten_perp_prep,cten_campmid_prep,
   "C. imbricatula\nB. perplexus-B. gregoryensis\nvvs. Middle Campanian","Cimbr_perp_to_midcamp")

cten_redo_campmid_overlap <- run_ecospat_dyn(cten_redo_prep,cten_campmid_prep,
   "C. imbricatula\nB. reduncus-B. scotti\nvvs. Middle Campanian","Cimbr_redo_to_midcamp")

cten_nebr_campup_overlap <- run_ecospat_dyn(cten_nebr_prep,cten_campup_prep,
   "C. imbricatula\nD. nebrascense-E. jenneyi\nvvs. Upper Campanian","Cimbr_nebr_to_upcamp")

cten_rees_campup_overlap <- run_ecospat_dyn(cten_rees_prep,cten_campup_prep,
   "C. imbricatula\nB. reesidei-B. eliasi\nvvs. Upper Campanian","Cimbr_rees_to_upcamp")
luci_nebr_campup_overlap <- run_ecospat_dyn(luci_nebr_prep,luci_campup_prep,
   "L. subundata\nD. nebrascense-E. jenneyi\nvvs. Upper Campanian","Lsubu_nebr_to_upcamp")
luci_rees_campup_overlap <- run_ecospat_dyn(luci_comp_prep,luci_campup_prep,
   "L. subundata\nB. compressus-B. cuneatus\nvvs. Upper Campanian","Lsubu_comp_to_upcamp")
chla_nebr_campup_overlap <- run_ecospat_dyn(chla_nebr_prep,chla_campup_prep,
"C. nebrascensis\nD. nebrascense-E. jenneyi\nvs. Upper Campanian","Cnebr_nebr_to_upcamp")

chla_rees_campup_overlap <- run_ecospat_dyn(chla_rees_prep,chla_campup_prep,
                                           "C. nebrascensis\nB. reesidei-B. eliasi\nvs. Upper Campanian","Cnebr_rees_to_upcamp")

chla_bacu_maastrlow_overlap <- run_ecospat_dyn(chla_bacu_prep,chla_maastrlow_prep,
                                            "C. nebrascensis\nB. baculus-B. grandis\nvs. Lower Maastrichtian","Cnebr_bacu_to_lowmaas")

chla_bacu_maastr_overlap <- run_ecospat_dyn(chla_bacu_prep,chla_maastr_prep,
                                            "C. nebrascensis\nB. baculus-B. grandis\nvs. Maastrichtian","Cnebr_bacu_to_maas")

#### Summarize Temporal Resolution Comparisons and Export Tables ####
# Combine the results of observed overlap
make_D_I_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$obs$D,data1[[2]]$obs$I))
  colnames(named1) <- c("species","Intervals_compared","D_val","I_val")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

# Outgroup Tables
conv_redu_campmid_D_I_row <- make_D_I_row(conv_redu_campmid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")
conv_nebr_campup_D_I_row <- make_D_I_row(conv_nebr_campup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")
bara_rees_campup_D_I_row <- make_D_I_row(bara_rees_campup_overlap,"I. barabini","B. reesidei-B. eliasi to Upper Campanian")
bara_nebr_campup_D_I_row <- make_D_I_row(bara_nebr_campup_overlap, "I. barabini", "D. nebrascense-E. jenneyi to Upper Campanian")

sage_rees_campup_D_I_row <- make_D_I_row(sage_rees_campup_overlap, "I. sagensis", "B. reesidei-B. eliasi to Upper Campanian")

sage_comp_campup_D_I_row <- make_D_I_row(sage_comp_campup_overlap, "I. sagensis", "B. compressus-B. cuneatus to Upper Campanian")

subl_redu_campmid_D_I_row <- make_D_I_row(subl_redu_campmid_overlap, "I. sublaevis", "B. reduncus-B. scotti to Mid Campanian")

subl_nebr_campup_D_I_row <- make_D_I_row(subl_nebr_campup_overlap, "I. sublaevis", "D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_campmid_D_I_row <- make_D_I_row(azer_macl_campmid_overlap, "I. azerbaydjanensis", "B. maclearni-B. sp. (smooth) to Mid Campanian")

oblo_rees_campup_D_I_row <- make_D_I_row(oblo_rees_campup_overlap, "I. oblongus", "B. reesidei-B. eliasi to Upper Campanian")

sask_redu_campmid_D_I_row <- make_D_I_row(sask_macl_campmid_overlap, "I. saskatchewanensis", "B. reduncus-B. scotti to Mid Campanian")

temp_res_I_D_results <-
  rbind(conv_redu_campmid_D_I_row, conv_nebr_campup_D_I_row, bara_rees_campup_D_I_row, bara_nebr_campup_D_I_row, sage_rees_campup_D_I_row, sage_comp_campup_D_I_row, subl_redu_campmid_D_I_row, subl_nebr_campup_D_I_row, azer_macl_campmid_D_I_row, oblo_rees_campup_D_I_row, sask_redu_campmid_D_I_row)

# Outgroup tables

conv_redu_campmid_D_I_row <- make_D_I_row(cten_perp_campmid_overlap, "C. imbricatula", "B. perplexus-B. gregoryensis to Mid Campanian")

conv_nebr_campup_D_I_row <- make_D_I_row(cten_redu_campmid_overlap, "C. imbricatula", "B. reduncus-B. scotti to Upper Campanian")
bara_rees_campup_D_I_row <- make_D_I_row(cten_nebr_campup_overlap,"C. imbricatula","D. nebrascense-E. jenneyi to Upper Campanian")

bara_nebr_campup_D_I_row <- make_D_I_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")

sage_rees_campup_D_I_row <- make_D_I_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")

sage_comp_campup_D_I_row <- make_D_I_row(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")

subl_redu_campmid_D_I_row <- make_D_I_row(chla_nebr_campup_overlap,"C. nebrascensis","D. nebrascense-E. jenneyi to Upper Campanian")

subl_nebr_campup_D_I_row <- make_D_I_row(chla_rees_campup_overlap,"C. nebrascensis","B. reesidei-B. eliasi to Upper Campanian")

azer_macl_campmid_D_I_row <- make_D_I_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")

oblo_rees_campup_D_I_row <- make_D_I_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")

temp_res_outgroup_I_D_results <- rbind(conv_redu_campmid_D_I_row,conv_nebr_campup_D_I_row,bara_rees_campup_D_I_row,bara_nebr_campup_D_I_row,sage_rees_campup_D_I_row,sage_comp_campup_D_I_row,
subl_redu_campmid_D_I_row,subl_nebr_campup_D_I_row,azer_macl_campmid_D_I_row,oblo_rees_campup_D_I_row)

# Combine the results of equivalency tests
make_exp_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.D,data1[[2]]$p.I,data1[[3]]$p.D,data1[[3]]$p.I))
  colnames(named1) <- c("species","Intervals_compared","Greater_Eq_p_D","Greater_Eq_p_I","Lower_Eq_p_D","Lower_Eq_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
conv_redu_campmid_eq_row <- make_exp_p_row(conv_redu_campmid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")

conv_nebr_campup_eq_row <- make_exp_p_row(conv_nebr_campup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")

bara_rees_campup_eq_row <- make_exp_p_row(bara_rees_campup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")

bara_nebr_campup_eq_row <- make_exp_p_row(bara_nebr_campup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")

sage_rees_campup_eq_row <- make_exp_p_row(sage_rees_campup_overlap,"I. sagensis","B. reesidei-B. eliasi to Upper Campanian")

sage_comp_campup_eq_row <- make_exp_p_row(sage_comp_campup_overlap,"I. sagensis","B. compressus-B. cuneatus to Upper Campanian")

subl_redu_campmid_eq_row <- make_exp_p_row(subl_redu_campmid_overlap,"I. sublaevis","B. reduncus-B. scotti to Mid Campanian")

subl_nebr_campup_eq_row <- make_exp_p_row(subl_nebr_campup_overlap,"I. sublaevis","D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_campmid_eq_row <- make_exp_p_row(azer_macl_campmid_overlap,"I. azerbaydjanensis","B. maclearni-B. sp. (smooth) to Mid Campanian")

oblo_rees_campup_eq_row <- make_exp_p_row(oblo_rees_campup_overlap,"I. oblongus","B. reesidei-B. eliasi to Upper Campanian")

sask_redu_campmid_eq_row <- make_exp_p_row(sask_macl_campmid_overlap,"I. saskatchewanensis","B. reduncus-B. scotti to Mid Campanian")

temp_res_eq_results <- rbind(conv_redu_campmid_eq_row,conv_nebr_campup_eq_row,bara_rees_campup_eq_row,bara_nebr_campup_eq_row,sage_rees_campup_eq_row,sage_comp_campup_eq_row,subl_redu_campmid_eq_row,subl_nebr_campup_eq_row,azer_macl_campmid_eq_row,oblo_rees_campup_eq_row,sask_redu_campmid_eq_row)
# Outgroup tables

`conv_redu_campmid_eq_row <- make_exp_p_row(cten_perp_campmid_overlap,"C. imbricatula","B. perplexus-B. gregoryensis to Mid Campanian")`

`conv_nebr_campup_eq_row <- make_exp_p_row(cten_redu_campmid_overlap,"C. imbricatula","B. reduncus-B. scotti to Upper Campanian")`

`bara_rees_campup_eq_row <- make_exp_p_row(cten_nebr_campup_overlap,"C. imbricatula","D. nebrascense-E. jenneyi to Upper Campanian")`

`bara_nebr_campup_eq_row <- make_exp_p_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")`

`sage_rees_campup_eq_row <- make_exp_p_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")`

`sage_comp_campup_eq_row <- make_exp_p_row(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")`

`subl_redu_campmid_eq_row <- make_exp_p_row(chla_nebr_campup_overlap,"C. nebrascensis","D. nebrascense-E. jenneyi to Upper Campanian")`

`subl_nebr_campup_eq_row <- make_exp_p_row(chla_rees_campup_overlap,"C. nebrascensis","B. reesidei-B. eliasi to Upper Campanian")`

`azer_macl_campmid_eq_row <- make_exp_p_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")`

`oblo_rees_campup_eq_row <- make_exp_p_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")`

`temp_res_outgroup_eq_results <- rbind(conv_redu_campmid_eq_row,conv_nebr_campup_eq_row,bara_rees_campup_eq_row,bara_nebr_campup_eq_row,sage_rees_campup_eq_row,sage_comp_campup_eq_row,subl_redu_campmid_eq_row,subl_nebr_campup_eq_row,azer_macl_campmid_eq_row,oblo_rees_campup_eq_row)}`
# Combine the results of equivalency tests of Dynamics (p-vals)

```r
make_exp_dyn_p_row <- function(data1,name,interval){
  named1 <-
  as.data.frame(cbind(name,interval,data1[[2]]$p.expansion,data1[[2]]$p.stability,data1[[2]]$p.unfilling,
  data1[[3]]$p.expansion,data1[[3]]$p.stability,data1[[3]]$p.unfilling))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Eq_p_expansion","Greater_Eq_p_stability","Greater_Eq_p_unfilling",
  "Lower_Eq_p_expansion","Lower_Eq_p_stability","Lower_Eq_p_unfilling")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_redu_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(conv_redu_campmid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")

conv_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(conv_nebr_campup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")

bara_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(bara_rees_campup_overlap,"I. barabini","B. reesidei-B. eliasi to Upper Campanian")

bara_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(bara_nebr_campup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")

sage_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(sage_rees_campup_overlap,"I. sagensis","B. reesidei-B. eliasi to Upper Campanian")

sage_comp_campup_eq_dyn_p_row <- make_exp_dyn_p_row(sage_comp_campup_overlap,"I. sagensis","B. compressus-B. cuneatus to Upper Campanian")

subl_redu_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(subl_redu_campmid_overlap,"I. sublaevis","B. reduncus-B. scotti to Mid Campanian")

subl_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(subl_nebr_campup_overlap,"I. sublaevis","D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(azer_macl_campmid_overlap,"I. azerbaydjanensis","B. maclearni-B. sp. (smooth) to Mid Campanian")
```

Supplementary Materials for Purcell and Myers, 20XX

678
oblo_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(oblo_rees_campup_overlap,"I. oblongus","B. reesidei-B. eliasi to Upper Campanian")

sask_redu_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(sask_macl_campmid_overlap,"I. saskatchewanensis","B. reduncus-B. scotti to Mid Campanian")

temp_res_eq_dyn_p_results <- 
  rbind(conv_redu_campmid_eq_dyn_p_row,conv_nebr_campup_eq_dyn_p_row,bara_rees_campup_eq_dyn_p_row,
        bara_nebr_campup_eq_dyn_p_row,sage_rees_campup_eq_dyn_p_row,sage_comp_campup_eq_dyn_p_row,
        subl_redu_campmid_eq_dyn_p_row,subl_nebr_campup_eq_dyn_p_row,azer_macl_campmid_eq_dyn_p_row,
        oblo_rees_campup_eq_dyn_p_row,sask_redu_campmid_eq_dyn_p_row)

# Outgroup tables

conv_redu_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(cten_perp_campmid_overlap,"C. imbricatula","B. perplexus-B. gregoryensis to Mid Campanian")

conv_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(cten_redu_campmid_overlap,"C. imbricatula","B. reduncus-B. scotti to Upper Campanian")

bara_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(cten_nebr_campup_overlap,"C. imbricatula","D. nebrascense-E. jenneyi to Upper Campanian")

bara_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")

sage_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")

sage_comp_campup_eq_dyn_p_row <- make_expDynPRow(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")

subl_redu_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(chla_nebr_campup_overlap,"C. nebrasensis","D. nebrascense-E. jenneyi to Upper Campanian")

subl_nebr_campup_eq_dyn_p_row <- make_exp_dyn_p_row(chla_rees_campup_overlap,"C. nebrasensis","B. reesidei-B. eliasi to Upper Campanian")
azer_macl_campmid_eq_dyn_p_row <- make_exp_dyn_p_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")

oblo_rees_campup_eq_dyn_p_row <- make_exp_dyn_p_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")


# Combine the results of similarity tests
make_sim_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[4]]$p.D,data1[[4]]$p.I,data1[[5]]$p.D,data1[[5]]$p.I))
  colnames(named1) <- c("species","Intervals_compared","Greater_Sim_p_D","Greater_Sim_p_I","Lower_Sim_p_D","Lower_Sim_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_redu_campmid_sim_row <- make_sim_p_row(conv_redu_campmid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")
conv_nebr_campup_sim_row <- make_sim_p_row(conv_nebr_campup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")
bara_rees_campup_sim_row <- make_sim_p_row(bara_rees_campup_overlap,"I. barabini","B. reesidei-B. eliasi to Upper Campanian")
bara_nebr_campup_sim_row <- make_sim_p_row(bara_nebr_campup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")
sage_rees_campup_sim_row <- make_sim_p_row(sage_rees_campup_overlap,"I. sagensis","B. reesidei-B. eliasi to Upper Campanian")

sage_comp_campup_sim_row <- make_sim_p_row(sage_comp_campup_overlap,"I. sagensis","B. compressus-B. cuneatus to Upper Campanian")

subl_redu_campmid_sim_row <- make_sim_p_row(subl_redu_campmid_overlap,"I. sublaevis","B. reduncus-B. scotti to Mid Campanian")

subl_nebr_campup_sim_row <- make_sim_p_row(subl_nebr_campup_overlap,"I. sublaevis","D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_campmid_sim_row <- make_sim_p_row(azer_macl_campmid_overlap,"I. azerbaydjanensis","B. maclearni-B. sp. (smooth) to Mid Campanian")

oblo_rees_campup_sim_row <- make_sim_p_row(oblo_rees_campup_overlap,"I. oblongus","B. reesidei-B. eliasi to Upper Campanian")

sask_redu_campmid_sim_row <- make_sim_p_row(sask_macl_campmid_overlap,"I. saskatchewanensis","B. reduncus-B. scotti to Mid Campanian")

temp_res_sim_results <-
rbind(conv_redu_campmid_sim_row,conv_nebr_campup_sim_row,bara_rees_campup_sim_row,
     bara_nebr_campup_sim_row,sage_rees_campup_sim_row,sage_comp_campup_sim_row,
     subl_redu_campmid_sim_row,subl_nebr_campup_sim_row,azer_macl_campmid_sim_row,
     oblo_rees_campup_sim_row,sask_redu_campmid_sim_row)

# Outgroup tables

conv_redu_campmid_sim_row <- make_sim_p_row(cten_perp_campmid_overlap,"C. imbricatula","B. perplexus-B. gregoryensis to Mid Campanian")

conv_nebr_campup_sim_row <- make_sim_p_row(cten_redu_campmid_overlap,"C. imbricatula","B. reduncus-B. scotti to Upper Campanian")

bara_rees_campup_sim_row <- make_sim_p_row(cten_nebr_campup_overlap,"C. imbricatula","D. nebrascense-E. jenneyi to Upper Campanian")
bara_nebr_campup_sim_row <- make_sim_p_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")

sage_rees_campup_sim_row <- make_sim_p_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")

sage_comp_campup_sim_row <- make_sim_p_row(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")

subl_redu_campmid_sim_row <- make_sim_p_row(chla_nebr_campup_overlap,"C. nebrascensis","D. nebrascense-E. jenneyi to Upper Campanian")

subl_nebr_campup_sim_row <- make_sim_p_row(chla_rees_campup_overlap,"C. nebrascensis","B. reesidei-B. eliasi to Upper Campanian")

azer_macl_campmid_sim_row <- make_sim_p_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")

oblo_rees_campup_sim_row <- make_sim_p_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")

temp_res_outgroup_sim_results <-
  rbind(conv_redu_campmid_sim_row,conv_nebr_campup_sim_row,bara_rees_campup_sim_row,
        bara_nebr_campup_sim_row,sage_rees_campup_sim_row,sage_comp_campup_sim_row,
        subl_redu_campmid_sim_row,subl_nebr_campup_sim_row,azer_macl_campmid_sim_row,
        oblo_rees_campup_sim_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_sim_dyn_p_row <- function(data1,name,interval){
  named1 <-
  as.data.frame(cbind(name,interval,data1[[4]]$p.expansion,data1[[4]]$p.stability,data1[[4]]$p.unfilling,
                 data1[[5]]$p.expansion,data1[[5]]$p.stability,data1[[5]]$p.unfilling))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Sim_p_expansion","Greater_Sim_p_stability","Greater_Sim_p_unfilling",
    "Lower_Sim_p_expansion","Lower_Sim_p_stability","Lower_Sim_p_unfilling")

  named1}
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))

return(named1)

conv_redu_cammpid_sim_dyn_p_row <- make_sim_dyn_p_row(conv_redu_cammpid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")

conv_nebr_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(conv_nebr_cammpup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")

bara_rees_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(bara_rees_cammpup_overlap,"I. barabini","B. reesidei-B. eliasi to Upper Campanian")

bara_nebr_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(bara_nebr_cammpup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")

sage_rees_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(sage_rees_cammpup_overlap,"I. sagensis","B. reesidei-B. eliasi to Upper Campanian")

sage_comp_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(sage_comp_cammpup_overlap,"I. sagensis","B. compressus-B. cuneatus to Upper Campanian")

subl_redu_cammpid_sim_dyn_p_row <- make_sim_dyn_p_row(subl_redu_cammpid_overlap,"I. sublaevis","B. reduncus-B. scotti to Mid Campanian")

subl_nebr_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(subl_nebr_cammpup_overlap,"I. sublaevis","D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_cammpid_sim_dyn_p_row <- make_sim_dyn_p_row(azer_macl_cammpid_overlap,"I. azerbaydjanensis","B. maclearni-B. sp. (smooth) to Mid Campanian")

oblo_rees_cammpup_sim_dyn_p_row <- make_sim_dyn_p_row(oblo_rees_cammpup_overlap,"I. oblongus","B. reesidei-B. eliasi to Upper Campanian")

sask_redu_cammpid_sim_dyn_p_row <- make_sim_dyn_p_row(sask_macl_cammpid_overlap,"I. saskatchewanensis","B. reduncus-B. scotti to Mid Campanian")
temp_res_sim_dyn_p_results <-
  rbind(conv_redu_campmid_sim_dyn_p_row,conv_nebr_campup_sim_dyn_p_row,bara_rees_campup_sim_dyn_p_row,
     bara_nebr_campup_sim_dyn_p_row,sage_rees_campup_sim_dyn_p_row,sage_comp_campup_sim_dyn_p_row,
     subl_redu_campmid_sim_dyn_p_row,subl_nebr_campup_sim_dyn_p_row,azer_macl_campmid_sim_dyn_p_row,
     oblo_rees_campup_sim_dyn_p_row,sask_redu_campmid_sim_dyn_p_row)

# Outgroup tables

conv_redu_campmid_sim_dyn_p_row <- make_sim_dyn_p_row(cten_perp_campmid_overlap,"C. imbricatula","B. perplexus-B. gregoryensis to Mid Campanian")
conv_nebr_campup_sim_dyn_p_row <- make_sim_dyn_p_row(cten_redu_campmid_overlap,"C. imbricatula","B. reduncus-B. scotti to Upper Campanian")
bara_nebr_campup_sim_dyn_p_row <- make_sim_dyn_p_row(cten_nebr_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")
bara_nebr_campup_sim_dyn_p_row <- make_sim_dyn_p_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")
sage_rees_campup_sim_dyn_p_row <- make_sim_dyn_p_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")
sage_comp_campup_sim_dyn_p_row <- make_sim_dyn_p_row(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")
subl_redu_campmid_sim_dyn_p_row <- make_sim_dyn_p_row(chla_nebr_campup_overlap,"C. nebrascensis","D. nebrascense-E. jenneyi to Upper Campanian")
subl_nebr_campup_sim_dyn_p_row <- make_sim_dyn_p_row(chla_rees_campup_overlap,"C. nebrascensis","B. reesidei-B. eliasi to Upper Campanian")
azer_macl_campmid_sim_dyn_p_row <- make_sim_dyn_p_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")
oblo_rees_campup_sim_dyn_p_row <- make_sim_dyn_p_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")
temp_res_outgroup_sim_dyn_p_results <-
  rbind(conv_redu_campmid_sim_dyn_p_row,conv_nebr_campup_sim_dyn_p_row,bara_rees_campup_sim_dyn
_p_row,
bara_nebr_campup_sim_dyn_p_row, sageres_campup_sim_dyn_p_row,sager_compcampup_sim_dyn_p_row,
subl_redu_campmid_sim_dyn_p_row,subl_nebr_campup_sim_dyn_p_row,azer_macl_campmid_sim_dyn_p_row,
  oblo_rees_campup_sim_dyn_p_row)

# Combine the results of niche dynamics
make_dynam_row <- function(data1,name,interval){
data <- as.data.frame(data1[[6]])
named1 <- as.data.frame(cbind(name,interval,t(data[,2])))
colnames(named1) <- c("species","Intervals_compared","Expansion","Stability","Unfilling")
named1[,c(3:5)] <- as.numeric(as.character(named1[,c(3:5)]))
return(named1)
}

conv_redu_campmid_dyn_row <- make_dynam_row(conv_redu_campmid_overlap,"I. convexus","B. reduncus-B. scotti to Mid Campanian")
conv_nebr_campup_dyn_row <- make_dynam_row(conv_nebr_campup_overlap,"I. convexus","D. nebrascense-E. jenneyi to Upper Campanian")

bara_rees_campup_dyn_row <- make_dynam_row(bara_rees_campup_overlap,"I. barabini","B. reesidei-B. eliasi to Upper Campanian")
bara_nebr_campup_dyn_row <- make_dynam_row(bara_nebr_campup_overlap,"I. barabini","D. nebrascense-E. jenneyi to Upper Campanian")
sage_rees_campup_dyn_row <- make_dynam_row(sage_rees_campup_overlap,"I. sagensis","B. reesidei-B. eliasi to Upper Campanian")
sage_comp_campup_dyn_row <- make_dynam_row(sage_comp_campup_overlap,"I. sagensis","B. compressus-B. cuneatus to Upper Campanian")
subl_redu_campmid_dyn_row <- make_dynam_row(subl_redu_campmid_overlap,"I. sublaevis","B. reduncus-B. scotti to Mid Campanian")

subl_nebr_campup_dyn_row <- make_dynam_row(subl_nebr_campup_overlap,"I. sublaevis","D. nebrascense-E. jenneyi to Upper Campanian")

azer_macl_campmid_dyn_row <- make_dynam_row(azer_macl_campmid_overlap,"I. azerbaydjanensis","B. maclearni-B. sp. (smooth) to Mid Campanian")

oblo_rees_campup_dyn_row <- make_dynam_row(oblo_rees_campup_overlap,"I. oblongus","B. reesidei-B. eliasi to Upper Campanian")

sask_redu_campmid_dyn_row <- make_dynam_row(sask_macl_campmid_overlap,"I. saskatchewanensis","B. reduncus-B. scotti to Mid Campanian")

temp_res_dyn_results <-
  rbind(conv_redu_campmid_dyn_row,conv_nebr_campup_dyn_row,bara_rees_campup_dyn_row,
        bara_nebr_campup_dyn_row,sage_rees_campup_dyn_row,sage_comp_campup_dyn_row,
        subl_redu_campmid_dyn_row,subl_nebr_campup_dyn_row,azer_macl_campmid_dyn_row,
        oblo_rees_campup_dyn_row,sask_redu_campmid_dyn_row)

# Outgroup tables

conv_redu_campmid_dyn_row <- make_dynam_row(cten_perp_campmid_overlap,"C. imbricatula","B. perplexus-B. gregoryensis to Mid Campanian")

conv_nebr_campup_dyn_row <- make_dynam_row(cten_redu_campmid_overlap,"C. imbricatula","B. reduncus-B. scotti to Upper Campanian")

bara_rees_campup_dyn_row <- make_dynam_row(cten_nebr_campup_overlap,"C. imbricatula","D. nebrascense-E. jenneyi to Upper Campanian")

bara_nebr_campup_dyn_row <- make_dynam_row(cten_rees_campup_overlap,"C. imbricatula","B. reesidei-B. eliasi to Upper Campanian")

sage_rees_campup_dyn_row <- make_dynam_row(luci_nebr_campup_overlap,"L. subundata","D. nebrascense-E. jenneyi to Upper Campanian")

sage_comp_campup_dyn_row <- make_dynam_row(luci_rees_campup_overlap,"L. subundata","B. reesidei-B. eliasi to Upper Campanian")
subl_redu_campmid_dyn_row <- make_dynam_row(chla_nebr_campup_overlap,"C. nebrascensis","D. nebrascense-E. jenneyi to Upper Campanian")

subl_nebr_campup_dyn_row <- make_dynam_row(chla_rees_campup_overlap,"C. nebrascensis","B. reesidei-B. eliasi to Upper Campanian")

azer_macl_campmid_dyn_row <- make_dynam_row(chla_bacu_maastrlow_overlap,"C. nebrascensis","B. baculus-B. grandis to Lower Maastrichtian")

oblo_rees_campup_dyn_row <- make_dynam_row(chla_bacu_maastr_overlap,"C. nebrascensis","B. baculus-B. grandis to Maastrichtian")

temp_res_outgroup_dyn_results <- rbind(conv_redu_campmid_dyn_row,conv_nebr_campup_dyn_row,bara_rees_campup_dyn_row,
                                      bara_nebr_campup_dyn_row,sage_rees_campup_dyn_row,sage_comp_campup_dyn_row,
                                      subl_redu_campmid_dyn_row,subl_nebr_campup_dyn_row,azer_macl_campmid_dyn_row,
                                      oblo_rees_campup_dyn_row)

## Write csv files of results
write.csv(temp_res_outgroup_I_D_results,file="tables/temp_res_outgroup_I_D_results.csv")
write.csv(temp_res_outgroup_eq_results,file="tables/temp_res_outgroup_eq_results.csv")
write.csv(temp_res_outgroup_sim_results,file="tables/temp_res_outgroup_sim_results.csv")
write.csv(temp_res_outgroup_dyn_results,file="tables/temp_res_outgroup_dyn_results.csv")
write.csv(temp_res_outgroup_eq_dyn_p_results,file="tables/temp_res_outgroup_eq_dyn_p_results.csv")
write.csv(temp_res_outgroup_sim_dyn_p_results,file="tables/temp_res_outgroup_sim_dyn_p_results.csv")

#### SPECIES WITHIN SAME INTERVAL COMPARISON ####

# Use above function to make plots of overlap comparison between spp and produce list of overlap metric results

conv_to_subl_redu_overlap <- run_ecospat_dyn(conv_redu_prep,subl_redu_prep,
                                           "I. convextus to I. sublaevis\nB. reduncus-B. scotti","Iconv_to_Isubl_redu")
conv_to_sask_redu_overlap <- run_ecospat_dyn(conv_redu_prep,sask_redu_prep,
   "I. convextus to I. saskatchewanensis\nB. reduncus-B. scotti","Iconv_to_Isask_redu_")

subl_to_sask_redu_overlap <- run_ecospat_dyn(subl_redu_prep,sask_redu_prep,
   "I. sublaevis to I. saskatchewanensis\nB. reduncus-B. scotti","Isubl_to_Isask_redu_")

conv_to_bara_nebr_overlap <- run_ecospat_dyn(conv_nebr_prep,bara_nebr_prep,
   "I. convextus to I. barabini\nD. nebrascense-E. jenneyi","Iconv_to_Ibara_nebr")

conv_to_subl_nebr_overlap <- run_ecospat_dyn(conv_nebr_prep,subl_nebr_prep,
   "I. convextus to I. sublaevis\nD. nebrascense-E. jenneyi","Iconv_to_Isubl_nebr")

bara_to_subl_nebr_overlap <- run_ecospat_dyn(bara_nebr_prep,subl_nebr_prep,
   "I. barabini to I. sublaevis\nD. nebrascense-E. jenneyi","Ibara_to_Isubl_nebr")

bara_to_sage_rees_overlap <- run_ecospat_dyn(bara_rees_prep,sage_rees_prep,
   "I. barabini to I. sagensis\nB. reesidei-B. eliasi","Ibara_to_Isage_rees")

bara_to_oblo_rees_overlap <- run_ecospat_dyn(bara_rees_prep,oblo_rees_prep,
   "I. barabini to I. oblongus\nB. reesidei-B. eliasi","Ibara_to_Ioblo_rees")

oblo_to_sage_rees_overlap <- run_ecospat_dyn(oblo_rees_prep,sage_rees_prep,
   "I. oblongus to I. sagensis\nB. reesidei-B. eliasi","Ioblo_to_Isage_rees")

#### Summarize Species Comparisons and Export Tables ####

# Combine the results of observed overlap
make_D_I_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$obs$D,data1[[2]]$obs$I))
  colnames(named1) <- c("species","Intervals_compared","D_val","I_val")
}
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
return(named1)
}

conv_to_subl_redu_D_I_row <- make_D_I_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")

conv_to_sask_redu_D_I_row <- make_D_I_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")

subl_to_sask_redu_D_I_row <- make_D_I_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_bara_nebr_D_I_row <- make_D_I_row(conv_to_bara_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")

conv_to_subl_nebr_D_I_row <- make_D_I_row(conv_to_subl_nebr_overlap,"I. convexus vs. I. sublaevis","B. reesidei-B. eliasi")

bara_to_subl_nebr_D_I_row <- make_D_I_row(bara_to_subl_nebr_overlap,"I. barabini vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_sage_rees_D_I_row <- make_D_I_row(bara_to_sage_rees_overlap,"I. barabini vs. I. sagensis","B. reesidei-B. eliasi")

bara_to_oblo_rees_D_I_row <- make_D_I_row(bara_to_oblo_rees_overlap,"I. barabini vs. I. oblongus","B. reesidei-B. eliasi")

oblo_to_sage_rees_D_I_row <- make_D_I_row(oblo_to_sage_rees_overlap,"I. oblongus vs. I. sagensis","B. reesidei-B. eliasi")

species_compare_I_D_results <- rbind(conv_to_subl_redu_D_I_row,conv_to_sask_redu_D_I_row,
                                      subl_to_sask_redu_D_I_row,conv_to_bara_nebr_D_I_row,
                                      conv_to_subl_nebr_D_I_row,
                                      bara_to_subl_nebr_D_I_row,
                                      bara_to_sage_rees_D_I_row,
                                      bara_to_oblo_rees_D_I_row,
                                      oblo_to_sage_rees_D_I_row)
# Combine the results of equivalency tests

```r
make_exp_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.D,data1[[2]]$p.I,data1[[3]]$p.D,data1[[3]]$p.I))
  colnames(named1) <- c("species","Intervals_compared","Greater_Eq_p_D","Greater_Eq_p_I","Lower_Eq_p_D","Lower_Eq_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}
```

```r
conv_to_subl_eq_row <- make_exp_p_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")
```

```r
conv_to_sask_eq_row <- make_exp_p_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")
```

```r
subl_to_sask_eq_row <- make_exp_p_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")
```

```r
conv_to_bara_eq_row <- make_exp_p_row(conv_to_bara_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")
```

```r
conv_to_subl_eq_row <- make_exp_p_row(conv_to_subl_nebr_overlap,"I. convexus vs. I. sublaevis","D. nebrascense-E. jenneyi")
```

```r
bara_to_subl_eq_row <- make_exp_p_row(bara_to_subl_nebr_overlap,"I. barabini vs. I. sublaevis","D. nebrascense-E. jenneyi")
```

```r
bara_to_sage_eq_row <- make_exp_p_row(bara_to_sage_rees_overlap,"I. barabini vs. I. sagensis","B. reesidei-B. eliasi")
```

```r
bara_to_oblo_eq_row <- make_exp_p_row(bara_to_oblo_rees_overlap,"I. barabini vs. I. oblongus","B. reesidei-B. eliasi")
```

oblo_to_sage_eq_row <- make_exp_p_row(oblo_to_sage_rees_overlap,"I. oblongus vs. I. sagensis","B. reesidei-B. eliasi")

species_compare_eq_results <- rbind(conv_to_subl_eq_row,conv_to_sask_eq_row,
                                      subl_to_sask_eq_row,conv_to_bara_eq_row,
                                      conv_to_subl_eq_row,
                                      bara_to_subl_eq_row,bara_to_sage_eq_row,
                                      bara_to_oblo_eq_row,oblo_to_sage_eq_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_exp_dyn_p_row <- function(data1,name,interval){
  named1 <-
  as.data.frame(cbind(name,interval,data1[[2]]$p.expansion,data1[[2]]$p.stability,data1[[2]]$p.unfilling,
                  data1[[3]]$p.expansion,data1[[3]]$p.stability,data1[[3]]$p.unfilling))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Eq_p_expansion","Greater_Eq_p_stability","Greater_Eq_p_unfilling",
    "Lower_Eq_p_expansion","Lower_Eq_p_stability","Lower_Eq_p_unfilling")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}
conv_to_subl_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")

conv_to_sask_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")

subl_to_sask_eq_dyn_p_row <- make_exp_dyn_p_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_bara_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_bara_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")
conv_to_subl_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_subl_nebr_overlap, "I. convexus vs. I. sublaevis", "D. nebrascense-E. jenneyi")

bara_to_subl_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_subl_nebr_overlap, "I. barabini vs. I. sublaevis", "D. nebrascense-E. jenneyi")

bara_to_sage_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_sage_rees_overlap, "I. barabini vs. I. sagensis", "B. reesidei-B. eliasi")

bara_to_oblo_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_oblo_rees_overlap, "I. barabini vs. I. oblongus", "B. reesidei-B. eliasi")

oblo_to_sage_eq_dyn_p_row <- make_exp_dyn_p_row(oblo_to_sage_rees_overlap, "I. oblongus vs. I. sagensis", "B. reesidei-B. eliasi")

species_compare_eq_dyn_p_results <- rbind(conv_to_subl_eq_dyn_p_row, conv_to_sask_eq_dyn_p_row, subl_to_sask_eq_dyn_p_row, conv_to_bara_eq_dyn_p_row, conv_to_subl_eq_dyn_p_row, bara_to_subl_eq_dyn_p_row, bara_to_sage_eq_dyn_p_row, bara_to_oblo_eq_dyn_p_row, oblo_to_sage_eq_dyn_p_row)

# Combine the results of similarity tests
make_sim_p_row <- function(data1, name, interval) {
  named1 <- as.data.frame(cbind(name, interval, data1[[4]]$p.D, data1[[4]]$p.I, data1[[5]]$p.D, data1[[5]]$p.I))
  colnames(named1) <- c("species", "Intervals_compared", "Greater_Sim_p_D", "Greater_Sim_p_I", "Lower_Sim_p_D", "Lower_Sim_p_I")
  named1[, c(3:4)] <- as.numeric(as.character(named1[, c(3:4)]))
  return(named1)
}

692
conv_to_subl_sim_row <- make_sim_p_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")

conv_to_sask_sim_row <- make_sim_p_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")

subl_to_sask_sim_row <- make_sim_p_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_bara_sim_row <- make_sim_p_row(conv_to_bara_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")

conv_to_subl_sim_row <- make_sim_p_row(conv_to_subl_nebr_overlap,"I. convexus vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_subl_sim_row <- make_sim_p_row(bara_to_subl_nebr_overlap,"I. barabini vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_sage_sim_row <- make_sim_p_row(bara_to_sage_rees_overlap,"I. barabini vs. I. sagensis","B. reesidei-B. eliasi")

bara_to_oblo_sim_row <- make_sim_p_row(bara_to_oblo_rees_overlap,"I. barabini vs. I. oblongus","B. reesidei-B. eliasi")

oblo_to_sage_sim_row <- make_sim_p_row(oblo_to_sage_rees_overlap,"I. oblongus vs. I. sagensis","B. reesidei-B. eliasi")

species_compare_sim_results <- rbind(conv_to_subl_sim_row,conv_to_sask_sim_row,
                                      subl_to_sask_sim_row,conv_to_bara_sim_row,
                                      conv_to_subl_sim_row,
                                      bara_to_subl_sim_row,bara_to_sage_sim_row,
                                      bara_to_oblo_sim_row,oblo_to_sage_sim_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_sim_dyn_p_row <- function(data1,name,interval){
  named1 <-
  as.data.frame(cbind(name,interval,data1[[4]]$p.expansion,data1[[4]]$p.stability,data1[[4]]$p.unfilling,
                 data1[[5]]$p.expansion,data1[[5]]$p.stability,data1[[5]]$p.unfilling))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Sim_p_expansion","Greater_Sim_p_stability","Greater_Sim_p_unfilling",
    "Lower_Sim_p_expansion","Lower_Sim_p_stability","Lower_Sim_p_unfilling")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_subl_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")

conv_to_sask_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")

subl_to_sask_sim_dyn_p_row <- make_sim_dyn_p_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_barab_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_barab_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")

conv_to_subl_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_subl_nebr_overlap,"I. convexus vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_subl_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_subl_nebr_overlap,"I. barabini vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_sage_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_sage_rees_overlap,"I. barabini vs. I. sagensis","B. reesidei-B. eliasi")

bara_to_oblo_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_oblo_rees_overlap,"I. barabini vs. I. oblongus","B. reesidei-B. eliasi")
oblo_to_sage_sim_dyn_p_row <- make_sim_dyn_p_row(oblo_to_sage_rees_overlap,"I. oblongus vs. I. sagensis","B. reesidei-B. eliasi")

species_compare_sim_dyn_p_results <- rbind(conv_to_subl_sim_dyn_p_row,conv_to_sask_sim_dyn_p_row,
subl_to_sask_sim_dyn_p_row,conv_to_bara_sim_dyn_p_row,
conv_to_subl_sim_dyn_p_row,
bara_to_subl_sim_dyn_p_row,bara_to_sage_sim_dyn_p_row,
bara_to_oblo_sim_dyn_p_row,oblo_to_sage_sim_dyn_p_row)

# Combine the results of niche dynamics
make_dynam_row <- function(data1,name,interval){
  data <- as.data.frame(data1[[6]])
  named1 <- as.data.frame(cbind(name,interval,t(data[,2])))
  colnames(named1) <- c("species","Intervals_compared","Expansion","Stability","Unfilling")
  named1[,c(3:5)] <- as.numeric(as.character(named1[,c(3:5)]))
  return(named1)
}

conv_to_subl_dyn_row <- make_dynam_row(conv_to_subl_redu_overlap,"I. convexus vs. I. sublaevis","B. reduncus-B. scotti")

conv_to_sask_dyn_row <- make_dynam_row(conv_to_sask_redu_overlap,"I. convexus vs. I. saskatchewanensis","B. reduncus-B. scotti")

subl_to_sask_dyn_row <- make_dynam_row(subl_to_sask_redu_overlap,"I. sublaevis vs. I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_bara_dyn_row <- make_dynam_row(conv_to_bara_nebr_overlap,"I. convexus vs. I. barabini","D. nebrascense-E. jenneyi")

conv_to_subl_dyn_row <- make_dynam_row(conv_to_subl_nebr_overlap,"I. convexus vs. I. sublaevis","D. nebrascense-E. jenneyi")
bara_to_subl_dyn_row <- make_dynam_row(bara_to_subl_nebr_overlap,"I. barabini vs. I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_sage_dyn_row <- make_dynam_row(bara_to_sage_rees_overlap,"I. barabini vs. I. sagensis","B. reesidei-B. eliasi")

bara_to_oblo_dyn_row <- make_dynam_row(bara_to_oblo_rees_overlap,"I. barabini vs. I. oblongus","B. reesidei-B. eliasi")

oblo_to_sage_dyn_row <- make_dynam_row(oblo_to_sage_rees_overlap,"I. oblongus vs. I. sagensis","B. reesidei-B. eliasi")

species_compare_dyn_results <- rbind(conv_to_subl_dyn_row,conv_to_sask_dyn_row,
                                     subl_to_sask_dyn_row,conv_to_bara_dyn_row,
                                     conv_to_subl_dyn_row,
                                     bara_to_subl_dyn_row,bara_to_sage_dyn_row,
                                     bara_to_oblo_dyn_row,oblo_to_sage_dyn_row)

## Write csv files of results
write.csv(species_compare_I_D_results,file="tables/species_compare_I_D_results.csv")
write.csv(species_compare_eq_results,file="tables/species_compare_eq_results.csv")
write.csv(species_compare_sim_results,file="tables/species_compare_sim_results.csv")
write.csv(species_compare_dyn_results,file="tables/species_compare_dyn_results.csv")
write.csv(species_compare_eq_dyn_p_results,file="tables/species_compare_eq_dyn_p_results.csv")
write.csv(species_compare_sim_dyn_p_results,file="tables/species_compare_sim_dyn_p_results.csv")

#### INOCERAMUS TO OUTGROUP COMPARISONS ####

cten_to_conv_redu_overlap <- run_ecospat_dyn(cten_redu_prep,conv_redu_prep,
                                           "C. imbricatula to I. convexus\nB. reduncus-B. scotti","Cimbr_to_Iconv_redu_")

cten_to_subl_redu_overlap <- run_ecospat_dyn(cten_redu_prep,subl_redu_prep,
                                           "C. imbricatula to I. sublaevis\nB. reduncus-B. scotti","Cimbr_to_Isubl_redu_")
cten_to_sask_redu_overlap <- run_ecospat_dyn(cten_redu_prep,sask_redu_prep,
    "C. imbricatula to I. saskatchewanensis
B. reduncus-B. scotti","Cimbr_to_Isask_redu_")

c腾_to_bara_nebr_overlap <- run_ecospat_dyn(cten_nebr_prep,bara_nebr_prep,
    "C. imbricatula to I. barabini
D. nebrascense-E. jenneyi","Cimbr_to_Ibara_nebr_")

c腾_to_conv_nebr_overlap <- run_ecospat_dyn(cten_nebr_prep,conv_nebr_prep,
    "C. imbricatula to I. convexus
D. nebrascense-E. jenneyi","Cimbr_to_Iconv_nebr_")

c腾_to_subl_nebr_overlap <- run_ecospat_dyn(cten_nebr_prep,subl_nebr_prep,
    "C. imbricatula to I. sublaevis
D. nebrascense-E. jenneyi","Cimbr_to_Isubl_nebr_")

chla_to_bara_nebr_overlap <- run_ecospat_dyn(chla_nebr_prep,bara_nebr_prep,
    "C. nebrascensis to I. barabini
D. nebrascense-E. jenneyi","Cnebr_to_Ibara_nebr_")

chla_to_conv_nebr_overlap <- run_ecospat_dyn(chla_nebr_prep,conv_nebr_prep,
    "C. nebrascensis to I. convexus
D. nebrascense-E. jenneyi","Cnebr_to_Iconv_nebr_")

chla_to_subl_nebr_overlap <- run_ecospat_dyn(chla_nebr_prep,subl_nebr_prep,
    "C. nebrascensis to I. sublaevis
D. nebrascense-E. jenneyi","Cnebr_to_Isubl_nebr_")

luci_to_bara_nebr_overlap <- run_ecospat_dyn(luci_nebr_prep,bara_nebr_prep,
    "L. subundata to I. barabini
D. nebrascense-E. jenneyi","Lsubu_to_Ibara_nebr_")

luci_to_conv_nebr_overlap <- run_ecospat_dyn(luci_nebr_prep,conv_nebr_prep,
    "L. subundata to I. convexus
D. nebrascense-E. jenneyi","Lsubu_to_Iconv_nebr_")
luci_to_subl_nebr_overlap <- run_ecospat_dyn(luci_nebr_prep, subl_nebr_prep,
                                           "L. subundata to I. sublaevis
D. nebrascense-E. jenneyi", "Lsubu_to_Isubl_nebr_")

luci_to_sage_comp_overlap <- run_ecospat_dyn(luci_comp_prep, sage_comp_prep,
                                          "L. subundata to I. sagensis
B. compressus-B. cuneatus", "Lsubu_to_Isage_comp")

chla_to_bara_rees_overlap <- run_ecospat_dyn(chla_rees_prep, bara_rees_prep,
                                         "C. nebrascensis to I. barabini
B. reesidei-B. eliasi", "Cnebr_to_Ibara_rees")

chla_to_sage_rees_overlap <- run_ecospat_dyn(chla_rees_prep, sage_rees_prep,
                                         "C. nebrascensis to I. sagensis
B. reesidei-B. eliasi", "Cnebr_to_Isage_rees")

chla_to_oblo_rees_overlap <- run_ecospat_dyn(chla_rees_prep, oblo_rees_prep,
                                         "C. nebrascensis to I. oblongus
B. reesidei-B. eliasi", "Cnebr_to_Ioblo_rees")

cnten_to_bara_rees_overlap <- run_ecospat_dyn(cten_rees_prep, bara_rees_prep,
                                         "C. imbricatula to I. barabini
B. reesidei-B. eliasi", "Cimbr_to_Ibara_rees")

cnten_to_sage_rees_overlap <- run_ecospat_dyn(cten_rees_prep, sage_rees_prep,
                                         "C. imbricatula to I. sagensis
B. reesidei-B. eliasi", "Cimbr_to_Isage_rees")

cnten_to_oblo_rees_overlap <- run_ecospat_dyn(cten_rees_prep, oblo_rees_prep,
                                         "C. imbricatula to I. oblongus
B. reesidei-B. eliasi", "Cimbr_to_Ioblo_rees")

#### Summarize Outgroup Comparisons and Export Tables ####

# Combine the results of observed overlap
make_D_I_row <- function(data1, name, interval){
  named1 <- as.data.frame(cbind(name, interval, data1[[2]]$obs$D, data1[[2]]$obs$I))
  colnames(named1) <- c("species", "Intervals_compared", "D_val", "I_val")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}
cten_to_conv_redu_D_I_row <- make_D_I_row(cten_to_conv_redu_overlap,"C. imbricatula vs. I. convexus","B. reduncus-B. scotti")

ctener_to_subl_redu_D_I_row <- make_D_I_row(cten_to_subl_redu_overlap,"C. imbricatula vs. I. sublaevis","B. reduncus-B. scotti")

ctener_to_sask_redu_D_I_row <- make_D_I_row(cten_to_sask_redu_overlap,"C. imbricatula vs. I. saskatchewanensis","B. reduncus-B. scotti")

ctener_to_bara_nebr_D_I_row <- make_D_I_row(cten_to_bara_nebr_overlap,"C. imbricatula vs. I. barabini","D. nebrascense-E. jenneyi")

ctener_to_conv_nebr_D_I_row <- make_D_I_row(cten_to_conv_nebr_overlap,"C. imbricatula vs. I. convexus","D. nebrascense-E. jenneyi")

ctener_to_subl_nebr_D_I_row <- make_D_I_row(cten_to_subl_nebr_overlap,"C. imbricatula vs. I. sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_D_I_row <- make_D_I_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I. barabini","D. nebrascense-E. jenneyi")

chla_to_conv_nebr_D_I_row <- make_D_I_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_D_I_row <- make_D_I_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sagensis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_D_I_row <- make_D_I_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")

luci_to_conv_nebr_D_I_row <- make_D_I_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")
luci_to_subl_nebr_D_I_row <- make_D_I_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_D_I_row <- make_D_I_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_D_I_row <- make_D_I_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini","B. reesidei-B. eliasi")

chla_to_sage_rees_D_I_row <- make_D_I_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")

chla_to_oblo_rees_D_I_row <- make_D_I_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

citen_to_bara_rees_D_I_row <- make_D_I_row(citen_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

citen_to_sage_rees_D_I_row <- make_D_I_row(citen_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")

citen_to_oblo_rees_D_I_row <- make_D_I_row(citen_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")

outgroup_compare_I_D_results <-
  rbind(cten_to_conv_redu_D_I_row,cten_to_subl_redu_D_I_row,cten_to_sask_redu_D_I_row,
         chla_to_bara_nebr_D_I_row,cten_to_conv_nebr_D_I_row,cten_to_subl_nebr_D_I_row,
         chla_to_bara_nebr_D_I_row,cten_to_conv_nebr_D_I_row,cten_to_subl_nebr_D_I_row,
         chla_to_conv_nebr_D_I_row,cten_to_subl_nebr_D_I_row,
         luci_to_bara_nebr_D_I_row,luci_to_conv_nebr_D_I_row,luci_to_subl_nebr_D_I_row,
         luci_to_sage_nebr_D_I_row,luci_to_conv_nebr_D_I_row,luci_to_subl_nebr_D_I_row,
         chla_to_bara_rees_D_I_row,cten_to_sage_rees_D_I_row,cten_to_subl_rees_D_I_row,
         chla_to_sage_rees_D_I_row,cten_to_oblo_rees_D_I_row,
         citen_to_bara_rees_D_I_row,cten_to_sage_rees_D_I_row,cten_to_subl_rees_D_I_row,cten_to_oblo_rees_D_I_row)

# Combine the results of equivalency tests
make_exp_p_row <- function(data1,name,interval){
    # Code for combining results
named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.D,data1[[2]]$p.I,data1[[3]]$p.D,data1[[3]]$p.I))
colnames(named1) <-
c("species","Intervals_compared","Greater_Eq_p_D","Greater_Eq_p_I","Lower_Eq_p_D","Lower_Eq_p_I")
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
return(named1)
}

cten_to_conv_redu_eq_row <- make_exp_p_row(cten_to_conv_redu_overlap,"C. imbricatula vs. I. convexus","B. reduncus-B. scotti")

cten_to_subl_redu_eq_row <- make_exp_p_row(cten_to_subl_redu_overlap,"C. imbricatula vs. I. sublaevis","B. reduncus-B. scotti")

cten_to_sask_redu_eq_row <- make_exp_p_row(cten_to_sask_redu_overlap,"C. imbricatula vs. I. saskatchewanensis","B. reduncus-B. scotti")

cten_to_bara_nebr_eq_row <- make_exp_p_row(cten_to_bara_nebr_overlap,"C. imbricatula vs. I. barabini","D. nebrascense-E. jenneyi")

cten_to_conv_nebr_eq_row <- make_exp_p_row(cten_to_conv_nebr_overlap,"C. imbricatula vs. I. convexus","D. nebrascense-E. jenneyi")

cten_to_subl_nebr_eq_row <- make_exp_p_row(cten_to_subl_nebr_overlap,"C. imbricatula vs. I. sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_eq_row <- make_exp_p_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I. barabini","D. nebrascense-E. jenneyi")

chla_to_conv_nebr_eq_row <- make_exp_p_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_eq_row <- make_exp_p_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sagensis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_eq_row <- make_exp_p_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")
luci_to_conv_nebr_eq_row <- make_exp_p_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")

luci_to_subl_nebr_eq_row <- make_exp_p_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_eq_row <- make_exp_p_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_eq_row <- make_exp_p_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini","B. reesidei-B. eliasi")

chla_to_sage_rees_eq_row <- make_exp_p_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")

chla_to_oblo_rees_eq_row <- make_exp_p_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

ceten_to_bara_rees_eq_row <- make_exp_p_row(cten_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

coten_to_sage_rees_eq_row <- make_exp_p_row(cten_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")

coten_to_oblo_rees_eq_row <- make_exp_p_row(cten_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")

outgroup_compare_eq_results <- rbind(cten_to_conv_redu_eq_row,cten_to_subl_redu_eq_row,cten_to_sask_redu_eq_row,
chla_to_bara_nebr_eq_row,cten_to_conv_nebr_eq_row,cten_to_subl_nebr_eq_row,
chla_to_bara_nebr_eq_row,chla_to_conv_nebr_eq_row,chla_to_subl_nebr_eq_row,
luci_to_bara_nebr_eq_row,luci_to_conv_nebr_eq_row,luci_to_subl_nebr_eq_row,
luci_to_sage_comp_eq_row,
chla_to_bara_rees_eq_row,chla_to_sage_rees_eq_row,chla_to_oblo_rees_eq_row,
# Combine the results of equivalency tests of Dynamics (p-vals)
make_exp_dyn_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.expansion,data1[[2]]$p.stability,data1[[2]]$p.unfilling,
    data1[[3]]$p.expansion,data1[[3]]$p.stability,data1[[3]]$p.unfilling))
  colnames(named1) <- c("species","Intervals_compared","Greater_Eq_p_expansion","Greater_Eq_p_stability","Greater_Eq_p_unfilling",
    "Lower_Eq_p_expansion","Lower_Eq_p_stability","Lower_Eq_p_unfilling")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

cnten_to_conv_redu_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_conv_redu_overlap,"C. imbricatula vs. I. convexus","B. reduncus-B. scotti")

cnten_to_subl_redu_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_subl_redu_overlap,"C. imbricatula vs. I. sublaevis","B. reduncus-B. scotti")

cnten_to_sask_redu_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_sask_redu_overlap,"C. imbricatula vs. I. saskatchewanensis","B. reduncus-B. scotti")

cnten_to_bara_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_bara_nebr_overlap,"C. imbricatula vs. I. barabini","D. nebrascense-E. jenneyi")

cnten_to_conv_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_conv_nebr_overlap,"C. imbricatula vs. I. convexus","D. nebrascense-E. jenneyi")

cnten_to_subl_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_subl_nebr_overlap,"C. imbricatula vs. I. sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I. barabini","D. nebrascense-E. jenneyi")
chla_to_conv_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")

luci_to_conv_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")

luci_to_subl_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_eq_dyn_p_row <- make_exp_dyn_p_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini","B. reesidei-B. eliasi")

chla_to_sage_rees_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")

chla_to_oblo_rees_eq_dyn_p_row <- make_exp_dyn_p_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

cten_to_bara_rees_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

cten_to_sage_rees_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")

cten_to_oblo_rees_eq_dyn_p_row <- make_exp_dyn_p_row(cten_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")
outgroup_compare_eq_dyn_p_results <-
rbind(cten_to_conv_redu_eq_dyn_p_row,cten_to_subl_redu_eq_dyn_p_row,cten_to_sask_redu_eq_dyn_p_row,
    chla_to_bara_nebr_eq_dyn_p_row,cten_to_conv_nebr_eq_dyn_p_row,cten_to_subl_nebr_eq_dyn_p_row,
    chla_to_bara_nebr_eq_dyn_p_row,chla_to_conv_nebr_eqDyn_p_row,chla_to_subl_nebr_eq_dyn_p_row,
    luci_to_bara_nebr_eq_dyn_p_row,luci_to_conv_nebr_eq_dyn_p_row,luci_to_subl_nebr_eq_dyn_p_row,
    luci_to_sage_comp_eq_dyn_p_row,
    chla_to_bara_rees_eq_dyn_p_row,chla_to_sage_rees_eq_dyn_p_row,chla_to_oblo_rees_eq_dyn_p_row,
    cten_to_bara_rees_eqDyn_p_row,cten_to_sage_rees_eq_dyn_p_row,cten_to_oblo_rees_eq_dyn_p_row)

# Combine the results of similarity tests
make_sim_p_row <- function(data1,name,interval){
    named1 <- as.data.frame(cbind(name,interval,data1[[4]]$p.D,data1[[4]]$p.I,data1[[5]]$p.D,data1[[5]]$p.I))
    colnames(named1) <-
    c("species","Intervals_compared","Greater_Sim_p_D","Greater_Sim_p_I","Lower_Sim_p_D","Lower_Sim_p_I")
    named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
    return(named1)
}

cten_to_conv_redu_sim_row <- make_sim_p_row(cten_to_conv_redu_overlap,"C. imbricatula vs. I. convexus","B. reduncus-B. scotti")

citen_to_subl_redu_sim_row <- make_sim_p_row(cten_to_subl_redu_overlap,"C. imbricatula vs. I. sublaevis","B. reduncus-B. scotti")

citen_to_sask_redu_sim_row <- make_sim_p_row(cten_to_sask_redu_overlap,"C. imbricatula vs. I. saskatchewanensis","B. reduncus-B. scotti")
cten_to_bara_nebr_sim_row <- make_sim_p_row(cten_to_bara_nebr_overlap,"C. imbricatula vs. I. barabini","D. nebrascense-E. jenneyi")

cten_to_conv_nebr_sim_row <- make_sim_p_row(cten_to_conv_nebr_overlap,"C. imbricatula vs. I. convexus","D. nebrascense-E. jenneyi")

cnten_to_subl_nebr_sim_row <- make_sim_p_row(cten_to_subl_nebr_overlap,"C. imbricatula vs. I. sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_sim_row <- make_sim_p_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I. barabini","D. nebrascense-E. jenneyi")

chla_to_conv_nebr_sim_row <- make_sim_p_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_sim_row <- make_sim_p_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_sim_row <- make_sim_p_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")

luci_to_conv_nebr_sim_row <- make_sim_p_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")

luci_to_subl_nebr_sim_row <- make_sim_p_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_sim_row <- make_sim_p_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_sim_row <- make_sim_p_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini", "B. reesidei-B. eliasi")

chla_to_sage_rees_sim_row <- make_sim_p_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")
chla_to_oblo_rees_sim_row <- make_sim_p_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

cten_to_bara_rees_sim_row <- make_sim_p_row(cten_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

cten_to_sage_rees_sim_row <- make_sim_p_row(cten_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")

cten_to_oblo_rees_sim_row <- make_sim_p_row(cten_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")

outgroup_compare_sim_results <-

# Combine the results of equivalency tests of Dynamics (p-vals)
make_sim_dyn_p_row <- function(data1,name,interval){
  named1 <-
as.data.frame(cbind(name,interval,data1[[4]]$p.expansion,data1[[4]]$p.stability,data1[[4]]$p.unfilling, data1[[5]]$p.expansion,data1[[5]]$p.stability,data1[[5]]$p.unfilling))
  colnames(named1) <-
c("species","Intervals_compared","Greater_Sim_p_expansion","Greater_Sim_p_stability","Greater_Sim_p_unfilling",
  "Lower_Sim_p_expansion","Lower_Sim_p_stability","Lower_Sim_p_unfilling")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
cten_to_conv_red_sim_dyna_p_row <- make_sim_dyn_p_row(cten_to_conv_red_overlap,"C. imbricatula vs. I. convexus","B. reduncus-B. scotti")

calen_to_subl_red_sim_dyna_p_row <- make_sim_dyn_p_row(calen_to_subl_red_overlap,"C. imbricatula vs. I. sublaevis","B. reduncus-B. scotti")

calen_to_sask_red_sim_dyna_p_row <- make_sim_dyn_p_row(calen_to_sask_red_overlap,"C. imbricatula vs. I. saskatchewanensis","B. reduncus-B. scotti")

calen_to_bara_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(calen_to_bara_nebr_overlap,"C. imbricatula vs. I. barabini","D. nebrascense-E. jenneyi")

calen_to_conv_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(calen_to_conv_nebr_overlap,"C. imbricatula vs. I. convexus","D. nebrascense-E. jenneyi")

calen_to_subl_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(calen_to_subl_nebr_overlap,"C. imbricatula vs. I. sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I. barabini","D. nebrascense-E. jenneyi")

chla_to_conv_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")

luci_to_conv_nebr_red_sim_dyna_p_row <- make_sim_dyn_p_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")
luci_to_subl_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_sim_dyn_p_row <- make_sim_dyn_p_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_sim_dyn_p_row <- make_sim_dyn_p_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini","B. reesidei-B. eliasi")

chla_to_sage_rees_sim_dyn_p_row <- make_sim_dyn_p_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")

chla_to_oblo_rees_sim_dyn_p_row <- make_sim_dyn_p_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

cilen_to_bara_rees_sim_dyn_p_row <- make_sim_dyn_p_row(cten_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

cilen_to_sage_rees_sim_dyn_p_row <- make_sim_dyn_p_row(cten_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")

cilen_to_oblo_rees_sim_dyn_p_row <- make_sim_dyn_p_row(cten_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")

outgroup_compare_sim_dyn_p_results <- rbind(cten_to_conv_redu_sim_dyn_p_row,cten_to_subl_redu_sim_dyn_p_row,cten_to_sask_redu_sim_dyn_p_row,
chla_to_bara_nebr_sim_dyn_p_row,cten_to_conv_nebr_sim_dyn_p_row,cten_to_subl_nebr_sim_dyn_p_row,
chla_to_bara_nebr_sim_dyn_p_row,chla_to_conv_nebr_sim_dyn_p_row,chla_to_subl_nebr_sim_dyn_p_row,
luci_to_bara_nebr_sim_dyn_p_row,luci_to_conv_nebr_sim_dyn_p_row,luci_to_subl_nebr_sim_dyn_p_row,
luci_to_sage_comp_sim_dyn_p_row,
chla_to_bara_rees_sim_dyn_p_row,chla_to_sage_rees_sim_dyn_p_row,chla_to_oblo_rees_sim_dyn_p_row,
cten_to_bara_rees_sim_dyn_p_row,cten_to_sage_rees_sim_dyn_p_row,cten_to_oblo_rees_sim_dyn_p_row)

# Combine the results of niche dynamics
make_dynam_row <- function(data1,name,interval){
    data <- as.data.frame(data1[[6]])
    named1 <- as.data.frame(cbind(name,interval,t(data[,2])))
    colnames(named1) <- c("species","Intervals_compared","Expansion","Stability","Unfilling")
    named1[,c(3:5)] <- as.numeric(as.character(named1[,c(3:5)]))
    return(named1)
}

cten_to_conv_redu_dyn_row <- make_dynam_row(cten_to_conv_redu_overlap,"C. imbricatula vs. I.
convexus","B. reduncus-B. scotti")

cten_to_subl_redu_dyn_row <- make_dynam_row(cten_to_subl_redu_overlap,"C. imbricatula vs. I.
sublaevis","B. reduncus-B. scotti")

cten_to_sask_redu_dyn_row <- make_dynam_row(cten_to_sask_redu_overlap,"C. imbricatula vs. I.
saskatchewanensis","B. reduncus-B. scotti")

cten_to_bara_nebr_dyn_row <- make_dynam_row(cten_to_bara_nebr_overlap,"C. imbricatula vs. I.
baborini","D. nebrascense-E. jenneyi")

cten_to_conv_nebr_dyn_row <- make_dynam_row(cten_to_conv_nebr_overlap,"C. imbricatula vs. I.
convexus","D. nebrascense-E. jenneyi")

cten_to_subl_nebr_dyn_row <- make_dynam_row(cten_to_subl_nebr_overlap,"C. imbricatula vs. I.
sublaevis","D. nebrascense-E. jenneyi")

chla_to_bara_nebr_dyn_row <- make_dynam_row(chla_to_bara_nebr_overlap,"C. nebrascensis vs. I.
baborini","D. nebrascense-E. jenneyi")
chla_to_conv_nebr_dyn_row <- make_dynam_row(chla_to_conv_nebr_overlap,"C. nebrascensis vs. I. convexus","D. nebrascense-E. jenneyi")

chla_to_subl_nebr_dyn_row <- make_dynam_row(chla_to_subl_nebr_overlap,"C. nebrascensis vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_bara_nebr_dyn_row <- make_dynam_row(luci_to_bara_nebr_overlap,"L. subundata vs. I. barabini","D. nebrascense-E. jenneyi")

luci_to_conv_nebr_dyn_row <- make_dynam_row(luci_to_conv_nebr_overlap,"L. subundata vs. I. convexus","D. nebrascense-E. jenneyi")

luci_to_subl_nebr_dyn_row <- make_dynam_row(luci_to_subl_nebr_overlap,"L. subundata vs. I. sublaevis","D. nebrascense-E. jenneyi")

luci_to_sage_comp_dyn_row <- make_dynam_row(luci_to_sage_comp_overlap,"L. subundata vs. I. sagensis","B. compressus-B. cuneatus")

chla_to_bara_rees_dyn_row <- make_dynam_row(chla_to_bara_rees_overlap,"C. nebrascensis vs. I. barabini","B. reesidei-B. eliasi")

chla_to_sage_rees_dyn_row <- make_dynam_row(chla_to_sage_rees_overlap,"C. nebrascensis vs. I. sagensis","B. reesidei-B. eliasi")

chla_to_oblo_rees_dyn_row <- make_dynam_row(chla_to_oblo_rees_overlap,"C. nebrascensis vs. I. oblongus","B. reesidei-B. eliasi")

cten_to_bara_rees_dyn_row <- make_dynam_row(cten_to_bara_rees_overlap,"C. imbricatula vs. I. barabini","B. reesidei-B. eliasi")

cten_to_sage_rees_dyn_row <- make_dynam_row(cten_to_sage_rees_overlap,"C. imbricatula vs. I. sagensis","B. reesidei-B. eliasi")
cten_to_oblo_rees_dyn_row <- make_dynam_row(cten_to_oblo_rees_overlap,"C. imbricatula vs. I. oblongus","B. reesidei-B. eliasi")

outgroup_compare_dyn_results <- rbind(cten_to_conv_redu_dyn_row,cten_to_subl_redu_dyn_row,cten_to_sask_redu_dyn_row,
                                      chla_to_bar_nebr_dyn_row,cten_to_conv_nebr_dyn_row,cten_to_subl_nebr_dyn_row,
                                      chla_to_bar_nebr_dyn_row,chla_to_conv_nebr_dyn_row,chla_to_subl_nebrDyn_row,
                                      luci_to_bar_nebr_dyn_row,luci_to_conv_nebr_dyn_row,luci_to_subl_nebr_dyn_row,
                                      luci_to_sage_comp_dyn_row,
                                      chla_to_bar_rees_dyn_row,chla_to_sage_rees_dyn_row,chla_to_oblo_rees_dyn_row,
                                      cten_to_bar_rees_dyn_row,cten_to_sage_rees_dyn_row,cten_to_oblo_rees_dyn_row)

## Write csv files of results
write.csv(outgroup_compare_I_D_results,file="tables/outgroup_compare_I_D_results.csv")
write.csv(outgroup_compare_eq_results,file="tables/outgroup_compare_eq_results.csv")
write.csv(outgroup_compare_sim_results,file="tables/outgroup_compare_sim_results.csv")
write.csv(outgroup_compare_dyn_results,file="tables/outgroup_compare_dyn_results.csv")
write.csv(outgroup_compare_eq_dyn_p_results,file="tables/outgroup_compare_eq_dyn_p_results.csv")
write.csv(outgroup_compare_sim_dyn_p_results,file="tables/outgroup_compare_sim_dyn_p_results.csv")

#### SPECIES TO GENUS COMPARISONS ####

conv_to_inoceram_redu_overlap <- run_ecospat_dyn(conv_redu_prep,Inoceramus_redu_prep,
                                                       "I. convextus to Inoceramus\nB. reduncus-B. scotti","Iconv_to_Inocer_redu_")

subl_to_inoceram_redu_overlap <- run_ecospat_dyn(subl_redu_prep,Inoceramus_redu_prep,
                                                       "I. sublaevis to Inoceramus\nB. reduncus-B. scotti","Isubl_to_Inocer_redu_")

sask_to_inoceram_redu_overlap <- run_ecospat_dyn(sask_redu_prep,Inoceramus_redu_prep,
                                                       "I. saskatchewanensis to Inoceramus\nB. reduncus-B. scotti","Isask_to_Inocer_redu_")
conv_to_inoceram_nebr_overlap <- run_ecospat_dyn(conv_nebr_prep, Inoceramus_nebr_prep,
  "I. convextus to Inoceramus\nD. nebrascense-E. jenneyi","Iconv_to_Inocer_nebr")

subl_to_inoceram_nebr_overlap <- run_ecospat_dyn(subl_nebr_prep, Inoceramus_nebr_prep,
  "I. sublaevis to Inoceramus\nD. nebrascense-E. jenneyi","Isubl_to_Inocer_nebr")

bara_to_inoceram_nebr_overlap <- run_ecospat_dyn(bara_nebr_prep, Inoceramus_nebr_prep,
  "I. barabini to Inoceramus\nD. nebrascense-E. jenneyi","Ibara_to_Inocer_nebr")

bara_to_inoceram_rees_overlap <- run_ecospat_dyn(bara_rees_prep, Inoceramus_rees_prep,
  "I. barabini to Inoceramus\nB. reesidei-B. eliasi","Ibara_to_Inocer_rees")

sage_to_inoceram_rees_overlap <- run_ecospat_dyn(sage_rees_prep, Inoceramus_rees_prep,
  "I. sagensis to Inoceramus\nB. reesidei-B. eliasi","Isage_to_Inocer_rees")

oblo_to_inoceram_rees_overlap <- run_ecospat_dyn(oblo_rees_prep, Inoceramus_rees_prep,
  "I. oblongus to Inoceramus\nB. reesidei-B. eliasi","Ioblo_to_Inocer_rees")

#### Summarize Species to Genus Comparisons and Export Tables ####

# Combine the results of observed overlap
make_D_I_row <- function(data1, name, interval){
  named1 <- as.data.frame(cbind(name, interval, data1[[2]]$obs$D, data1[[2]]$obs$I))
  colnames(named1) <- c("species", "Intervals_compared", "D_val", "I_val")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_inoceram_redun_D_I_row <- make_D_I_row(conv_to_inoceram_redun_overlap,"I. convexus","B. reduncus-B. scotti")
subl_to_inoceram_redu_D_I_row <- make_D_I_row(subl_to_inoceram_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")

sask_to_inoceram_redu_D_I_row <- make_D_I_row(sask_to_inoceram_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_inoceram_nebr_D_I_row <- make_D_I_row(conv_to_inoceram_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")

subl_to_inoceram_nebr_D_I_row <- make_D_I_row(subl_to_inoceram_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_nebr_D_I_row <- make_D_I_row(bara_to_inoceram_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_rees_D_I_row <- make_D_I_row(bara_to_inoceram_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_rees_D_I_row <- make_D_I_row(sage_to_inoceram_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_rees_D_I_row <- make_D_I_row(oblo_to_inoceram_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_I_D_results <- rbind(conv_to_inoceram_redu_D_I_row,subl_to_inoceram_redu_D_I_row,
                                sask_to_inoceram_redu_D_I_row,conv_to_inoceram_nebr_D_I_row,
                                subl_to_inoceram_nebr_D_I_row,bara_to_inoceram_nebr_D_I_row,
                                bara_to_inoceram_rees_D_I_row,sage_to_inoceram_rees_D_I_row,
                                oblo_to_inoceram_rees_D_I_row)

# Combine the results of equivalency tests
make_exp_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.D,data1[[2]]$p.I,data1[[3]]$p.D,data1[[3]]$p.I))
  colnames(named1) <- c("species","Intervals_compared","Greater_Eq_p_D","Greater_Eq_p_I","Lower_Eq_p_D","Lower_Eq_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
conv_to_inoceram_redu_eq_row <- make_exp_p_row(conv_to_inoceram_redu_overlap, "I. convexus", "B. reduncus-B. scotti")
subl_to_inoceram_redu_eq_row <- make_exp_p_row(subl_to_inoceram_redu_overlap, "I. sublaevis", "B. reduncus-B. scotti")
sask_to_inoceram_redu_eq_row <- make_exp_p_row(sask_to_inoceram_redu_overlap, "I. saskatchewanensis", "B. reduncus-B. scotti")
conv_to_inoceram_nebr_eq_row <- make_exp_p_row(conv_to_inoceram_nebr_overlap, "I. convexus", "D. nebrascense-E. jenneyi")
subl_to_inoceram_nebr_eq_row <- make_exp_p_row(subl_to_inoceram_nebr_overlap, "I. sublaevis", "D. nebrascense-E. jenneyi")
bara_to_inoceram_nebr_eq_row <- make_exp_p_row(bara_to_inoceram_nebr_overlap, "I. barabini", "D. nebrascense-E. jenneyi")
bara_to_inoceram_rees_eq_row <- make_exp_p_row(bara_to_inoceram_rees_overlap, "I. barabini", "B. reesidei-B. eliasi")
sage_to_inoceram_rees_eq_row <- make_exp_p_row(sage_to_inoceram_rees_overlap, "I. sagensis", "B. reesidei-B. eliasi")
oblo_to_inoceram_rees_eq_row <- make_exp_p_row(oblo_to_inoceram_rees_overlap, "I. oblongus", "B. reesidei-B. eliasi")
gen_to_sp_eq_results <- rbind(conv_to_inoceram_redu_eq_row, subl_to_inoceram_redu_eq_row, conv_to_inoceram_nebr_eq_row, subl_to_inoceram_nebr_eq_row, bара_to_inoceram_nebr_eq_row, bара_to_inoceram_rees_eq_row, sage_to_inoceram_rees_eq_row, oblo_to_inoceram_rees_eq_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_exp_dyn_p_row <- function(data1, name, interval) {
  named1 <-
as.data.frame(cbind(name, interval, data1[[2]]$p.expansion, data1[[2]]$p.stability, data1[[2]]$p.unfilling,}
data1[[3]]$p.expansion,data1[[3]]$p.stability,data1[[3]]$p.unfilling))

colnames(named1) <-
c("species","Intervals_compared","Greater_Eq_p_expansion","Greater_Eq_p_stability","Greater_Eq_p_unfilling",
"Lower_Eq_p_expansion","Lower_Eq_p_stability","Lower_Eq_p_unfilling")
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
return(named1)
}

conv_to_inoceram_redu_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_inoceram_redu_overlap,"I. convexus","B. reduncus-B. scotti")
subl_to_inoceram_redu_eq_dyn_p_row <- make_exp_dyn_p_row(subl_to_inoceram_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")
sask_to_inoceram_redu_eq_dyn_p_row <- make_exp_dyn_p_row(sask_to_inoceram_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")
conv_to_inoceram_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_inoceram_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")
subl_to_inoceram_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(subl_to_inoceram_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")
bara_to_inoceram_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_inoceram_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")
bara_to_inoceram_rees_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_inoceram_rees_overlap,"I. barabini","B. reesidei-B. eliasi")
sage_to_inoceram_rees_eq_dyn_p_row <- make_exp_dyn_p_row(sage_to_inoceram_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")
oblo_to_inoceram_rees_eq_dyn_p_row <- make_exp_dyn_p_row(oblo_to_inoceram_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_eq_dyn_p_results <-
rbind(conv_to_inoceram_redu_eq_dyn_p_row,subl_to_inoceram_redu_eq_dyn_p_row,
sask_to_inoceram_redu_eq_dyn_p_row,conv_to_inoceram_nebr_eq_dyn_p_row,
# Combine the results of similarity tests
make_sim_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[4]]$p.D,data1[[4]]$p.I,data1[[5]]$p.D,data1[[5]]$p.I))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Sim_p_D","Greater_Sim_p_I","Lower_Sim_p_D","Lower_Sim_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_inoceram_redu_sim_row <- make_sim_p_row(conv_to_inoceram_redu_overlap,"I. convexus","B. reduncus-B. scotti")
subl_to_inoceram_redu_sim_row <- make_sim_p_row(subl_to_inoceram_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")
sask_to_inoceram_redu_sim_row <- make_sim_p_row(sask_to_inoceram_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")
conv_to_inoceram_nebr_sim_row <- make_sim_p_row(conv_to_inoceram_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")
subl_to_inoceram_nebr_sim_row <- make_sim_p_row(subl_to_inoceram_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")
bara_to_inoceram_nebr_sim_row <- make_sim_p_row(bara_to_inoceram_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")
bara_to_inoceram_rees_sim_row <- make_sim_p_row(bara_to_inoceram_rees_overlap,"I. barabini","B. reesidei-B. eliasi")
sage_to_inoceram_rees_sim_row <- make_sim_p_row(sage_to_inoceram_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")
oblo_to_inoceram_rees_sim_row <- make_sim_p_row(oblo_to_inoceram_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_sim_results <- rbind(conv_to_inoceram_redu_sim_row,subl_to_inoceram_redu_sim_row,
sask_to_inoceram_redu_sim_row,conv_to_inoceram_nebr_sim_row,
subl_to_inoceram_nebr_sim_row,bara_to_inoceram_nebr_sim_row,
bara_to_inoceram_rees_sim_row,sage_to_inoceram_rees_sim_row,
oblo_to_inoceram_rees_sim_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_sim_dyn_p_row <- function(data1,name,interval){
  named1 <-
as.data.frame(cbind(name,interval,data1[[4]]$p.expansion,data1[[4]]$p.stability,data1[[4]]$p.unfilling,
                 data1[[5]]$p.expansion,data1[[5]]$p.stability,data1[[5]]$p.unfilling))
colnames(named1) <-
c("species","Intervals_compared","Greater_Sim_p_expansion","Greater_Sim_p_stability","Greater_Sim_p_unfilling",
       "Lower_Sim_p_expansion","Lower_Sim_p_stability","Lower_Sim_p_unfilling")
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
return(named1)
}

conv_to_inoceram_redu_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_inoceram_redu_overlap,"I.
convexus","B. reduncus-B. scotti")
subl_to_inoceram_redu_sim_dyn_p_row <- make_sim_dyn_p_row(subl_to_inoceram_redu_overlap,"I.
sublaevis","B. reduncus-B. scotti")
sask_to_inoceram_redu_sim_dyn_p_row <- make_sim_dyn_p_row(sask_to_inoceram_redu_overlap,"I.
saskatchewanensis","B. reduncus-B. scotti")
conv_to_inoceram_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_inoceram_nebr_overlap,"I.
convexus","D. nebrascense-E. jenneyi")
subl_to_inoceram_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(subl_to_inoceram_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_inoceram_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_rees_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_inoceram_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_rees_sim_dyn_p_row <- make_sim_dyn_p_row(sage_to_inoceram_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_rees_sim_dyn_p_row <- make_sim_dyn_p_row(oblo_to_inoceram_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_sim_dyn_p_results <-
  rbind(conv_to_inoceram_redu_sim_dyn_p_row, subl_to_inoceram_redu_sim_dyn_p_row,
        sask_to_inoceram_redu_sim_dyn_p_row, conv_to_inoceram_nebr_sim_dyn_p_row,
        subl_to_inoceram_nebr_sim_dyn_p_row, bara_to_inoceram_nebr_sim_dyn_p_row,
        bara_to_inoceram_rees_sim_dyn_p_row, sage_to_inoceram_rees_sim_dyn_p_row,
        oblo_to_inoceram_rees_sim_dyn_p_row)

# Combine the results of niche dynamics
make_dynam_row <- function(data1,name,interval){
  data <- as.data.frame(data1[[6]])
  named1 <- as.data.frame(cbind(name,interval,t(data[,2])))
  colnames(named1) <- c("species","Intervals_compared","Expansion","Stability","Unfilling")
  named1[,c(3:5)] <- as.numeric(as.character(named1[,c(3:5)]))
  return(named1)
}

conv_to_inoceram_redu_dyn_row <- make_dynam_row(conv_to_inoceram_redu_overlap,"I. convexus","B. reduncus-B. scotti")

subl_to_inoceram_redu_dyn_row <- make_dynam_row(subl_to_inoceram_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")
sask_to_inoceram_redu_dyn_row <- make_dynam_row(sask_to_inoceram_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_inoceram_nebr_dyn_row <- make_dynam_row(conv_to_inoceram_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")

subl_to_inoceram_nebr_dyn_row <- make_dynam_row(subl_to_inoceram_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_nebr_dyn_row <- make_dynam_row(bara_to_inoceram_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_rees_dyn_row <- make_dynam_row(bara_to_inoceram_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_rees_dyn_row <- make_dynam_row(sage_to_inoceram_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_rees_dyn_row <- make_dynam_row(oblo_to_inoceram_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

geno_to_sp_dyn_results <- rbind(conv_to_inoceram_redu_dyn_row,subl_to_inoceram_redu_dyn_row,
sask_to_inoceram_redu_dyn_row,conv_to_inoceram_nebr_dyn_row,
subl_to_inoceram_nebr_dyn_row,bara_to_inoceram_nebr_dyn_row,
bara_to_inoceram_rees_dyn_row,sage_to_inoceram_rees_dyn_row,
oblo_to_inoceram_rees_dyn_row)

## Write csv files of results

write.csv(geno_to_sp_I_D_results,file="tables/geno_to_sp_I_D_results.csv")
write.csv(geno_to_sp_eq_results,file="tables/geno_to_sp_eq_results.csv")
write.csv(geno_to_sp_sim_results,file="tables/geno_to_sp_sim_results.csv")
write.csv(geno_to_sp_dyn_results,file="tables/geno_to_sp_dyn_results.csv")
write.csv(geno_to_sp_eq_dyn_p_results,file="tables/geno_to_sp_eq_dyn_p_results.csv")
write.csv(geno_to_sp_sim_dyn_p_results,file="tables/geno_to_sp_sim_dyn_p_results.csv")
##### RUN PCA OF MORE THAN 2 SPP TO GENUS #####

## Extract values to matrix for each raster stack of env proxy variables:

```r
redu_taxa_list <- list(conv_redu_prep, subl_redu_prep, sask_redu_prep, Inoceramus_redu_prep)
nebr_taxa_list <- list(conv_nebr_prep, subl_nebr_prep, bara_nebr_prep, Inoceramus_nebr_prep)
rees_taxa_list <- list(bara_rees_prep, sage_rees_prep, oblo_rees_prep, Inoceramus_rees_prep)
```

```r
name_list_redu <- c("I. convexus","I. sublaevis", "I. saskatchewanensis","Inoceramus")
name_list_nebr <- c("I. convexus","I. sublaevis","I. barabini","Inoceramus")
name_list_rees <- c("I. barabini","I. sagensis", "I. oblongus","Inoceramus")
```

# Create function to tabulate PCA scores for plotting spp and genus information
# Function will only take a list of prepped occurrence/raster stack data
# and output a list of 1-Joined (global) environment pts, 2-pts for each spp/gen,
# and 3-the convex hull index values for creating polygons around spp/gen
```r
run_gen_spp_pca <- function(occ_raster_list,name_list){

bg_score_list <- list() # make list to put bg scores into

for (i in 1:length(occ_raster_list)){

    # Get scores for each environment's backgrounds
    first_bg<- getValues(occ_raster_list[[i]][[2]])

    ## Clean out missing values:
    first_bg <- first_bg[complete.cases(first_bg), ]

    bg_score_list[[i]] <- first_bg
}
```
## Combined global environment:

```r
joined_bg <- list.rbind(bg_score_list)
```

# Run PCA analysis on combine BG pts (all env data)
```r
pca_joined_bg <- dudi.pca(joined_bg, center = TRUE,
                        scale = TRUE, scannf = FALSE, nf = 2)

joined_bg_scores <- pca_joined_bg$li # get the pca scores
```

```r
occ_loc_list <- list() # Make list to put occ locations in PCA into
for (i in 1:length(occ_raster_list)){
    # Map occurrence data into the 2d ordination (have to coerce spatialPointsDataFrame into data.frame)
    # Explicitly match the colnames so only using the right ones
    # Only selected the li element (doesn't include others)

    first_occ_scores <-
        suprow(pca_joined_bg,
              data.frame(occ_raster_list[[i]][[1]][, colnames(joined_bg)])$li

    occ_loc_list[[i]] <- first_occ_scores
}
```

# Add column with spp/gen names to occ location list:
```r
for (i in 1:length(occ_raster_list)) {
    occ_loc_list[[i]]$tax <- name_list[i]
}
```

# Bind all the taxa localities together
occ_loc_df <- list.rbind(occ_loc_list)

bg_loc_list <- list() # make list to put bg locations in PCA into

for (i in 1:length(occ_raster_list)){

  # Map BG data to the 2d ordination
  first_bg_scores <- suprow(pca_joined_bg, bg_score_list[[i]])

  bg_loc_list[[i]] <- first_bg_scores

}

# NOT USED ANYMORE, BUT MAYBE STILL WANT TO KEEP IN CASE
# hull_list <- list() # make list to hold convex hulls
#
# for (i in 1:length(occ_raster_list)){
# #
# # Make convex hulls to put around points for each spp/genus
# # first_hull <- chull(occ_loc_list[[i]])
# # first_hull <- c(first_hull,first_hull[1]) # Get list of index values for making a polygon
# #
# # hull_list[[i]] <- first_hull
# #
# # }

final_list <- list(joined_bg_scores,occ_loc_df)

return(final_list)
# Use above function to make list of PCA information for plotting in each interval
redu_taxa_pca_list <- run_gen_spp_pca(redu_taxa_list, name_list_redu)
nebr_taxa_pca_list <- run_gen_spp_pca(nebr_taxa_list, name_list_nebr)
rees_taxa_pca_list <- run_gen_spp_pca(rees_taxa_list, name_list_rees)

# Use listed outputs from above function run on list of spp/gen to make plots:

library(devtools)
library(ggConvexHull)

pdf("Figures/pca_genus_to_spp_figures.pdf")

# REDU

ggplot(redu_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
    geom_point(data=redu_taxa_pca_list[[1]],aes(x=Axis1, y=Axis2),color='grey25') + # global env pts
    geom_convexhull(alpha = 0.3, aes(fill = tax,color = tax), linewidth = 1) +
    scale_fill_manual(values = c("#56B4E9", "#E69F00","#F0E442","#999999")) +
    scale_color_manual(values = c("#56B4E9", "#E69F00","#F0E442","#999999")) +
    ggtitle("PCA of Inoceramus genus and Species in B. reduncus-B. scotti") +
    theme_light() +
    theme(legend.title = element_blank())

# NEBR

ggplot(nebr_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
    geom_point(data=nebr_taxa_pca_list[[1]],aes(x=Axis1, y=Axis2),color='grey25') + # global env pts
    geom_convexhull(alpha = 0.3, aes(fill = tax,color = tax), linewidth = 1) +
    scale_fill_manual(values = c("#009E73", "#56B4E9", "#F0E442","#999999")) +
    scale_color_manual(values = c("#009E73", "#56B4E9", "#F0E442","#999999")) +
    ggtitle("PCA of Inoceramus genus and Species in D. nebrascense-E. jenneyi") +
    theme_light() +
    theme(legend.title = element_blank())
# REES

```r
ggplot(rees_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
ggplot(rees_taxa_pca_list[[1]],aes(x=Axis1, y=Axis2),color='grey25') + # global env pts
geom_convexhull(alpha = 0.3, aes(fill = tax, color = tax), linewidth = 1) +
scale_fill_manual(values = c("#009E73", "#CC79A7", '#0072B2', "#999999")(values = c("#009E73", "#CC79A7", '#0072B2', "#999999")+

ggtitle("PCA of Inoceramus genus and Species in B. reesidei-B. eliasi") +
theme_light() +
theme(legend.title = element_blank())
```
dev.off()

#### SPECIES TO GENUS (ALL OCC, EVEN NO SPP ID) COMPARISONS ####

```r
conv_to_inoceram_all_redu_overlap <- run_ecospat_dyn(conv_redu_prep,Inoceramus_all_redu_prep,
                                          "I. convextus to Inoceramus\nB. reduncus-B. scotti","Iconv_to_Inocer_all_redu_"")

subl_to_inoceram_all_redu_overlap <- run_ecospat_dyn(subl_redu_prep,Inoceramus_all_redu_prep,
                                          "I. sublaevis to Inoceramus\nB. reduncus-B. scotti","Isubl_to_Inocer_all_redu_"")

sask_to_inoceram_all_redu_overlap <- run_ecospat_dyn(sask_redu_prep,Inoceramus_all_redu_prep,
                                          "I. saskatchewanensis to Inoceramus\nB. reduncus-B. scotti","Isask_to_Inocer_all_redu_"")

conv_to_inoceram_all_nebr_overlap <- run_ecospat_dyn(conv_nebr_prep,Inoceramus_all_nebr_prep,
                                          "I. convextus to Inoceramus\nD. nebrascense-E. jenneyi","Iconv_to_Inocer_all_nebr")

subl_to_inoceram_all_nebr_overlap <- run_ecospat_dyn(subl_nebr_prep,Inoceramus_all_nebr_prep,
                                          "I. sublaevis to Inoceramus\nD. nebrascense-E. jenneyi","Isubl_to_Inocer_all_nebr")
```
bara_to_inoceram_all_nebr_overlap <- run_ecospat_dyn(bara_nebr_prep, Inoceramus_all_nebr_prep, 
  "I. barabini to Inoceramus\nD. nebrascense-E. jenneyi","Ibar_to_Inocer_all_nebr")

bara_to_inoceram_all_rees_overlap <- run_ecospat_dyn(bara_rees_prep, Inoceramus_all_rees_prep, 
  "I. barabini to Inoceramus\nB. reesidei-B. eliasi","Ibar_to_Inocer_all_rees")

sage_to_inoceram_all_rees_overlap <- run_ecospat_dyn(sage_rees_prep, Inoceramus_all_rees_prep, 
  "I. sagensis to Inoceramus\nB. reesidei-B. eliasi","Isage_to_Inocer_all_rees")

oblo_to_inoceram_all_rees_overlap <- run_ecospat_dyn(oblo_rees_prep, Inoceramus_all_rees_prep, 
  "I. oblongus to Inoceramus\nB. reesidei-B. eliasi","Ioblo_to_Inocer_all_rees")

#### Summarize Species to Genus (all occ) Comparisons and Export Tables ####

# Combine the results of observed overlap

make_D_I_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$obs$D,data1[[2]]$obs$I))
  colnames(named1) <- c("species", "Intervals_compared", "D_val", "I_val")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_inoceram_all_redu_D_I_row <- make_D_I_row(conv_to_inoceram_all_redu_overlap,"I. convexus","B. reduncus-B. scotti")

subl_to_inoceram_all_redu_D_I_row <- make_D_I_row(subl_to_inoceram_all_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")

sask_to_inoceram_all_redu_D_I_row <- make_D_I_row(sask_to_inoceram_all_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")
conv_to_inoceram_all_nebr_D_I_row <- make_D_I_row(conv_to_inoceram_all_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")

subl_to_inoceram_all_nebr_D_I_row <- make_D_I_row(subl_to_inoceram_all_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_nebr_D_I_row <- make_D_I_row(bara_to_inoceram_all_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_rees_D_I_row <- make_D_I_row(bara_to_inoceram_all_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_all_rees_D_I_row <- make_D_I_row(sage_to_inoceram_all_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_all_rees_D_I_row <- make_D_I_row(oblo_to_inoceram_all_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_I_D_results <- rbind(conv_to_inoceram_all_redu_D_I_row,subl_to_inoceram_all_redu_D_I_row,
                                sask_to_inoceram_all_redu_D_I_row,conv_to_inoceram_all_nebr_D_I_row,
                                subl_to_inoceram_all_nebr_D_I_row,bara_to_inoceram_all_nebr_D_I_row,
                                bara_to_inoceram_all_rees_D_I_row,sage_to_inoceram_all_rees_D_I_row,
                                oblo_to_inoceram_all_rees_D_I_row)

# Combine the results of equivalency tests
make_exp_p_row <- function(data1,name,interval){
  named1 <- as.data.frame(cbind(name,interval,data1[[2]]$p.D,data1[[2]]$p.I,data1[[3]]$p.D,data1[[3]]$p.I))
  colnames(named1) <- c("species","Intervals_compared","Greater_Eq_p_D","Greater_Eq_p_I","Lower_Eq_p_D","Lower_Eq_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_inoceram_all_redu_eq_row <- make_exp_p_row(conv_to_inoceram_all_redu_overlap,"I. convexus","B. reduncus-B. scotti")
subl_to_inoceram_all_redu_eq_row <- make_exp_p_row(subl_to_inoceram_all_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")

sask_to_inoceram_all_redu_eq_row <- make_exp_p_row(sask_to_inoceram_all_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_inoceram_all_nebr_eq_row <- make_exp_p_row(conv_to_inoceram_all_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")

subl_to_inoceram_all_nebr_eq_row <- make_exp_p_row(subl_to_inoceram_all_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_nebr_eq_row <- make_exp_p_row(bara_to_inoceram_all_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_rees_eq_row <- make_exp_p_row(bara_to_inoceram_all_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_all_rees_eq_row <- make_exp_p_row(sage_to_inoceram_all_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_all_rees_eq_row <- make_exp_p_row(oblo_to_inoceram_all_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_eq_results <- rbind(conv_to_inoceram_all_redu_eq_row,subl_to_inoceram_all_redu_eq_row,
                                 sask_to_inoceram_all_redu_eq_row,conv_to_inoceram_all_nebr_eq_row,
                                 subl_to_inoceram_all_nebr_eq_row,bara_to_inoceram_all_nebr_eq_row,
                                 bara_to_inoceram_all_rees_eq_row,sage_to_inoceram_all_rees_eq_row,
                                 oblo_to_inoceram_all_rees_eq_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_exp_dyn_p_row <- function(data1,name,interval){
  named1 <-
  as.data.frame(cbind(name,interval,data1[[2]]$p.expansion,data1[[2]]$p.stability,data1[[2]]$p.unfilling,
                  data1[[3]]$p.expansion,data1[[3]]$p.stability,data1[[3]]$p.unfilling))
  colnames(named1) <-
  c("species","Intervals_compared","Greater_Eq_p_expansion","Greater_Eq_p_stability","Greater_Eq_p_unfillin
  g",
      "Lower_Eq_p_expansion","Lower_Eq_p_stability","Lower_Eq_p_unfilling")
}
named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
return(named1)

conv_to_inoceram_all_redu_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_inoceram_all_redu_overlap, "I. convexus", "B. reduncus-B. scotti")
subl_to_inoceram_all_redu_eq_dyn_p_row <- make_exp_dyn_p_row(subl_to_inoceram_all_redu_overlap, "I. sublaevis", "B. reduncus-B. scotti")
sask_to_inoceram_all_redu_eq_dyn_p_row <- make_exp_dyn_p_row(sask_to_inoceram_all_redu_overlap, "I. saskatchewanensis", "B. reduncus-B. scotti")
conv_to_inoceram_all_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(conv_to_inoceram_all_nebr_overlap, "I. convexus", "D. nebrascense-E. jenneyi")
subl_to_inoceram_all_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(subl_to_inoceram_all_nebr_overlap, "I. sublaevis", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_nebr_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_inoceram_all_nebr_overlap, "I. barabini", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_rees_eq_dyn_p_row <- make_exp_dyn_p_row(bara_to_inoceram_all_rees_overlap, "I. barabini", "B. reesidei-B. eliasi")
sage_to_inoceram_all_rees_eq_dyn_p_row <- make_exp_dyn_p_row(sage_to_inoceram_all_rees_overlap, "I. sagensis", "B. reesidei-B. eliasi")
oblo_to_inoceram_all_rees_eq_dyn_p_row <- make_exp_dyn_p_row(oblo_to_inoceram_all_rees_overlap, "I. oblongus", "B. reesidei-B. eliasi")

gen_to_sp_eq_dyn_p_results <- rbind(conv_to_inoceram_all_redu_eq_dyn_p_row, subl_to_inoceram_all_redu_eq_dyn_p_row,
                                       sask_to_inoceram_all_redu_eq_dyn_p_row, conv_to_inoceram_all_nebr_eq_dyn_p_row,
                                       subl_to_inoceram_all_nebr_eq_dyn_p_row, bar_to_inoceram_all_nebr_eq_dyn_p_row,
                                       bara_to_inoceram_all_rees_eq_dyn_p_row, sage_to_inoceram_all_rees_eq_dyn_p_row,
                                       oblo_to_inoceram_all_rees_eq_dyn_p_row)
# Combine the results of similarity tests

```r
def make_sim_p_row(data1, name, interval) {
  named1 <- as.data.frame(cbind(name, interval, data1[[4]]$p.D, data1[[4]]$p.I, data1[[5]]$p.D, data1[[5]]$p.I))
  colnames(named1) <- c("species","Intervals Compared","Greater_Sim_p_D","Greater_Sim_p_I","Lower_Sim_p_D","Lower_Sim_p_I")
  named1[,c(3:4)] <- as.numeric(as.character(named1[,c(3:4)]))
  return(named1)
}

conv_to_inoceram_all_redu_sim_row <- make_sim_p_row(conv_to_inoceram_all_redu_overlap, "I. convexus", "B. reduncus-B. scotti")
subl_to_inoceram_all_redu_sim_row <- make_sim_p_row(subl_to_inoceram_all_redu_overlap, "I. sublaevis", "B. reduncus-B. scotti")
sask_to_inoceram_all_redu_sim_row <- make_sim_p_row(sask_to_inoceram_all_redu_overlap, "I. saskatchewanensis", "B. reduncus-B. scotti")
conv_to_inoceram_all_nebr_sim_row <- make_sim_p_row(conv_to_inoceram_all_nebr_overlap, "I. convexus", "D. nebrascense-E. jenneyi")
subl_to_inoceram_all_nebr_sim_row <- make_sim_p_row(subl_to_inoceram_all_nebr_overlap, "I. sublaevis", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_nebr_sim_row <- make_sim_p_row(bara_to_inoceram_all_nebr_overlap, "I. barabini", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_rees_sim_row <- make_sim_p_row(bara_to_inoceram_all_rees_overlap, "I. barabini", "B. reesidei-B. eliasi")
sage_to_inoceram_all_rees_sim_row <- make_sim_p_row(sage_to_inoceram_all_rees_overlap, "I. sagensis", "B. reesidei-B. eliasi")
oblo_to_inoceram_all_rees_sim_row <- make_sim_p_row(oblo_to_inoceram_all_rees_overlap, "I. oblongus", "B. reesidei-B. eliasi")

gen_to_sp_sim_results <- rbind(conv_to_inoceram_all_redu_sim_row, subl_to_inoceram_all_redu_sim_row,
  ...)
```
sask_to_inoceram_all_redu_sim_row, conv_to_inoceram_all_nebr_sim_row, 
subl_to_inoceram_all_nebr_sim_row, barab_to_inoceram_all_nebr_sim_row, 
bara_to_inoceram_all_rees_sim_row, sage_to_inoceram_all_rees_sim_row, 
oblo_to_inoceram_all_rees_sim_row)

# Combine the results of equivalency tests of Dynamics (p-vals)
make_sim_dyn_p_row <- function(data1, name, interval)
{
  named1 <-
  as.data.frame(cbind(name, interval, data1[[4]]$p.expansion, data1[[4]]$p.stability, data1[[4]]$p.unfilling, 
                 data1[[5]]$p.expansion, data1[[5]]$p.stability, data1[[5]]$p.unfilling))
  colnames(named1) <-
  c("species", "Intervals_compared", "Greater_Sim_p_expansion", "Greater_Sim_p_stability", "Greater_Sim_p_unfilling", 
    "Lower_Sim_p_expansion", "Lower_Sim_p_stability", "Lower_Sim_p_unfilling")
  named1[, c(3:4)] <- as.numeric(as.character(named1[, c(3:4)]))
  return(named1)
}

conv_to_inoceram_all_redu_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_inoceram_all_redu_overlap, "I. convexus", "B. reduncus-B. scotti")
subl_to_inoceram_all_redu_sim_dyn_p_row <- make_sim_dyn_p_row(subl_to_inoceram_all_redu_overlap, "I. sublaevis", "B. reduncus-B. scotti")
sask_to_inoceram_all_redu_sim_dyn_p_row <- make_sim_dyn_p_row(sask_to_inoceram_all_redu_overlap, "I. saskatchewanensis", "B. reduncus-B. scotti")
conv_to_inoceram_all_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(conv_to_inoceram_all_nebr_overlap, "I. convexus", "D. nebrascense-E. jenneyi")
subl_to_inoceram_all_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(subl_to_inoceram_all_nebr_overlap, "I. sublaevis", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_nebr_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_inoceram_all_nebr_overlap, "I. barabini", "D. nebrascense-E. jenneyi")
bara_to_inoceram_all_rees_sim_dyn_p_row <- make_sim_dyn_p_row(bara_to_inoceram_all_rees_overlap, "I. barabini", "B. reesidei-B. eliasi")
sage_to_inoceram_all_rees_sim_dyn_p_row <- make_sim_dyn_p_row(sage_to_inoceram_all_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_all_rees_sim_dyn_p_row <- make_sim_dyn_p_row(oblo_to_inoceram_all_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_sim_dyn_p_results <-
rbind(conv_to_inoceram_all_redu_sim_dyn_p_row,subl_to_inoceram_all_redu_sim_dyn_p_row,
sask_to_inoceram_all_redu_sim_dyn_p_row,conv_to_inoceram_all_nebr_sim_dyn_p_row,
subl_to_inoceram_all_nebr_sim_dyn_p_row,bara_to_inoceram_all_nebr_sim_dyn_p_row,
bara_to_inoceram_all_rees_sim_dyn_p_row,sage_to_inoceram_all_rees_sim_dyn_p_row,
oblo_to_inoceram_all_rees_sim_dyn_p_row)

# Combine the results of niche dynamics
make_dynam_row <- function(data1,name,interval){
  data <- as.data.frame(data1[[6]])
  named1 <- as.data.frame(cbind(name,interval,t(data[,2])))
  colnames(named1) <- c("species","Intervals_compared","Expansion","Stability","Unfilling")
  named1[,c(3:5)] <- as.numeric(as.character(named1[,c(3:5)]))
  return(named1)
}

conv_to_inoceram_all_redu_dyn_row <- make_dynam_row(conv_to_inoceram_all_redu_overlap,"I. convexus","B. reduncus-B. scotti")

subl_to_inoceram_all_redu_dyn_row <- make_dynam_row(subl_to_inoceram_all_redu_overlap,"I. sublaevis","B. reduncus-B. scotti")

sask_to_inoceram_all_redu_dyn_row <- make_dynam_row(sask_to_inoceram_all_redu_overlap,"I. saskatchewanensis","B. reduncus-B. scotti")

conv_to_inoceram_all_nebr_dyn_row <- make_dynam_row(conv_to_inoceram_all_nebr_overlap,"I. convexus","D. nebrascense-E. jenneyi")
subl_to_inoceram_all_nebr_dyn_row <- make_dynam_row(subl_to_inoceram_all_nebr_overlap,"I. sublaevis","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_nebr_dyn_row <- make_dynam_row(bara_to_inoceram_all_nebr_overlap,"I. barabini","D. nebrascense-E. jenneyi")

bara_to_inoceram_all_rees_dyn_row <- make_dynam_row(bara_to_inoceram_all_rees_overlap,"I. barabini","B. reesidei-B. eliasi")

sage_to_inoceram_all_rees_dyn_row <- make_dynam_row(sage_to_inoceram_all_rees_overlap,"I. sagensis","B. reesidei-B. eliasi")

oblo_to_inoceram_all_rees_dyn_row <- make_dynam_row(oblo_to_inoceram_all_rees_overlap,"I. oblongus","B. reesidei-B. eliasi")

gen_to_sp_dyn_results <- rbind(conv_to_inoceram_all_redu_dyn_row,subl_to_inoceram_all_redu_dyn_row,
sask_to_inoceram_all_redu_dyn_row,conv_to_inoceram_all_nebr_dyn_row,
subl_to_inoceram_all_nebr_dyn_row,bara_to_inoceram_all_nebr_dyn_row,
bara_to_inoceram_all_rees_dyn_row,sage_to_inoceram_all_rees_dyn_row,
oblo_to_inoceram_all_rees_dyn_row)

## Write csv files of results
write.csv(gen_to_sp_I_D_results,file="tables/gen_all_to_sp_I_D_results.csv")
write.csv(gen_to_sp_eq_results,file="tables/gen_all_to_sp_eq_results.csv")
write.csv(gen_to_sp_sim_results,file="tables/gen_all_to_sp_sim_results.csv")
write.csv(gen_to_sp_dyn_results,file="tables/gen_all_to_sp_dyn_results.csv")
write.csv(gen_to_sp_eq_dyn_p_results,file="tables/gen_all_to_sp_eq_dyn_p_results.csv")
write.csv(gen_to_sp_sim_dyn_p_results,file="tables/gen_all_to_sp_sim_dyn_p_results.csv")

##### RUN PCA OF MORE THAN 2 SPP TO GENUS #####

## Extract values to matrix for each raster stack of env proxy variables:
redu_taxa_list <- list(conv_redu_prep,subl_redu_prep,sask_redu_prep,Inoceramus_all_redu_prep)
nebr_taxa_list <- list(conv_nebr_prep, subl_nebr_prep, bara_nebr_prep, Inoceramus_all_nebr_prep)

rees_taxa_list <- list(bara_rees_prep, sage_rees_prep, oblo_rees_prep, Inoceramus_all_rees_prep)

name_list_redu <- c("I. convexus", "I. sublaevis", "I. saskatchewanensis", "Inoceramus")
name_list_nebr <- c("I. convexus", "I. sublaevis", "I. barabini", "Inoceramus")
name_list_rees <- c("I. barabini", "I. sagensis", "I. oblongus", "Inoceramus")

# Create function to tabulate PCA scores for plotting spp and genus information
# Function will only take a list of prepped occurrence/raster stack data
# and output a list of 1-Joined (global) environment pts, 2-pts for each spp/gen,
# and 3-the convex hull index values for creating polygons around spp/gen
run_gen_spp_pca <- function(occ_raster_list, name_list) {
  bg_score_list <- list() # make list to put bg scores into

  for (i in 1:length(occ_raster_list)) {
    # Get scores for each environment's backgrounds
    first_bg <- getValues(occ_raster_list[[i]][[2]])

    ## Clean out missing values:
    first_bg <- first_bg[complete.cases(first_bg),]

    bg_score_list[[i]] <- first_bg
  }

  ## Combined global environment:
  joined_bg <- list.rbind(bg_score_list)
# Run PCA analysis on combine BG pts (all env data)

```r
pca_joined_bg <- dudi.pca(joined_bg, center = TRUE,
                         scale = TRUE, scannf = FALSE, nf = 2)
joined_bg_scores <- pca_joined_bg$li  # get the pca scores
```

```r
occ_loc_list <- list()  # Make list to put occ locations in PCA into
for (i in 1:length(occ_raster_list)){
  # Map occurrence data into the 2d ordination (have to coerce spatialPointsDataFrame into data.frame)
  # Explicitly match the colnames so only using the right ones
  # Only selected the li element (doesn't include others)
  first_occ_scores <-
    suprow(pca_joined_bg,
           data.frame(occ_raster_list[[i]][[1]][, colnames(joined_bg)])$li
           )
  occ_loc_list[[i]] <- first_occ_scores
}
```

# Add column with spp/gen names to occ location list:
for (i in 1:length(occ_raster_list)) {
  occ_loc_list[[i]]$tax <- name_list[i]
}

# Bind all the taxa localities together
occ_loc_df <- list.rbind(occ_loc_list)

bg_loc_list <- list()  # make list to put bg locations in PCA into
for (i in 1:length(occ_raster_list)) {

    # Map BG data to the 2d ordination
    first_bg_scores <- suprow(pcaJoinedBG, bg_score_list[[i]])$li

    bg_loc_list[[i]] <- first_bg_scores

}

# NOT USED ANYMORE, BUT MAYBE STILL WANT TO KEEP IN CASE
# hull_list <- list() # make list to hold convex hulls
#
# for (i in 1:length(occ_raster_list)) {
#
#    # Make convex hulls to put around points for each spp/genus
#    first_hull <- chull(occ_loc_list[[i]])
#    first_hull <- c(first_hull, first_hull[1]) # Get list of index values for making a polygon
#
#    # hull_list[[i]] <- first_hull
#    #
# }

final_list <- list(joined_bg_scores, occ_loc_df)

return(final_list)
}

# Use above function to make list of PCA information for plotting in each interval
redu_taxa_pca_list <- run_gen_spp_pca(redu_taxa_list, name_list_redu)
nebr_taxa_pca_list <- run_gen_spp_pca(nebr_taxa_list, name_list_nebr)
rees_taxa_pca_list <- run_gen_spp_pca(rees_taxa_list, name_list_rees)

# Use listed outputs from above function run on list of spp/gen to make plots:

library(devtools)
library(ggConvexHull)

pdf("Figures/pca_genus_all_to_spp_figures.pdf")
# REDU
ggplot(redu_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
  geom_point(data=redu_taxa_pca_list[[1]], aes(x=Axis1, y=Axis2), color='grey25') + # global env pts
  geom_convexhull(alpha = 0.3, aes(fill = tax, color = tax), linewidth = 1) +
  scale_fill_manual(values = c("#56B4E9", "#E69F00", "#009E73", "#999999")) +
  scale_color_manual(values = c("#56B4E9", "#E69F00", "#009E73", "#999999")) +
  ggtitle("PCA of Inoceramus genus and Species in B. reduncus-B. scotti") +
  theme_light() +
  theme(legend.title = element_blank())

# NEBR

ggplot(nebr_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
  geom_point(data=nebr_taxa_pca_list[[1]], aes(x=Axis1, y=Axis2), color='grey25') + # global env pts
  geom_convexhull(alpha = 0.3, aes(fill = tax, color = tax), linewidth = 1) +
  scale_fill_manual(values = c("#56B4E9", "#E69F00", "#009E73", "#999999")) +
  scale_color_manual(values = c("#56B4E9", "#E69F00", "#009E73", "#999999")) +
  ggtitle("PCA of Inoceramus genus and Species in D. nebrascense-E. jenneyi") +
  theme_light() +
  theme(legend.title = element_blank())

# REES

ggplot(rees_taxa_pca_list[[2]], aes(x=Axis1, y=Axis2)) + # first spp/genus pts
library(ggplot2)

setwd("C:/Users/ceara/Documents/Province Project/ENM_Analysis_Code-and-Files")

## Duration vs D in interval comparisons

dat <- read.csv("csv_for_dat_analysis/duration_d_stability_vals.csv")

# fit linear regression model to dataset and view model summary
dat_model <- lm(d~ratio, data=dat)

summary(dat_model)
names(summary(dat_model))

r <- summary(dat_model)$adj.r.squared

## Mean Duration vs mean D in interval

dat_mean <- read.csv("csv_for_dat_analysis/mean_duration_d_stability_vals.csv")
# Fit linear regression model to dataset and view model summary

```r
mean_dat_model <- lm(d_mean~ratio, data=dat_mean)
r_mean_dat <- summary(mean_dat_model)$adj.r.squared
```

# Export corr plots

```r
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

dat$intervals_compared<-sub("to", "to \n ", dat$intervals_compared)
dat_mean$t_comparison<-sub("to", "to \n ", dat_mean$t_comparison)

pdf(file = "Figures/corr_plots_duration_overlap.pdf",width = 2.75591, height = 2.75591 )

```r
  ggplot(data = dat) +
  geom_point(aes(x = ratio, y = d, color = intervals_compared), shape = 19, size = 0.3) +
  geom_smooth(method='lm', mapping = aes(x = ratio, y = d), size=0.3) +
  theme_light() +
  labs(x = "Duration Ratio (biozone : substage/stage)", y = "D", color = "Intervals Compared") +
  scale_colour_manual(values=cbPalette) +
  theme(plot.title = element_text(size=9, hjust=0.5),
        axis.title.x = element_text(size=9, hjust=0.5),
        axis.title.y = element_text(size=9, vjust=-1.5),
        axis.text.x = element_text(size=6, vjust=2),
        axis.text.y = element_text(size=6, angle=90, vjust=-0.5),
        legend.position="bottom",
        legend.text = element_text(size=6),
        legend.title= element_blank(),
        legend.spacing.x = unit(-0.15, 'cm'),
        legend.spacing.y = unit(-0.15, 'cm'),
        legend.box.margin=margin(-12,15,-10,0)) +
  annotate(geom="text", x=0.4, y=0.9, size = 3, label=paste("R^2 == ", round(r, digits = 3)), parse=TRUE) +
  guides(color = guide_legend(byrow=TRUE, nrow = 4))
```

ggplot(data = dat_mean) +
geom_point(aes(x = ratio, y = d_mean, color = t_comparison), shape = 19, size = 0.3) +
geom_smooth(method='lm', mapping = aes(x = ratio, y = d_mean), size=0.3) +
theme_light() +
labs(x = "Duration Ratio (biozone : substage/stage)", y = "Mean D", color = "Intervals Compared") +
scale_colour_manual(values=cbPalette) +
theme(plot.title = element_text(size=9, hjust=0.5),
      axis.title.x = element_text(size=9, vjust=2.5),
      axis.title.y = element_text(size=9, vjust=-1.5),
      axis.text.x = element_text(size=6, vjust=2),
      axis.text.y = element_text(size=6, angle=90, vjust=-0.5),
      legend.position="bottom",
      legend.text = element_text(size=6),
      legend.title= element_blank(),
      legend.spacing.x = unit(-0.15, 'cm'),
      legend.spacing.y = unit(-0.15, 'cm'),
      legend.box.margin=margin(-12,15,-10,0)) +
annotate(geom="text", x=0.3, y=1, size = 3, label=paste("R^2 == ", round(r_mean_dat, digits = 3)),
         parse=TRUE) +
guides(color = guide_legend(byrow=TRUE,nrow = 4))

dev.off()

## Duration vs mean D in interval (spp comparisons)
dat_spp <- read.csv("spp_d_stability_vals.csv")

ggplot(data = dat_spp) +
geom_point(aes(x = short_dur, y = d, color = species)) +
geom_smooth(method='lm', mapping = aes(x = short_dur, y = d))

# fit linear regression model to dataset and view model summary
model <- lm(d ~ short_dur, data=dat_spp)
summary(model)

## Duration ratio vs mean D in interval

dat <- read.csv("duration_d_stability_vals.csv")

# plot the data
plot(dat$ratio, dat$d, pch = 19, col = "lightblue")

# Add regression line
abline(lm(dat$d ~ dat$ratio), col = "red", lwd = 3)

# Pearson correlation
text(paste("Correlation: ", round(cor(dat$ratio, dat$d), digits = 2)), x = 0.25, y = 0.78)

## Duration ratio vs mean D in interval

dat_mean <- read.csv("mean_duration_d_stability_vals.csv")

# plot the data
plot(dat_mean$ratio, dat_mean$d_mean, pch = 19, col = "lightblue")

# Add regression line
abline(lm(dat_mean$d_mean ~ dat_mean$ratio), col = "red", lwd = 3)

# Pearson correlation
## Duration vs mean D in interval (spp comparisons)

```r
dat_spp <- read.csv("spp_d_stability_vals.csv")

# plot the data
plot(dat_spp$short_dur, dat_spp$d, pch = 19, col = "lightblue")

# Add regression line
abline(lm(dat_spp$d ~ dat_spp$short_dur), col = "red", lwd = 3)

# Pearson correlation
text(paste("Correlation: ", round(cor(dat_spp$short_dur, dat_spp$d, digits = 2), x = 1.3, y = 0.7))
```

---

**Note:** The code above is an R script that demonstrates the correlation between short duration (short_dur) and mean D (d) for species comparisons. The script reads a CSV file named `spp_d_stability_vals.csv`, plots the data, adds a regression line, and calculates the Pearson correlation coefficient.