COMATULID CRINOIDS IN THE FOSSIL RECORD: METHODS AND RESULTS FOR THE EXTREMELY IMPERFECT

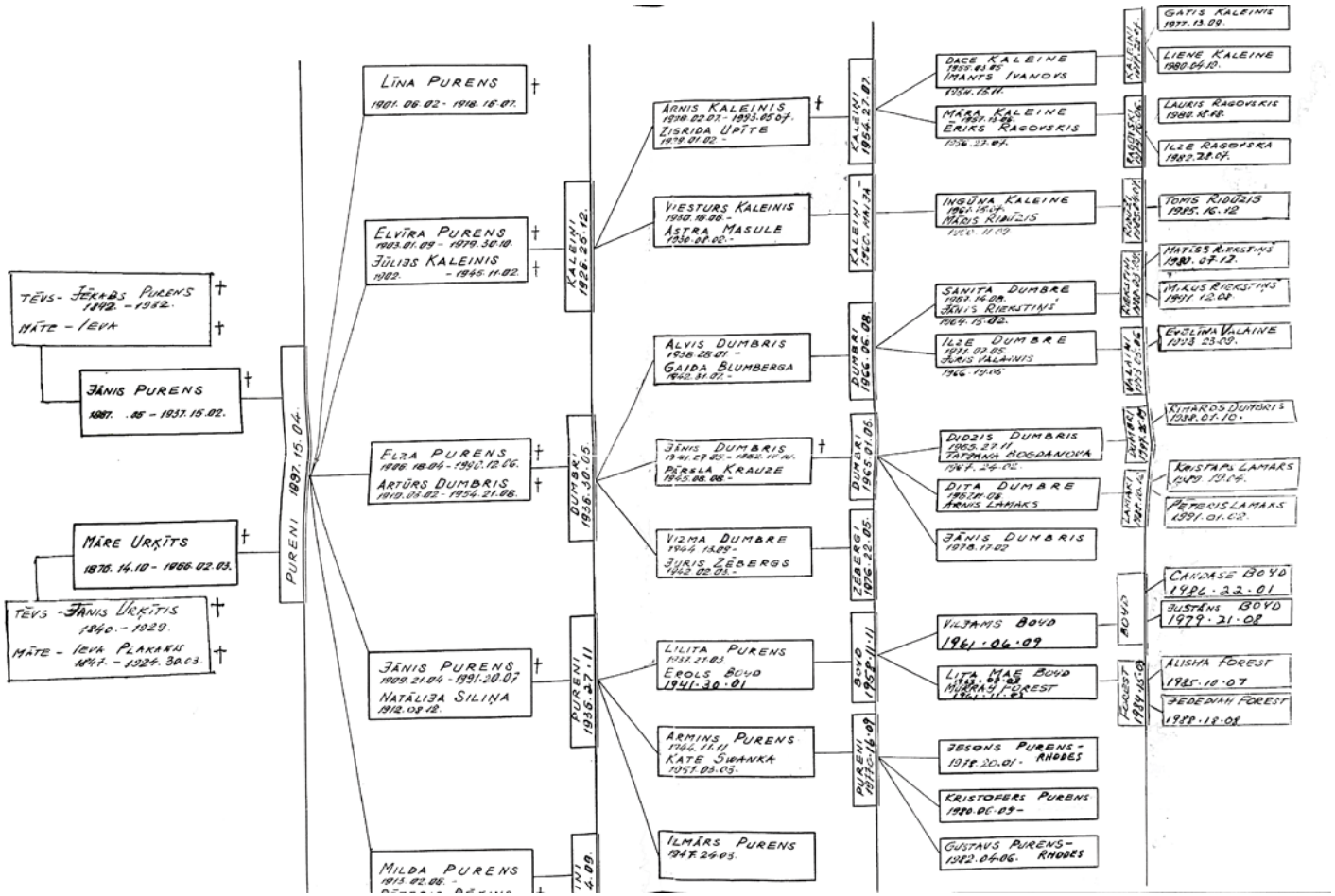
by

Kristopher J. S. Purens

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Geology)
in the University of Michigan
2014

Doctoral Committee:

Professor Tomasz K. Baumiller, Chair
Emeritus Professor Gerald R. Smith
Associate Professor Jeffrey A. Wilson
Associate Research Scientist Miriam L. Zelditch



The evolutionary tree of life is an unbroken chain stretching from billions of years ago to the present day. The most proximal part is pictured here.

DEDICATION

Dedicated to my grandparents: Janis and Natalia Purens, Ben and Josephine Swanka.

ACKNOWLEDGMENTS

I would like to thank many for their support and encouragement. This work would not have been possible without mentorship and countless assistance from T. Baumiller. Committee members M. Zelditch, G. Smith, and J. Wilson have provided advice and discussion. Thanks to my fellow Baumiller lab members A. Janevski, V. Syverson, and M. Veitch for support and assistance. Additional discussion was provided by C. Messing, L. Hemery, D. Miller, D. Meyer, and the University of Michigan Paleontology Seminar participants. Access to museum specimens was assisted by D. Pawson, J. Jagt, J. Bleaker, and M. Eléaume.

On a personal level, Abigail Purens has been a steadfast supporter through good and bad, and without her I could not have succeeded. Many other friends have been tremendously important for my continued progress, including Daniel Poon, Emile Moacdieh, Andrea Jokisaari, Becca Anderson, Joe Anderson, Per Wiger, Tom Rose, Ivan Ludmer, Megan Regel, Lauren Jacobs and many others. Of particular note is Meredith Petrie's near daily help in staying focused. My brothers, Jason and Gus, and my mother, Kate, have always been supportive, and I thank them for their continued presence and encouragement. My father, Armins, provided me with copies of many of S. J. Gould's books — which was greatly inspirational.

Funding support was provided by the Department of Earth and Environmental Sciences. NSF grant 1036393 (T. K. Baumiller, PI) — Collaborative Research: Assembling the Echinoderm Tree of Life was an important source of funding, as well as NSF grant EAR 0824793and National Geographic Society NGS 8505-08.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

LIST OF APPENDICES

ABSTRACT

Paleontologists' attempts to understand patterns of evolutionary change have always been influenced by biases in both the fossil record and collection/description. This dissertation has two basic goals: to better understand biases, and to better understand patterns of evolutionary change. Comatulid crinoids, the most abundant modern crinoids, have a depauperate described fossil record, with an order of magnitude lower generic diversity reported from any stage than that described from the modern ocean. While comatulids have generally been described as having a poor fossil record, the nature of that record is unclear, and little work has attempted to understand potential sources of bias. Two methods are used to address potential bias in the described record: morphological analysis of comatulid centrodorsals to test the role of differential material used for modern and paleontological taxonomic descriptions, and a capture-mark-recapture (CMR) method to understand regional and temporal difference in detection rate. To address patterns of morphological change, we examine the disparity of comatulid centrodorsals, both within species, and for the whole group through geologic time.

Investigation of intraspecific variation of comatulid centrodorsals finds no evidence of lumping bias in fossil taxonomic descriptions. However, analysis using CMR found substantial differences in detection rate both regionally and temporally. Extremely low detection rates, especially in the non-European Cenozoic, mean that there are substantial biases that hide a significant increase in comatulid diversity over the Cenozoic. Collection efforts should be undertaken to better understand whether the bias is in the fossil record, or in efforts to collect and describe comatulid material from the Cenozoic.

Previous efforts to describe the pattern of morphological change in crinoids have shown a pattern of rapid expansion into morphospace followed by stasis. Such a pattern can be explained by a decrease in rates of evolution, or as constraints preventing further dispersion into morphospace. These explanations have differing signatures on subclade disparity and homoplasy. A lowering rate of change scenario should mean homoplasy is rare and subclade disparity to be low, while the constraints scenario should result in homoplasy being more common and subclade disparity being high. We found evidence for common homoplasy and high subclade disparity for the comatulids. This supports constraints as the explanation for the pattern of early expansion into morphospace followed by stasis.

CHAPTER I

Introduction

Charles Darwin famously quipped that the fossil record was extremely imperfect, and it is no secret that only a tiny fraction of life's history is captured therein (Darwin 1859). His motivation for noting the imperfection was to excuse the relative lack of transitional forms known to science— essentially, he was attempting to work within the confines of available data to understand the history of life on Earth. This dissertation focuses addresses both the imperfection of the fossil record, and the patterns and processes of evolutionary change.

Imperfection of the fossil record— While the fossil record is imperfect, our knowledge of what exists within that fossil record is imperfect as well. A common theme through recent work into the nature of the fossil record is that deficiencies therein can be rigorously addressed. A great deal has been learned from studies that seek to understand taphonomic processes for crinoids, including tumbling experiments (Baumiller 2003, Gorzelak and Salamon 2013) and biostratinomic investigations (Meyer and Meyer 1986), as well as attempts to quantify completeness (Foote 1999). These studies have generally found that crinoids are taphonomically similar to other echinoderms.

Collections, sampling, and description can also create biases within the scientific literature. Sampling biases have also been a subject of intense interest, but research into these has generally focused on regional biases, sampling effects, and ways to fairly compare assemblages. Less studied aspects include the ease with which species can be recognized, compared to the modern, and quantification of regional biases in the fossil record.

Patterns and processes of evolutionary change— The origin of morphological diversity, or disparity, has been a major focus of the paleobiology revolution. G. G. Simpson famously postulated that the sudden appearance of new forms in the fossil record was the result of "quantum evolution", a brief period of rapid evolution from one adaptive zone to another (Simpson 1953). Interest in the origin of disparity increased as methods to quantify morphospace (Raup 1962) became easier with increased computing power. Gould's seminal *Wonderful Life* (1986) piqued interest in patterns of morphological diversity, with claims of the importance of contingency and that unique processes acting early in group's histories limited what morphologies could arise later. One important set of papers that studied the disparity of crinoids (Foote 1994, 1996, 1999) found patterns of early maximal disparity with little net change afterwards. This pattern has proven to be common across many groups (Hughes et al. 2013), seemingly supporting the importance of "quantum evolution" and contingency. Other studies have found that high rates early in group's history are rare (Harmon et al. 2010), and the apparently different results between these different studies have yet to be reconciled.

Comatulid crinoids— This dissertation is focused on one taxonomic group, the comatulid crinoids. One of the most powerful tools in the paleontologist's toolkit is using modern analogues to understand ancient organisms: the present as a key to the past. Crinoids are some of the most abundant organisms during the Paleozoic, but have much lower ecological significance today. It is perhaps not surprising then, that paleontologists have a disproportionate interest in modern crinoids. However, modern crinoids are not perfect analogues for Paleozoic crinoids. The most important distinction is that the most abundant crinoids today, the comatulids, are stalkless and mobile (Fig. 1.1). Stalked crinoid forms can be abundant today, but generally only at depths below 100 meters (Meyer and Macurda 1977). Mobility in comatulids is hypothesized to be a response to increased predation pressure (Meyer and Macurda 1977, Baumiller et al. 2010). Modern comatulids are gracile compared to Paleozoic crinoids, with a greatly reduced calyx and

delicate arms. One key element in comatulids is the centrodorsal, an aboral structural element that is found where the stalk is found in the typical Paleozoic crinoid. The centrodorsal is a cup-like interface between the arm and cirri. It is the largest single element in a comatulid, and is the most common described fossil element for the comatulids.

The post-Paleozoic crinoid diversity paradox—One major motivation for this work is a desire to understand an unusual pattern (Fig. 1.2). Comatulids vastly outnumber other crinoids in today's oceans, but show lower apparent diversity from the fossil record. Comatulids account for around 80% of modern crinoid diversity and less than 50% of fossil crinoid diversity over the post-Paleozoic. Two basic possibilities seem apparent: an extraordinary late Cenozoic radiation for comatulids, or a bias that disproportionately prevents comatulids from entering the scientific literature.

The *Treatise on Invertebrate Paleontology* (Moore and Teichart 1978) — the definitive compendium of fossil invertebrate morphology — reveals a striking difference in material described for Paleozoic and comatulid crinoids. While the most common Paleozoic crinoid fossils are disarticulated elements, articulated specimens are common enough to fill the *Treatise*. The comatulid's pages in the *Treatise* are distinctly different: page after page of centrodorsals, with occasional modern specimens used to illustrate whole organism morphology. This material problem isn't limited to the *Treatise*- it is evident in museum collections as well (Fig. 1.3). This leads to an obvious question: does the material used to define species create a serious bias between the fossil record and the modern?

Chapter II tests if centrodorsals alone are sufficient to differentiate closely related species of comatulids. If centrodorsals are not a robust source of taxonomic information, species described by paleontologists may not be equivalent to species defined by neontologists. If a single fossil genus is equivalent to ~10 modern genera, the diversity paradox disappears. To answer this, a detailed morphological analysis of comatulid

centrodorsals intraspecific variation was performed. The results, presented in chapter two, suggest that this is a not a serious problem.

Chapter III focuses on a different style of analysis, measuring disparity, or morphological diversity of comatulid centrodorsals. Given the observation from chapter II that centrodorsals are taxonomically informative, we measure the morphological patterns centrodorsals shows through geologic time. One common pattern observed in many taxa is high apparent rates of morphological evolution early in a clade's history, followed by a slower infilling of taxonomic richness. We find that comatulid centrodorsal disparity was high even near the origin of the group, in the Jurassic, and has been fairly consistent since then. Understanding what processes create this kind of pattern is important, and a framework is used that helps bridge the gap between observations by paleontologists and those by neontologists for patterns of morphological evolution.

Chapter IV focuses on understanding the pattern of detection biases through time and space for the comatulids. There are many potential biases that could prevent comatulid species from entering the paleontological literature, but they would generally fall under two categories: failure to enter the fossil record; or failure to be found, recognized and described by paleontologists. Differentiating between these two might be difficult, but one solution is to understand how detection rates differ between areas of high paleontologist effort versus low areas of effort. Using a capture-mark-release technique, we test the patterns of time and space to see if the record is heterogeneous or consistent— poorly sampled, or well sampled. If the regions and time intervals where paleontologists have spent the most effort show fairly complete sampling, it bodes well that the gap between modern and fossil diversity can be improved upon. Additionally, this method lets us estimate how many undescribed taxa may have once existed – a way to understand if the comatulids did undergo a significant Cenozoic radiation, as well as a guide for where future efforts should be directed.

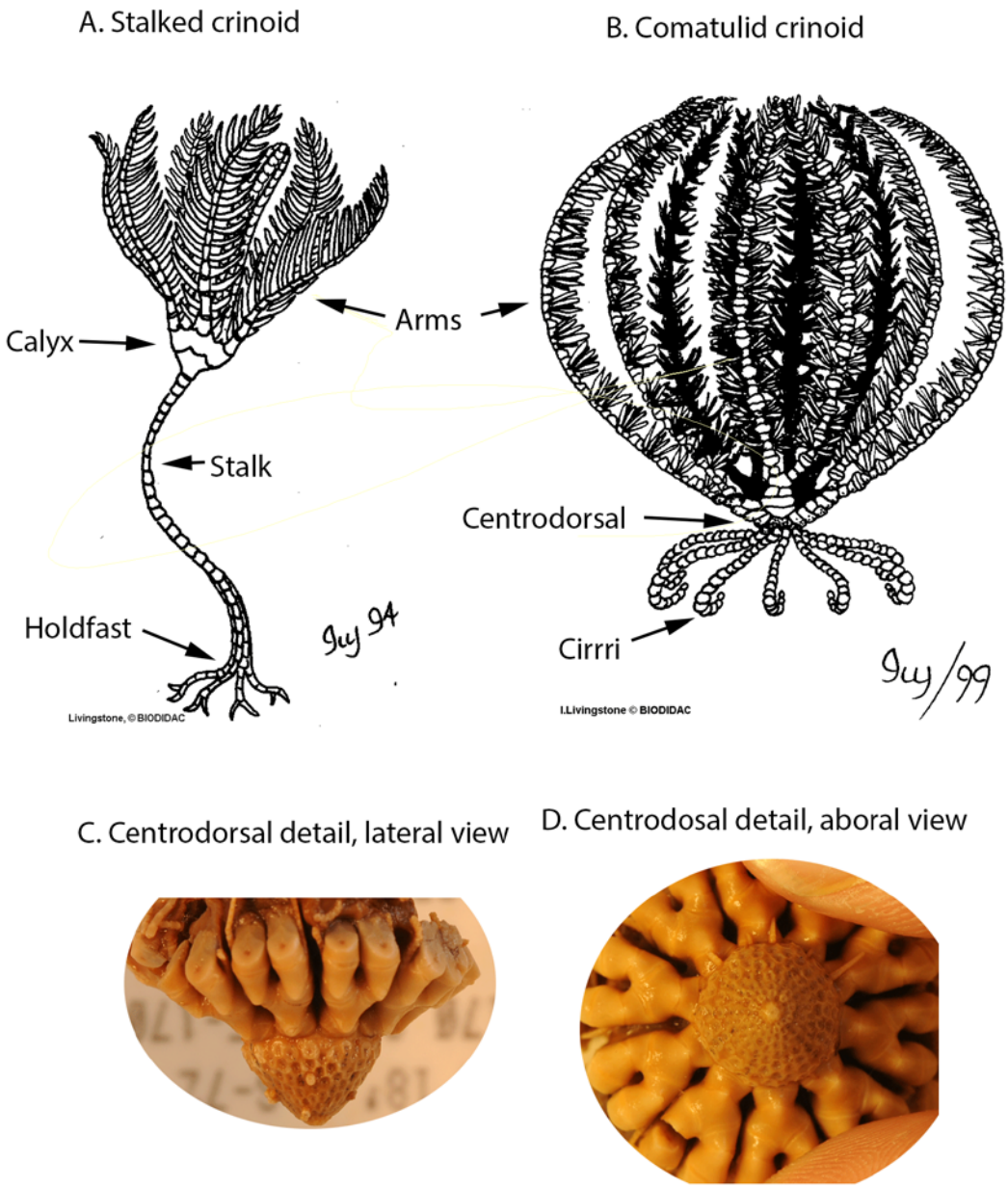


Figure 1.1 Generalized crinoid morphology. The typical Paleozoic crinoid is stalked (A), while the most abundant crinoids today, the comatulids (B) are unstalked. Located where the stalk would be attached to the calyx in the stalked crinoid is a comatulid's centrodorsal (C, D). The centrodorsal is a key structural element that connects arms to cirri.

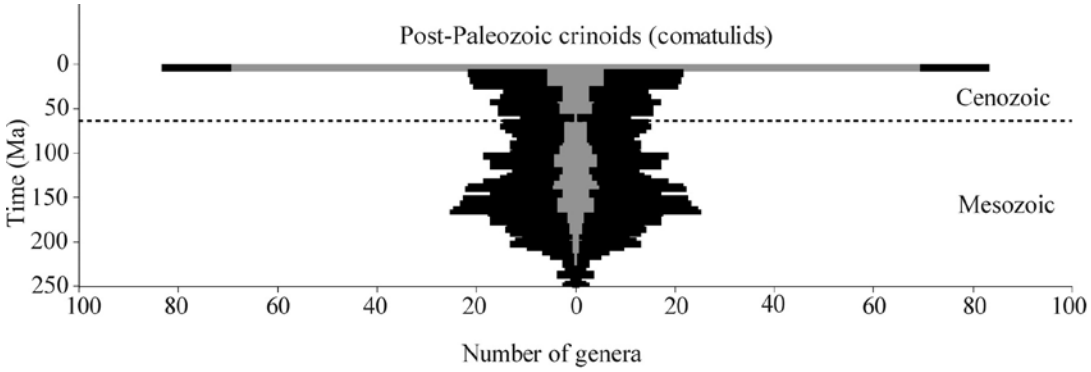


Figure 1.2 Post-Paleozoic crinoid diversity, adapted from Janevski and Baumiller 2010. Comatulid diversity is depicted in grey while non-comatulid diversity is depicted in black. Comatulids show a large increase from the fossil record to the modern, while non-comatulid fossil disparity is commensurate with the modern. Understanding this pattern is a primary focus of this dissertation.

Figure 1.3 Modern crinoids are abundant in thousands of whole specimens available (left). In contrast, paleontological collections can be rather sparse (right), with much less material available. These photos are from the USNM's Marine Invertebrate and Paleobiology collections, respectively.

CHAPTER II

Taxonomic Bias Does Not Explain Low Fossil Diversity In Comatulid Crinoids

Abstract

This work seeks to understand the impact that taxonomic bias may exert on the diversity history of the comatulid crinoids. While neontologists can use a whole organism for taxonomic description, paleontologists focus on only one element, the centrodorsal, the element most often described for fossil comatulids. With complete specimens available, one might expect that neontologists are able to discriminate more species, resulting in a bias that would result in a lower apparent diversity of fossil versus extant crinoids. However, neontologists generally do not use many of the centrodorsal characters available for taxonomic description that are exploited by paleontologists, potentially biasing upwards diversity of fossil relative to extant crinoids, provided centrodorsals are a rich source of information.

In this study, centrodorsal shape of modern and fossil comatulid species were measured using quantitative methods that can be applied uniformly to both groups. Two different methods applied to centrodorsal shape, disparity and finite mixture analysis, reveal no obvious bias of over- or under-splitting of Recent versus fossil comatulid species. Interestingly, the methods identified a putative modern cryptic species within an extant species complex that also finds support in molecular data. With no evidence found of taxonomic bias driving the diversity record of fossil comatulids, sampling and preservation are the likely sources of bias producing the 10-fold higher diversity of extant over fossil comatulids.

Introduction

"All the king's horses and all the king's men
Couldn't put Humpty together again"

-Traditional nursery rhyme

The problem of fidelity in the fossil record has long been a specter haunting the interpretation of paleontological data. Biases in the fossil record caused by differences in material available for taxonomic assessment are important in many groups, but can be particularly acute for organisms that disarticulate rapidly upon death. Just as the king's men labored to reconstruct Humpty Dumpty, paleontologists have long labored to describe and reconstruct ancient organisms based on remains in varying degrees of disaggregation and completeness. As a result of the partial material available to paleontologists, many characters used to differentiate extant taxa, such as soft parts and behavior, are rare or unobservable in the fossil record. Even in cases where preservable hard parts are the basis for taxonomic description, much information can be lost to taphonomic processes. In the case of common, complete disarticulation, taxonomic work must focus on what is available. Focusing taxonomic efforts on a single, highly identifiable element, is one strategy used extensively in study of comatulid crinoids. This paper assesses the comparability between taxonomic descriptions generated from whole specimens of modern organisms, versus those described from fragmentary fossil material of comatulid crinoids.

One of the major tasks undertaken by paleontologists over the past centuries has been to catalogue biological diversity through geologic time, culminating in efforts to describe relative diversity throughout the Phanerozoic (Sepkoski et al. 1981, Alroy et al. 2008). Comatulids, the most diverse extant crinoids, are stalkless and mobile, ranging worldwide from the abyssal ocean depths to shallow reefs. The diversity record of post-

Paleozoic crinoids reveals a striking pattern (Fig. 2.1): whereas the non-comatulid crinoids show similar levels of diversity in the fossil and modern records, the diversity of the comatulids jumps by about an order of magnitude between any fossil time bin and the modern. Such a pattern demands explanation.

This paper focuses on one potential bias that may account for such a dramatic difference in diversity: differences in taxonomic practice between neontologists and paleontologists. New species of living comatulids are generally described from whole specimens using taxonomic characters present on arms, pinnules, and cirri. In contrast, because articulated fossil comatulids are very rare, fossil comatulids are generally described using the centrodorsal (CD) element alone. The CD serves as the interface between the comatulid's cirri and arms, making it a key structural element. CDs are easily recognizable, and often the largest single plate in the comatulid. CDs are homologous to proximal columnals in stalked crinoids. Taxonomic characters used to describe fossil material are generally the size and shape of the whole CD; number, size and arrangement of cirral scars; and shape and size of the oral cavity. Descriptions of fossil comatulid taxa, relying heavily on the small number of characters of the CD, and lacking many morphological characters available to neontologists, might be expected to result in recognition of fewer taxa among fossil comatulids. Is it possible that the sharp rise between fossil and Recent comatulid diversity could be the result of the disjunct character sets used for taxonomic descriptions?

Therefore, the question addressed here is whether a single paleontological species represents multiple modern species, i.e., whether CD disparity corresponds tightly with modern species, or even corresponds with modern taxonomy at all. If much of the difference between fossil and modern diversity can be explained by this type of taxonomic bias, there should be several indicators that support this. In this paper, two methods are used for testing this hypothesis: morphospace volume (disparity), and finite mixture modeling. If a single fossil species is equivalent to many modern species, one

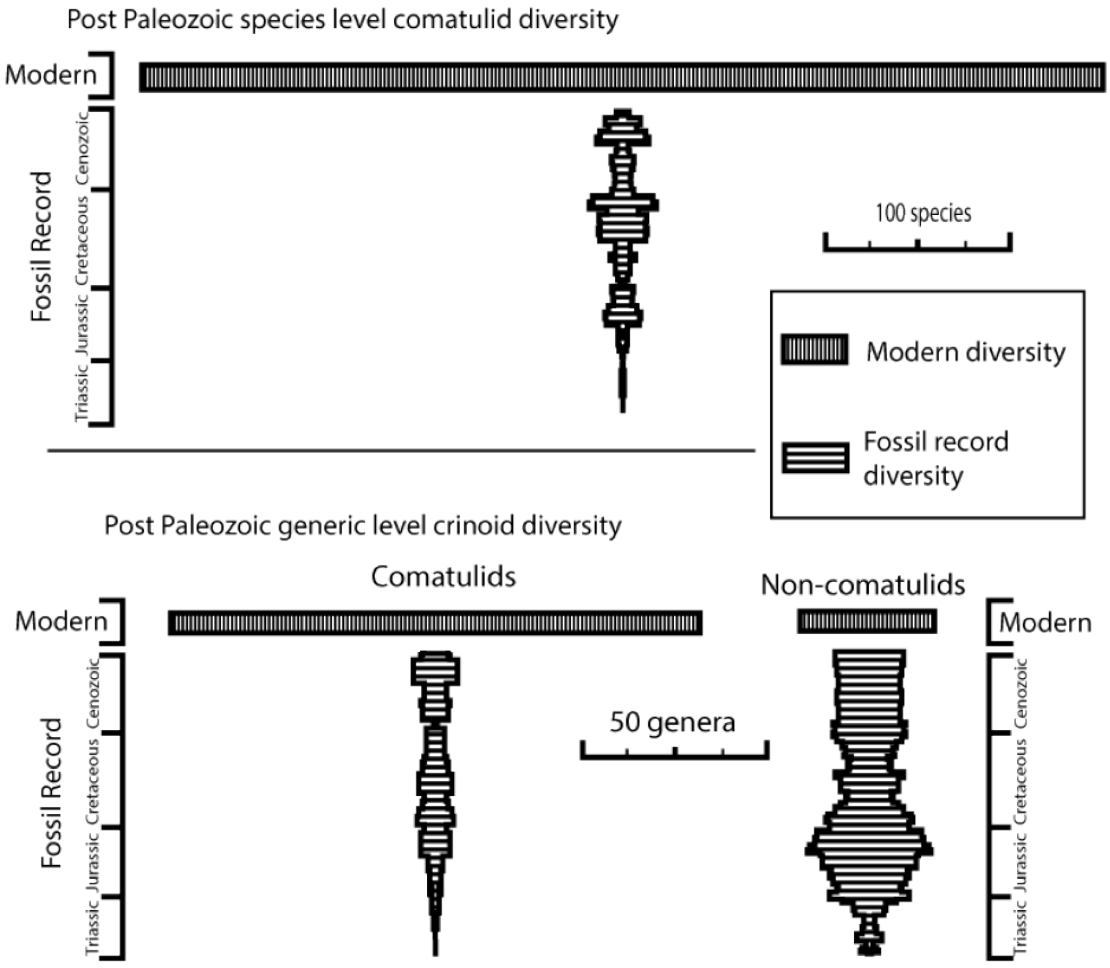


Figure 2.1 Diversity through time for post-Paleozoic crinoids. At the species level, comatulids have an average of ~15 times more diversity in the Recent compared to the fossil record. At the generic level, comatulids (bottom left) show the same qualitative pattern as at the species level, while non-comatulids (right) show a much more modest increase into the modern (Janevski and Baumiller 2010).

would expect the disparity of the fossil taxon to cover more morphospace than any of the modern taxa under any evolutionary model excepting strict stasis (Fig. 2.2: A, B). The increase in morphospace coverage as more species are subsumed will not be linear, but will instead depend upon the difference between the means of the taxa. If fossil species have a disparity (Fig. 2.2: C) that is different than the range suggested by modern taxa,

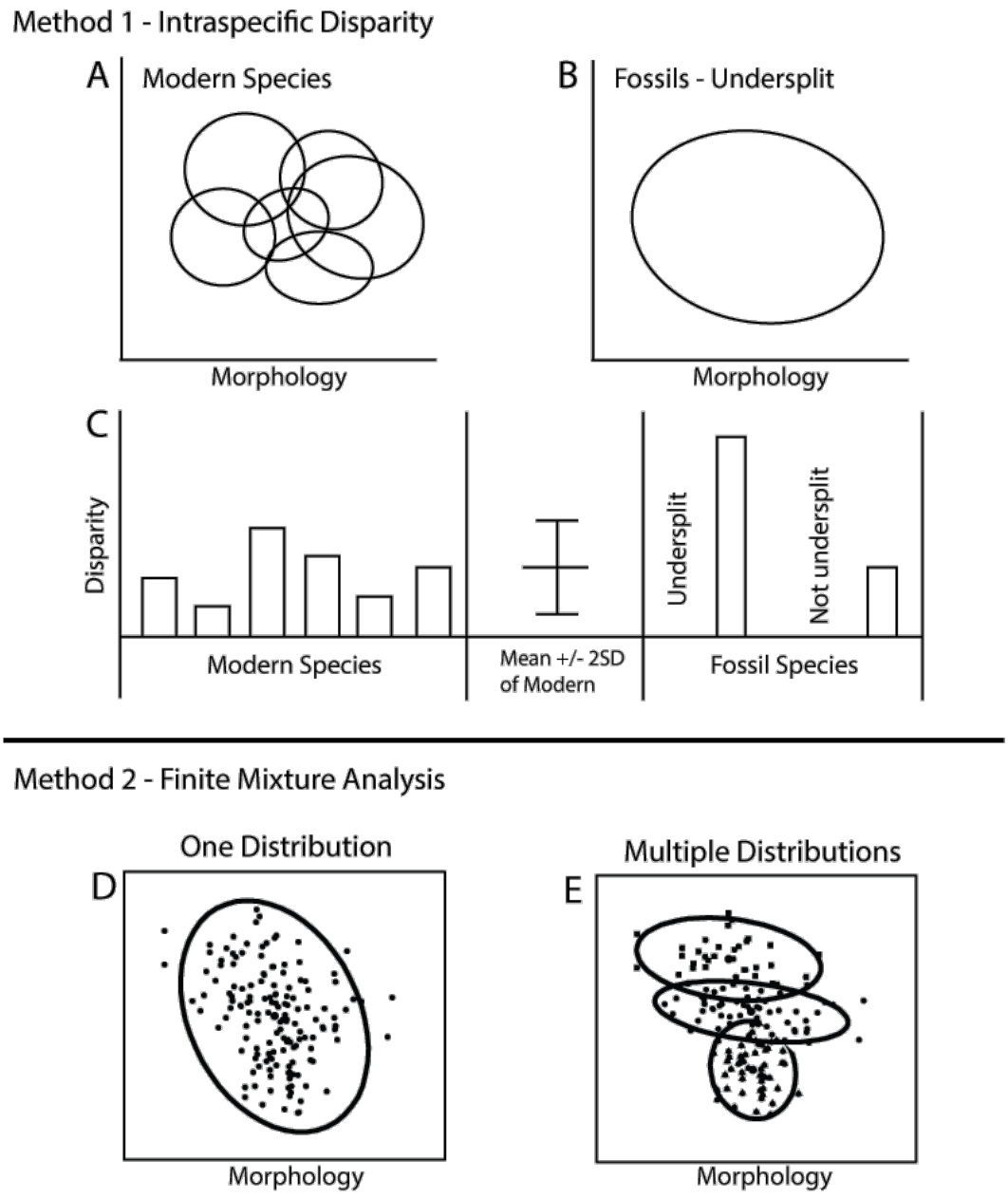


Figure 2.2 Explanation of tests to identify taxonomic bias. Given a high modern diversity of comatulids (A), one possible explanation is that (B) a fossil taxon is undersplit relative to the modern taxa. If fossil taxa are undersplit, they should cover more morphospace and therefore have greater disparity (C). The second method used, finite mixture analysis, takes a data set and estimates parameters for density functions to explain the data. This method tests whether one distribution (D) or multiple distributions (E) best explain the observed data. If fossil taxa are undersplit relative to the modern, we expect multiple distributions to be common when analyzed using FMA.

that is evidence that they are not equivalent units. The second technique is finite mixture analysis (FMA). This method seeks to explain the data as a number of normally distributed groups with varying means, covariances, and volumes (Fig. 2.2: D, E). While a measure of disparity can only detect bulk changes in morphospace occupied, finite mixture analysis can identify other differences in morphospace distribution such as varying means, covariances, and volumes. If analysis with these two techniques shows fossil species to have more groups than would be expected based on their current accepted taxonomy, it is evidence of hidden diversity in those taxa, and thus provides evidence of taxonomic bias.

Material and Methods

A total of 585 CDs from eight modern and three fossil species were included in this study (Table 2.1). Specimens were selected from museum collections with a preference toward larger sample sizes, and toward CDs from which cirri were already missing, in order to minimize degrading museum collections. Selection was in no way based on perceived patterns of morphological variation within the selected species, and as such they should be a random sample. All specimens within a given lot were assessed for usability. Specimens with abrasive wear or damage were removed from the sample. Specimens were also excluded if they would require extensive preparation before photography and data collection, such as removing attached cirri from modern specimens, or cemented grains obscuring the outlines of fossil specimens. Photographs were taken from aboral, lateral, and, if possible, oral views. Specimens were held in place with a clip, and rotated in order to provide consistent alignment for photography.

CD shape was measured using standard geometric morphometric (GM) methods (Zelditch et al. 2012). A total of four landmarks and fifteen semi-landmarks in lateral view were digitized using TPSDIG2 software (Fig. 2.3). These landmarks describe the overall lateral view shape of the CD, and were selected based on applicability across

Table 2.1 Specimens and sources. NBCN: Naturalis Biodiversity Center, Leiden, Netherlands; MNMN: Muséum National d'Histoire Naturelle, Paris, France; NMNH: National Museum of Natural History, Smithsonian, Washington, D.C., U.S.A.; NMM: Natuurhistorisch Museum, Maastricht, Netherlands.

Species	Sample size	Museum	Locality
Modern			
Comactinia meridionalis	49	NBCN	Caribbean
Florometra mawsoni	108	MNHN, NMNH	Antarctic Ocean
Florometra serratissima	50	MNHN, NMNH	Northern Pacific
Hathrometra tenella	10	NMNH	Martha's Vineyard, MA
Leptometra celtica	21	NBCN	Mediterranean Sea
Promachoensis kerguelensis	168	MNHN, NMNH	Antarctic Ocean
Psathyometra fragilis	7	NMNH	Monterey Bay, CA
Tropiometra carinata	24	NBCN	Atlantic Ocean
Fossils			
Jaekelometra belgica	62	NMM	Maastricht, Netherlands
Jaekelometra concava	54	NMM	Maastricht, Netherlands
Semiometra impressa	18	NMM	Maastricht, Netherlands

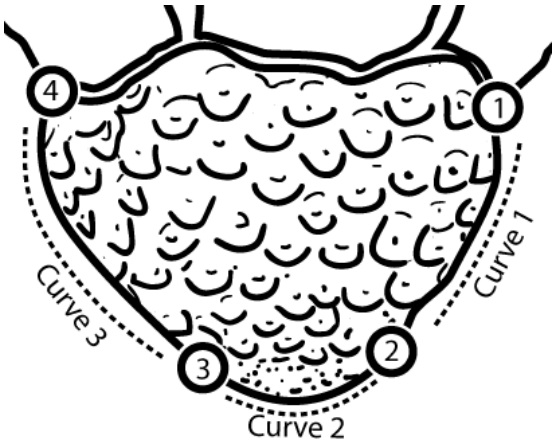


Figure 2.3 Landmark diagram. Four landmarks and three curves were used in this study. Landmark one lies in the midpoint of a radial and CD. Landmarks two and three lie at bottom edge of last cirral scar along margin. Landmark four is at intersection of interradial and CD. Semi landmark curves follow curvature of surface in plane as defined by landmarks.

Comatulida. After digitizing, data were analyzed in R (R Development Core Team 2005). Data were aligned by Generalized Procrustes Analysis (GPA) (Gower 1975, Rohlf and Slice 1990), with semi-landmarks aligned using the minimum Procrustes distance criterion. Superimposition was done using the package geomorph (Adams and Otárola-Castillo 2013). Superimposed GM data are provided as Supplementary Material.

Disparity was calculated based on the aligned morphometric data for each species (Zelditch et al. 2012). Disparity is measured as the variance of shape, computed by summing the variances over all the superimposed coordinates. 1000 bootstrap replicates were calculated for each taxon in order to determine how sampling may affect the results. For both modern and fossil specimens, the mean and +/- two standard deviations were calculated to describe the range of expected disparity. A one tailed Welsh's *t* test was conducted to test the hypothesis that the fossil mean disparity is larger than the modern mean disparity. This test is similar to the Student's *t* test with a correction for the samples potentially having unequal variance.

The package Mclust (Fraley and Raftery 1999) was used to generate normal mixture models in order to identify groups within the data. This approach differs from other methods such as discriminant function analysis and CVA in not requiring *a priori* classifications. There are four parameter classes in the models: (1) mean, (2) volume, (3) shape of the distribution, and (4) orientation of groups within the multivariate space. Models are selected according to an information criterion, penalizing the models for each parameter used. Among the competing models, the one with the lowest score loses the least information and is therefore preferred over the others. To begin with, FMA was conducted on the full sample of each species in order to test if fossil species showed evidence of multimodality beyond that seen in modern species using the Bayesian Information Criterion (BIC).

Subsequent analysis was conducted by grouping together species of the same genera with sample sizes >100, namely *Promachocrinus*, *Florometra*, and *Jaekelometra*.

This allows us to test the method's efficiency at detecting species in the same genus, as well as employing small sample size corrected Akaike information criterion (AICc) alongside BIC. These two criteria differ in the weights accorded to penalties; the AICc has a smaller penalty for adding parameters, compared to BIC, up to a limit. As the number of parameters approaches sample size (*n*), the penalty for AICc goes up dramatically, putting a hard limit on model size but allowing for more complex models up to that limit. This also means that the AICc does not work well for large models at small sample sizes, becoming extremely conservative or undefined, and as such this method is only used on those groups with sufficient sample size to allow detection of ~10 groups, which is adequate to explain much of the diversity jump between the fossil record and the modern. In general, the BIC generally does not choose too large a model whereas the AICc generally does not choose too small a model. Using the range of components from BIC:AICc should cover the range from type I to type II errors (Vrieze 2012).

Because of the nature of the models used here, dimension reduction was necessary in order to reduce the number of parameters. Principal component analysis (PCA) was conducted on each group subjected to this procedure. Three principal components (PC) were used for both BIC and AICc model fitting procedures, as a compromise between including as much data as possible and allowing a larger number of components to be present in the mixture models. This corresponds to the most common number of components selected by the broken stick model (Frontier 1976). In order to ensure that results were not overly influenced by number of PCs selected, comparison FMAs were also conducted through a broad range of PCs as the input. In general, the number of groups found was highest near the number of components indicated by the broken stick model. The comparative results between taxa were consistent across a broad range of PCs, so only the results based on three PCs are presented here.

Testing that this method can differentiate species in the same genus provides evidence that it is an effective tool for identifying hidden diversity within the fossil

record. In order to allow direct comparison between models at the genus level, each group was subsampled with replacement at $n = 116$, the lowest sample size among the taxa used; 1000 subsample replicates were done for each taxon; and the number of groups for each optimal model was recorded. For each taxon, calculations were performed to see how often the bootstrapped models of the other taxa resulted in fewer, equal, or more groups. Full results and a more complete explanation of this method are given in Appendix A. Bootstrapped finite mixture analyses of full samples of *Promachocrinus* and *Florometra* were also performed. These analyses are identical to the method described above, but instead of using the sample size of 116, they used the full sample sizes for each taxon. The results from these can be compared with those for the subsampled data to understand the importance of sample size for analysis of this kind. Chi-square tests were performed for *Florometra* and *Jaekelometra*, comparing the distribution of *a priori* identifications to the FMA model with an equal number of groups. If the chi-squared value is significant, it suggests a relationship between the *a priori* groups and the components identified by the finite mixture models. Lastly, in order to understand whether more groups are present in different genera, a method was developed to determine if a greater, lesser, or equal number of groups are found via FMA between two groups that had been resampled. By summing the number of times that another taxon has an equal, greater, or lesser number of groups, a direct comparison between taxa can be made. This method is fully described in Appendix A.

Results

Disparity of species is reported in Fig. 2.4. The disparities of modern and fossil taxa broadly overlap. The species with the highest disparity is the modern *Promachocrinus kerguelensi*s, at 0.016. The ones with the lowest disparity are *Comactinia meridionalis* at 0.004 and *Tropiometra carinata* at 0.005, both from the Recent. The CDs of these species with low disparity would be described qualitatively as

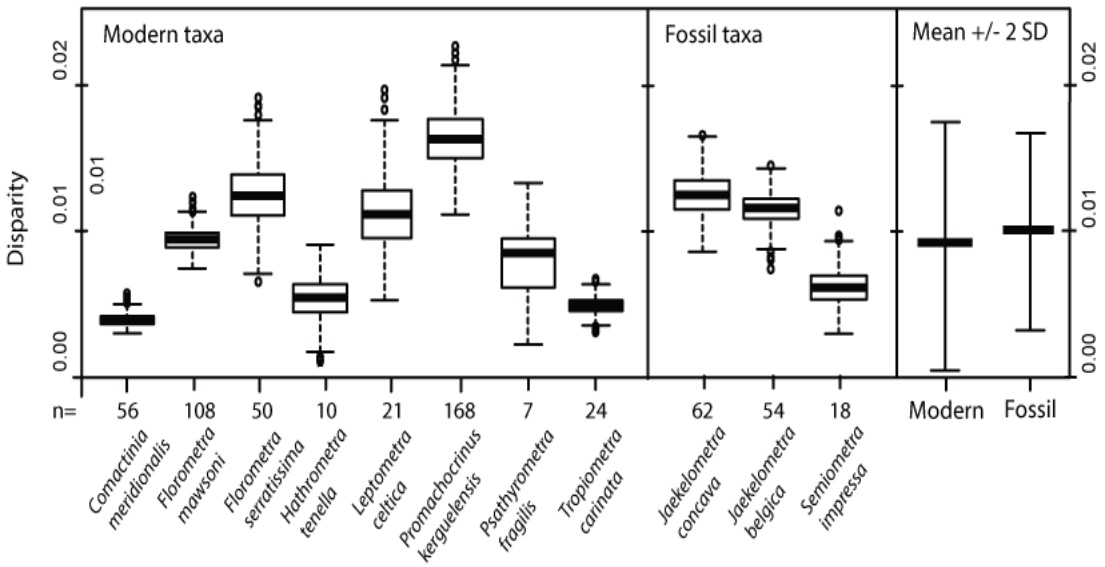


Figure 2.4 Intraspecific disparity for modern and fossil comatulids. For each species, disparity was bootstrapped 1000 times, displayed as Tukey boxplots (Tukey 1977). All fossil taxa fall within the range of disparities represented in the moden taxa. Welch two sample *t*-test is not significant for the test that fossil taxa have larger disparity than modern taxa ($p = 0.32$).

flat pentagonal disks, while the CDs of the species with highest disparity are tall and conical. Additionally, the range of bootstrap results is smallest for these two flat pentagonal CDs compared to the conical forms. The two fossil *Jaekelometra* species have very similar disparity, with broadly overlapping bootstrap distributions. Mean disparity of the fossil taxa, at 0.010, is slightly higher than the mean of the modern taxa at 0.009. Welsh's two sample *t* -test is not significant for the hypothesis that the disparity of the fossil taxa is larger than the disparity of the modern taxa.

BIC results for FMA and mean shape for species is shown in Fig. 2.5. For the majority of modern species, and all fossil species, one group was the preferred solution to the FMA. The two species with lowest sample size, *H. tenella* and *P. fragilis*, with respective $n = 10$ and $n = 7$, had equivocal results with support for multiple groups. For all others, support for the preferred model was very strong, with a difference in BIC > 10.

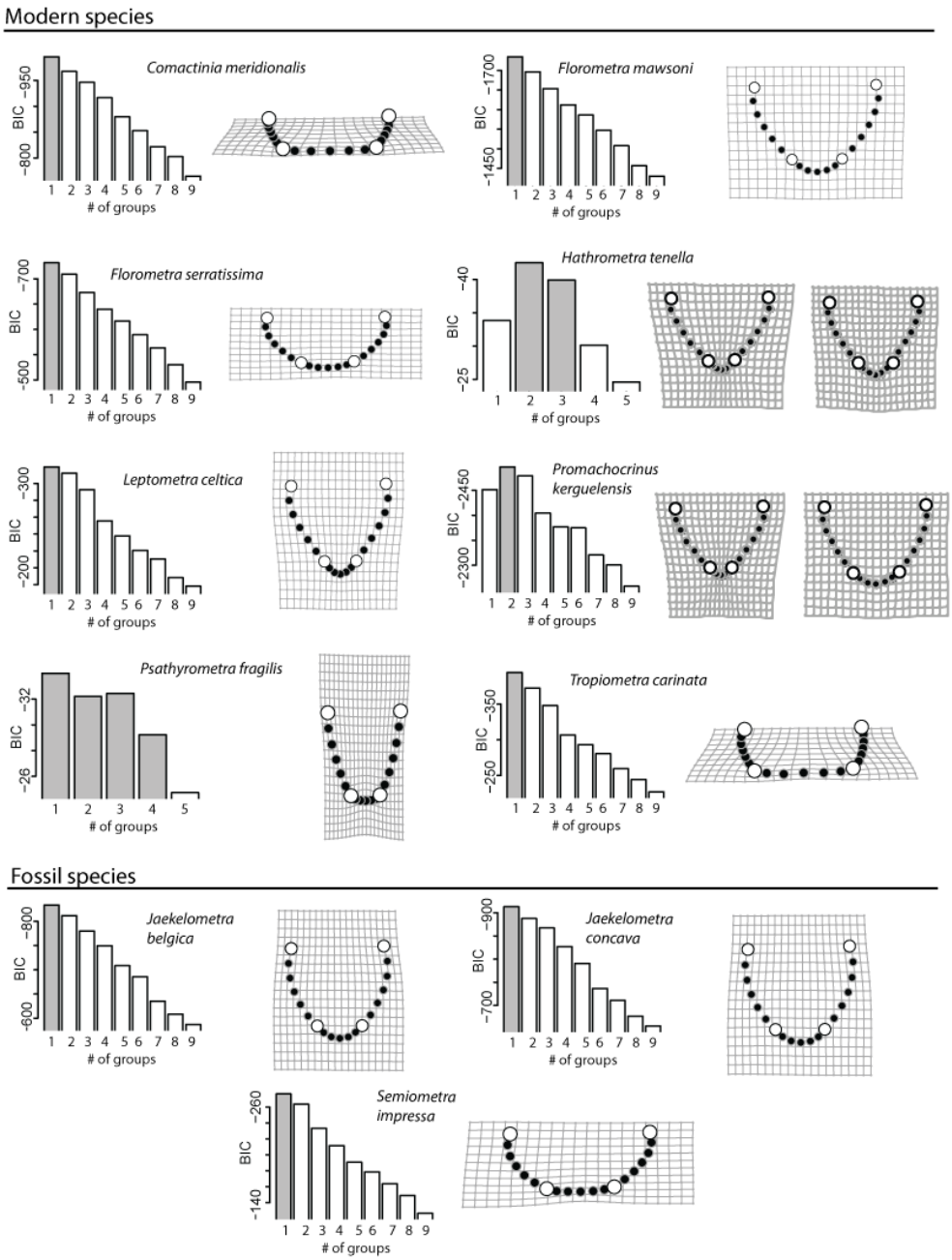


Figure 2.5 Results of finite mixture analysis by species and mean shapes. For each species in this study, the BIC by number of groups is displayed. The best supported model is the one with lowest BIC value. Shaded bars indicate models with probability of support >0.05. High number of groups indicates multi-modality and suggests hidden diversity. Also pictured are thin-plate spline deformation grids of mean shape for each group in the best supported FMA model. One group is preferred in the majority of modern species analyzed, and in all fossil species. Of the three modern species that show support for multiple groups, two have extremely low sample sizes.

P. kerguelensis is unique among species with a large sample size in showing clear support for two groups.

A summary of the results from FMA subsampled to $n = 116$ and with full sample size are presented in Table 2.2. For *Florometra* and *Jaekelometra*, the number of groups found overlaps with the number of described species included in this study. Two groups found within *Florometra* correspond closely to *F. serratissima* and *F. mawsoni*, with 155/158 (98.1%) correctly grouped. In comparison, when two groups are identified within *Jaekelometra*, only 70/116 (60.3%) corresponded to the *a priori* groups. Chi-square tests suggest a strong correspondence between the *Florometra* species and groups found via FMA, and no evidence of such correspondence for *Jaekelometra*. In this case, a random distribution for *Florometra* is rejected at $p = 6.3 \times 10^{-37}$, while it is not rejected for *Jaekelometra* at $p = 0.73$. Surprisingly, *Promachocrinus keurgeulensis*, with but a single described modern species, is classified into a number of groups ranging from 2, using BIC at $n = 116$, to 4, with AICc at $n = 175$. Both *Promachocrinus* and *Florometra* contained more groups than *Jaekelometra* in over 90% of bootstrapped model runs, while *Promachocrinus* had equal or more groups than *Florometra* in 90% of model runs (Table 2.3).

Discussion

Taxonomic considerations — The high disparity and multiple FMA groups within *P. kerguelensis* are of interest, suggesting hidden diversity within this species. Recent molecular phylogenetics (Hemery et al. 2012) shows *P. kerguelensis* comprises ~seven mitochondrially distinct lineages, and at least two lineages based on nuclear genomics. Additional molecular phylogenetics show *Florometra mawsoni* nested within this group (Hemery et al. 2013). These cryptic lineages are candidates for description as species, but attempts to find morphological characters to distinguish them have not succeeded.

Table 2.2 Taxonomic diversity vs. number of groups in preferred finite mixture analysis.

	Jaekelometra	*Promachocrinus*		*Florometra*	
A priori species richness	2				
Sample size	116	116	175	116	158
Mean components BIC	1.08	2.51	2.64	2.13	2.32
Median components BIC	1	2	3	2	2
Mean components AICc	1.93	2.83	3.97	2.55	3.50
Median components AICc	2	3	4	2	3

Table 2.3 Summary of which taxa have more groups based on FMA. Summary of relationships that occur in 90%+ of bootstrap model runs chosen via BIC. Table A.3 in the appendix shows the full data from which these results are calculated. *Promachocrinus* and *Florometra* appear to have higher richness than *Jaekelometra*. This is the opposite of what would be expected based on inclusion of two *Jaekelometra* species and one *Promachocrinus* species, suggesting that taxonomy is biased towards increased diversity in the fossil record.

	Florometra	*Promachocrinus*
Jaekelometra	*J.* < *F.*	*J.* < *P.*
Promachocrinus	*P.* >= *F.*	-

Morphological traits of color, pattern, and number of arm pairs were recorded, but all traits were distributed across the haplotype network with one exception: all individuals identified with six radials belong to a small cluster of related haplotypes within Hemery et al.'s (2012) clade D, even though individuals with other numbers of radials were also present within that clade. Other numbers of radials, from seven to 11, were present across the *Promachocrinus* haplotype network. Specimens with five radials seem to be exclusive to *F. mawsoni*. The lack of other morphological characters beyond CD shape to differentiate these clades is stunning and of great concern to all taxonomists looking to describe species based on morphological features.

However, the *Promachocrinus* species complex also provides an opportunity to test the methods used in this study. The results for *Promachocrinus* are interesting for both methods used here, both having the largest single disparity, and FMA finding more

groups than the traditional taxonomy. Recalculating mean modern disparity +/- 2 SD without *Promachocrinus* results in a range of 0.008 +/- 0.006. All other modern and fossil specimens fall within that range, with *Promachocrinus* having significantly higher disparity (one sided $Z = 2.53$, $p = 0.0057$). Similarly, FMA shows between two and four groups within *Promachocrinus*, which is consistent with hidden diversity within that taxon. If *Promachocrinus* is used as an example of hidden diversity, both methods used here find a significant difference between it and other modern taxa. *Promachocrinus* provides evidence that our methods can detect lumped diversity, a pattern not seen in the fossil taxa studied.

Our methods show no support for the distinction of the two nominal fossil *Jaekeolometra* species. No correspondence was found between CD shape and the species defined by traditional taxonomy. Additional attempts to recover *J. belgica* and *J. concava* as distinct species with our data were unsuccessful; no difference was found in shape via permutation ANOVA (Oksanen et al. 2007), and the optimal solution using FMA on centroid size also recovered only one group. Jagt's (1999) warning that "... *J. belgica* and *J. concava* [may] represent but a single biological species" seems to be correct. An alternative, that they are real biological species, can still be correct if either (1) there are no differences in CD shape (as it was measured here) between the two species, or (2) our tests have insufficient power to differentiate shapes. It is possible that other CD characters beyond the analysis of CD shape measured in only one orientation would reveal multiple species. Other characters used in CD taxonomy include traits related to shape in dorsal/ventral orientation, plus cirral scar number, size and patterning, lumen size and shape, etc. However, there seems to be little that differentiates *J. belgica* and *J. concava* except for size. This example provides evidence that the fossil record may be oversplit relative to the modern, a surprising result. However, with only one example, generalizing seems premature. In order to test if the results were unique to our method or dataset, FMA was implemented for a linear measurement data set previously published

for these *Jaekelometra* species (Jagt 1999). One group was the optimal result achieved for this data set using centrodorsal height, and height + width. Given Jagt's unique warning that these two putative species may represent a single biological species, there is little evidence that this is a general pattern, but care may be warranted.

Explaining the Observed Diversity Pattern — Using both disparity and FMA, fossil taxa appear to be comparable to modern taxa, with no evidence of undersplitting in the fossil record. If taxonomic bias is not the cause of the dramatic increase in diversity from fossil to modern, there are several alternatives to consider including (1) centrodorsals provide no taxonomic information (2) there is a biologically real jump in diversity (3), random sampling error, or (4) non taxonomic biases related to sampling or preservation.

If centrodorsals were useless for taxonomic purposes, we would not expect to be able to differentiate between our modern species using centrodorsals characters. However, that is not the case, with centrodorsals from the various modern species clearly different from each other. Pairwise MANOVA tests for different shape + size revealed that there were significant differences in means for every pairwise combination. Centrodorsals differ between modern species, and therefore do contain taxonomic information.

A real ten-fold increase in diversity over a short geologic time interval would be extraordinary, but is not supported by molecular phylogenetics. A time-calibrated phylogeny of comatulids (Rouse et al. 2013) shows a Triassic origination of the Comatulida and provides evidence for an increased molecular substitution rate in one family, the Comasteridae. This group's radiation began 22 ± 1 Ma, with the ages of the included species averaging more than 9 Ma. Eight genera of Comasteridae were all inferred to have originated at least five million years ago, while only three genera have even been described from the fossil record (Howe 1942, Sieverts 1933, Vadasz 1914).

Given this discrepancy between molecular and fossil data, the absence in the fossil record suggests that biases are driving the pattern.

The possibility that mere chance would produce such a pattern is also not supported. One can compare the range of diversities observed per time bin in the fossil record and calculate the probability that the modern comatulid diversity of ~540 species (Messing 1997) is different than the fossil record. A *Z*-test rejects a null hypothesis that modern diversity is pulled from the same sample as fossil diversity ($Z = 26.7$, $p \approx 0$). It seems that the most likely explanation for the missing diversity in the fossil record is some combination of systematic biases. Additionally, these biases must be ones that affect the comatulids but not other extant crinoids, whose fossil record and modern diversity are commensurate. Broadly, other biases fit into two categories: 1) recognizable diversity exists in the fossil record, but has not entered the literature either due to lack of sampling or lack of description, or 2) diversity is not recorded in the fossil record due to the vagaries of preservation.

There are several lines of evidence that suggest sampling is a significant factor. One likely candidate is related to geographic factors. Some localities, such as the late Cretaceous chalks of Europe, are known for having abundant fossil comatulids, but comatulids are described from few other localities. In fact, 51 of 55 comatulid fossil localities entered into the Paleobiology Database (accessed June 24, 2014) were found in Europe. While the PBDB does not currently include all described fossil localities, the pattern here is typical of the literature. In contrast, modern comatulids have a global distribution, with relatively low diversity (3 genera) in the modern equivalents to the European fossil record such as the Mediterranean Sea. It seems likely that the concentration of material from Europe is a result of the historic concentration of scientific effort there, as well as the ease of collecting comatulid material from the unconsolidated late Cretaceous chalk beds. Also of note is the lack of fossil localities from the western Pacific, which is the area of maximal comatulid diversity today. Sampling bias as a

significant factor in the diversity record is supported by recent taxonomy in other undersampled regions. For instance, nine genera have been described from the fossil record of New Zealand recently (Eagle 2001, 2008). Therefore, undersampling appears to be a contributor to the observed diversity jump for the comatulids.

Another approach to the problem of sampling is to compare alpha diversity of well-sampled, environmentally similar localities. For instance, the generic diversity in soft sediment environments is similar for well-studied localities in both the modern and fossil records. Twelve living genera are reported from Lizard Island, Australia in the Recent (Messing et al. 2006). This is similar to nine genera reported from the Chattian of New Zealand, and nine described genera from the Danian of Denmark. Similar diversity in these environments suggests that comatulid diversity is comparable. This comparability seems to limit the possibility that comatulid diversity never entered the fossil record, at least from certain environments. Nevertheless, the top candidate for high diversity is shallow reef environments where the majority of modern comatulids occur (Messing 1997). Hotspots of comatulid diversity such as the tropical Indo-West Pacific, with ~150 shallow water species, and the north coast of Papua New Guinea, with over 100 species (Messing 1994), exceed the diversity of any fossil time bin. The fossil records of these regions are not well sampled, and it is not clear how much record will be there if sampling is attempted.

Reef environments may be especially problematic for preservation of comatulids because they are high energy environments with unique taphonomic properties. Rapid disarticulation leads to a host of other biostratinomic effects (Meyer and Meyer 1986), characterized by high rates of abrasion (Bromley 1990, p280, Folk and Robles 1964), dispersal of elements, size sorting by hydrodynamic properties, as well as higher rates of cementation than lagoonal environments (Scoffin 1992). Mixing with non-comatulid material means that centrodorsals are “needles in the haystack”. Centrodorsals are but a single one of the ~10,000 major ossicles of a comatulid, and comatulids comprise

somewhere in the range of 0.5-4% of material in reef environments (Meyer and Meyer 1986): between 250,000 and 2,000,000 grains might need to be inspected in order to identify a single centrodorsal. This is an extraordinary amount of effort, and some beneficial, coincidental hydrodynamic sorting or similar process might be necessary for recognition of comatulids from reef environments. These factors suggest that recoverable diversity from reef environments may not be as complete as that from the soft sediment environments. At present, it is not possible to weigh the evidence for the two hypotheses that could explain the missing diversity in the fossil record. Most likely, both biases contribute to that missing diversity. Certainly some diversity is not captured due to the vagaries of sampling, and it seems very likely that some may not be preserved at all, especially in reef environments.

Conclusion

'Must a name mean something?' Alice asked doubtfully.

'Of course it must,' Humpty Dumpty said with a short laugh

-Lewis Carroll, *Alice's Adventures in Wonderland*

Sometimes, heroic efforts to reassemble disarticulated material can be successful, such as Gislén's (1934) many-month effort to reconstruct a comatulid from a statistical analysis of thousands of individual elements. Assuming relatively unbiased preservation and a normal distribution of body sizes, he statistically reconstructed comatulid arm branching patterns by estimating the relative number of arm elements bracketing branches. He then was able to estimate a number of different arm branching patterns, as well as testing articulation patterns. Gislén did indeed put his crinoids back together again.

Gislén's effort, a level worthy of all the king's men, is not a simple solution to the problem of low fidelity of the fossil record to actual diversity; his efforts required many months of work for a relatively limited result. In Lewis Carroll's telling, Alice inquired as to whether a name must mean something – or perhaps in our case, a species. Specifically, the question at hand is whether a species described by a paleontologist means the same type of grouping as one described by a neontologist. With an order of magnitude difference in generic diversity between Recent and fossil comatulids, one should examine whether taxonomic bias may be the principal cause of this difference. Multiple lines of evidence explored here provide no evidence of taxonomic bias as the cause of this difference in diversity.

Modern workers can be convinced to examine CDs more closely, and molecular phylogenetics provides an important tool for understanding diversity in modern comatulids and also for understanding how diversity in the fossil record compares to it. Larger sample sizes and more morphological data could help bridge the diversity gap, but limits will quickly be reached on what can be differentiated based on the available material. Hemery (2012) used over a thousand specimens of *Promachocrinus kerguelensis* in her molecular study, and GM requires well-preserved specimens, but hundreds of CDs are not usually available for the paleo-taxonomist. The number of specimens available therefore might present a serious problem for species' identification in the fossil record.

The order of magnitude difference in diversity between Recent and fossil comatulids appears to be, at least in part, caused by bias. However, the results of this study suggest that the bias is not due either to the loss of information from fragmentation or to differences in taxonomic practice applied to living species and those known only from the fossil record. Therefore, that bias must be accounted for elsewhere; raising the question, is the diversity of the past under sampled or unpreserved? With the diversity hotspot of the Western Pacific clearly undersampled, paleontologists must turn in that

direction with an eye toward description before we can know how much diversity is unpreserved.

Acknowledgements

This work would not have been possibly without mentorship and countless other assistance from T. Baumiller. Gratitude is also due to M. Zelditch for many discussions about methods as well as critical comments on the manuscript. A. Janevski, V. Syverson, C. Messing, L. Hemery, D. Miller and the University of Michigan Paleontology Seminar provided comments, discussion and expertise that improved this manuscript. Access to specimens was assisted by D. Pawson, J. Jagt, J. Bleaker, and M. Eléaume, without which this work would not be possible. Thanks to A. Purens and M. Petrie for continued encouragement. Funding support was provided by National Science Foundation grants (DEB1036393, EAR 0824793), as well as the National Geographic Society (NGS 8505-08).

CHAPTER III

Constraints Control Disparity Of Comatulid Crinoids

Abstract

A morphometric analysis of the comatulids, the most diverse group of extant crinoids, reveals a pattern of rapid increase in morphospace occupation followed by stability to the Recent. Overall disparity within the comatulids was higher in this study than previously reported. Hypotheses to explain this pattern include decreasing rates of within-lineage change or constraints at the edges of comatulid morphospace. Since these hypotheses predict different patterns of disparity within subclades, superfamily disparity was compared to whole sample disparity. The majority of superfamilies show similar disparity to the whole sample, supporting constraints as the primary control on comatulid morphospace dispersion. If constraints are of primary importance in comulatid evolution, common techniques such as phylogenetic reconstruction may have limited usefulness for reconstructing comatulid evolutionary history. Our results suggest that constructing a reliable phylogeny for fossil comatulids may depend on finding characters that show slow rates of change.

Introduction

Paleontologists' attempts to describe and understand the history of life on Earth are heavily influenced by the vagaries of preservation that affect the groups studied. Comatulid crinoids are a group with an incredibly sparse fossil record relative to their high diversity today. Given their generally poor preservation, the question is what information can be extracted from their fossil record. One set of questions that can be

addressed relatively robustly, even with a minimal fossil record, are those related to morphological patterns through time.

Today's comatulids are a diverse and successful group, found throughout the world's oceans, with over 540 species described (Messing 1997). Their fossil record, in comparison, is very poor, with over an order of magnitude less diversity reported in any time bin (Janevski and Baumiller 2010). It is unlikely that a Holocene radiation of such a magnitude is real (Rouse et al. 2013, Summer et al. 2014, Chapter II). In addition to their low fossil diversity, comatulids are characterized by the low quality of fossil material, with articulated specimens virtually absent. As a consequence of the latter, fossil comatulids are described almost entirely on the basis of one element, the centrodorsal (CD). This means that morphological traits for most other body parts are not available, or are incredibly difficult to approximate (Gislén 1934). In order to understand the evolutionary history of comatulids using fossils, one is therefore forced to rely on the CD. Given these constraints, this paper seeks to describe patterns of evolution using character-based disparity, which is tractable with the sample sizes and material available.

A great deal of work has been done to understand patterns of morphological diversity, or disparity, through time (Gould 1991, Foote 1994, Foote 1997b, Fortey et al. 1996, Erwin 2007, Harmon et al. 2010, Hughes et al. 2013). Four general patterns of morphological diversification that we wish to consider here are: (1) an increasing cone of disparity, (2) an early increase in disparity followed by stability, (3) a shift in morphospace occupation over time, and (4) a loss of morphospace occupation through time (Fig. 3.1). These patterns relate to fundamental questions regarding modes of evolution, ecology and the controls on morphology of organisms through time. Important questions include the degree to which different constraints control evolutionary change, rates of evolution, and the appearance of morphologic innovations. A pattern that has been commonly recognized is pattern 2: maximal morphological disparity early in clade's history (Hughes et al. 2013). Hypotheses to explain such a pattern include slowing rates

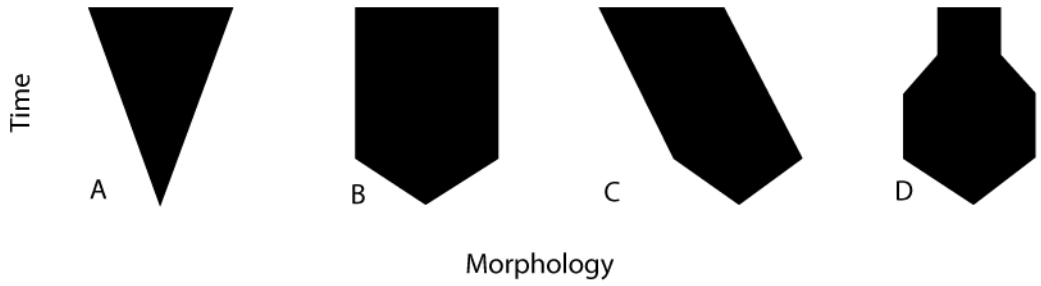


Figure 3.1 Some general patterns of morphospace occupation through time tested for in this study: (A) a gradual increase in morphospace occupied, (B) a rapid initial increase in morphospace occupied followed by stasis, (C) a shift in mean morphology, and (D) a decrease in morphospace occupied.

of morphological evolution and boundaries that constrain expansion into morphospace (Foote 1996, 1999). One method that has been used to differentiate among similar hypotheses involves the relative disparity within subclades (Harmon et al. 2010), as the predictions for disparity within these subgroups vary depending on the evolutionary processes that create the pattern. One problem is that both the data and methods used in these investigations vary widely, and it is not clear if discrete data collected from fossils should generate patterns similar to phylogenetic comparative analysis of continuous traits.

Considering the differences in data and methods, a reasonable first step is to examine the expected pattern under the different proposed models. In order to address this, Fig. 3.2 shows simulations of several common models and data types that have been previously studied. One key finding (Harmon et al. 2010) is that several models of continuous character evolution — early burst and single/multiple stationary peaks — can produce patterns of maximal early disparity. Additionally, two of the continuous models can produce clearly discrete data: multiple peak, and when applied to multivariate data, early burst. The random walk and single stationary peak models can be discretized via assignment to arbitrary levels, but it seems unlikely that this is a general explanation for

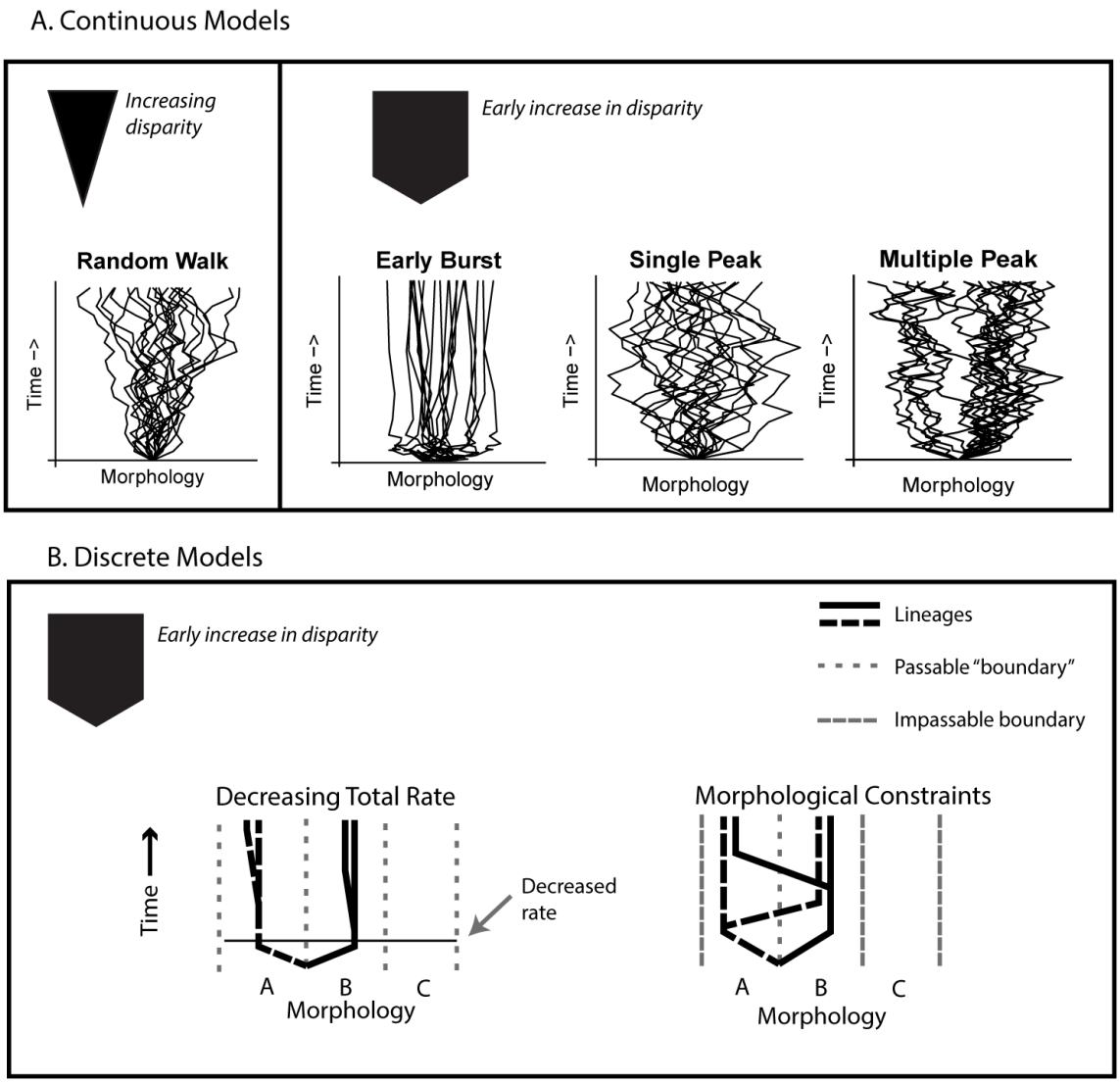


Figure 3.2 Patterns of morphospace change under continuous and discrete models. Models of continuous characters used to investigate disparity through time include the random walk, early burst, single stationary peak, and multiple peak models. A—Early burst models on continuous data have been shown to be rare (Harmon et al. 2010), suggesting that decrease in absolute rates is rare. Morphological constraints, as modeled under the single and multiple peak models, suggest more variation within lineages. B— Similarly, hypothesis to explain the early increase in disparity based on discrete paleontological data also make different predictions about the relative amount of variation within lineages.

discrete data. Therefore, given a discrete character set and the pattern of maximal early disparity, we consider two alternatives: 1) further spread into morphospace is limited because rates of evolution slow dramatically or 2) further spread into morphospace is limited because of boundaries on morphological evolution.

Methods

Data. —Thirty-one centrodorsal characters were assembled via a literature search of the *Treatise on Invertebrate Paleontology* (Moore and Teichert 1978) and several other taxonomic sources (Clark 1915; Rasmussen 1961; Jagt 1999). A sample of taxa from the *Treatise* was selected and characters were scored. These taxa were supplemented by a sample from the taxonomic literature, drawn from a database on comatulid occurrences, to equalize samples to n=10 in each time bin. The final specimen list is reported in Table 3.1. Missing characters were scored as N/A. Discrete, continuous, and meristic characters were used. Discrete characters were coded as {0,1} while meristic and continuous characters were scaled to minimal and maximal values of 0 and 1, respectively. All subsequent analyses were done in R (R Development Core Team 2012).

In order to provide a broad overview through time, the time bins selected were Jurassic, Cretaceous, Cenozoic excluding Recent, and Recent. Although comatulids are believed to have originated in the Triassic, only two genera have been reported from this period; thus, the Triassic was not explicitly included in this study, even though we know that disparity at the group's origin was necessarily low. Using coarse time bins reduced the number of pairwise comparisons that must be made, and also increased sample size per bin, providing higher statistical power than would be achieved with finer temporal resolution. The goal in sampling was to produce a consistent, unbiased, and comparable sample from each time bin.

Disparity Metrics —Many methods have been proposed for quantifying the disparity of discrete characters (Foote 1994, 1999; Ciampaglio et al. 2001). For this

Table 3.1 List of species included in this study. An = Antedonacea, Co= Comasteracea, Ma = Mariametracea, No = Notocrinacea, Pa = Paracomatulacea, So = Solanocrinitacea, and Tr = Tropiometracea. J = Jurassic, K = Cretaceous, Ce = Cenozoic excepting Recent, R = Recent.

	Species	superfamily	time bin
1	*Archaeometra koprivnicensis*	So	J
2	*Burdigalocrinus lorioli*	So	J
3	*Comatulina beaugrandi*	So	J
4	*Palaeocomaster guirandi*	So	J
5	*Paracomatula helvetica*	Pa	J
6	*Pterocoma pennata*	Tr	J
7	*Rhodanometra lorioli*	Tr	J
8	*Solanocrinites costatus*	So	J
9	*Solanocrinites lambertsi*	So	J
10	*Thiollericrinus heberti*	So	J
11	*Coelometra campichei*	So	K
12	*Decameros ricordeanus*	So	K
13	*Glenotremites paradoxus*	No	K
14	*Hertha mystica*	An	K
15	*Jaekelometra meijeri*	Pa	K
16	*Loriolometra retzii*	No	K
17	*Placometra laticirra*	Tr	K
18	*Pseudoantedon icauensis*	So	K
19	*Remesimetra discoidalis*	No	K
20	*Semiometra impressa*	No	K
21	*Amphorometra bruennichi*	Tr	Ce
22	*Bruennichometra granulata*	Tr	Ce
23	*Cypelometra iheringi*	Tr	Ce
24	*Discometra rhodanica*	Ma	Ce
25	*Hertha plana*	An	Ce
26	*Himerometra caldwellensis*	Ma	Ce
27	*Microcrinus conoideus*	An	Ce
28	*Palaeantedon caroliniana*	An	Ce
29	*Palaeantedon soluta*	An	Ce
30	*Stenometra pellati*	Tr	Ce
31	*Atelecrinus balanoides*	Pa	Re
32	*Comactinia echinoptera*	Co	Re
33	*Comatella nigra*	Co	Re
34	*Cyllometra manca*	Ma	Re
35	*Eudiocrinus ornatus*	Ma	Re
36	*Himerometra martensi*	Ma	Re
37	*Perometra diomedeae*	An	Re

38	*Pontiometra andersoni*	Ma	Re
39	*Psathyrometra fragilis*	An	Re
40	*Zygometra comata*	Ma	Re

study, three measures of disparity were used in order to ensure that observed patterns were not incidental to the chosen metric: Manhattan distance, Euclidean distance, and sum of variance. These three should effectively capture the range of evolutionary patterns that are of interest.

Disparity calculated from Manhattan distance is the average number of different character states between specimens in the sample, standardized for missing data. In this study, it is generally equivalent to the distance method of Foote (1999), Gower's similarity coefficient (Gower 1971), or pairwise dissimilarity (Ciampaglio et al. 2001). This method is attractive because it has been shown to have low sensitivity to sample sizes, missing data, and number of characters used (Ciampaglio et al. 2001).

Euclidean disparity is the sum of Euclidean distances between all specimens divided by the number of specimens. This method is more sensitive to detecting evolutionary patterns than the Manhattan disparity, but it is also more sensitive to number of individuals, missing data, and number of characters (Ciampaglio et al. 2001). Given its increased sensitivity, statistical tests using this metric are more likely to detect differences between groups, but at an increased risk of false positives. Pairing this distance metric with that of the Manhattan disparity therefore gives us a range of powers and sensitivity to type I versus type II errors. If the results of both metrics agree, it is strong evidence that the observed patterns are robust.

Sum of variances is calculated by summing the variances of each character over the sample. While the sum of variances is more sensitive to the numbers of specimens and characters included, it responds differently than other measures in detecting an elongation in one direction with a concurrent contraction in another (Ciampaglio et al. 2001). It is therefore included to help detect shifts in morphospace occupation that are not

associated with a shift in morphospace volume or mean shape. If sum of variances increases as distance measures decrease, it is an indicator of a shift in morphospace occupation that can be examined in more detail, and not necessarily a change in morphospace volume.

Principal coordinate ordination (PCO) was performed on the Manhattan distance matrix in order to assist with visualization. PCO, an eigen-analysis based rotation, is preferred to principal components analysis because of its superior treatment of missing data and ability to handle different distance metrics (Lofgren et al. 2003). PCO axes are similar to those of principal components analysis, which are linearly uncorrelated variables that contain descending amounts of the sample's variation. This allows for visualization of complex data sets with a minimum amount of information loss.

Changes in morphospace volume—Increasing morphospace volume through time would indicate the occupation of new areas in morphospace, whereas decreasing volume would indicate occupied morphospace has contracted. For each of the three disparity metrics, difference in total disparity was tested between successive time bins using 1000 bootstrap replicates, or resampled with replacement. A significant difference in disparity between time bins indicates a change in morphospace occupation over time. Superfamilies were also tested to see if their disparities were differed more than expected by sampling from the whole population. At each sample size, 1000 bootstrap replicates estimating disparity of the whole sample were generated and compared to the observed superfamily disparity. If the disparity of superfamilies fell outside the 95% confidence interval (CI), these were deemed to be significantly different from the entire group.

Changes in morphospace occupation—Permutation MANOVA (Oksanen et al. 2007) was used to test for differences in mean centrodorsal morphology between time bins and superfamilies, and for an interaction between superfamilies and time. This method permutes individuals against a distance matrix, testing for larger distances between groups than expected by chance, thus allowing tests using any distance metric

and without requiring dimension reduction. Pairwise MANOVA was then performed on the first three PCO axes for those independent variables that were significantly different, in order to determine which subgroups were driving the overall differences. In order to correct *p* values for multiple hypotheses, the false discovery rate (FDR) method of Benjamini and Hochberg (1995) was used. This method controls for multiple tests based on the proportion of false positives (type I errors), versus the number of rejected null hypotheses, or discoveries. FDR methods are more powerful than family-wise error rate corrections such as the Bonferroni correction, at the expense of a higher possibility of type I errors. This is an appropriate *a posteriori* test to identify pairwise differences after a statistically significant MANOVA.

Results

Results of the PCO by time are displayed in Fig. 3.3. The first component accounts for 36.5% of the variance, and the second component for 24.6%. Associated with high values on the first coordinate are irregular cirral columns, low numbers of cirral socket rows, and lack of radial and interradial ridges (Table 3.2). Highly weighted on the second coordinate is possession of a cavernous oral cavity, shape in lateral view, and several characters associated with specific morphologies on the oral and aboral faces. By eye, there did not appear to be any clear trend in morphospace occupation through time.

After their origin in the Triassic (Rouse et al. 2011), the disparity of comatulid centrodorsals increased rapidly such that by the Jurassic it reached a level that did not change significantly in the three subsequent time bins (Fig. 3.4). Using both Manhattan and Euclidean disparity, a small, non-significant decrease characterizes each successive time bin through to the Recent. Using sum of variance, the pattern was slightly more equivocal, with insignificantly higher disparity in the Cretaceous and the Recent. The agreement of all three methods suggests that the pattern of no net change is robust.

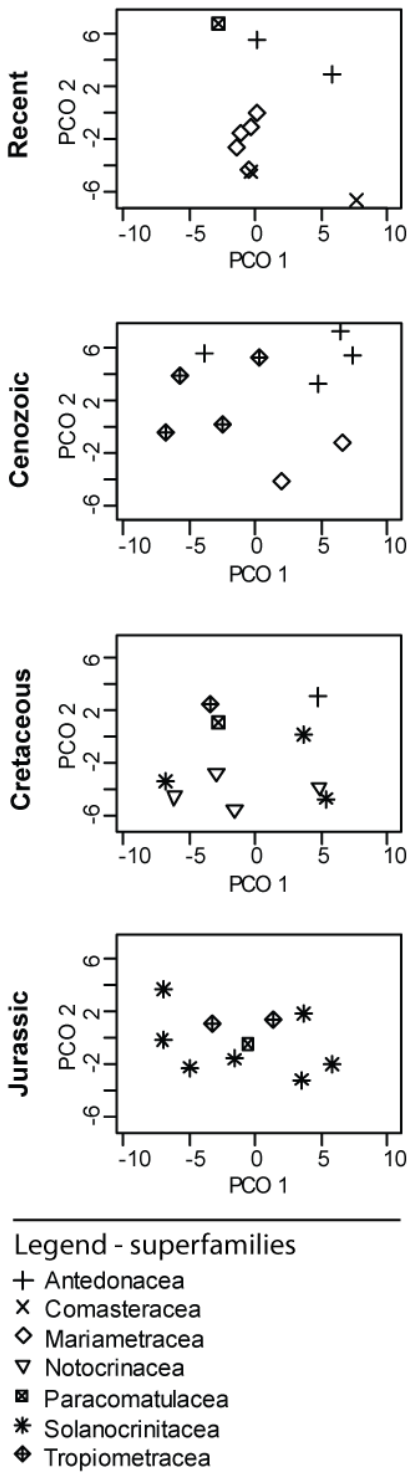


Figure 3.3 Morphospace occupation by time and superfamily using principal coordinate ordination. There were no significant differences in total volume between time bin or in shape between time bin. Key for individuals is in Appendix B.1 (Fig. B.1).

Table 3.2 Characters with correlation coefficient > 0.60 on first two principal coordinate axes.

1st principal coordinate character	character #	correlation
irregular columns	4	-0.83
# socket rows	11	-0.83
smooth oral surface	28	0.78
radial ridges present	9	-0.73
interradial ridges present	8	-0.61
2nd principal coordinate character	**character #**	**correlation**
cavernous oral cavity	22	0.92
radial pits present	25	-0.66
basal rod furrows reach CD edge	29	0.64
lateral shape	1	-0.64
dorsal star presence	20	-0.61
dorsal area size	13	-0.61

Permutational MANOVA of the distance matrix (Table 3.3) shows significantly different mean shapes for superfamilies ($p = 0.001$), no significant differences between time bins ($p = 0.36$) and no significant interaction between time bins and superfamilies ($p = 0.18$). Superfamily explains 31.6% of the observed variance in morphology. Pairwise tests for differences between superfamilies (Table 3.4) show that 10 of 21 pairwise comparisons reveal a difference in mean morphology. The pairwise differences are driven by morphology within the Antedonacea, Mariametracea, and Tropiometracea. All significant pairwise differences involve at least one of these superfamilies. Two superfamilies (Fig. 3.5), the Antedonacea and Mariametracea, have significantly less disparity than the whole. Disparity for the other superfamilies lies within the 95% bootstrap CI of the whole sample, which means they are not significantly different.

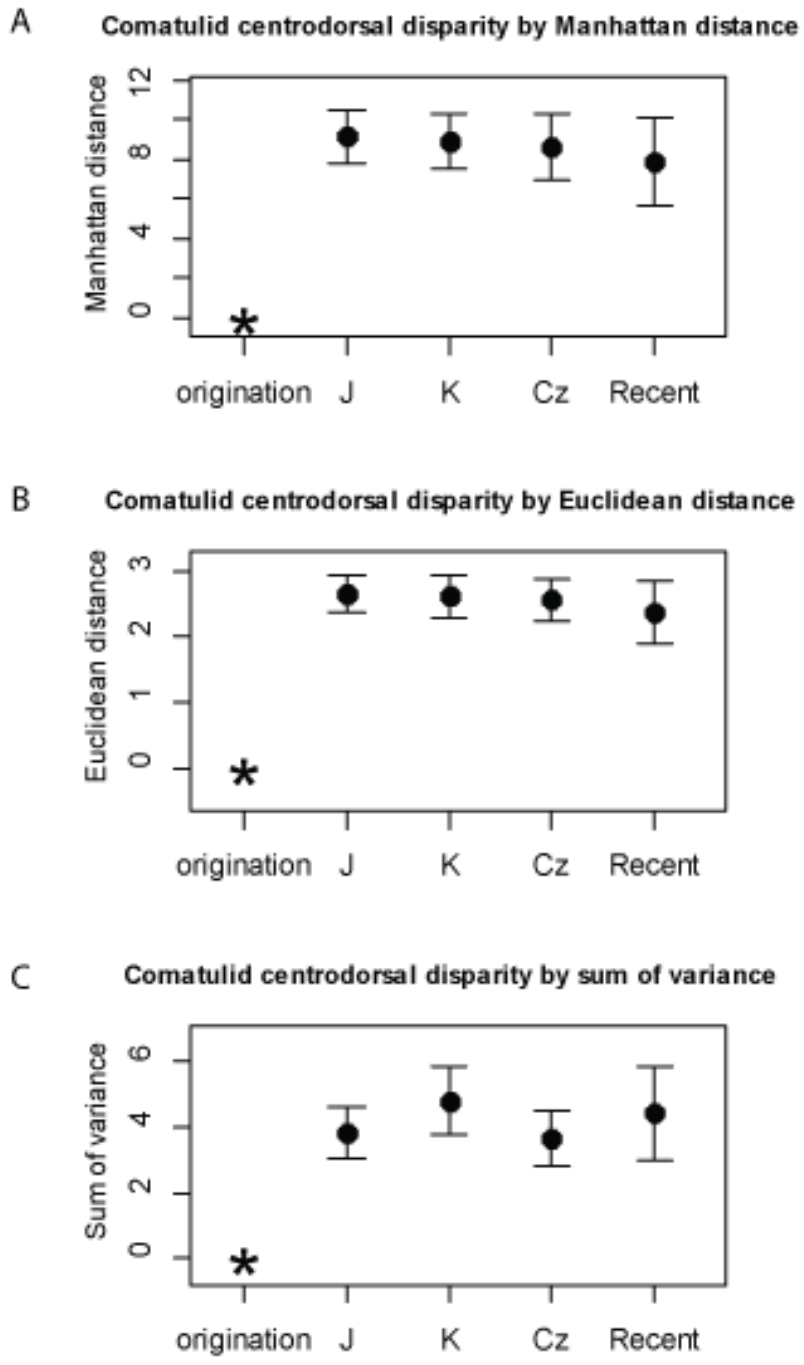


Figure 3.4 Comatulid centrodorsal disparity through time using three different metrics. Error bars are 95% bootstrap confidence intervals. At origination, which occurred at some time during the Triassic, any group will have approximately zero disparity by these methods. No statistically significant changes in volume were detected between time bins after the initial increase.

Table 3.3 Permutation MANOVA of Euclidean distance matrix for effect of time bin and superfamily on mean morphology. While morphologies differ between superfamilies, there is not a change in shape by time. *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.0005$

	Df	Sums of Sqs.	Mean S. Sqs.	*F*	R^2	Pr(>*F*)
Superfamilies	6	53.22	8.87	2.62	0.32	0.001 **
Time Bin	3	10.69	3.56	1.05	0.06	0.374
Sfams. * Time	5	20.00	4.00	1.18	0.12	0.187
Residuals	25	84.67	3.39		0.50	
Total	39	168.59			1.00	

Discussion

Morphological disparity of comatulid crinoid centrodorsals increased soon after the origin of the group in the Triassic, an expansion that must have been very rapid because, according to a recent time-calibrated molecular phylogeny, comatulids originated 208 +/- 40 Ma (Rouse et al. 2011). This apparently high rate of early morphological expansion starkly contrasts subsequent time intervals in which little net change occurs. Our results support a pattern of early maximal disparity, and high superfamily disparity suggests that rates of evolution are high for centrodorsals, even though no further increase in disparity occurs.

The results presented here are consistent with a previous study covering post-Paleozoic crinoids (Foote 1999), even though there were differences in methods and data utilized. Foote's data consisted of all post-Paleozoic crinoids, not just the comatulids. In this study, analysis was undertaken at the superfamily level, differing from Foote's work. Moreover, here only a single element, the centrodorsal, was examined, whereas Foote looked at whole-organism traits; thus the two data sets are largely non-overlapping. Of Foote's 90 total characters, only 22 showed variation within the comatulids. Of those 22 traits, only 8 overlap with the 31 included in this study. Several of the 8 overlapping characters are further modified in this study. For example, we describe CD shape with 3

Table 3.4 Pairwise mean Manhattan distances (diagonal and above) between superfamilies and *p*-values to test different mean morphologies (below diagonal). Manhattan distances are the mean number of differences between taxa, of 31 total characters. Pairwise comparisons using MANOVA on the first three PCO axes reveal many significant differences in morphology corrected for multiple tests (Benjamini & Hochberg, 1995), even with the small sample sizes available here. *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.0005$.

		Antedonacea	Comasteracea	Mariametracea	Notocrinacea	Paracomatulacea	Solanocrinitacea	Tropiometracea
	N=	7	2	7	4	3	10	7
Antedonacea	7	6.3	11.2	10.4	10.3	7.2	9.2	7.0
Comasteracea	2	0.002 **	2.8	6.1	10.6	12.5	11.3	11.4
Mariametracea	7	4e-5 ***	0.46	4.8	11.9	10.4	10.2	11.0
Notocrinacea	4	0.002 **	0.65	0.23	5.0	8.7	6.9	9.2
Paracomatulacea	3	0.30	0.52	0.04 *	0.30	7.5	8.6	6.8
Solanocrinitacea	10	0.01 *	0.16	0.002 **	0.39	0.48	9.4	8.1
Tropiometracea	7	0.03 *	0.03 *	0.002 **	0.03 *	0.91	0.13	6.6

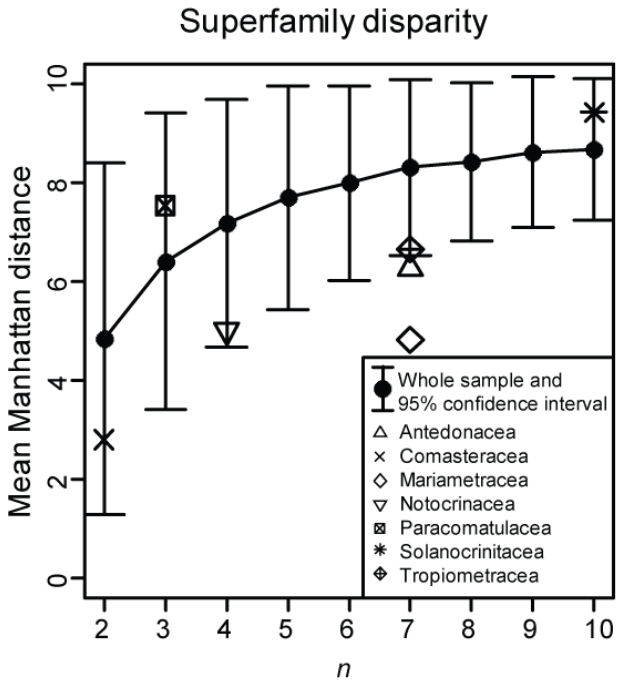


Figure 3.5 Superfamily disparity vs. whole sample disparity. The Antedonacea and Mariametracea show significantly less disparity than the 95% bootstrap CI with 1000 replicates, generated by sampling all specimens with replacement. Results are similar using Euclidean distance and sums of variances.

continuous characters, whereas Foote treated it with 1 ordered and 2 unordered characters. Another important difference is that this study includes Recent taxa, while Foote's data extended only through the Eocene. Despite these differences, the overall pattern observed here is qualitatively similar to that described by Foote, with an early expansion into morphospace followed by stability.

Another difference between the results of this and Foote's (1999) study is the magnitude of comatulid disparity. Measuring disparity in a comparable way, as the average number of differences between taxa divided by the number of characters, the total disparity for comatulids in this study is 0.311 +/- 0.021 (95% CI) whereas using Foote's data it is 0.086+/- 0.009 (95% CI), a significantly smaller value. This is largely due to the fact that Foote's data include many characters that do not vary within the

comatulids, biasing that disparity downwards. Restricting Foote's data to the 22 characters that vary within the comatulids increases the disparity calculated from his data set to 0.275 +/- 0.028 (95% CI), a value that is not significantly different from the one calculated from the data set used here (two-sided bootstrap test of means, $p = 0.105$).

Clearly, the choice of characters can dramatically affect results. This highlights at least two problems with discrete character disparity methods. First, as seen in Foote's (1999) study, differences in absolute disparity between groups may reflect biases related to recognition of morphological characters between those groups instead of biological processes. Second, characters that are either inapplicable or missing lead to some level of incommensurability. This problem can only be avoided for pairwise comparisons with exact matches of applicable and non-missing characters. Foote (1999) demonstrated that random weighting of characters does not bias the results, but the two problems mentioned above can be non-random, for example by taxon, preservation state, or effort in identifying characters to include in analysis. Comparison of absolute disparities from differing datasets should therefore only be undertaken with extreme caution. Avoiding inapplicable characters is the only clearly guaranteed practice for avoiding such biases. Foote (1999) found that Paleozoic crinoids display a wider range of morphological designs than post-Paleozoic crinoids, and there certainly are morphological traits that were present during the Paleozoic that are not present in the post-Paleozoic. However, the converse may also be true: many CD and cirral morphologies are not known from the Paleozoic. It is not clear to what degree the disarticulated state of most fossil comatulids biases our ability to recognize morphological disparity, but there are reasons to expect that it is significant.

The poor fossil record of comatulids, by far the most diverse crinoids today, means that much of their fossil morphology is unknown to science (Donovan 1991, Baumiller 2003, Gorzelak and Salamon 2013, Chapter II). Some morphological characters unavailable from the fossil record are reported in the Recent, such as unequal

arm length in *Comatula rotalaria* (Messing et al. 2006) and possession of as many as ten radials in *Promachocrinus kerguelensis* (Carpenter 1879, Hemery et al. 2012). Neither of these morphologies was recorded as varying by Foote (1999) within the comatulids. If such characters were found in fossil comatulids, they would contribute to post-Paleozoic crinoid disparity. With only a handful of articulated fossil comatulids known, and the vast majority of descriptions consisting only of disarticulated elements, it is likely that many comatulid morphological traits that once existed remain unknown. For comatulids, making claims about whole body morphological disparity when only a handful of preserved specimens display those morphologies is likely to be a high-error enterprise.

Explanations for the observed pattern—An early spread into morphospace followed by a gradual increase in diversity has been a common pattern described for fossils. Foote (1994, 1996, 1999) has repeatedly shown such a pattern in various groups of crinoids, in both the Paleozoic and post-Paleozoic. Other groups that show a similar pattern include dinosaurs (Brusatte et al. 2008), arthropods (Briggs 1992; Wills et al. 1994; Lofgren et al. 2003), crustaceans (Wills 1998), angiosperm pollen (Lupia 1999), priapulids (Wills 2010), cetaceans (Slater et al. 2010) and ecological "carnivores" (Wesley-Hunt 2005). Three models of morphological change that have been quantitatively tested are Brownian motion (BM), single stationary peak (SSP), and early burst (EB) (Harmon et al. 2008). The SSP and EB models make differing predictions for patterns of disparity between subclades and the overall group. Subclades exhibit low variation versus the whole in the EB model, but more variation under the SSP model, when contrasted to the BM model. Both SSP and EB are consistent with the evolutionary pattern observed in the fossil record of comatulids, with disparity peaking early and staying at a high level, but they partition disparity differently among subclades. SSP is consistent with Foote's hypothesis of morphological boundaries, or constraints around an adaptive peak, while the explanations of lower speciation rates or smaller morphological changes per speciation event produce patterns more consistent with the EB model. One

consideration is that Harmon et al 2010 only tested a few simple models; more complex models such as multiple stationary peaks (MSP) may be more realistic. However, MSP models have more parameters and thus increase model complexity, pushing the boundaries of what is detectable using the comparative method. Other models, such as single moving peak and multiple moving peaks might better match expectations of evolution near adaptive peaks. However, as the increasing number of parameters necessary to fit these models may limit our ability to detect them, the simple SSP model may end up preferred. An evolutionary process where taxa follow multiple moving adaptive peaks traveling across morphospace, bounded at some level, could easily appear to produce a pattern consistent with the SSP model when a subset of data is examined. However, these various models partition disparity differently between subclades, it seems reasonable to differentiate among them by comparing disparity within subclades, in the case of our data superfamilies, to overall disparity.

The Antedonacea and Mariametracea show less disparity than the whole sample, while the Comasteracea, Notocrinacea, Paracomatulacea, Solanocrinitacea, and Tropiometracea show disparity similar to that of the whole sample; this result is more consistent with morphological boundaries (SSP) than decreasing rates of change (EB). Molecular studies incorporate estimates of clade age, which is not included in this study, and which could potentially explain lower within-group disparity for several superfamilies. Rouse et al. (2013), for instance, report an origination time for crown Mariametracea of 53 +/- 3 Ma, which might explain the relatively lower disparity observed in that group. Analysis of comatulid superfamilies is also problematic because we lack a robust understanding of their phylogenetic relationships. In fact, it is not even clear whether they represent clades, as molecular work by Rouse et al. (2013) shows mixed results for the robustness of the traditional taxa. That report shows Mariametracea and Comasteracea as monophyletic, and other groups such as the Antedonacea and Tropiometracea as poly- or paraphyletic. If these taxonomic groups are defined on

phenetic similarity, then that could create patterns of disparity that are not related to any evolutionary processes. A high-quality phylogeny of fossil comatulids would greatly benefit efforts to understand these patterns.

Unfortunately, phylogenetic analyses based on the CD shape, the most readily available element of fossil comatulids, may not be very effective if CD evolutionary rates are high. A study of intraspecific disparity (Chapter 2) shows that while CD shape can be used to differentiate sister species, it is less useful for higher taxa due to a high rate of shape evolution. Such a high rate of CD evolution within comatulids would have serious implications for what can be gleaned of their evolutionary history, given the rarity of articulated whole specimens. And even if the latter were available, recent work by Summers et al. (2014) suggests that whole-body morphological traits in the Comatulida are also highly labile. Thus reconstructing a reliable phylogeny of fossil comatulids will require traits that evolve at both high and low rates, just as molecular phylogenies are based on genes that evolve at both high and low rates.

Conclusions

1. The pattern of morphospace occupation for the comatulids is one of early expansion followed by constant disparity. No statistically significant changes in volume or mean shape were detected between the time bins in this study. This is consistent with other investigations of comatulid morphospace occupation (Foote 1999).

2. The overall amount of disparity within comatulid CDs is higher than expected from previous work (Foote 1999),: post-Paleozoic crinoids are as morphologically diverse as those in the Paleozoic.

3. There are significant differences in mean shape between many comatulid superfamilies, and the Antedonacea and Mariametracea show less disparity than expected by chance. Lower disparity within these groups compared to the whole suggests a decreased rate of evolution, while the ones with similar disparity to the whole group

support constraints as the cause of the disparity pattern. Differentiating between these two hypotheses is therefore equivocal, but the balance of evidence (five versus two) favors constraints as cause.

4. The observed pattern of low disparity within the Antedonacea and Mariametracea could also be caused by 1) the lack of time calibration, since a young group has less time to diverge into morphospace, or 2) superfamilies based on phenetic similarities rather than phylogeny. The solution to these problems would be to analyze disparity over a reliable time-calibrated phylogeny. Such a phylogeny does not exist for fossil comatulids. Analysis of modern specimens might therefore be the best option.

5. If the SSP model applies to comatulid CDs, constructing a reliable morphological phylogeny may be problematic, especially over the entire Post-Paleozoic. One recent paper by Summers et al. (2014) found whole body characters for the Comatulidae to be highly labile and to demonstrate significant homoplasy. If the same is true of CD characters, reconstructing relationships of fossil comatulids may be exceptionally difficult.

Acknowledgements

Many thanks are due to V. Syverson, M. Zelditch, and the University of Michigan Paleontology Seminar. Access to specimens was assisted by D. Pawson, J. Jagt, J. Bleaker, and M. Eléaume. Funding support was provided by National Science Foundation grants (DEB1036393, EAR 0824793), National Geographic Society (NGS 8505-08), and a University of Michigan Department of Earth and Environmental Sciences Turner student award.

CHAPTER IV

Failure To Account For Regional And Temporal Differences In Detection Underestimates Diversity And Overestimates Certainty For Comatulid Crinoids

Abstract

The most abundant modern crinoids, the comatulids, have much smaller apparent diversity in the fossil record than other extant crinoids. Hypotheses for this inconsistency include a late Cenozoic radiation of the comatulids, or low rates of detection for fossil comatulids. One estimate (Foote 1999) of crinoid detection rates over the Paleozoic and Mesozoic suggests a probability of about 0.4 per genus per ~5 MY, which seems inconsistent with the low observed fossil diversity of comatulids in the Cenozoic. That study assumed homogeneity in detection rate between sub-taxa as well as in time and space, which may not be true for comatulids. In order to differentiate between our hypotheses, we employ a capture-mark-recapture (CMR) method. The CMR method allows for simultaneous estimation of both genus duration and detection rate, and allows for complex models that vary these by time, space, taxon, and more. The CMR model used here was able to classify uncertainty, while we show other common methods were biased in unpredictable ways. We find over an order of magnitude variability in detection rate for comatulids through time and space, and support for a significant increase in comatulid diversity in the Cenozoic.

Introduction

One major focus of paleontologists over the past three decades has been the patterns of biological diversity through time (e.g., Sepkoski et al. 1981; Valentine 1985; Alroy et al. 2001; Alroy 2008). Efforts to characterize temporal patterns of diversity have been intertwined with analyses of the fossil record's quality, to reduce biases in observed diversity patterns. Of particular interest are several studies that show high completeness and consistency for various fossil taxa (Foote and Sepkoski 1999; Valentine et al. 2006). In this paper, we investigate the patterns of completeness and coverage for one group that is often considered to have a poor fossil record, the comatulid crinoids, to understand more clearly how this group's diversity has changed over time.

Methods used to measure the quality of the fossil record include those based on taxon occurrences (Foote and Raup 1996; Solow and Smith 1996; Foote and Sepkoski 1999; Alroy 2010; Liow and Nichols 2010), inferences derived from phylogenies (Benton et al. 2000), and the proportions of extant organisms described in the fossil record (Foote and Sepkoski 1999; Valentine et al. 2006). One common goal of these studies is to measure completeness, a term that generally includes both the proportions of once-living organisms that have entered the fossil record and how that proportion changes over stratigraphic intervals. Two parameters are important for a taxon becoming known to science: the duration of time that taxon was extant, and the probability of detecting it in any given interval. For this paper, we focus on the use of capture-mark-recapture models (CMR) (Connolly and Miller 2001a; Liow 2010). We also compare the results of the CMR method to results obtained using other common metrics for the comatulid crinoids, in order to provide a better understanding of how these methods relate. This comparison seems especially important for a group such as the comatulids, which have been described as having a depauperate fossil record.

Comatulid crinoids are a successful group of post-Paleozoic echinoderms, with a global marine distribution and ~125 described modern genera. This group has been said to have a poor fossil record (Meyer and Meyer 1986; Donovan 1991; Baumiller and Gazdzicki 1996), contrasting with the record of Paleozoic non-comatulid crinoids, which shows them to be some of the most abundant organisms during the Carboniferous (Ausich 1997). Various explanations have been proposed to explain comatulids' poor representation in the fossil record, including taxonomic lumping, identification problems caused by disarticulation, the poor preservation potential of reef environments (Meyer and Meyer 1986), as well as simply neglect by paleontologists (Howe 1942; Oyen and Portell 2001). In contrast, several analyses (Foote and Sepkoski 1999) using differing techniques suggest that the preservation potential for crinoids is similar to that of many other fossil taxa. A literal reading of the record of fossil crinoids (Fig. 4.1) shows a rise to high diversity in the mid- to late Paleozoic, when they were a dominant part of the Paleozoic ecosystem. This is followed by a sharp decline leading to near disappearance at the Permian-Triassic, and a subsequent rebound to ~20% of Paleozoic diversity by the Late Triassic, which was maintained through the rest of the Mesozoic and Cenozoic. Although crinoids rebounded to moderate diversities following the P-Tr, they never achieved anywhere near their Paleozoic importance in post-Paleozoic ecosystems.

The Paradox of Comatulid Diversity

While crinoids as a whole are a relatively minor component of the modern fauna, the comatulid crinoids can be locally abundant and relatively diverse, with ~4 times (WoRMS Editorial Board 2014) as much generic diversity as all other extant crinoid groups. Comatulids are distinct from other crinoids in that they are stalkless, and thus possess unsurpassed mobility (Janevski and Baumiller 2010). They are considered a part of the modern evolutionary fauna (Sepkoski 1981). This increased mobility in comatulids is hypothesized to be a response to increased predation pressure (Meyer and Macurda

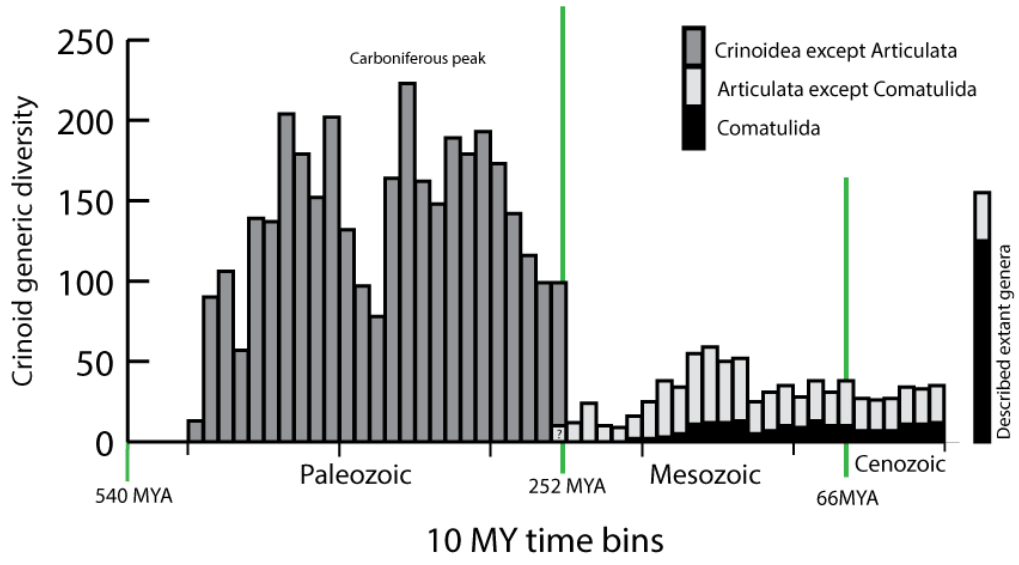


Figure 4.1 Crinoid diversity through time. Range-through crinoid diversity compiled from several sources (Webster 2003; Moore and Teichart 1978; Hess et al. 2011). Paleozoic crinoid diversity is far higher than that recorded in the fossil record of the post-Paleozoic, but modern crinoid diversity (WoRMS Editorial Board 2014) is similar to mean crinoid diversity in the Paleozoic. This pattern is being driven by the high diversity of comatulids today, a group with a depauperate fossil record.

1977; Baumiller et al. 2010). However, the fossil record shows less than half as many comatulids as non-comatulids through the post-Paleozoic (Fig. 4.1). Are crinoids in the post-Paleozoic one-quarter as diverse as they were in the Paleozoic, as read from described fossil diversity? Or are crinoids in the post-Paleozoic closer to two-thirds as diverse, comparing modern crinoid diversity to Paleozoic diversity?

This disjunction between the fossil diversity of comatulids compared to non-comatulids, and the present diversity of comatulids, could be explained in a number of ways, two of which we examine here. First, the comatulids may have undergone a recent rapid diversification, and we would therefore not expect to see high diversity through much of their fossil record. Second, a bias in the detection rate of comatulids could lead

to underestimating their historic diversity relative to non-comatulids. With the modern crinoid diversity of ~160 genera nearing the peak Paleozoic diversity of around ~220 crinoid genera, understanding the patterns of preservation for the comatulids will inform our understanding of the Paleozoic fossil record as well. It is an open question as to which biases may affect the observed diversity pattern of Paleozoic crinoids.

The capture-mark-recapture method has been applied to paleontological data for other groups (Nichols and Pollock 1983; Connolly and Miller 2001b; Liow 2013; Liow and Finarelli 2014) to understand fossil record quality, as well as patterns of diversity, extinction and origination. The CMR method uses records of occurrences and absences to estimate parameters for survival per unit time, ϕ, and overall detection probability, p, for a given taxon. An example is provided in Fig. 4.2). Because CMR has historically been used to analyze ecological data, the terminology used for CMR methods can differ from what is common in paleontology. The variables and terms used in this paper are described in Table 4.1. In addition to detection probability and extinction rates, we provide calculations for genus duration, estimates of true diversity, as well as the number of originations expected over the comatulids' history. CMR allows for the analysis and comparison of sophisticated models for these parameters, and is able to accommodate changes through time, space, cohort, and taxon. In this study, we compare simple models that lump together data with a model that accounts for differences in preservation and detection through time and space.

Methods

A comatulid fossil occurrence database was generated by searching the primary literature for any mention of comatulids. In addition to taxonomic information, the database includes stratigraphic information, locality, and other additional information of import. Taxonomy was updated based on the most recent available information (WoRMS

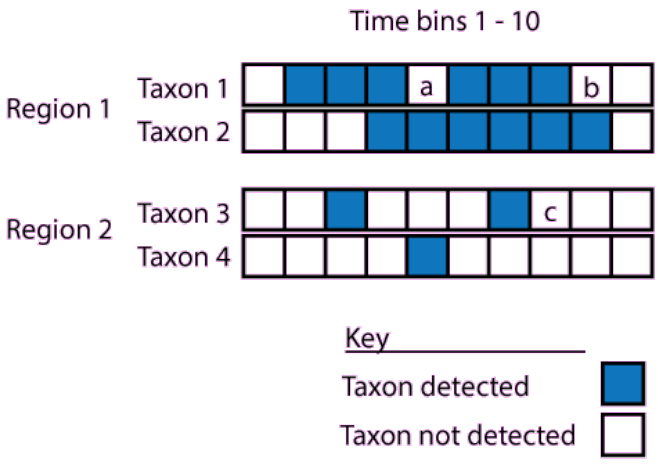


Figure 4.2 Demonstration of CMR method. This example shows two taxa in two regions. We can conclude that taxon 1 was extant during the bin marked 'a' since it was extant both before and after. The number of bins between first and last occurrences allows estimation of detection rate, *p*. Two explanations for the taxon's absence in stage 'b' are that it either was not detected, or went extinct at the end of the previous stage. The probability of survival per time unit time, φ, is the inverse of the extinction rate per unit time, and can be estimated from the time between the taxon's first and last appearances. In our example, the taxa in region 2 have shorter apparent duration, but also appear to have lower *p*. Thus, taxon 3's absence from stage 'c' is more likely to be due to lack of detection, and not extinction, compared to taxon 1 in stage 'b'. The CMR method estimates *p* and φ using all available data via a maximum likelihood equation, as well as providing error bars around those parameters.

Table 4.1 Glossary for CMR method

Variable with equation	Explanation
s_t	Number of taxa observed in a stage.
$\bar{s}$	Mean number of taxa found over a number of stages.
P	Detection probability, genus/stage.
ϕ	Genus survival rate per million years.
$\varepsilon = 1 - \phi$	Average extinction rate genus/million years.
$\text{genus duration} = \frac{}{\log_{\varphi} 0.5}$	Median genus duration in million years. Median is used because under this model, genus duration is –right skewed.
$\hat{S} = \bar{s} / p$	Estimated true diversity in a time bin.
$\hat{B} = S_{t+1} - S_t(1 - \varepsilon)$	Number of taxa originating per time bin.
$\sum_{t} \hat{B}$	Total number of taxa that originated through time.

Editorial Board 2014). Only fossil occurrences were used in this analysis. A total of 247 occurrences are in our database, of which 187 were resolved to a stage.

In addition to the CMR results, we also include four other analyses for comparative purposes. First, we calculated the proportion of modern diversity that is known from the fossil record– a simple, widely-reported measure of observed fossil record quality (Foote and Raup 1996; Foote and Sepkoski 1999; Valentine et al. 2006). In addition to worldwide diversity, we also tested several taxon lists against described diversity in the fossil record in order to determine if there is a general trend of preservation by region. Second, we included data from another independent analysis of the minimum diversity through time on a time-calibrated molecular phylogeny (Rouse et al. 2012). This is the sum of genera that are inferred to be alive at any time, based on the branch lengths of the taxa contained therein, presented as a minimum estimate for diversity through time. The third analysis is the detection rate analysis of Alroy (2008). This metric is calculated based on the number of taxa that are "part-timers" and "three-timers". Three-timers are those taxa that are found in three successive time bins, whereas part-timers are found in the first and last time bin, but not in the middle. This metric, calculated as part-timers divided by part-timers + three timers, has a number of very attractive properties (see Alroy, 2008), but can perform poorly with a sparse record. For instance, Alroy's method avoids biases caused by exceptional preservation in a single time bin, but it also means that exceptional preservation is not used as information to inform the model. Lastly, we calculated the *FreqRat* statistic described by Foote and Raup (1996). This method is a calculation based on the frequency of taxa that extend through one *f*(1), two *f*(2), or three *f*(3) intervals, calculated as $f(2)^{\wedge 2}/(f(1)*f(3))$. It is shown to be resistant to violation of assumptions and quite resistant to variations in underlying distributions. For cases of variation in preservation probability, as appears to be the case for comatulids, it accurately reflects the "effective" preservation probability.

With *FreqRat* calculated, we then calculate an accurate completeness for the whole group (Foote 1997 Appendix 1).

Using the CMR method, only taxa resolved to the stage level were included. The only exception to this stage-level requirement was for the Pliocene epoch, where we combined the two recognized stages to provide some coverage for the Pliocene, since those occurrences were not resolved to the stage level. However, excluding occurrences during the Pliocene had very little effect on our results. A significant factor in the data is the regionality of the described fossil comatulids; most fossil comatulid occurrences are European (163 of the 187 occurrences with stage level resolution, or 87%, were of European origin). Therefore, a regional factor was included in the analysis in order to investigate the differences between European and non-European detection rates.

Analysis was conducted using the RMark (Laake 2013) interface to the MARK program (White and Burnham 1999). All analyses were performed in the program R for statistical computing (R Development Core Team 2012). Potential parameters explored included shifts in φ and p over time, directional change through time, as well as differences in these parameters according to region. Occurrences were grouped temporally at the stage level, with one exception as described above. Potential parameters were initially selected using an iterative method, and models were assessed using Akaike's information criterion corrected for sample size, commonly referred to as AICc (Akaike 1998, Hurvich and Tsai 1989). AICc is a measure of the relative quality of a statistical model given a set of data that takes the number of parameters into account, as each parameter added to a model will increase model fit while increasing model complexity. AICc includes a penalty based on the number of parameters added to the model to prevent over fitting; each added parameter adds $2k(k+1)/(n-k-1)$ (where n is sample size and k is number of parameters) to the AIC score, in addition to the penalty included in AIC of 2 for each parameters. The difference in AICc (ΔAICc) yields scores

for all models compared to the one with the lowest AICc. A model with ΔAICc of 2 is substantially worse, and one with ΔAICc of 8 has less than 1% relative support.

Stages differ in duration, which may generate non-biological biases in our data. We considered several treatments to address this issue. For φ, we compared two scenarios. The first had probability of survival between each stage modeled as equal, ignoring the differences in duration. The second modeled φ based on the duration between stage midpoints in millions of years. Models fit with the extinction rate per million years were nearly identical, with all parameter estimates broadly overlapping with those estimated with extinction rate per stage; for example, one typical example resulted in ΔAICc 0.18, which translates to a relative weight of 91%. In other words, the results were almost completely independent of which way time was divided for φ. Therefore, we used the per million years treatment, as it eased calculation of derived parameters. For *p*, the concern was that longer stages would have higher detection probabilities. We compared models with a search intensity factor of stage duration in million years, to one that did not treat stage durations for *p*. Comparison of one strongly performing, typical model resulted in more support for not treating stage durations, with ΔAICc 1.78. Therefore, no correction for unequal stage duration was employed for *p,* while extinction probability was modeled based on millions of years between stage midpoints.

The iterative process was implemented as follows: First, a basic model was run (Table 4.2), followed by a set of models that were identical except for the addition of one candidate parameter. The parameter with the highest explanatory power, as measured by AICc, was retained for the next iteration. In the final iteration, two parameters proved almost equally useful, and therefore both were retained. The iterative process ended when further parameters showed evidence of over fitting as measured by AICc, or provided different detection probability for single stages. Four time bins, and two regions, were shown to be useful for modeling *p,* while two time bins and two regions were selected for ϕ (Table 4.3). In order to simplify communication, when referring to results from those

Table 4.2 Model results when not accounting for differences in time or region. p = detection probability/stage, ε = extinction rate/million years, $\dot{S}$ = mean diversity over time.

	Parameter estimate	LCI	UCI
p	0.26	0.19	0.34
ε	0.03	0.04	0.02
Genus duration	26	17	39
$\dot{S}$	9.3	12.5	7.1

Table 4.3 Time bins used for CMR model.

Range (MYA)	First stage	Last Stage	Internal reference
227-141.2	Norian	Berriasian	Triassic/Jurassic
141.2-85.8	Valanginian	Coniacian	Early/Mid Cretaceous
86.3-61.6	Santonian	Danian	Late Cretaceous
61.6-2.6	Selandian	Pliocene	Cenozoic

time bins we use the terms Triassic/Jurassic, Early/Mid Cretaceous, Late Cretaceous, and Cenozoic, even though the optimally modeled time bins did not align perfectly with the geologic periods.

Following the procedures outlined above, an automated model selection procedure was used, generating models using all available combinations of candidate parameters as well as their interaction terms. Model averaging based on the AICc was used to generate a single estimate for each parameter (Burnham and Anderson 2002, Lukacs et al. 2009). Parameter values were averaged and weighted by relative AICc support. This process reduces bias and increases precision of parameter estimates (Burnham and Anderson 2002). The estimates of real diversity and generic longevity were then calculated based on these averaged parameter values. The models were generated using the MuMin package in R (Barton 2011).

Although the four assumptions underlying the CMR method, as listed in Liow and Nichols (2010, p.87) may appear to cause inherent bias in these sampling parameters, we

felt that the nature of the question this study is testing adequately addressed the shortcomings in this model: 1) The first assumption is that there are equal encounter probabilities for taxa after the initial encounter. This is certainly biased; more common taxa are more likely to be encountered to begin with, and in subsequent bins. Since our primary question is the differing encounter probabilities, we are explicitly testing for this effect through time and space, and the uncertainty added by differences within taxa in one environment should be handled adequately with confidence intervals around parameter estimates. The second assumption, 2) of equal extinction probabilities for all taxa encountered, is similarly treated by the model averaging and confidence intervals. It is not clear how violating these first two assumptions would generate patterns that might be misleading. 3) Sampling intervals are short compared to the interval over which extinction is estimated is not an issue because this study is not seeking to carefully measure extinction rates; the model averaging method does account for variation in extinction rate through time. Lastly, assumption 4) of independent detection and extinction between taxa might be important for specialist taxa (Liow and Nichols 2010), but there is no *a priori* reason to believe that comatulids are linked in such a way. The most important concern is that error in p and ϕ are correlated and can bias each other; we address this concern in the discussion. While this may affect the magnitude of our results, it should not affect the trends observed.

From the averaged parameters, derivative variables were calculated (Table 4.1). Genus duration is calculated based on a fixed chance of probability of extinction per million years, yielding an exponential decay curve, where the median taxon duration is shorter than the mean taxon duration. For this right-skewed statistic, we report median taxon duration as the time period over which half of the genera will go extinct. $\hat{S}$, the estimated true diversity, is calculated as the observed diversity divided by detection rate. Analyzed at the stage level generates a highly volatile estimated true diversity. For instance, several stages in the Cenozoic have no reports of comatulids. This method

would therefore estimate a nonsensical diversity of 0. Thus, the mean diversity over each of the time bins was used, calculated as the average diversity per stage. The number of total originations was calculated by adding the change in diversity between time bins and the origination rate necessary to replace periodic extinctions. This estimate of the total number of genera that have ever lived provides an estimate of the number of genera that remain unknown to science.

Results

Of the ~125 genera of comatulids that are extant, we found reports for nine of those genera in the paleontological literature, making total coverage ~7%. Of the 16 genera of comatulids reported from the Antarctic today, only *Notocrinus* has a reported fossil occurrence in our database, suggesting ~6% coverage. Of 14 genera reported from Davies Reef, Great Barrier Reef, Australia (Bradbury et al. 1987), none are reported from the fossil record. Of 11 genera from the Red Sea and surroundings, only one, or 9%, is reported from the fossil record (Hellal 2012). Of 3 genera recorded as extant in the Mediterranean Sea, only one is found in the fossil record, and thus coverage is 33% (Tortonese 1980). The Red Sea and Mediterranean Sea are most comparable to the European rock record estimates, and while those two have higher percentages of extant taxa described from the fossil record than the other two regions, the differences do not approach statistical significance. Completeness of fossil comatulid diversity is not only bad overall, but is deficient regionally.

The results of Alroy's detection probability method are reported in Fig. 4.3. They suggest a high rate of detection in the Late Cretaceous, and a poor or undefined p during other time bins. This method requires both three-timers and part-timers in order to calculate a detection probability. One tradeoff is that it calculated a separate detection probability for each stage. The pattern is virtually identical when run solely on European data; in contrast, the non-European data set produces entirely undefined results,

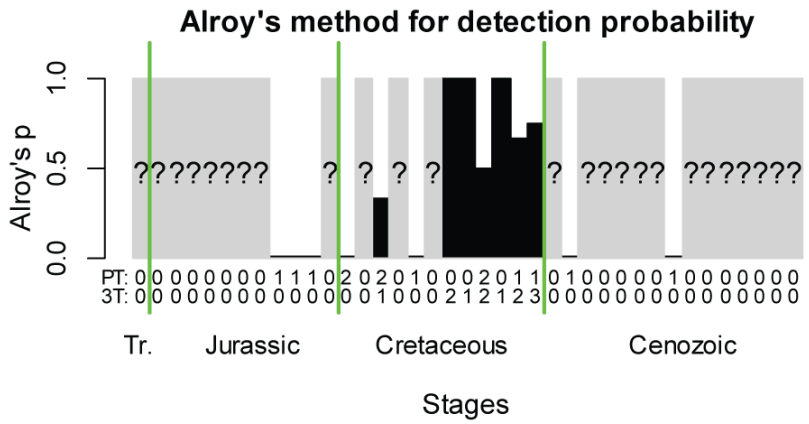


Figure 4.3 Detection rate using Alroy's (2008) method. This method is undefined at most stages, due to the lack of 3-timers over much of the range. This method suggests high detection rate in the late Cretaceous, and low detection rate in the remaining times. The pattern is virtually identical to this when run solely on European data, while the non-European data set results entirely in undefined results. PT = part-timer's per stage, 3T = 3-timers per stage.

demonstrating that the pattern is driven by the relatively good record in Europe. Summed across all time bins, the results of this method suggest a detection probability per genus per stage of 0.54, which is similar to the results for crinoids reported in other studies (Foote and Raup 1996, Foote and Sepkoski 1999).

For our data, *FreqRat* (Foote 1997) is 0. This is due to a lack of genera that survived through exactly two bins. This is likely due to chance, and inflating the number of two-bin genera to be equal to the number of three-bin genera results in a *FreqRat* of 0.16. Calculating the completeness, or proportion of fossil taxa that are known, is calculated per Foote (1997, Appendix 1). Using the detection rate per stage, we can calculate the mean number of stages from the mean duration of genera in million years calculated from the CMR data (Table 4.4), using the average stage length of 5.2 MY. Mean duration in stages is 9.75 based on the Cenozoic of Europe, the highest duration calculated using the CMR method from our data. Completeness calculated via this method, assuming a *FreqRat* of 0.16, is 0.81, while completeness with *FreqRat* of 0 is 0.

Table 4.4 Detection probability (p)

	European			
Time Bin				
Mya	227-141.2	141.2-85.8	86.3-61.6	61.6-2.6
estimate	0.226	0.369	0.605	0.090
lower 95% CI	0.101	0.224	0.364	0.037
upper 95% CI	0.432	0.543	0.804	0.200

	Non-European			
Time Bin				
Mya	227-141.2	141.2-85.8	86.3-61.6	61.6-2.6
estimate	0.048	0.093	0.209	0.037
lower 95% CI	0.001	0.005	0.024	0.007
upper 95% CI	0.629	0.675	0.740	0.167

This is obviously a very broad range and almost certainly includes the real completeness, but which also does very little to constrain it.

Another independent method of assessing historic diversity is to examine inferred diversity using a time-calibrated molecular phylogeny. One recent example (Rouse et al. 2013) that included less than a third of the ~125 described genera, with only 35 comatulid genera, yielding a conservative estimate, suggests a minimum diversity of 22 genera 30 million years ago and 8 genera 65 million years ago. Of their 35 genera, all are inferred to have been extant during the Pleistocene. This shows that the low coverage found is not simply the result of a very recent radiation of the comatulids. Only two comatulid genera are recorded in our database from the Pleistocene, suggesting a maximal detection rate of 0.057%. Both of these genera are represented in Rouse's phylogeny.

Estimated global diversity using CMR modeling is reported in Fig. 4.4 (upper). The most likely real diversity, *S*, is substantially higher than the diversity detected via the range-through method in all time bins. The most likely real diversity of comatulids rises from ~9 in the Triassic/Jurassic to ~33 in the Cenozoic. This supports a hypothesis of increasing comatulid diversity over the Cenozoic. For most stages, less diversity is

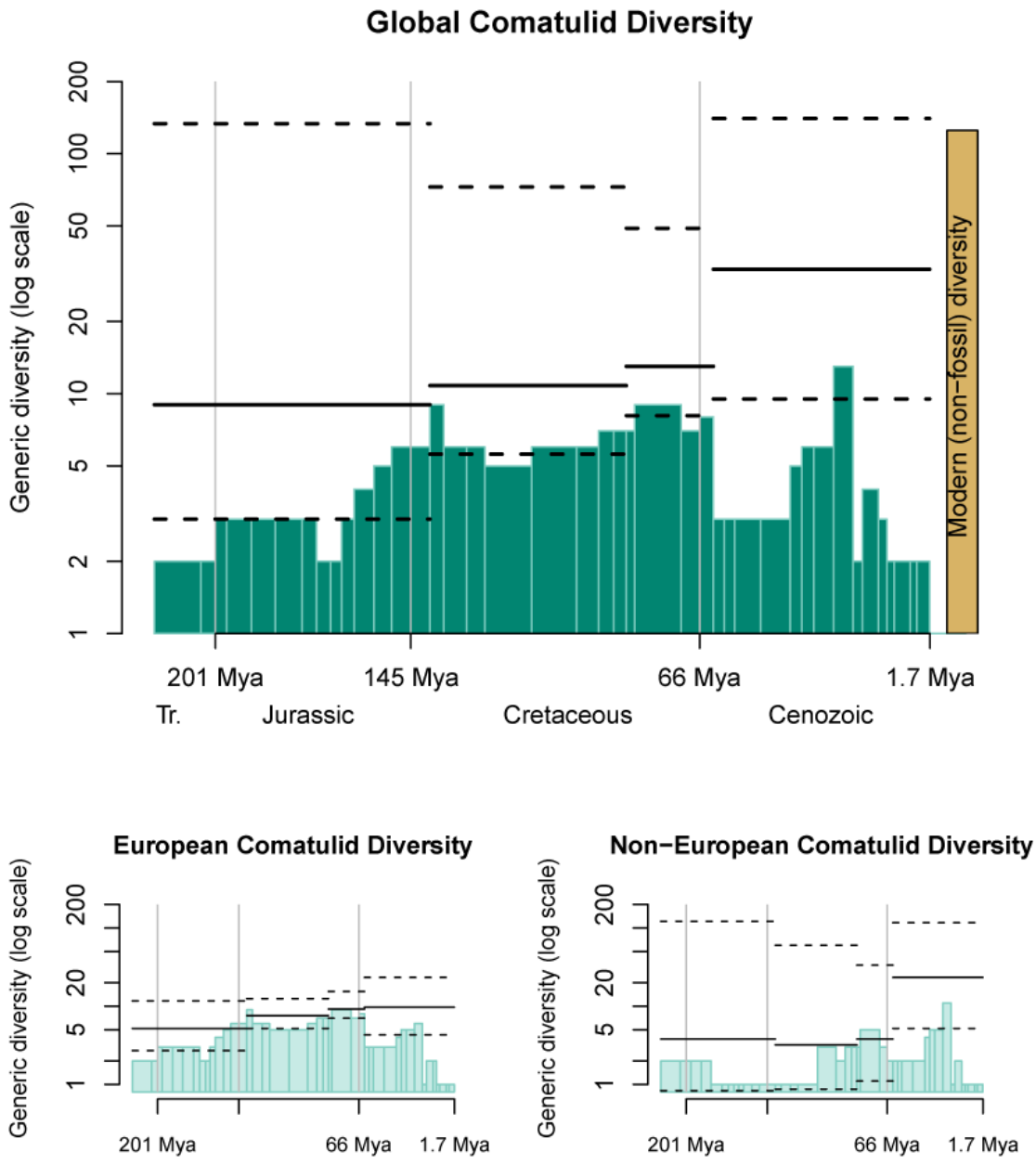


Figure 4.4 Detection rate using Alroy's (2008) method. This method is undefined at most stages, due to the lack of 3-timers over much of the range. This method suggests high detection rate in the late Cretaceous, and low detection rate in the remaining times. The pattern is virtually identical to this when run solely on European data, while the non-European data set results entirely in undefined results. PT = part-timer's per stage, 3T = 3-timers per stage.

present via the range-through method than is estimated at the lower bound of the confidence interval calculated by the CMR method. Estimates at the upper confidence intervals for the CMR method rival modern diversity during the Triassic/ Jurassic and Cenozoic (Fig. 4.4), but these large error bars are the result of poorly constrained parameters, due to the tiny amount of data available.

Breaking down the CMR results by region, European diversity (Fig. 4.4.B) is fairly tightly constrained through the entire post-Paleozoic, with a gentle increase in diversity from ~5 genera in the Triassic/Jurassic to ~10 genera in the Cenozoic. Pre-Cretaceous estimated true diversity matches well with diversity via the range-through method, suggesting that most genera have been described. Incompleteness increases after the end of the Danian in Europe.

In contrast, non-European diversity is poorly constrained through most of the post-Paleozoic. The most constrained time period is during the Late Cretaceous, with estimated true diversity between 1 and 33, and a most likely diversity of 4. This increases significantly during the Cenozoic, with estimated true diversity then ranging from 5 to 117, and most likely diversity at 23 genera. This nearly 6-fold increase in diversity is unique in the results, and suggests a significant Cenozoic radiation for the comatulids.

Detection probability (Table 4.4) shows large heterogeneity over space/time, ranging between 0.605 in the European region during the late Cretaceous to 0.037 in non-European regions in the Cenozoic. Detection probability in Europe is consistently higher than for the rest of the world, likely reflecting both search effort as well as peculiarities of regional geology such as the presence of chalk deposits and shallow epeiric seas. Confidence intervals for Europe are significantly smaller than for non-Europe. Survival rate per million years ranges from 0.95 to 0.98 (Table 4.5), which equates to median generic life spans of 21 MY and 49 MY respectively (Table 4.6). This heterogeneity in record quality will introduce significant bias in efforts to summarize diversity through time and other calculations based on the fossil record.

Table 4.5 Survival rate per million years (φ)

	Europe		Non-European	
MYA	227-61.6	61.6-2.6	227-61.6	61.6-2.6
estimate	0.978	0.980	0.953	0.973
lower CI	0.963	0.943	0.677	0.835
upper CI	0.986	0.993	0.995	0.996

Table 4.6 Median generic duration in millions of years.

Mya	227 – 61.6	61.6 – 2.6
Europe	30.6	33.8
Non-European	14.4	25.1

Coverage based on estimated true diversity is presented in Table 4.7. Using the estimated true diversity from the CMR method, our results show that 67% of expected European comatulid genera have been described, while only 21% of non-European fossil genera have been described. These results suggest very poor coverage for most of the world. This method assumes that all unknown fossil genera are as likely to be found as the known genera. Given that common genera are more likely to be captured, this estimate is likely very conservative.

Discussion

Understanding and accounting for biases is a crucial component.of many paleontological studies. In this paper, we attempt to differentiate between two explanations for the low diversity of comatulids reported from the fossil record: either they are detected at a lower rate than other crinoids, or a Cenozoic diversification event left little time for them to enter the fossil record. What we have found is support for both alternatives: comatulid diversity did appear to rise in

Table 4.7 Predicted/described originations and completeness by region.

	predicted originations	described diversity	proportion known
European	44.7	30	0.67
Non-European	86.7	18	0.21

the Cenozoic, and the comatulid detection rate in the Cenozoic is exceptionally low. Additionally, the heterogeneous nature of comatulid detection unpredictably biases various metrics that are used to understand the quality of the fossil record. It is not clear if the low detection rate for comatulids is driven by failure to entry into the fossil record, or if a lack of collection effort prevents occurrences from being recorded in the scientific literature.

Much work in the past decade has focused on the adequacy of the fossil record, with many studies reporting apparently high completeness for many marine invertebrate groups (Foote and Raup 1996, Foote and Sepkoski 1999, Valentine et al. 2006). These studies estimate the probabilities of detection for crinoids per genus per ~5 MY time intervals at ~0.38 and ~0.5, and state that 50% of extant crinoid families have a described fossil record. These values are similar to the per-genus per-stage detection rate estimated using Alroy's method reported here of 0.54. Foote's (1997) *FreqRat,* a measure of detection probability, is 00 for our data, but reasonable assumptions about sampling suggest it may be closer to 0.16. These methods all make differing assumptions, but have generally been described as resistant to biases in assumptions (Alroy 2010, Foote 1997). Their disagreement is troubling, but seems to be caused by the extreme nature of the data set used here, necessitating the use of the CMR method in order to provide an appropriate model with appropriate model selection techniques.

Using the CMR method, the detection rate is heterogeneous in both time and space, peaking briefly in the European region at 0.605 per genus per stage in the Late Cretaceous, with worldwide detection rates ranging from 0.04 to 0.21 per stage.

Valentine et al. 2006 also found heterogeneity in the well-preserved Bivalvia, with "small body size, reactive shell structures, commensal or parasitic habit, deep-sea distribution, narrow geographic range, restriction to regions exposing few Neogene marine sediments, or recent date of formal taxonomic description" linked to low detection rate, suggesting that heterogeneity is also present in well preserved taxa. One important note is that detection probability will depend on numerous factors, from preservation potential, to existence in exposed rock, as well as collector effort and interest. Differentiating between these causes for variation detection probability is outside the scope of this paper but should be pursued.

The heterogeneity that we document can lead to misleading statistics regarding completeness (Foote 1997). A simple thought experiment based on these results demonstrates how heterogeneity can result in misleading completeness statistics. Assume two time bins with equal diversity, one of which has 100% completeness, whereas the other has 0% completeness in their respective fossil records. Calculating a simple completeness metric based on the information contained in the fossil record would show 100% completeness, a deceptive result. This highlights the spottiness of the fossil record, and suggests that averaging completeness metrics through time and space may systematically overestimate completeness in the fossil record. This problem will also result in overconfidence in other metrics, such as Good's *u* (1953), used to estimate coverage for the shareholder quorum subsampling method (Alroy 2010). While the method presented here would not solve the problem for the extreme example described above, even a very small amount of data, as in our non-European data set, allows for estimates of the uncertainty. The completeness metric of Foote (1997) also does not give reliable results for our data set, even though simulations showed it to be resistant to the problem of heterogeneity in taxon preservation. Relatively small sample sizes are certainly an influence, but edge effects (Alroy 2010) may also be significant for our data set. Further work should investigate heterogeneity in detection rate for other taxa, as well

as implications for studies that attempt to summarize patterns through time using taxonomic lists.

One particularly interesting result of this study is the high detection rate and comparatively good record of comatulids in the late Cretaceous of Europe. This coincides with the existence of extensive, shallow mid-continental seas. Epeiric seas were also much more pervasive during the Mississippian, also known as the "Age of crinoids." It is probably not a coincidence that the presence of globally widespread carbonate ramps is linked to that high diversity, and one hypothesis suggests environmental causes for that high diversity (Kammer and Ausich 2006). The evidence presented here suggests that an environment similar to that found during the Mississippian is linked to an order of magnitude higher detection rate in the late Cretaceous of Europe. It seems likely that the environmental hypothesis and increased detection rates are difficult to differentiate. Relatively high diversity observed in the Silurian, for instance, might be dampened by a lower chance of detection. Thus, it is not clear whether the Mississippian was truly the "Age of crinoids" and not the "Age of high crinoid detection rates." The relationship between diversity and bias in the rock record has long been a subject of discussion in the context of global biodiversity, and this conversation is not over. The method presented here is a significant step forward in that it can separate real patterns from the happenstance of preservation biases.

The overprint of taphonomic bias on diversity is not unique to comatulid crinoids. One study on another group of echinoderms, the cidaroid sea urchins (Greenstein 1992) used a different approach to understand taphonomic effects. That study focused on degree of disarticulation through time, and therefore is representative of the quality of material found in the fossil record, contrasting with the study herein in which the lack of detection due to material availably or description were not discriminated. Greenstein (1992) found the most articulated cidaroid specimens in the Middle Jurassic through end Cretaceous, and the Eocene/Oligocene. Our results find higher detection rates overlapping with the

increase in articulated specimens from the Middle Jurassic through end Cretaceous, but found low detection rates coinciding with the Eocene/Oligocene peak observed by Greenstein. If similar biostratinomic and diagenetic processes operate for the comatulids and the cidaroids, it would suggest that there is material from the Eocene/Oligocene in the fossil record that is yet to be described. Under this scenario, the extremely low detection rate for the comatulids during the Eocene/Oligocene is largely due to a failure to enter the literature, as opposed to a failure to enter the fossil record.

Our most likely estimated true diversity for the Cenozoic is still only one-third of the reported modern diversity. There are at least three reasons that our estimate may be too low. First, CMR methods assume an equal chance of capture for each taxon, but the taxa most likely to be detected are those that are most abundant to begin with. Uncommon taxa are probably underrepresented using this method, which therefore can reduce the estimated diversity relative to the real diversity. Second, the CMR method simultaneously calculates detection probability and survival rate. Considering that an absence of a taxon outside its range could be explained by either no detection or extinction before that time period, the detection and extinction parameters are inversely related. In our results, the modeled lower survival rate for non-European compared to European comatulids is counterbalanced by a higher detection rate. If survival rates for the non-European comatulids are the same as those in Europe, our diversity estimates are low. Lastly, there may be a bias related to material used to describe fossil comatulids. Taxonomy of fossil comatulids is based almost exclusively on the morphology of one element, the centrodorsal. One recent study suggests this might not be a major concern because centrodorsals are shown to be adequate for differentiating closely related cryptic species (Chapter II, above).

One proposal for dealing with differences in preservation potential between groups is to conduct separate analyses for each higher taxon of interest (Alroy 2010). That would certainly be an improvement over lumping groups that have different

preservation modes, but our results suggest that differences in preservation through time and space can also generate misleading patterns. It is not clear when or how one should subdivide through time, space, and taxonomic hierarchy. The evidence for preservational heterogeneity presented here suggests that merely dividing into higher level taxonomic groups may not be sufficient.

Conclusions

Biases in the observed fossil record, caused by peculiarities of both taphonomy and collection, can hinder all paleontologists. It has been observed that when fossil material is present, the quality of that record seems to be high, but much work still needs to be done to understand what material is present, and what has been described. We address these potential biases for one group, in order to understand how they can create an overprint on diversity estimates. We find that biases obscure a real pattern, a troubling result.

Two hypotheses for the cause of low comatulid diversity reported in the fossil record were investigated; low detection rates versus a recent diversification event. Our results support both of these hypotheses, indicating that over the Cenozoic, detection rates are low and that the group did appear to undergo a taxonomic radiation, consistent with that observed from molecular phylogenies. Inconsistent detection rates for the comatulids mean that several metrics that attempt to summarize quality of the fossil record may be biased. Despite a poor record, we can constrain comatulid diversity through time by using both absence and presence data to guide our investigation. Comatulid diversity was likely lower during the Mesozoic than diversity observed today. The capture-mark-recapture method used here should be applicable to other taxa, and would provide significant insights into patterns of preservation through time and space.

CHAPTER V

Conclusion

Understanding the imperfect nature of the fossil record, and working within the limitations imposed by it has been a major theme of the paleobiology revolution. One recurring observation has strong signals overwhelming biases. The comatulid crinoids have a very imperfect fossil record, one that seems to push the limit of signal strength. A straightforward reading of the observed fossil record suggests that the comatulid underwent a massive Holocene radiation. However, it appears that especially low levels of detection in the Cenozoic obscure a gentler rise in their diversity. The results presented here suggest that peak morphological diversity was reached in the Jurassic, paving the way for a more casual increase in diversity. Heterogeneity in the detection rate is strong enough to obscure, but not completely hide, the record of comatulid diversity.

The dissertation began by addressing a common problem in paleontology: less complete specimens than are available to neontologists. While the lack of soft parts in the fossil record is a common example of this bias, taxonomic work on disarticulated specimens is a problem faced by those studying many organisms with many discrete skeletal elements. Two methods were introduced that could help address similar problems in other groups: comparing intraspecific disparity and finite mixture analysis. In addition to the proximal question, these methods proved adequate for detection of cryptic diversity in a modern group that were genetically distinct, but with little morphological difference at the whole body level. This suggests that those undertaking taxonomy on whole

comatulids might find centrodorsals, a part often neglected, informative. More generally, these methods should useful for detecting cryptic species and settling issues related to anagenetic events in the fossil record. The results indicate that centrodorsal shape is a good source of information for taxonomic work. From this, we conclude that the fossil record of comatulids is not depauperate simply because of an inability to separate taxa based on the material available to taxonomists. However, we also observed that some distantly related species had very similar centrodorsal shapes. This leads into chapter three, where we address centrodorsal disparity through time.

Chapter three focused on how centrodorsal morphology has evolved over the group's history. Recent work has shown that traditional morphologically defined comatulid taxa, from families to genera, are polyphyletic on molecular phylogenies. This is troubling for paleontologists, with only morphology available, but this may have broader implications for processes of morphological evolution. A common pattern seen in studies of morphological disparity in the fossil record is one of rapid explosion of morphologies, followed by stability. This pattern is well established for crinoids. This is tied into the concept of an adaptive radiation, and George Gaylord Simpson's observation that a great deal of morphological change seems to accumulate early in many group's history. However, the observed pattern of early morphological diversification can be explained either as a slowdown in absolute rates, or a slowdown in observed rates. If distantly related comatulids can evolve to have the same shape, even after having diverged into morphospace, it suggests that rates of morphological evolution may be high even though no new morphological territory is reached. This seems to be the case, with high disparity within the traditional comatulid superfamilies found in this study. So, while chapter two showed that centrodorsals may be adequate for differentiating closely related species, chapter three suggests that reconstructing higher level taxonomic structure may be more difficult.

Having shown centrodorsals to be adequate for taxonomy at the species level, and having provided evidence that comatulid morphospace has been filled since the Jurassic, we turned in Chapter four to understanding the quality of the observed fossil record for comatulids. What we discovered is that the observed record shows great heterogeneity, and is quite bad excepting the Cretaceous of Europe. Worldwide, the observed Cenozoic record of comatulids is especially bad, with a detection rate low enough to drive completeness near single digits. Additionally, total diversity during the Cretaceous seems to be significantly lower than that during the Cenozoic. The exceptionally poor record over the Cenozoic hides this real increase in diversity, creating the illusion of equal diversity through the fossil record of comatulids. An important question is where future efforts should be focused. Centrodorsals are but a single ossicle of the ~10,000 that make up a typical comatulid, which means that finding them can take significant effort. Awareness of what centrodorsals are, and that they may be diagnostic even though they are single disarticulated elements, may not be high among workers sorting through Cenozoic sediment.

High detection rates in the Cenozoic of Europe could be explained by either better environmental conditions for comatulid preservation or by greater collection and description efforts by paleontologists, or perhaps by both. If better preservation was the primary reason for the higher observed detection rates, namely the presence of relatively shallow, widespread, epicontinental sea, then it is entirely possible that much of comatulid diversity through time will not be discoverable. However, there are a few reasons to believe the problem is related to the observations of comatulids, and not their existence in the fossil record to begin with. First, efforts to describe material from the rest of the world have been successful, with recent work describing many new comatulids from the fossil record of the Pacific, and observations that material from North America is not being properly studied. There are also reasons to believe that the Cretaceous of Europe did have environmental reasons for superior preservation; the chalk sediments

typical of the late Cretaceous of Europe are diagnostic of an environment that should be excellent from comatulid preservation, and those chalks are not widespread through time and space The relatively lower detection rates in the Jurassic and Cenozoic of Europe may be more realistic expectations for much of the non-European world.

This dissertation has been focused on understanding an incomplete fossil record, including methods to quantify the quality of the record, to understand what we can learn from the record, and what may be lost. In chapter II, we focused on whether fragmentary material is enough to identify comatulid species; it appears to be adequate. In chapter III, we focused on morphological changes through time for comatulid centrodorsals, confirming a common pattern of early expansion into morphospace followed by stasis; the results here suggest a high rate of evolution with constraints that limit further expansion, which may make differentiating higher level comatulid taxa difficult. In chapter IV , we attempted to understand the causes of the jump in comatulid diversity from the observed fossil record to the recent; we found evidence that low detection rate and a Cenozoic radiation contributed. Overall, we have learned that there is much to be learned even with an imperfect record to work with.

APPENDIX A

Use of FMA for GM data

Finite mixture analysis (GM) is a powerful method, but care must be used when choosing data input into the FMA models. The assumptions inherent to any data set must be considered, as those assumptions strongly bias how well various models fit that data set. Geometric morphometric (GM) data have properties that must be accounted for in analysis, such as non-independence of landmarks due to superimposition. One concern is the use of diagonal models for principal component data. Principal components analysis (PCA) rotates the data such that maximal variance is expressed along each successive principal component. Each datum in a principal component data set obtained via generalized procrustes analsysis is necessarily dependent on the others, as while they are statistically orthogonal they are not independent. Additionally, the rotation applied during PCA should minimize the covariance between dimensions. This prior rotation is not accounted for in the models assessed by MCLUST. Including this class of models, but accounting for the parameters associated with the PCA rotation, results in low enough BIC and AICc values that they would be rejected. As such, they were not included in analysis. Results additionally indicated that the “EEE” (ellipsoidal, equal volume, shape, and orientation) and the spherical “E” and “V” models tended towards over splitting, and thus were excluded. In the end, only the full “VVV” (ellipsoidal varying volume, shape and orientation) model was used, in order to maximize consistency when comparing groups. I recommend that future attempts to use FMA on GM data use only the “VVV” model in order to prevent errant results.

In order to assess the role of sampling in the results, 1000 bootstrap replicas were undertaken with the three genera with highest sample sizes. Results for BIC are reported in Table 4 and via AICc in Table 5. Optimal number of groups approximates a Poisson distribution under both AICc and BIC results. AICc results average higher and with more spread for all three genera, but the single highest result are found in BIC scores, with one optimal score of 6 groups for *Promachocrinus*.

In order to test whether one group has more than another, a method has been devised for comparison using bootstrapped samples. Tested here are whether the number of groups between our three genera are greater, smaller, or equal. For each model run, the average proportion of model runs of the comparison taxa that match your inequality will be the proportion of times that inequality holds.

Example.--Let us test whether the number of groups observed for *Promachocrinus* is equal to the number of groups observed for *Jaekelometra* using the BIC results, drawing data from Table 4. In this equation, *g* is number of groups and *gmax* is the maximum number of groups from bootstrapped model runs.

$\sum_{g=1}^{gmax}$ proportion of species 1 model results with g groups $*$ proportion of species 2 model results with g groups (g)=

$$\sum \frac{940}{1000} * \frac{16}{1000} + \frac{48}{1000} * \frac{513}{1000} + \frac{10}{1000} * \frac{389}{1000} + \frac{1}{1000} * \frac{55}{1000} + \frac{1}{1000} * \frac{8}{1000} + \frac{0}{1000} * \frac{1}{1000} =$$

odds that # of groups are the same for *Jaekelometra* and *Promachocrinus* = 0.043617

Table A.1 Optimal number of groups over 1000 bootstrap replicas scored by BIC, *n* = 116.

	# of groups					
Genera						
Jaekelometra	940	48	10	1	1	0
Promachocrinus	16	513	389	55	8	1
Florometra	2	873	118	7	0	0

Table A.2 Optimal number of groups over 1000 bootstrap replicas scored by AICc, n = 116.

# of groups	1	2	3	4	5	6
Genera						
Jaekelometra	416	298	230	54	3	0
Promachocrinus	0	313	551	126	10	0
Florometra	0	520	412	65	3	0

Table A.3 Equalities between fossil genera. See text for details of method. This table shows the probability that the number of components in one taxon is equal, larger than, or smaller than the number of components in another taxon. High proportions indicate likely relationships.

	Proportion of results	
equalities	BIC	AICc
Jaekelometra = *Florometra*	0.045	0.253
Jaekelometra = *Promachocrinus*	0.044	0.227
Promachocrinus = *Florometra*	0.494	0.398
inequalities		
Florometra > *Jaekelometra*	0.944	0.575
Florometra > *Promachocrinus*	0.083	0.188
Jaekelometra > *Florometra*	0.011	0.172
Jaekelometra > *Promachocrinus*	0.008	0.121
Promachocrinus > *Florometra*	0.405	0.414
Promachocrinus > *Jaekelometra*	0.930	0.653

APPENDIX B

Description of characters and individual morphospace occupation

1. Height of centrodorsal (CD) divided by width of CD. Minimum = 0.76, maximum = 5.14.

2. In lateral view, length from center of CD to edge along radial groove divided by length from center of CD to edge midway between radial grooves. Minimum = -0.5, maximum = 0.21.

3. Centrodorsal columnal fusion. Fused = 0, unfused = 1.

4. Cirral scar sockets arranged in columns. Columns = 1, irregular = 0.

5. Total number of cirral scars. Minimum = 10, maximum = 90.

6. Number of cirral scar columns. Minimum = 5, maximum = 20.

7. Cirral socket scars touching. Touching = 1, mixed = .5, separated = 0.

8. Interradial ridges between sockets. Present = 1, absent = 0.

9. Radial ridges stretching from oral face to aboral face, lined up with radial intersection. Present = 1, absent = 0.

10. Depth of cirral scar sockets. Depth > 0.3 of width =1, shallow =0.

11. Cirral scar socket arranged in rows. Present = 1, absent = 0.

12. Presence of fulcral ridges in cirral socket scars. Present = 1, absent = 0.

13. Diameter of dorsal surface divided by maximum CD diameter. Minimum = 0, maximum = 0.95.

14. Dorsal area concavity. Concave = 0, flat = .5, convex = 1

15. Dorsal area shape. Round = 1, other = 0.

16. Texture of dorsal area. Smooth = 1, other =0

17. Granulated dorsal surface. Present = 1, absent = 0.

18. Synarthrial articulation on ventral surface. Present = 1, absent = 0.

19. Cirral scars present on ventral surface. Present = 1, absent = 0.

20. Dorsal star. Present = 1, absent = 0.

21. Width of oral cavity divided by CD width. Greater than 0.35 = 1, < 0.35 = 0.

22. Oral cavity cavernous. Present = 1, absent = 0.

23. Interradial buttresses within oral cavity. Present = 1, absent = 0.

24. Shape of oral cavity. Round = 1, other = 0.

25. Radial pits on oral face. Present = 1, absent = 0.

26. Concavity of oral face. Concave = 1, flat or irregular = 0.

27. Coelomic impressions on furrows of oral face of centrodorsal. Present = 1, absent = 0.

28. Texture on oral face. Smooth = 1, otherwise = 0.

29. Basal rod furrows intersect edge of oral surface. Present = 1, absent = 0.

30. Proximal end of basal rod burrows rounded. Present = 1, absent = 0.

31. Basal rod furrow forks. Present = 1, absent =0.

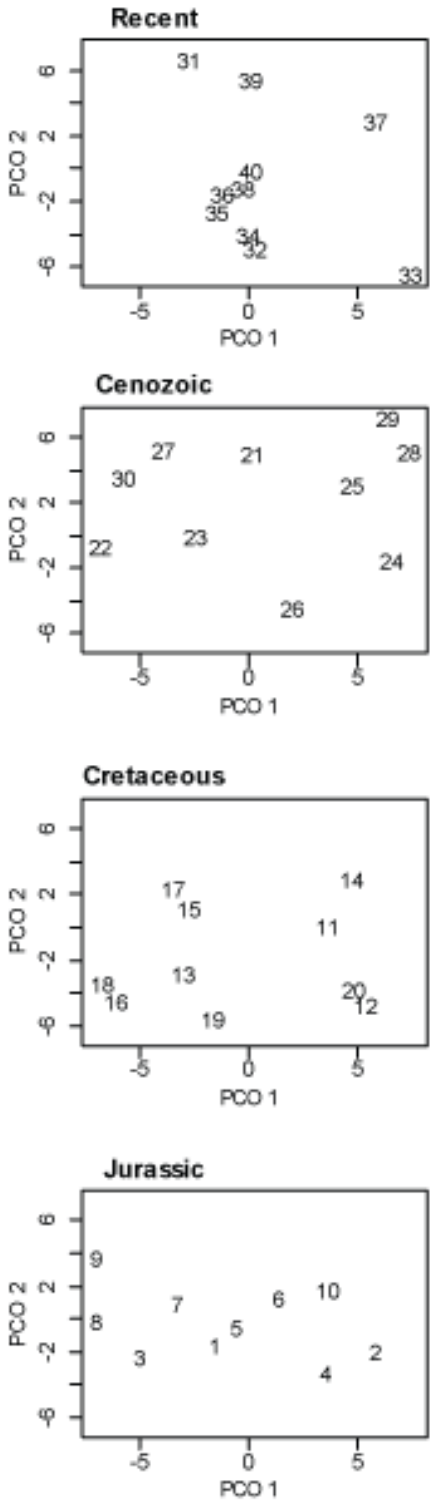


Figure B.1 Morphospace through time by individual. Specimens from Fig. 3.2 are identified in the same order as in Table 3.1.

REFERENCES

Adams, D. C., and E. Otárola-Castillo. 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution 4:393–399.

Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. Pp.199–213 *in* Selected Papers of Hirotugu Akaike. Springer.

Alroy, J. 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. Quantitative methods in paleobiology. Paleontological Society Papers 16:55–80.

Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezusko, M. Foote, F. T. Fursich, T. A. Hansen, S. M. Holland, L. C. Ivany, D. Jablonski, D. K. Jacobs, D. C. Jones, M. A. Kosnik, S. Lidgard, S. Low, A. I. Miller, P. M. Novack-Gottshall, T. D. Olszewski, M. E. Patzkowsky, D. M. Raup, K. Roy, J. J. Sepkoski, M. G. Sommers, P. J. Wagner, and A. Webber. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. Proceedings of the National Academy of Sciences 98:6261–6266.

Alroy, J., M. Aberhan, D. J. Bottjer, M. Foote, F. T. Fürsich, P. J. Harries, A. J. W. Hendy, S. M. Holland, L. C. Ivany, W. Kiessling, M. A. Kosnik, C. R. Marshall, A. J. McGowan, A. I. Miller, T. D. Olszewski, M. E. Patzkowsky, S. E. Peters, L. Villier, P. J. Wagner, N. Bonuso, P. S. Borkow, B. Brenneis, M. E. Clapham, L. M. Fall, C. A. Ferguson, V. L. Hanson, A. Z. Krug, K. M. Layou, E. H. Leckey, S. Nürnberg, C. M. Powers, J. A. Sessa, C. Simpson, A. Tomašových, and C. C. Visaggi. 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science 321:97–100.

Ausich, W. I. 1997. Regional encrinites: a vanished lithofacies. Paleontological events: stratigraphic, ecologic and evolutionary implications. Columbia University Press, New York:509–519.

Barton, K. 2011. MuMIn: Multi-model inference. http://cran.r-project.org/web/packages/MuMIn/index.html

Baumiller, T. K., and A. Gazdzicki. 1996. New crinoids from the Eocene La Meseta formation of Seymour island, Antarctic peninsula. Palaeontologia Polonica 55:101–116.

Baumiller, T. 2003. Experimental and biostratinomic disarticulation of crinoids: taphonomic implications. Echinoderm Research 2001: proceedings of the 6th European Conference on Echinoderm Research, Banyuls-sur-mer, 3-7 September 2001

Baumiller, T, Salamon, M. A., Gorzelak, P., Mooi, R., Messing, C. G., and Gahn, F. J.. 2010. Post-Paleozoic crinoid radiation in response to benthic predation preceded the Mesozoic marine revolution. Proceedings of the National Academy of Sciences 107:5893–5896

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological):289–300.

Benton, M. J., M. A. Wills, and R. Hitchin. 2000. Quality of the fossil record through time. Nature 403:534–537.

Bradbury, R. H., R. E. Reichelt, D. L. Meyer, and R. A. Birtles. 1987. Patterns in the distribution of the crinoid community at Davies Reef on the central Great Barrier Reef. Coral Reefs 5:189–196.

Briggs, D. E., R. A. Fortey, and M. A. Wills. 1992. Morphological disparity in the Cambrian. Science 256:1670–1673.

Bromley, R. 1990. Trace Fossils: Biology and Taphonomy: UnwinHyman. London, xi.

Brusatte, S. L., M. J. Benton, M. Ruta, and G. T. Lloyd. 2008. The first 50 Myr of dinosaur evolution: macroevolutionary pattern and morphological disparity. Biology Letters 4:733–736.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer.

Carpenter, P. H. 1879. Preliminary Report upon the Comatulae of the “Challenger” Expedition. Proceedings of the Royal Society of London 28:383–395.

Carroll, L. 1917. Through the looking glass: And what Alice found there. Rand, McNally.

Ciampaglio, C. N., M. Kemp, and D. W. McShea. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. Paleobiology 27:695–715.

Clark, A. H. 1915. A monograph of the existing crinoids. US Government Printing Office.

Connolly, S. R., and A. I. Miller. 2001a. Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. Paleobiology 27:751–767.

Connolly, S. R., and A. I. Miller. 2001b. Global Ordovician faunal transitions in the marine benthos: proximate causes. Paleobiology 27:779–795.

Darwin, C. 1859: On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life. J. Murray.

Donovan, S. K. 1991. The taphonomy of echinoderms: Calcareous multi-element skeletons in the marine environment. The processes of fossilization:241–269. Eagle, M. K. 1993. A new fossil isocrinid crinoid from the Late Oligocene of Waitete Bay, northern Coromandel. Records of the Auckland Institute and Museum 30:1–12.

Eagle, M. K. 2001. New fossil crinoids (Articulata: Comatulida) from the late Oligocene of Waitete Bay, Northern Coromandel Peninsula, New Zealand. Records of the Auckland Museum 37:81.

———. 2008. New comatulid crinoids from the Meyers pass limestone member (Waitakian (Chattian)) of the Pentland hills and Hurstlea, south Canterbury, New Zealand. Records of the Auckland Museum 45:101–129.

Erwin, D. H. 2007. Disparity: morphological pattern and developmental context. Palaeontology 50:57–73.

Folk, R. L., and R. Robles. 1964. Carbonate sands of Isla Perez, Alacran Reef Complex, Yucatan. The Journal of Geology:255–292.

Foote, M. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. Paleobiology 20:320–344.

———. 1996. Ecological Controls on the Evolutionary Recovery of Post-Paleozoic Crinoids. Science 274:1492–1495.

———. 1997a. Estimating taxonomic durations and preservation probability. Paleobiology 23:278–300.

——— 1997b. The evolution of morphological diversity. Annual Review of Ecology and Systematics:129–152.

——— 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. Paleobiology 25:1–116.

Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Foote, M., and J. J. Sepkoski. 1999. Absolute measures of the completeness of the fossil record. Nature 398:415–417.

Fortey, R., D. E. Briggs, and M. A. Wills. 1996. The Cambrian evolutionary “explosion”: decoupling cladogenesis from morphological disparity. Biological Journal of the Linnean Society 57:13–33.

Fraley, C., and A. E. Raftery. 1999. MCLUST: Software for model-based cluster analysis. Journal of Classification 16:297–306.

Frontier, S. 1976. Étude de la décroissance des valeurs propres dans une analyse en composantes principales: Comparaison avec le moddle du bâton brisé. Journal of Experimental Marine Biology and Ecology 25:67–75.

Gislén, T. 1934. A reconstruction problem, analysis of fossil comatulids from N. America, with a survey of all known types of comatulid arm-ramifications. H. Ohlsson.

Good, I. J. 1953. The Population Frequencies of Species and the Estimation of Population Parameters. Biometrika 40:237–264.

Gorzelak, P., and M. A. Salamon. 2013. Experimental tumbling of echinoderms—taphonomic patterns and implications. Palaeogeography, Palaeoclimatology, Palaeoecology 386:569–574.

Gould, S. J. 1986. Wonderful life: the Burgess Shale and the nature of history. Random House.

Gould, S. J. 1991.The disparity of the Burgess Shale arthropod fauna and the limits of cladistic analysis: why we must strive to quantify morphospace. Paleobiology 17:411–423.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. Biometrics:857–871.

Gower, J. C. 1975. Generalized procrustes analysis. Psychometrika 40:33–51.

Greenstein, B. J. 1992. Taphonomic bias and the evolutionary history of the family Cidaridae (Echinodermata: Echinoidea). Paleobiology 18:50–79.

Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, and T. J. Near. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Hellal, A. M. 2012. Taxonomic Study on the Feather Stars (Crinoidea: Echinodermata) from Egyptian Red Sea Coasts and Suez Canal, Egypt. Open Journal of Marine Science 02:51–57.

Hemery, L. G., M. Roux, N. Améziane, and M. Eléaume. 2013. High-resolution crinoid phyletic inter-relationships derived from molecular data. Cahiers de Biologie Marine 54:511–523.

Hemery, L., M. Eléaume, V. Roussel, N. Améziane, C. Gallut, D. Steinke, C. Cruaud, A. Couloux, and N. Wilson. 2012. Comprehensive sampling reveals circumpolarity and sympatry in seven mitochondrial lineages of the Southern Ocean crinoid species *Promachocrinus kerguelensis* (Echinodermata). Molecular Ecology 21:2502–2518.

Howe, H. V. 1942. Neglected Gulf Coast Tertiary microfossils. AAPG Bulletin 26:1188–1199.

Hess, H., C. G. Messing, and W. I. Ausich. 2011. Treatise on Invertebrate Paleontology, Part T (Revised): Volume 3: Echinodermata 2: (Crinoidea). The University of Kansas, Paleontological Institute, Lawrence, Kansas.

Hurvich, C. M., and C. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Hughes, M., S. Gerber, and M. A. Wills. 2013. Clades reach highest morphological disparity early in their evolution. Proceedings of the National Academy of Sciences 110:13875–13879.

Jagt, J. W. 1999. Late Cretaceous-Early Palaeogene echinoderms and the K/T boundary in the southeast Netherlands and northeast Belgium–Part 2: Crinoids. Scripta Geologica 116:59–255.

Janevski, G. A., and T. K. Baumiller. 2010. Could a stalked crinoid swim? A biomechanical model and characteristics of swimming crinoids. Palaios 25:588–596.

Kammer, T. W., and W. I. Ausich. 2006. The “Age of Crinoids”: a Mississippian biodiversity spike coincident with widespread carbonate ramps. Palaios 21:238–248.

Laake, J. L. 2013. RMark. An R interface for analysis of capture-recapture data with MARK. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center.

Liow, L. H. 2013. Simultaneous estimation of occupancy and detection probabilities: an illustration using Cincinnatian brachiopods. Paleobiology 39:193–213.

Liow, L. H., and J. A. Finarelli. 2014: A dynamic global equilibrium in carnivoran diversification over 20 million years. Proceedings of the Royal Society of London B: Biological Sciences 281.

Liow, L. H., and J. D. Nichols. 2010. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. Pp.81–94 *in* Quantitative Methods in Paleobiology.

Lofgren, A. S., R. E. Plotnick, and and P. J. Wagner. 2003. Morphological diversity of Carboniferous arthropods and insights on disparity patterns through the Phanerozoic. Paleobiology 29:349–368.

Lukacs, P. M., K. P. Burnham, and D. R. Anderson. 2009. Model selection bias and Freedman’s paradox. Annals of the Institute of Statistical Mathematics 62:117–125.

Lupia, R. 1999. Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. Paleobiology 25:1–28.

Messing, C. G. 1994. Comatulid crinoids (Echinodermata) of Madang, Papua New Guinea, and environs: diversity and ecology. Echinoderms through time. Balkema, Rotterdam:237–243.

———. 1997. Living comatulids. Paleontological Society Papers 3:3–30.

Messing, C. G., D. Meyer, U. Siebeck, L. Jermiin, D. Vaney, and G. W. Rouse. 2006. A modern soft-bottom, shallow-water crinoid fauna (Echinodermata) from the Great Barrier Reef, Australia. Coral Reefs 25:164–168.

Meyer, D. L., and D. B. Macurda Jr. 1977. Adaptive radiation of the comatulid crinoids. Paleobiology 3:74–82.

Meyer, D. L., and K. B. Meyer. 1986. Biostratinomy of Recent crinoids (Echinodermata) at Lizard Island, Great Barrier Reef, Australia. Palaios:294–302.

Moore, R. C., and C. Teichert. 1978. Treatise on Invertebrate Paleontology, Part T, Echinodermata. Vol. 2.

Nichols, J. D., and K. H. Pollock. 1983. Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture-recapture models. Paleobiology 9:150–163.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests. 2007. The vegan package. Community ecology package.

Oyen, C. W., and R. W. Portell. 2001. Diversity patterns and biostratigraphy of Cenozoic echinoderms from Florida. Palaeogeography, Palaeoclimatology, Palaeoecology 166:193–218.

Paleobiology Database. www.pbdb.org. 2014.

R Development Core Team. 2012. R: A language and environment for statistical computing.

Rasmussen, H. W. 1961. A monograph on the Cretaceous Crinoidea. E. Munksgaard.

Raup, D. M. 1962. Computer as aid in describing form in gastropod shells. Science 138:150–152.

Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Biology 39:40–59.

Rouse, G. W., L. S. Jermiin, N. G. Wilson, I. Eeckhaut, D. Lanterbecq, T. Oji, C. M. Young, T. Browning, P. Cisternas, and L. E. Helgen. 2013. Fixed, free, and fixed: the fickle phylogeny of extant Crinoidea (Echinodermata) and their Permian–Triassic origin. Molecular phylogenetics and evolution 66:161–181.

Scoffin, T. 1992. Taphonomy of coral reefs: a review. Coral Reefs 11:57–77.

Sepkoski Jr, J. J. 1981. A factor analytic description of the Phanerozoic marine fossil record. Paleobiology 7:36–53.

Sepkoski, J. J., R. K. Bambach, D. M. Raup, and J. W. Valentine. 1981. Phanerozoic marine diversity and the fossil record. Nature 293:435–437.

Sieverts, H. 1933. Drepanocrinus Jaekel, ein Synonym von Roveacrinus Douglas, und ein neuer Vertreter dieser Gattung aus der deutschen Kreide. Zentralblatt Für Mineralogie, Geologie Und Paläontologie B:54–59.

Simpson, G. G. 1953. Major features of evolution. Columbia University Press, New York.

Slater, G. J., S. A. Price, F. Santini, and M. E. Alfaro. 2010. Diversity versus disparity and the radiation of modern cetaceans. Proceedings of the Royal Society B: Biological Sciences 277:3097–3104.

Solow, A. R., and W. Smith. 1997. On fossil preservation and the stratigraphic ranges of taxa. Paleobiology 23:271–277.

Summers, M., C. G. Messing, and G. W. Rouse. 2014. Phylogeny of Comatulidae (Echinodermata: Crinoidea: Comatulida): A new classification and an assessment of morphological characters for crinoid taxonomy. Molecular Phylogenetics and Evolution.

Tortonese, E. 1980. Review of present status of knowledge of the Mediterranean Echinoderms. Echinoderms: Present and Past:141–149.

Tukey, J. W. 1977. Exploratory data analysis. Addison-Wesley.

Vadasz, E. 1914. Regenerationserscheinungen an fossilen Echinoiden. Centralblatt Für Mineralogie Geologie Und Paläontologie etc:283–288.

Valentine, J. W. 1985. Phanerozoic diversity patterns: profiles in macroevolution. p.

Valentine, J. W., D. Jablonski, S. Kidwell, and K. Roy. 2006. Assessing the fidelity of the fossil record by using marine bivalves. Proceedings of the National Academy of Sciences 103:6599–6604.

Vrieze, S. I. 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological methods 17:228.

Webster, G. D. 2003. Bibliography and index of Paleozoic crinoids, coronates, and hemistreptocrinoids, 1758-1999. Geological Society of America, Boulder, Colo.

Wesley-Hunt, G. D. 2005. The morphological diversification of carnivores in North America. Paleobiology 31:35–55.

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird study 46:S120–S139.

Wills, M. A. 1998. Crustacean disparity through the Phanerozoic: comparing morphological and stratigraphic data. Biological Journal of the Linnean Society 65:455–500.

——— 1998. Cambrian and Recent disparity: the picture from priapulids. Paleobiology 24:177-199.

Wills, M. A., D. E. Briggs, and R. A. Fortey. 1994. Disparity as an evolutionary index: a comparison of Cambrian and Recent arthropods. Paleobiology 20:93–130.

WoRMS Editorial Board. 2014. World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2014-10-15.

Zelditch, M. L., D. L. Swiderski, and H. D. Sheets. 2012. Geometric morphometrics for biologists: a primer. Academic Press.