1 **Detecting comatulid crinoid cryptic species in the fossil record**

2 Kristopher J. S. Purens¹

3 *University of Michigan Museum of Paleontology*

4 *1109 Geddes Avenue*

5 *Ann Arbor, MI 48109-1079, U.S.A.*

6 *mail: krisrhodes@gmail.com*

(1) *Present address: 1701 Kipling Street, Houston, TX 77098*

Abstract--

This work assesses the impact that taxonomic bias may exert on the richness 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 richness 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, this would bias upwards richness of fossil relative to extant crinoids, provided centrodorsals are a rich source of information.

In this study, centrodorsal shape of Recent and fossil comatulid species was measured using quantitative methods that can be applied uniformly to both groups. Two different methods applied to centrodorsal shape—disparity and finite mixture modeling—reveal no obvious bias of over- or under-splitting of Recent versus fossil comatulid species. Interestingly, the methods identified high richness within a putative extant species complex, supported with molecular data. Lacking evidence of taxonomic bias influencing the richness record of fossil comatulids, sampling and preservation are the likely sources of bias producing the 10-fold higher richness of extant over fossil comatulids.

Keywords: comatulid crinoid; geometric morphometrics; disparity; taxonomic bias; finite mixture models; cryptic species

1. 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 major concern for 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 after 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 comatulid crinoids, total disarticulation is very common—a single skeletal element, the centrodorsal, is the largest and most durable element. Consequently, the centrodorsal has been the source of taxonomic characters for description of most fossil comatulids. 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 species richness through geologic time, culminating in efforts to describe relative richness throughout the Phanerozoic (Sepkoski et al. 1981, Alroy et al. 2008). Comatulids,

the most diverse extant crinoids, are stalkless and mobile (Fig. 1 A), ranging worldwide from the abyssal ocean depths to shallow reefs. The richness record of post-Paleozoic crinoids reveals a striking pattern (Fig. 2), whereas the non-comatulid crinoids have similar levels of richness in the fossil record and Recent, the richness of the comatulids jumps by more than an order of magnitude between any fossil time bin and the Recent. Described generic richness from any time bin during the Paleozoic is similar to that described in the Recent. It is not clear what is driving the relatively lower crinoid richness during the post-Paleozoic in the fossil record, but such a pattern demands explanation.

Taxonomic treatment of fragmentary organisms varies greatly across taxa, and influences efforts to describe richness in broad spectrum groups including vertebrates, plants, and echinoderms. This treatment of fragmentation is taxon specific and inconsistent, ranging from little acknowledgement to active discussion of treatment. For crinoids, acknowledgement of the problem of disarticulation dates to Clark (1915), where he indicated that a primary motivation for his detailed description of Recent comatulid radials was to provide context for identification in the fossil record. The difficulty of identifying crinoid species from fragmentary material has been a concern of many, such as Howe (1942) who highlighted the problem for comatulids and Moore & Jeffords (1968), who said using disassociated crinoid remains for some taxonomic purposes is "... impossibl[e]...". However, Moore & Jeffords (1968) were optimistic that the diversity described from disarticulated specimens would one day far exceed that described from whole specimens. Moore & Jeffords expressed hope that studies would validate the use of disarticulated elements for taxonomy. Since then, taxonomy has been conducted on the overwhelmingly disassociated comatulid crinoid material, notably the comprehensive treatments of Rasmussen (1961) and Moore & Teichert (1978). Discussion of the relative taxonomic utility of disarticulated elements has largely dropped out of the

scientific conversation, with the implicit assumption that it is effective. It is important to note that, among the crinoids, only the fossil comatulids are described primarily from fragmented material, and only the comatulids have the large discrepancy in richness between the Recent and fossil record. Clearly, the effectiveness of disarticulated elements for taxonomy must be readdressed.

This paper focuses on this potential bias that may account for the dramatic discrepancy in richness: differences in taxonomic practice between neontologists and paleontologists. These differences in taxonomic practice arise from different availability of material for taxonomic description. 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 (Fig 1 B-D). The centrodorsal serves as the interface between the comatulid's cirri and arms, making it a key structural element. Centrodorsals are easily recognizable, and are commonly the largest single plate in the comatulid. Centrodorsals are homologous to proximal columnals in stalked crinoids. Taxonomic characters used to describe fossil material are generally the size and shape of the whole centrodorsal, 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 centrodorsal, 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 richness 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 centrodorsal disparity corresponds tightly with modern

species, or even corresponds with modern taxonomy at all. We can consider two hypotheses related to taxonomy that could explain the observed pattern, (1) the presence of undescribed cryptic species in the fossil record, or (2) oversplitting of extant taxa. There are several lines of evidence that suggest (1) is preferred: first, the previously described widespread belief that taxonomic efforts on disarticulated crinoid material is problematic and secondly studies that have generally found good agreement between comatulid species and molecular phylogenetics (Hemery et al. 2013, Rouse et al. 2013). We can test support for both of these hypothesis by testing for systematic, measureable differences between fossil species and Recent species. In this paper, two methods are used: morphospace volume (disparity) and finite mixture models. Additionally, if the null hypothesis of random results is best supported, we can consider other potential explanations for the observed pattern.

If a single fossil species is equivalent to many Recent species, one 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. 3: A, B). The increase in morphospace coverage as more species are subsumed will not be linear, but instead will depend upon the difference between the means of the taxa. If fossil species have a disparity (Fig. 3: C) that is different than the range suggested by modern taxa, that is evidence that they are not equivalent units. The second technique is finite mixture modeling (FMM). This method seeks to explain the data as a number of normally distributed groups with varying means, covariances, and volumes (Fig. 3: D, E). Whereas a measure of disparity can only detect bulk changes in morphospace occupied, finite mixture modeling can identify other differences in morphospace distribution described by their varying means, covariances, and volumes. If analysis with these two techniques indicates fossil species to have more groups than would be expected based on their current accepted taxonomy, it is evidence of hidden richness in

those taxa, and thus provides evidence of taxonomic bias.

2. Material and Methods

A total of 585 centrodorsals from eight Recent and three fossil species were included in this study (Table 1). Specimens were selected from museum collections with a preference toward larger sample sizes, and toward centrodorsals from which cirri were already missing, in order to minimize damaging museum specimens. Selection was in no way based on perceived patterns of morphological variation within the selected species, therefore, specimens studied should be considered 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 Recent 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.

Centrodorsal 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. 4). These landmarks describe the overall lateral view shape of the centrodorsal, and were selected based on applicability across the Comatulida. After digitizing, data were analyzed in R (R Development Core Team 2005). Data were aligned by generalized Procrustes analysis (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 geometric morphometric 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 Welch's *t* test was conducted to test the hypothesis that the fossil mean disparity is larger than the Recent mean disparity. This test is similar to the Student's *t* test with a correction for the samples potentially having unequal variance.

The most common distribution of continuous measurements of biological data is the normal distribution (see Sokal and Rohlf 1995, Ch. 6), and finite mixture modeling searches for normally distributed groups within given data. Two normal distributions with differing means, even ones that overlap significantly, can be identified given sufficient sample size, because they create a distinctly multi-modal distribution when combined. In other words, an observed data set could be explained by mixing multiple normal distributions, and these distributions in turn reflect morphological differences between (for example) several species. A character with an overlapping multi-modal distribution may not be useful for taxonomic purposes due to the inability to differentiate individuals in the overlapping values, but still can provide information about the underlying processes that created the distributions. Using multidimensional data can increase the power to differentiate peaks, which is one of the reasons that geometric morphometric data is being used here. A more complete discussion regarding use of geometric morphometric data for finite mixture modeling is presented in appendix A, as there are some technical concerns that must be addressed.

The finite mixture modeling method works by iteratively searching for distributions that best

explain the data using a maximum likelihood algorithm. For each set of geometric morphometric data models were fit that attempted to explain that data as a result of one distribution, two distributions, and so on, up to a total of ten distributions. These various models were then compared using model selection methods to determine how many groups were best supported by the data. Model selection methods were either Bayesian information criterion (BIC) or Akaike's information criterion, corrected for sample size (AICc). These criteria allow comparison of models, testing to see which are best supported, weighing model fit versus number of parameters included in the model. The models here explain the observed multivariate geometric morphometric data, allowing for differences in volume, covariance structure, shape, and position for each group.

Finite mixture modeling is used here to answer two questions:

1. How many distributions best explain centrodorsal shape for Recent and fossil species. If fossils are undersplit, we expect models with more distributions to be preferred, compared to Recent species. This was undertaken on all species, regardless of sample size. This method is described fully in appendix B.

2. Does genus A have more groups than genus B? This method was only undertaken on the three genera with sample size > 100 . This was accomplished in two different ways; first, with the whole sample for each genus and second, with subsampling of groups, at sample size of 116, equal to the number of individuals in the genus with the smallest sample. This investigation at the generic level is undertaken in order to further investigate the relationship between genera and centrodorsal shape. A complete explanation of this method is described in appendix C.

3. Results

Disparity of species is reported in Fig. 5. The disparities of Recent and fossil taxa broadly overlap. The species with the highest disparity is the modern *Promachocrinus kerguelensis*, 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 centrodorsals of these species with low disparity would be described qualitatively as flat pentagonal disks, whereas the centrodorsals of the species with highest disparity are tall and conical. Additionally, the range of bootstrap results is smallest for these two flat pentagonal centrodorsals 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 Recent taxa.

BIC results for finite mixture modeling and mean shape for species are shown in Fig. 6. For the majority of modern species, and all fossil species, one group was the preferred solution to the finite mixture modeling. 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 , indicating relative support for the multiple group hypotheses at less than 0.001. *P. kerguelensis* is unique among species with a large sample size in showing clear support for two groups.

A summary of the results from finite mixture modeling subsampled to $n = 116$ with full sample size are presented in Table 2. For *Florometra* and *Jaekelometra*, the number of groups 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 taxonomic species. Chi-square tests suggest a strong correspondence between the *Florometra* species and groups found via finite mixture modeling, and no evidence of such correspondence for

211 *Jaekelometra*. In this case, a random distribution for *Florometra* is rejected at $p = 6.3 \times 10^{-37}$, whereas it
 212 is not rejected for *Jaekelometra* at $p = 0.73$. Surprisingly, *Promachocrinus keurgenensis*, with but a single
 213 described modern species, is classified into a number of groups ranging from 2, using BIC at $n =$
 214 116, to 4, with AICc at $n = 175$. Both *Promachocrinus* and *Florometra* contained more groups than
 215 *Jaekelometra* in over 90% of bootstrapped model runs, whereas *Promachocrinus* had equal or more
 216 groups than *Florometra* in 90% of model runs (Table 3).

217 4. Discussion

218 4.1 Taxonomic considerations — The high disparity and multiple finite mixture modeling groups
 219 within *P. keurgenensis* are of interest, suggesting hidden richness within this species. Recent molecular
 220 phylogenetics (Hemery et al. 2012) indicate *P. keurgenensis* comprises ~seven mitochondrially distinct
 221 lineages, and at least two lineages based on nuclear genomics. Additional molecular phylogenetics
 222 nest *Florometra mawsoni* within this group (Hemery et al. 2013). These cryptic lineages are candidates
 223 for description as species, but attempts to find morphological characters to distinguish them have
 224 not succeeded. Morphological traits of color, pattern, and number of arm pairs were recorded, but
 225 all traits were distributed across the haplotype network with one exception: all individuals identified
 226 with six radials belong to a small cluster of related haplotypes within Hemery et al.'s (2012) clade D,
 227 even though individuals with other numbers of radials were also present within that clade. Other
 228 numbers of radials, from seven to 11, were present across the *Promachocrinus* haplotype network.
 229 Specimens with five radials seem to be exclusive to *F. mawsoni*. The lack of other morphological
 230 characters beyond centrodorsal shape to differentiate these clades is of great concern to taxonomists
 231 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 finite mixture modeling finding more groups than the traditional taxonomy. Recalculating mean modern disparity ± 2 standard deviations 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, finite mixture modeling shows between two and four groups within *Promachocrinus*, which is consistent with hidden richness within that taxon. If *Promachocrinus* is used as an example of hidden richness, both methods used here indicate a significant difference between it and other modern taxa. *Promachocrinus* provides evidence that our methods can detect lumped richness, a pattern not found in the fossil taxa studied.

Our methods indicate no support for the distinction of the two nominal fossil *Jaekelometra* species. No correspondence was present between centrodorsal 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 present in shape via permutation ANOVA (Oksanen et al. 2007), and the optimal solution using finite mixture modeling 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 centrodorsal 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 centrodorsal characters beyond the analysis of centrodorsal shape measured in only one orientation would reveal multiple species. Other characters used in centrodorsal 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 in this case, the fossil record may be oversplit relative to the modern, which is contrary to the author's expectations prior to beginning the study. However, with only one example, generalization is premature. In order to test if the results were unique to our method or dataset, finite mixture modeling 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 both centrodorsal height, and height + width. Given Jagt's 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.

4.2 Explaining the Observed Diversity Pattern —Using both disparity and finite mixture modeling, fossil taxa appear to be comparable to modern taxa, with no chronic evidence of undersplitting in the fossil record and limited evidence of oversplitting in the Recent. If taxonomic bias is not the cause of the dramatic increase in richness from fossil to modern, there are several alternatives to consider including (1) centrodorsals provide no taxonomic information (2) there is a real biologic increase in richness (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 centrodorsal 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 and size revealed that there were significant differences in means for every pairwise combination. Centrodorsals differ among modern species and, therefore, do contain taxonomic information.

A real ten-fold increase in richness over a short geologic time interval would be of great

interest— however this is not supported by molecular phylogenetics. A time-calibrated phylogeny of comatulids (Rouse et al. 2013) indicates 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, but 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 non-preservation 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 richnesses per time bin in the fossil record and calculate the probability that the modern comatulid richness of ~540 species (Messing 1997) is different than the fossil record. A Z-test rejects a null hypothesis that modern richness is pulled from the same sample as fossil richness ($Z = 26.7, p \approx 0$). It seems that the most likely explanation for the missing richness in the fossil record is some combination of biases. Additionally, these biases must be ones that affect the comatulids but not other extant crinoids, whose fossil record and modern richness are comparable. Broadly, other biases fit into two categories: 1) recognizable richness exists in the fossil record, but has not entered the literature either due to lack of sampling or lack of description, or 2) richness is not recorded in the fossil record due to the vagaries of preservation.

4.3 Influence of sampling on richness — 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 in Europe. Whereas 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 richness (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 richness today. Sampling bias as a significant factor in the richness 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 richness 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 richness 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 richness in these environments suggests that comatulid richness is comparable. This comparability seems to limit the possibility that comatulid richness never entered the fossil record, at least from certain environments. Nevertheless, the top candidate for high richness is shallow reef environments where the majority of modern comatulids occur (Messing 1997). Hotspots of comatulid richness 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 richness 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 biostratigraphic effects (Meyer and Meyer 1986), characterized by high rates of abrasion (Bromley 1990, 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). Given these numbers, between 250,000 and 2,000,000 grains might need to be inspected in order to identify a single centrodorsal. Some beneficial, coincidental hydrodynamic sorting or similar process might be necessary for recognition of comatulids from reef environments. These factors suggest that recoverable richness 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 richness in the fossil record. Most likely, both biases contribute to that missing richness. Certainly some richness 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.

5. 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 are 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 richness; his efforts required many months of work for a relatively limited result. 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 richness 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 support of taxonomic bias as the cause of this difference in richness.

Modern workers should be convinced to examine centrodorsals more closely, and molecular phylogenetics provides an important tool for understanding richness in modern comatulids and also for understanding how richness in the fossil record compares to it. Larger sample sizes and more morphological data could help bridge the richness gap, but limits will quickly be reached on what can be differentiated based on the available material. Hemery et al. (2012) used more than a thousand specimens of *Promachocrinus kerguelensis* in her molecular study, and geometric

morphometrics requires well-preserved specimens. However, hundreds of centrodorsals are not usually available for the paleontologist. Therefore, the number of specimens available might present a serious problem for species' identification in the fossil record.

The order of magnitude difference in richness 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 richness of the past undersampled or unpreserved? With the richness hotspot of the Western Pacific clearly undersampled, paleontologists must turn in that direction with an eye toward description before we can know how much richness is unpreserved.

Acknowledgements

This work would not have been possible 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, J. Wilson, G. Smith, W. DiMichele, M. Foote, and two anonymous reviewers 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. Puren and M. Petrie for continued encouragement. Funding support was provided by National Science Foundation grants (DEB1036393, EAR 0824793), the National Geographic Society (NGS 8505-08), and a Turner award from the University of Michigan Department of Earth and Environmental Sciences.

Literature Cited

- 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.
- 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.
- Bromley, R. 1990. *Trace Fossils: Biology and Taphonomy*. UnwinHyman. London, xi.
- Carroll, L. 1917. *Through the looking glass: And what Alice found there*. Rand, McNally.
- Clark, A. H. 1915. A monograph of the existing crinoids. *Bulletin - United States National Museum* 82. Washington :Smithsonian Institution Press.
- 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.

- 414 Folk, R. L., and R. Robles. 1964. Carbonate sands of Isla Perez, Alacran Reef Complex, Yucatan.
415 The Journal of Geology 72(3):255–292.
- 416 Fraley, C., and A. E. Raftery. 1999. MCLUST: Software for model-based cluster analysis. Journal of
417 Classification 16:297–306.
- 418 Frontier, S. 1976. Étude de la décroissance des valeurs propres dans une analyse en composantes
419 principales: Comparaison avec le moddle du bâton brisé. Journal of Experimental Marine
420 Biology and Ecology 25:67–75.
- 421 Gislén, T. 1934. A reconstruction problem, analysis of fossil comatulids from N. America, with a
422 survey of all known types of comatulid arm-ramifications. Acta Universitatis Lundensis, 30:
423 1-59
- 424 Gower, J. C. 1975. Generalized procrustes analysis. Psychometrika 40:33–51.
- 425 Hemery, L. G., M. Roux, N. Améziiane, and M. Eléaume. 2013. High-resolution crinoid phyletic
426 inter-relationships derived from molecular data. Cahiers de Biologie Marine 54:511–523.
- 427 Hemery, L., M. Eléaume, V. Roussel, N. Améziiane, C. Gallut, D. Steinke, C. Cruaud, A. Couloux,
428 and N. Wilson. 2012. Comprehensive sampling reveals circumpolarity and sympatry in seven
429 mitochondrial lineages of the Southern Ocean crinoid species *Promachocrinus kerguelensis*
430 (Echinodermata). Molecular Ecology 21:2502–2518.
- 431 Hess, H., C. G. Messing, and W. I. Ausich. 2011. Treatise on Invertebrate Paleontology, Part T
432 (Revised): Volume 3: Echinodermata 2: (Crinoidea). The University of Kansas,
433 Paleontological Institute, Lawrence, Kansas.
- 434 Howe, H. V. 1942. Neglected Gulf Coast Tertiary microfossils. AAPG Bulletin 26:1188–1199.
- 435 Jagt, J. W. 1999. Late Cretaceous-Early Palaeogene echinoderms and the K/Tboundary in the
436 southeast Netherlands and northeast Belgium - Part 2: Crinoids. Scripta Geologica 116:59–

- 437 255.
- 438 Janevski, G. A., and T. K. Baumiller. 2010. Could a stalked crinoid swim? A biomechanical model
439 and characteristics of swimming crinoids. *Palaios* 25:588–596.
- 440 Messing, C. G. 1994. Comatulid crinoids (Echinodermata) of Madang, Papua New Guinea, and
441 environs: diversity and ecology. *Echinoderms through time*. Balkema, Rotterdam:237–243.
- 442 ---. 1997. Living comatulids. *Paleontological Society Papers* 3:3–30.
- 443 Messing, C. G., D. Meyer, U. Siebeck, L. Jermiin, D. Vaney, and G. W. Rouse. 2006. A modern soft-
444 bottom, shallow-water crinoid fauna (Echinodermata) from the Great Barrier Reef,
445 Australia. *Coral Reefs* 25:164–168.
- 446 Meyer, D. L., and K. B. Meyer. 1986. Biostratigraphy of Recent crinoids (Echinodermata) at Lizard
447 Island, Great Barrier Reef, Australia. *Palaios*:294–302.
- 448 Moore, R. C. and R. M. Jeffords. 1968. Classification and nomenclature of fossil crinoids based on
449 studies of dissociated parts of their columns. *The University of Kansas Paleontological*
450 *Contributions* 46(9):1-86.
- 451 Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests.
452 2013. *vegan: Community Ecology Package*. <http://CRAN.R-project.org/package=vegan>.
- 453 Paleobiology Database. 2014. Accessed 24 June 2014. <https://paleobiodb.org/>.
- 454 R Development Core Team. 2008. *R: A language and environment for statistical computing*. R
455 Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- 456 Rasmussen, H. W. 1961. A monograph on the Cretaceous Crinoidea. *Biologiske skrifter*, 12(1).
- 457 Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal
458 superimposition of landmarks. *Systematic Biology* 39:40–59.

- 459 Rouse, G. W., L. S. Jermiin, N. G. Wilson, I. Eeckhaut, D. Lanterbecq, T. Oji, C. M. Young, T.
460 Browning, P. Cisternas, and L. E. Helgen. 2013. Fixed, free, and fixed: the fickle phylogeny
461 of extant Crinoidea (Echinodermata) and their Permian–Triassic origin. *Molecular*
462 *phylogenetics and evolution* 66:161–181.
- 463 Scoffin, T. 1992. Taphonomy of coral reefs: a review. *Coral Reefs* 11:57–77.
- 464 Sepkoski, J. J., R. K. Bambach, D. M. Raup, and J. W. Valentine. 1981. Phanerozoic marine diversity
465 and the fossil record. *Nature* 293:435–437.
- 466 Sieverts, H. 1933. *Drepanocrinus* Jaekel, ein Synonym von *Roveacrinus* Douglas, und ein neuer Vertreter
467 dieser Gattung aus der deutschen Kreide. *Zentralblatt Für Mineralogie, Geologie Und*
468 *Paläontologie* B:54–59.
- 469 Sokal, R. R., and F. J. Rohlf. 1995. *Biometry: the principles and practice of statistics in biological*
470 *research*. Freeman, New York.
- 471 Tukey, J. W. 1977. *Exploratory data analysis*. Addison-Wesley.
- 472 Vadasz, E. 1914. Regenerationserscheinungen an fossilen Echinoiden. *Centralblatt Für Mineralogie*
473 *Geologie Und Paläontologie etc.* 1914:283–288.
- 474 Vrieze, S. I. 2012. Model selection and psychological theory: a discussion of the differences between
475 the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).
476 *Psychological methods* 17:228.
- 477 Webster, G. D. 2003. *Bibliography and index of Paleozoic crinoids, coronates, and*
478 *hemistreptocrinoids, 1758-1999*. Geological Society of America, Boulder, Colorado.
- 479 Zelditch, M. L., D. L. Swiderski, and H. D. Sheets. 2012. *Geometric morphometrics for biologists: a*
480 *primer*. Academic Press.
- 481

Figure Captions

Figure 1 — A typical comatulid (A, after Clark 1915, plate 24) showing the lack of a stalk, allowing mobility. Homologous to a stalk is the centrodorsal (B-D), which serves to interface the arms and cirri. The centrodorsal is the largest single element in the comatulid, with a variety of shapes including conical (B, *Promachocrinus kerguelensis*, USNM 31356), pentameral and tall (C, *Psathyrometra fragilis* USNM 35784), to rounded and buttonlike (D, *Florometra mawsoni*, USNM 35902).

Figure 2 — Crinoid richness through time. Range-through crinoid generic richness compiled from several sources (Webster 2003, Moore & Teichert 1978, Hess et al. 2011). Paleozoic crinoid richness is far higher than that recorded in the fossil record of the post-Paleozoic, but Recent crinoid richness is similar to mean crinoid richness in the Paleozoic. This pattern results from a relatively tiny number of described comatulids in the fossil record, while post-Paleozoic non-comatulids seem to be well represented. When examined at the species level, the difference between Recent and fossil comatulid richness is exacerbated, to a 15x difference (Janevski & Baumiller 2010). It is not clear what causes this differential change in comatulid versus non-comatulid richness. One candidate explanation is the nearly exclusive use of centrodorsals to describe fossil comatulids.

Figure 3— Explanation of tests to identify taxonomic bias. Given a high richness of comatulids today (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 modeling, 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 finite mixture modeling.

Figure 4— Landmark diagram. Four landmarks and three curves were used in this study. Landmark one lies in the midpoint of a radial and centrodorsal. Landmarks two and three lie at bottom edge of last cirral scar along margin. Landmark four is at intersection of interr radial and centrodorsal. Semi landmark curves follow curvature of surface in plane as defined by landmarks.

Figure 5— 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 modern taxa. Welch two sample *t*-test is not significant for the test that fossil taxa have larger disparity than modern taxa ($p = 0.32$).

Figure 6— Results of finite mixture modeling 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 richness. Also pictured are thin-plate spline deformation grids of mean shape for each group in the best supported finite mixture modeling 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.

Figure 1

FIGURES

Figure 2

Figure 3
Method 1 - Intraspecific Disparity

Method 2 - Finite Mixture Analysis

Figure 4

Figure 5

Figure 6 Species

Fossil species

1 Table 1 – Specimens and sources. NBCN: Naturalis Biodiversity Center, Leiden, Netherlands;
2 MNMN: Muséum National d'Histoire Naturelle, Paris, France; NMNH: National Museum of
3 Natural History, Smithsonian, Washington D. C., U. S. A.; NMM: Natuurhistorisch Museum,
4 Maastricht, Netherlands.

5

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

6

7 Table 2: Taxonomic diversity vs. number of groups for genera in FMA.

	<u><i>Jaekelometra</i></u>	<u><i>Promachocrinus</i></u>		<u><i>Florometra</i></u>	
A priori species richness	2	1		2	
Sample size	<u>116</u>	<u>116</u>	<u>175</u>	<u>116</u>	<u>158</u>
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

8

9 Table 3: Summary of which taxa have more groups based on FMA. Summary of relationships that
 10 occur in 90%+ of bootstrap model runs chosen via BIC. Table 6 in the appendix shows the full data
 11 from which these results are calculated. *Promachocrinus* and *Florometra* appears to have higher richness
 12 than *Jaekelometra*. This is the opposite of what would be expected based on inclusion of two
 13 *Jaekelometra* species and one *Promachocrinus* species, suggesting that taxonomy is biased towards
 14 increased diversity in the fossil record.

15

	<u><i>Florometra</i></u>	<u><i>Promachocrinus</i></u>
<i>Jaekelometra</i>	$J. < F.$	$J. < P.$
<i>Promachocrinus</i>	$P. \geq F.$	-

16

17

18

19 Table 4: Optimal number of groups over 1000 bootstrap replicas scored by BIC, $n = 116$.

# of groups	1	2	3	4	5	6
<u>Genus</u>						
<i>Jaekelometra</i>	940	48	10	1	1	0
<i>Promachocrinus</i>	16	513	389	55	8	1
<i>Florometra</i>	2	873	118	7	0	0

20

21 Table 5: Optimal number of groups over 1000 bootstrap replicas scored by AICc, $n = 116$.

# of groups	1	2	3	4	5	6
<u>Genus</u>						
<i>Jaekelometra</i>	416	298	230	54	3	0
<i>Promachocrinus</i>	0	313	551	126	10	0
<i>Florometra</i>	0	520	412	65	3	0

22

23 Table 6: Equalities between fossil genera. This table shows the probability that the number of
 24 components in one taxon is equal, larger than, or smaller than the number of components in
 25 another taxon. High proportions indicate likely relationships.

26

<u>equalities</u>	Proportion of results	
	<u>BIC</u>	<u>AICc</u>
<i>Jaekelometra</i> = <i>Florometra</i>	0.045	0.253
<i>Jaekelometra</i> = <i>Promachocrinus</i>	0.044	0.227
<i>Promachocrinus</i> = <i>Florometra</i>	0.494	0.398
<u>inequalities</u>		
<i>Florometra</i> > <i>Jaekelometra</i>	0.944	0.575
<i>Florometra</i> > <i>Promachocrinus</i>	0.083	0.188
<i>Jaekelometra</i> > <i>Florometra</i>	0.011	0.172

<i>Jaekelometra</i> > <i>Promachocrinus</i>	0.008	0.121
<i>Promachocrinus</i> > <i>Florometra</i>	0.405	0.414
<i>Promachocrinus</i> > <i>Jaekelometra</i>	0.930	0.653

27

28

Appendix A

Use of FMM for GM data.— 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.

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 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 FMMs 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. Based on these results, it is recommended that the broken stick method be used as an initial guideline for number of PCs to use.

FMM is a powerful method, but care must be used when choosing data input into the FMM models. The assumptions inherent to any data set must be considered, as those assumptions strongly

bias how well various models fit that data set. GM data have properties that must be accounted for in analysis, such as number of dimensions and non-independence of landmarks due to superimposition. PCA rotates the data such that maximal variance is expressed along each successive principal component, thus concentrating information into fewer dimensions. However, each datum in a principal component data set obtained via GPA is necessarily dependent on the others, and 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. The diagonal class of models used by MCLUST searches for models that are in many ways similar to the resulting rotation from a PCA, which means that the single mixture diagonal model are unfairly benefited by the use of principal component data. Including this class of models, but accounting for the “free” parameters associated with the PCA rotation, results in rejection of those models. 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.

Model selection for FMM— Model selection methods such as Akaike’s information criterion, corrected for sample size (AICc) and Bayesian information criterion (BIC) have become more common in the paleontological literature but some may be unfamiliar. These methods seek to allow direct comparison between differing models, allowing models with differing numbers of parameters to be fairly compared. 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 FMM 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).

Appendix B

To begin with, FMM 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 BIC. AICc was not useful with some of the very small sample sizes, and as such wasn't included.

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*. A general trend discovered in testing was the median number of groups found during resampling procedures was often more than that found using just the original data. This has not been pursued further in this paper but is an interesting pattern that warrants further inquiry. It should also be noted that while a single model was strongly preferred on any given wrong ($\Delta \text{AICc/BIC} > 8$), the resampling procedures show much more variation in number of groups in supported models. A single Mclust model run using a single criterion such as BIC appears to be overly confident in supporting the best models. Using the range from AICc:BIC, or bootstrapping both appear to give

very reasonable error bars with good consistency to minor changes in data set or number of PC's used.

Appendix C

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 g_{max} is the maximum number of groups from bootstrapped model runs.

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

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

groups are the same for *Jaekelometra* and *Promachocrinus* = 0.043617