

1 Introduction

One of the main goals within paleontology is the elucidation of large-scale evolutionary patterns as well as gaining insights into the underlying mechanisms. There are three primary methodologies utilized to explore evolutionary dynamics through time: reconstructing phylogenetic relationships, compiling taxonomic biodiversity, and quantifying morphologic diversity. The first two metrics largely focus on one particular aspect of evolution, speciation. Phylogenetics attempts to reconstruct the splitting of taxonomic groups into discrete units. Although the data collected for phylogenetic reconstructions can be co-opted for other purposes, the primary aim is identifying the relationships between taxa. Biodiversity is driven by the rate of origination of new taxa as well as the rate of extinction. Obviously, biodiversity is a complex metric that is biased by a number of different factors but it is largely capturing the splitting and truncation of evolutionary lineages. The quantification of morphological diversity (i.e. disparity), captures many aspects of evolutionary change including speciation, the degree of change during speciation, and the amount of anatomical change within a lineage. In addition, the metric captures extinction events as well as the properties of extinction, such as selectivity towards particular morphological forms or features. Given the amount of information contained within this metric, studies examining disparity are work intensive and time consuming, there are many potential biases, and the resulting patterns are often difficult to interpret. However, this methodology provides a broad and encompassing view of evolutionary change within and between lineages through time.

The quantification of organismal form has long been recognized as a valuable tool to explore evolutionary dynamics (Thompson, 1917; 1942). However, the mathematical exploration of anatomy only became commonplace with the rise of computational ability given the inherent mathematical complexity of organismal form (Raup, 1962). These computational methods of quantifying organismal form were first applied to fossil organisms in a series of papers describing the coiling properties in mollusks including snails and ammonoids (Raup, 1962; Raup and Michelson, 1965; Raup, 1966; Raup, 1967). Raup and Michelson (1965) quantified the geometry of shell coiling based on four distinctive properties and constructed a multidimensional space of theoretical forms. This morphospace of coiling shells could then be described in terms of both theoretically plausible forms as well as those observed in nature, allowing the formulation of hypotheses on the factors limiting the realized forms within morphospace. The areas occupied by different taxonomic groups were plotted to examine the different constraining factors and shared evolutionary trajectories between related taxa (Raup, 1967).

In the following decades, paleontologists increasingly stressed the importance of detailed examinations of the diversity of organismal form in addition to biodiversity. Initially, studies attempted to explore disparity by contrasting the biodiversity at multiple taxonomic ranks, that is, using class or phylum diversity as a proxy for disparity. The use of taxonomic rank as a proxy for disparity was applied broadly by Valentine (1969), exploring the taxonomic and ecological structure of the marine benthos through the Phanerozoic. Valentine (1969) found an inverse relationship between higher- and lower-order taxonomic diversity in marine organisms, which he attributed to an increased ecological specialization. This specialization canalized morphology and prevented large departures and morphological innovations later in the Phanerozoic. Yochelson (1978; 1979) used class-level diversity as a proxy for disparity, interpreting the origin of new classes as the appearance of major anatomical changes. These morphological shifts thus enabled adaptive radiations into new ecological habits. Following this, Jaanusson (1981) suggested that morphological innovation occurred stepwise, with each advancement crossing functional thresholds enabling progressive diversifications. Jaanusson (1981) provided multiple examples, such as shifts in brachiopod dentition, which qualitatively suggested a disjunction between taxonomic and morphological diversity and highlighted the importance of disparity in studies of macroevolution. Runnegar (1987) further defined disparity as ‘the amount of difference between related phyla, classes, species, individuals, proteins, genes etc’ (p. 41). Furthermore, Runnegar (1987) suggested that disparity was best explored qualitatively to capture the innovative changes, such as those discussed by Jaanusson (1981).

Alternatively, many studies attempted to apply quantitative approaches to estimations of disparity. Overall, these methods lagged behind taxonomic-based approaches for many reasons. There are many features within an organism that could potentially be quantified and several different metrics to attempt to capture the range of forms being characterized. For instance, Cherry et al. (1982) explored disparity within 184 vertebrate taxa based on a relatively small set of linear measurements. They found an equitable degree of morphological variation within genera of amphibians, lizards, and mammals. However, the equivalence of taxonomic level and morphological variation broke down at higher taxonomic ranks, which questions the direct use of class or phylum diversity as a meaningful proxy of morphological diversity (Cherry et al., 1982). Therefore, the need for quantifying morphology independent of relying on diversity of higher taxonomic biodiversity became apparent. This quantification of body plans was popularized with the re-description of the Burgess Shale fauna and its importance in understanding the evolution of animal body plans. Gould (1989; 1991)

wrote extensively regarding the hypothesis that disparity followed a very different pattern from diversity through the Phanerozoic and proposed an early peak (Cambrian) in diversity of body plans. Gould (1991, p. 441) stated ‘the claim for greater early disparity cannot be confidently established until we develop quantitative techniques for the characterization of morphospace and its differential filling through time’. This work spurred the large-scale collection of morphological data to test these proposed patterns of macroevolution.

Given their diversity and unique morphologies within the Burgess Shale, arthropods were often used to test the idea of an early peak in morphological disparity. Briggs et al. (1992) constructed a data matrix of 134 characteristics to analyze Cambrian and modern arthropods and found equitable levels of disparity. Based on this, they concluded that the view of Cambrian disparity was clouded by an artefact of taxonomy and the pull of the unusual. However, they also noted that the filling of the morphospace occurred rapidly, such that the Cambrian Explosion was dampened but still a pivotal event in the history of life (Briggs et al., 1992). This work was followed by a similar study on priapulids, in which Wills (1998) found a decrease in disparity between the morphologically distinctive (i.e. non-overlapping in morphospace) Cambrian and modern worms. Again, this both strengthened the hypothesis of an initial burst of morphological innovation in the Cambrian and challenged the decimation of disparity through time (Wills, 1998).

The focus of disparity studies on the Burgess Shale fauna following the publication of Gould’s *Wonderful Life* (1989) de-emphasized the importance of echinoderms because of their relative scarcity within the Burgess Shale fauna (Sprinkle and Collins, 2006; Zhao et al., 2010). However, echinoderms soon became model organisms for the study of morphological disparity. Echinoderms are ideal subjects for studies in disparity in that they were and are ecologically and taxonomically diverse, in addition to the expansive range of morphological features and body plans within the phylum (Paul and Smith, 1984). Importantly, most echinoderms are highly skeletonized, which produces fossils that are character-rich with potentially relatively little external morphology lost from taphonomic processes compared to other phyla (Brett et al., 1997; Deline and Thomka, 2017). Finally, echinoderms have an incredible richness of higher-order taxonomic groups in the early Paleozoic compared to today (Sumrall and Wray, 2007). This early burst of body plans places echinoderms as a group most likely to conform to the initial peak in disparity hypothesis.

Early studies of echinoderm morphological disparity confirmed the diversity and complexity of the phylum. Foote (1991) compiled a geometric morphometric data set of Paleozoic blastoids and found a significant disconnect between

taxonomic and morphologic diversity. Blastoids steadily increased in disparity from the Silurian through the Permian, highlighting their continued morphological plasticity and capacity for innovation. Overall, this provided a counterargument to the patterns seen in arthropods and priapulids with Permian blastoids occupying a broader range of morphospace right before the extinction of the clade (Foote, 1991). Foote (1992) then explored blastozoan evolution by constructing a discrete character matrix. This study showed that the ratio of disparity to diversity was highest in the Cambrian, but disparity grew steadily through the early Paleozoic peaking in the late Ordovician (Foote, 1992). This corroborated the hypothesis of an initial explosion of disparity in the Cambrian paired with continued evolutionary innovation within the clade. These early studies showed the potential for morphological studies within echinoderms as well as the capacity for continued morphological exploration and innovation within echinoderms. These studies also stressed the need to examine disparity beyond the Cambrian in many different echinoderm groups, using different methodologies, and across variable taxonomic scales (Foote, 1997).

Over the past 30 years, studies of echinoderm disparity have stretched across the phylum to explore different methods for estimating trends in disparity, potential biases in the quantification of morphology, and how taphonomy can alter these perceived trends. Studies of echinoderm disparity have highlighted the rapid initial morphological diversification within clades but also the importance of continued constraint through time, particularly following mass extinctions or during faunal turnover events. In addition, multiple studies have explored the underlying developmental and biologic factors enabling and constraining morphology through time. Deline et al. (2018) placed echinoderms within a broader framework of metazoan morphology. This study used the work of Ax (1996; 2000; 2003) to construct a morphospace of extant metazoans, which was then expanded to include a snapshot of the diversity of Cambrian animals. Overall, echinoderms cover a small area within metazoan disparity (Figure 1A). Even though sampling was correlated with genus-level taxonomic diversity, echinoderms were relatively undersampled, thus reducing their apparent morphological importance. This was further emphasized by focusing on the modern (six clades) and Cambrian (nine genera), neither of which highlight the peak in echinoderm body-plan diversity in the Late Ordovician (e.g. peak in class-level diversity). Furthermore, Deline et al. (2020) compiled an expansive data set of early Paleozoic echinoderm morphology recovering four major body plans during the initial explosion of echinoderm morphology (Figure 1B). In addition, many studies have explored morphological patterns within echinoderms at the class or subclass level (e.g. Lefebvre et al., 2006; Deline et al.,

Echinoderm Morphological Disparity

5

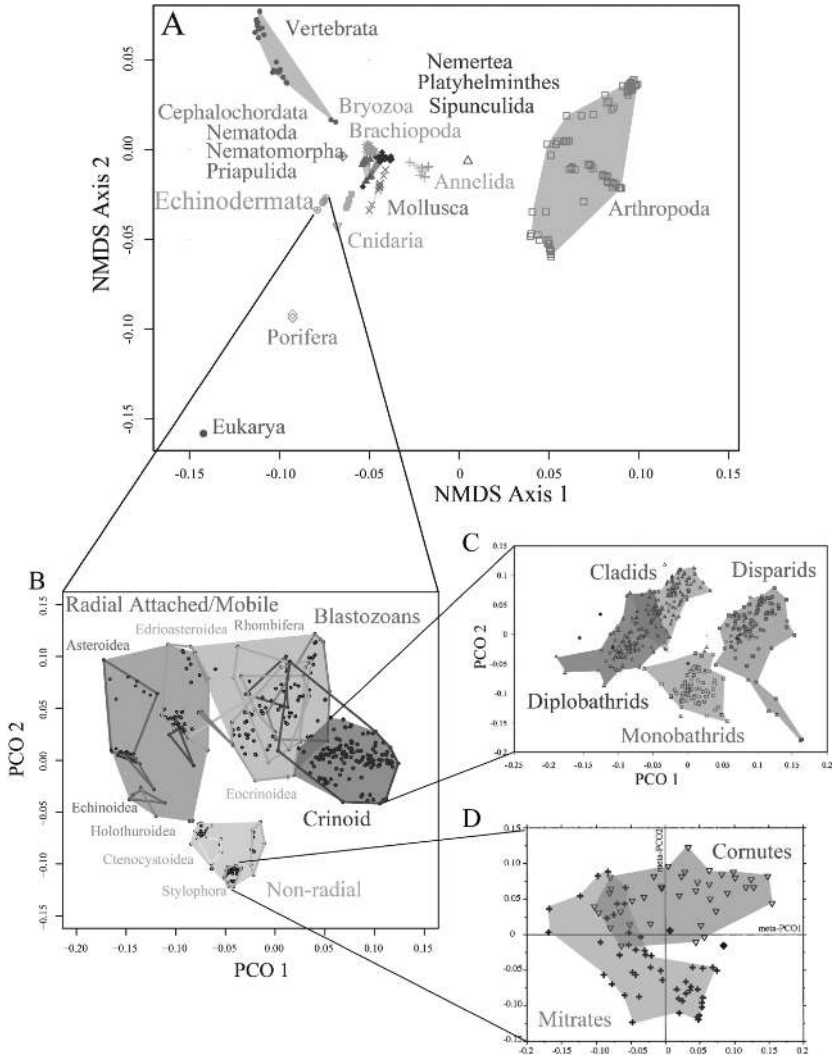


Figure 1 The quantification of echinoderm morphology at multiple taxonomic scales from kingdom (metazoans) to phylum to class (crinoids and stylophorans). Morphospaces modified from Lefebvre et al. (2006), Deline et al. (2012), Deline et al. (2018), and Deline et al. (2020).

2012; Figure 1C, D) documenting detailed patterns of morphological change through time. These studies across different taxonomic levels allow an expansive perspective of echinoderm morphological evolution within and beyond the phylum.

Foote (1997) reviewed the progress of studies of morphological disparity and highlighted future directions. Studies of echinoderm disparity have helped to fill

some of the prominent gaps in our knowledge. However, there are still ample opportunities to utilize morphology to better understand the evolution of echinoderms. The current review aims to explore the diverse methods that have been utilized to quantify morphological evolution in echinoderms as well as the choices and biases that alter perceived trends in disparity through time. Given the diversity of methods and experimental designs, several prominent patterns have been repeatedly found which will be discussed along with their important developmental and macroevolutionary implications. Finally, promising directions of study utilizing echinoderm morphology will be highlighted.

2 Methods of Quantifying Morphology

The quantification of morphology and estimates of disparity can be accomplished using multiple techniques (Hopkins and Gerber, 2017). Broadly, these methods can be broken down into two major groups: morphometric approaches utilizing continuous measures and categorical approaches utilizing discrete characters.

Morphometric Methods

Morphometric approaches have the benefit of being more intuitive in that the metrics directly correspond to easily visualized shape parameters. Common morphometric methods include the direct measurement of features (i.e. traditional morphometrics), comparisons of the positions of distinctive landmarks (i.e. geometric morphometrics), or the quantification of the overall outline of features or the entire organism (Webster and Sheets, 2010). The choice of the morphometric methodology depends on the organism, taxonomic scale of study, and hypothesis being addressed. Obviously, the level of anatomical detail that can be captured decreases with an increase in taxonomic scope of study. In addition, certain body plans lend themselves to specific methodologies. For example, the overall body of flattened taxa such as stylophorans or cinctans could be successfully characterized with outline or landmark analysis. Overall, all of these methods have been utilized in the study of echinoderm disparity.

Lefebvre et al. (2006) explored the morphological diversity within stylophorans using a traditional morphometric approach of directly measuring features. This group contains a wide array of characteristics and large variability in overall body plans. Therefore, this is an enticing group to explore morphologically, but the puzzling anatomy makes quantification difficult. The low profile of the theca allowed a two-dimensional quantification of the overall body shape as well as the geometry of individual plates. The direct comparison of individual thecal elements requires a study at lower taxonomic rank and careful analysis of

potential plate homologies across the group. To accurately quantify shape, they applied measurements such as circularity of the theca and individual plates as well as the relative area of individual plates or features compared with the area of the entire theca. Using ratios of features effectively removed body size from the analysis to obtain a clearer measure of the overall form. All told, this produced over 70 variables for both the upper and lower surface of the theca that accurately quantified body shape as well as the constituent pieces, which would not have been possible with other methods such as Procrustes-based landmark analysis (Lefebvre et al., 2006).

Landmark-based geometric morphometrics has been utilized many times to differentiate species, assess ontogenetic change, or document changes in disparity through time. As with all of the methodologies discussed, the a priori choices are pivotal to the results of the study. Ideally, each individual landmark should represent an easy-to-recognize homologous point (Figure 2). Within echinoderms, the clearest landmarks would appear at plate junctions, such that the placement of the landmark is accurate and reproducible. This level of anatomical similarity across samples again requires a lower taxonomic breadth of study, but this methodology allows for the easy visualization of the forms

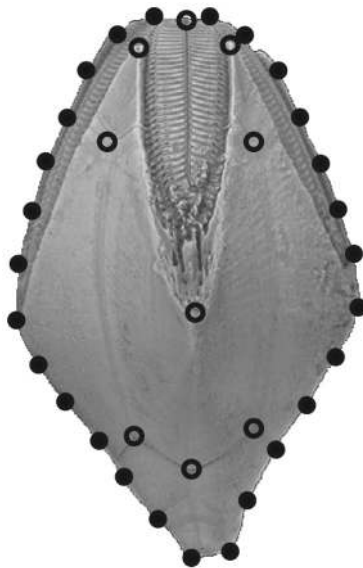


Figure 2 A comparison between landmarks defined by specific plate junctions (open circles) as used by Foote (1992) and semi-landmarks used to define a thecal outline (closed circles) as used by MacLeod (2015) shown on the blastoid *Pentremites meganae* (Atwood and Sumrall 2012).

being studied. Foote (1991) used this approach to study disparity in blastoids through the Paleozoic. This methodology is well-suited for this clade in that the reduction in thecal plates makes the recognition of plate homology easier. Landmarks were selected across the theca at plate junctions from three different perspectives: basal, oral surface, and theca profile. These landmarks were standardized to eliminate differences in size to exclusively compare the three-dimensional shape of the theca. Eble (2000) used similar methods to explore differences in disparity between sister clades of atelostomate echinoids. In this case, landmarks were chosen across the apical system, ambulacra, body outline, as well as the peristome and periproct. Similar to the blastoid study, this provided a three-dimensional geometric perspective of the echinoid test independent of size. In addition, the use of mostly non-plate junction landmarks allows for a broader taxonomic comparison. Importantly, Eble (2000) notes that while this methodology captures significant aspects of echinoid morphology, it is not exhaustive. Aspects of the spine tubercles, overall plating, and ambulacra pore structures were not characterized. Obviously, all of these methods quantify morphology based on a subset of features, but it is worthwhile to consider whether the patterns being characterized are representative of the overall morphology or exclusively the features being examined. Studies in other taxonomic groups (i.e. trilobites) have indicated that morphology of one body region can be used as a proxy for overall morphology (Hopkins, 2017), but that is likely far from universal and needs to be further tested in other taxonomic groups (see Deline and Ausich, 2017).

In many echinoderm groups, the body plating is extraordinarily variable such that the recognition of homologous points might be limited to the body openings, which would be morphologically uninformative. In these cases, methods characterizing the body shape or outline might be more applicable than the identification of distinctive landmarks. Characterizing the shape of a body or feature can be accomplished through tracing the feature, defining the shape through Fourier analysis, or by placing semi-landmarks (Figure 2) around the outline that can be compared between specimens (Webster and Sheets, 2010). MacLeod (2015) used a combined landmark and outline analysis method to characterize morphology within the eocrinoid *Gogia*. Gogiid eocrinoids are difficult to characterize morphologically because of a lack of distinctive and homologous thecal plates and a propensity for severe taphonomic alteration (e.g. crushing). MacLeod (2015) characterized gogiid morphology using four ‘distinctive landmarks’ that defined the aboral cup as well as a series of semi-landmarks that defined the outline of the lower theca. Although there are significant issues documenting standard orientations in gogiids that lack perfect radial symmetry, this method could potentially capture the shift in body outline

from the stalk to the theca in a broad sense. Using these landmarks separately and in combination, MacLeod (2015) suggested he was able to extract biologically meaningful information from the characterized individuals such as ontogenetic shifts in thecal morphology. Although more research is warranted, this study highlights that morphometrics can potentially be applied broadly, even within groups that would initially appear to be poor candidates for this type of analysis. However, as with other morphometric techniques, the taxonomic breadth can be fairly limited in that thecal or other outlines constructed with different underlying plates would not be a meaningful comparison.

In addition to these primary methods, there are other methodologies that could yield promising results in the characterization of morphology and assessing patterns of disparity through time. Echinoderm skeletons that grow through a combination of plate growth and addition allow mathematical modelling of growth and development (Zachos and Sprinkle, 2011). Detailed characterization of different growth models allows for the potential quantification of developmental morphology assessing a different aspect of body form. A similar approach was utilized by Hoyal Cuthill and Hunter (2020), that quantified crinoid calyx morphology based on graph theory. This method enables the compilation of theoretical forms and their structural implications, thus providing a way of assessing the functional implications of complex plating patterns (Hoyal Cuthill and Hunter, 2020). Geometric form can also be characterized in much greater detail with the construction of three-dimensional digital models with GIS-based methodologies (e.g. Sheffield et al., 2012) or x-ray tomography (e.g. Zamora et al., 2012). However, the increased level of morphological detail in these methods often comes with a greater financial cost (e.g. specialized software, increased computation requirements, or equipment) as well as requiring significantly more time and effort, which would severely limit the scope of studies utilizing these techniques.

Character-Based Methods

If the aim of the study is broader than can be achieved through the use of morphometric methods, or if taphonomic alteration makes those methods inappropriate, morphology can then be quantified through the use of discrete characters. Ideally, characters are chosen to cover all aspects of the organism's morphology, including convergent traits and autapomorphies. Many recent studies have utilized cladistic data sets for this purpose, but most character suites constructed for that purpose focus on phylogenetically important traits and intentionally avoid the confounding influence of homoplastic traits or

phylogenetically uninformative autapomorphies. The exclusion of these aspects of overall morphology would then create similar issues to those discussed earlier in regard to landmarks, potentially skewing the morphological patterns away from whole-organism characterization. Although, comparing character states captures aspects of morphology, especially at broader taxonomic scales, the meaning of the distances between organisms becomes murkier. With morphometrics, the resulting differences are easy to visualize and potentially equitable in the aspects of morphology they capture, but each discrete character is unique in terms of morphology being described. In addition, the genetic or developmental processes leading to the changes in geometric shape are potentially more direct than with the appearance of a novel trait, a change in character state, or shifts in characters broadly describing shape. Therefore, the distances between taxa using discrete characters should be seen as amalgamations of vastly different types of characters with unclear relationships to the underlying developmental mechanisms. That in no way invalidates the use of discrete characters but it necessitates caution in the interpretation of the constructed morphospaces in that the meaning of distances are then more abstract and likely to have affine rather than metric properties (Mitteroecker and Huttegger, 2009; Huttegger and Mitteroecker, 2011).

Foote (1992) constructed a character suite describing morphology within blastozoan echinoderms. Characters were chosen without regard to their known or presumed phylogenetic importance. However, he notes that there is often significant overlap between characters used for the purposes of capturing disparity and deciphering phylogenetic relationships. In total, Foote (1992) used 65 binary and multistate characters that described features across the entire body from attachment structures to characteristics of the brachioles. In coding the characters, Foote (1992) focused on topological position rather than homology to attempt to avoid an overt phylogenetic signal. Obviously, this choice has ramifications in the resulting patterns of disparity. If features like stalks evolve multiple times and were thus coded as independent features, that would increase the estimation of disparity, even though the two stalks are similar topologically and functionally such that they should occupy similar areas of morphospace. This again highlights the importance of caution in the straightforward use of cladistic data sets for the purpose of describing disparity.

Following this work, Foote (1994a; 1994b; 1995a; 1995b; 1999), in a series of papers, explored the morphology of crinoids throughout the Phanerozoic. Even though this is a lower taxonomic group, the high biodiversity, long geologic history, and diversity of forms within crinoids presented unique challenges. The wide array of forms within crinoids presents the dichotomy of being an ideal group to study patterns of disparity while at the same time