Why should we investigate the morphological disparity of plant clades?

Jack W. Oyston, Martin Hughes, Sylvain Gerber, and Matthew A. Wills

**Background** Disparity refers to the morphological variation in a sample of taxa, and is distinct from diversity or taxonomic richness. Diversity and disparity are fundamentally decoupled; many groups attain high levels of disparity early in their evolution, while diversity is still comparatively low. Diversity may subsequently increase even in the face of static or declining disparity by increasingly fine sub-division of morphological ‘design’ space (morphospace). Many animal clades reached high levels of disparity early in their evolution, but there have been few comparable studies of plant clades, despite their profound ecological and evolutionary importance. This study offers a prospective and some preliminary macroevolutionary analyses.

**Methods** Classical morphometric methods are most suitable when there is reasonable conservation of form, but lose traction where morphological differences become greater (e.g. in comparisons across higher taxa). Discrete character matrices offer one means to compare a greater diversity of forms. This study explores morphospaces derived from eight discrete data sets for major plant clades, and discusses their macroevolutionary implications.

**Key Results** Most of the plant clades in this study show initial, high levels of disparity that approach or attain the maximum levels reached subsequently. These plant clades are characterized by an initial phase of evolution during which most regions of their empirical morphospaces are colonized. Angiosperms, palms, pines and ferns show remarkably little variation in disparity through time. Conifers furnish the most marked exception, appearing at relatively low disparity in the latest Carboniferous, before expanding incrementally with the radiation of successive, tightly clustered constituent sub-clades.

**Conclusions** Many cladistic data sets can be repurposed for investigating the morphological disparity of plant clades through time, and offer insights that are complementary to more focused morphometric studies. The unique structural and ecological features of plants make them ideally suited to investigating intrinsic and extrinsic constraints on disparity.

**Key words:** Disparity, Embryophyta, morphological diversity, morphospace, angiosperms, conifers, ferns, macroevolution, clad shapes, developmental robustness, species diversity.

**INTRODUCTION**

The number of species within higher taxa, or within clades of a similar age (Magallon and Sanderson 2001), is hugely variable, even for sister groups diverging (by definition) at the same time. While rates and patterns of extinction are clearly influential, some clades appear much more adept at sub-dividing niche space and speciating than others, even in comparison with their closest relatives. Some groups foster enormous radiations in diversity despite maintaining conservative bodyplans and displaying only modest morphological variety relative to that in their parent clades. Insects, as the best example, have a highly constrained body organization (a fixed number of appendages and tagmata) relative to other groups of arthropods (cf. crustaceans and branchiopods in particular), yet constitute over half of all described arthropod species (Mayhew, 2007). Similarly, beetles display remarkably conservative organization within insects, despite their notoriously high contribution to global species richness (Erwin, 1997). There is no necessary relationship, therefore, between the number of species within a group (species richness or diversity) and its morphological diversity.

Indeed, there are suggestions that a constrained and entrenched bodyplan might actually be conducive to higher diversity (Rabosky, 2012).

In order to study the relationship between species richness and bodyplan conservation, we need to quantify both diversity and morphological variety or disparity for large groups. Methods for studying diversity are well established (Peet, 1974; Gotelli and Colwell, 2001; Benton, 2009; Ezard et al., 2011; Mayhew et al., 2012), but approaches for quantifying disparity are less familiar, particularly in the botanical literature (Chartier et al., 2014). While it is possible and informative to study diversity and disparity across clades within the extant biota (or, indeed, in any time slice), insights into the dynamics of their interaction are most fruitfully gained by investigating the trajectories of clades throughout their evolution. Most studies to date have focused on animals (Foote, 1994, 1997; Moyne and Neige, 2007; Hughes et al., 2013), but the long evolutionary history (Wellman, 2014) and rich fossil record of land plants (embryophytes) make them ideally suited for comparison. Diversity patterns through time within vascular plants have been studied for many years, typically deriving from...
species-level compilations of origins and extinctions (Knoll et al., 1979; Niklas et al., 1980; Lidgard and Crane, 1990; Kovach and Batten, 1993; Cascales-Miñana and Cleal, 2012, 2014). Results have differed in some details (Niklas and Tiffney 1994), but are broadly consistent in showing (a) a radiation of pteridophytes and gymnosperms in the Late Devonian-Carboniferous; (b) a gymnosperm-dominated flora in the earlymid Mesozoic of comparatively constant diversity; and (c) a mid-late Cretaceous to Tertiary diversity increase, due primarily to the radiation of the angiosperms. The presence of novel morphological features within this group raised the question of whether phases of embryophyte diversification could be explained by the acquisition of ‘key innovations’ within angiosperms (Endress, 2001), seed plants (Rudall and Bateman, 2007) and early land plants (Bateman et al., 1998; Renzaglia et al., 2000). Advances in plant phylogenetics have revealed that the timing of many plant radiations does not match the first appearance of hypothesized innovations (Sanderson and Donoghue, 1994; Davies et al., 2004; Vamosi and Vamosi, 2010), implying instead that the evolution of suites of characters over an extended period of time may enable diversification (Donoghue, 2009). The hunt for specific drivers has shifted to focus either on competitive interactions, for example between plants and herbivores (Agrawal, 2007; Futuyma and Agrawal, 2001), or on environmental factors such as climatic change (McElwain et al., 1999; Beerling et al., 2001; Willis and Niklas, 2004; Beerling and Berner, 2005; Feild and Arens, 2007; Boyce et al., 2009; Willis and McElwain, 2013).

In marked contrast to diversity, for which temporal patterns have been investigated for many years (Crane et al., 1994; Kenrick and Crane, 1997; Lupia, 1999; Schneider et al., 2004; Soltis and Soltis, 2004; Crisp and Cook, 2011; Donoghue, 2008), there have been only a handful of studies on the morphological disparity of plants (Lupia, 1999; Niklas, 1999; Chartier et al., 2014). Often these studies have focused on a specific aspect of plant morphology such as leaves (Boyce and Knoll, 2002; Boyce, 2005) or vascular systems (Wilson and Knoll, 2010; Feild et al., 2011). Disparity analyses have furnished an important means of assessing macroevolutionary patterns in animals for some years, and we believe that their application to plants would be equally insightful.

**Aims**

This paper has two primary aims. The first is to provide an overview of the methods used to quantify morphological disparity, with particular emphasis on their application to plant evolution. We contrast concepts of disparity with those of diversity or species richness, and explain how exploring both trajectories through time can shed light on the evolutionary dynamics of clades. Morphological disparity is usually quantified with reference to the axes of some form of morphospace; an n-dimensional space in which the distances between species or other operational taxonomic units are proportional to some measure of the morphological distances between them. We therefore distinguish between theoretical and empirical morphospaces and discuss their relative advantages and disadvantages for the study of plants. We also explore a variety of potential data sources and consider their relative merits. Particular emphasis is given to character-based empirical methods, which have proved broadly applicable to animal clades at a wide range of taxonomic levels (Hughes et al., 2013), but have yet to be utilized in plants. The second objective is to demonstrate the application of these methods to a select number of published character matrices for major plant groups. We compare and contrast the observed patterns of disparity through time with those seen in animals, and offer a prospectus for future studies of plant disparity.

**What is disparity and why should we study it?**

The macroevolution of any major clade through deep time can be characterized in a number of ways. There is perennial interest in how diversity changes (Sepkoski et al., 1981; Sepkoski, 1997; Sepkoski and Miller, 1998), particularly with regards to how species and higher taxonomic richness responds to major physical or biotic changes such as mass extinctions, the opening up of new habitats or the origination of other major groups. Equally fundamentally, we may wish to know how the constituent taxa of a clade are related, and may use phylogeny to better inform the patterns above. Increasingly, however, palaeobiologists are also focusing on the manner in which groups diversified morphologically to give rise to new body-plans or architectures (Fortey et al., 1996). The range or variance of morphological form across species or other taxa is usually referred to as ‘morphological variety’, ‘morphological disparity’ or simply ‘disparity’ in context. Disparity is therefore a property of a sample of taxa rather than of individual species, and is also measured relative to some set of quantifiable variables. Trajectories of disparity through time are often different from patterns of species and higher taxonomic diversity, and are also difficult to predict from phylogeny.

Although all morphological variety is generated within the context of a phylogeny, diversity and disparity are fundamentally decoupled (Foote, 1991; Fortey et al., 1996, 1997; Moyne and Neige, 2007). Large samples of morphologically very similar species typically have much lower disparity than small groups of morphologically highly dissimilar species. Specifically, numerous basal groups of animals show levels of disparity greater than or equal to their more diverse, derived counterparts (Fig. 1) (Foote, 1992, 1994, 1997; Wills et al., 1994; Wills, 1998) although exceptions exist (Benson et al., 2012). At a coarse level, higher taxonomic diversity (e.g. numbers of orders or classes) tends to be a better proxy for disparity than numbers of species or genera (Foote 1990). Plots of relative disparity through time are therefore often used alongside plots of diversity in order to understand the dynamics of clade evolution more fully.

Much of the initial impetus for quantifying levels of disparity came from claims about the evolutionary significance of the fossils from the Middle Cambrian Burgess Shale (Whittington, 1985; Conway Morris, 1989). In particular, it was claimed (Gould, 1989) that the range of morphological variety amongst Cambrian arthropods was far greater than that realized at any time subsequently; an argument couched (at least initially) in the perceived higher taxonomic status (i.e. sub-phylum or class) of many Burgess Shale genera. Gould (1991) subsequently pronounced an ‘inverted iconography’ model for the evolution of
life. An initial phase of experimentation and looser constraint on bodyplan evolution was posited to yield early maximal disparity, followed by a phase of winnowing in which most bodyplans were lost and the survivors consolidated and canalized. Subsequent evolution would typically yield few new bodyplans, but would see increases in diversity, i.e. increasing numbers of variations (species) upon a more limited number of constrained themes. However, empirical studies of marine invertebrates found that the disparity of Cambrian and Recent faunas was essentially equivalent (Briggs et al., 1992; Wills et al., 1994, 2012; Fortey et al., 1996) (Fig. 2). Subsequent studies have examined the disparity of clades at numerous successive time intervals, often demonstrating relatively high early disparity even while diversity is low (Foote, 1992, 1994; Wills, 1998). Recently, this approach has been applied to a larger data set of exclusively fossil animal clades (Hughes et al., 2013). The shape of the disparity profile of a clade through time can be summarized as a centre of gravity index (CG). Clades with precisely symmetrical patterns through time have indices of 0-50, those with higher levels of disparity early in their history have values <0-5 (bottom heavy), while those peaking late tend to >0-50 (top heavy). In a sample of 98 extinct clades that did not go extinct coincident with one of the ‘Big Five’ (Hallam and Wignall, 1997; Bambach, 2006) mass events, there was a significant bias towards bottom heaviness and early high disparity. Groups persisting to the present tend to have top-heavy profiles; not least because they are artificially truncated by the Recent. Those disappearing coincident with one of the ‘Big Five’ mass events also tend to be top heavy, and for similar reasons.

Other research agendas have become increasingly important within particular clades. One is the extent to which bodyplans are modular, and comprise units within which changes are relatively tightly correlated, but between which there is greater flexibility (Klingenberg et al., 2004; Monteiro and Nogueira, 2010; Cooper et al., 2010; Drake and Klingenberg, 2010). Another is the extent to which developmental vs. environmental factors constrain bodyplans over evolutionary time (Allen et al., 2008; Anderson et al., 2011). Increasingly, there is also interest in quantifying functional disparity, notably in fish and basal tetrapods (Friedman, 2010; Anderson et al., 2013).

Why study the disparity of plants?

In contrast to animals, there have been few studies investigating the morphological disparity of plant clades. It is possible that the patterns in plants may differ from those in animals; both the trends observed in statistical samples of clades, and the overall pattern of disparity through time for the group as a whole. In this latter context, it may be informative to compare plots of ordinal diversity through time (compiled from Benton, 1993), insofar as counts of higher taxa afford a very rough approximation to disparity (Fig. 3). Animals reach relatively high levels of ordinal diversity relatively early in their history; commensurate with the patterns revealed in explicit studies of disparity. The pattern observed in vascular plants differs markedly. Even accounting for the much later origin of vascular plants compared with animals, plants show a much more...
gradual increase in ordinal diversity, reaching 50% of their maximum relatively late in their evolutionary history. Plants show ordinal diversity increases in three discrete phases: (1) the Late Devonian, corresponding to the initial radiation of pteridophytes and gymnosperms; (2) a smaller increase at the start of the Cretaceous, coincident with the appearance of the angiosperms; and (3) a Late Cretaceous increase, corresponding to the appearance of many modern angiosperm groups (Niklas and Tiffney, 1994).

Ordinal diversity profiles (Fig. 3) suggest that vascular plants have fewer fundamentally different modes of morphological organization than animals, and acquired novel bodyplans more gradually. Strikingly, plants appear to be relatively unperturbed by the mass extinction events that were catastrophic for animals; or at least the recovery of plants was rapid enough to mask any significant diversity decreases in the fossil record (Rees, 2002; McElwain and Punyasena, 2007; Cascales-Miñana and Cleal, 2014). Plants therefore appear to have greater resilience to certain types of ecological disturbance than animals (Cascales-Miñana and Cleal, 2012); a surprising inference given that many aspects of plant morphology are thought to be tightly mechanically and physiologically constrained to optimize photosynthetic efficiency and structural support (Niklas and Kerchner, 1984).

Even relatively simple optimization models with a small number of variables can produce the diverse spectrum of habits and gross phenotypes seen across plant groups (Farnsworth and Niklas, 1995; Niklas, 1999) (Fig. 4); ecological disturbance may actually serve as a driver for increasing phenotypic diversity.
There are many approaches to quantifying morphology (Moore and Moser, 1995; Chapman and Rasskin-Gutman, 2001; Friedman and Williams, 2003). The genetic controls on leaf shape (Langlade et al., 2005; Chitwood et al., 2014) and compound leaf structures are gradually being better understood (Klingenberg et al., 2012). Leaf shape appears to be correlated with shoot morphology (Lacroix et al., 2003; Jeune et al., 2006), although the importance of selective, functional and historical constraints in the evolution of these hierarchical systems is poorly understood (Burns et al., 2008). Floral morphology, despite having long been recognized as a critical component of angiosperm disparity (Stebbins, 1951), has received relatively little attention until recently (Whibley et al., 2006; Stournaras et al., 2013; Chartier et al., 2014). Similar considerations apply to the architecture of inflorescences (Prusinkiewicz et al., 2007). Other work has investigated the evolution and possible adaptive value of different types of pollen (Lupia, 1999; Ressayre and Godelle, 2002) as well as physiological properties in the conductive vessels of major seed plant groups (Wilson and Knoll, 2010). Rather than attempt to assess disparity from large collations of morphological data, more holistic approaches tend to consider habit and gross architecture (Niklas and Kerchner, 1984; Niklas, 1999; Silva and Batalha, 2011).

The decoupling of diversity and disparity within higher plant clades appears every bit as great as that within animal groups (Yu et al., 2014). For example, the true grasses (Poaceae) and the bromeliads (Bromeliaceae) are both families of angiosperms in the order Poales. However, the true grasses are represented by about 10 000 species (The Plant List, http://www.theplantlist.org/) of varying size but relatively limited floral disparity, while the bromeliads contain just over 3000 species but show huge variation in inflorescence morphology (Benzing, 2000; Sajo et al., 2004). It is clear that a complete picture of plant disparity cannot be captured by focusing exclusively on the disparity of specific structures (as there is strong scale dependence) or by using diversity as a proxy. Holistic approaches that use a broad suite of characters sampled over large numbers of taxa will probably constitute the best way of quantifying plant disparity at macro-evolutionary scales. Here, we take some preliminary steps in this direction for a sample of higher plant clades.

**Types of data**

There are many approaches to quantifying morphology (Moore and Moser, 1995; Chapman and Rasskin-Gutman, 2001;
Lockwood et al., 2002), and the most suitable usually depends upon the application and the question being addressed. Where the forms being compared are broadly similar (e.g. typically species within genera or families), a variety of morphometric approaches can be used to derive sets of continuous variables describing shape and shape change, usually with some implicit standardization for size and orientation (Rohlf and Marcus, 1993; Adams et al., 2004) (Fig. 5). Three-dimensional, landmark-based approaches operate by identifying biologically (or functionally) homologous points (e.g. intersections between homologous structures) across all of the species or higher taxa (hereafter ‘operational taxonomic units’ or OTUs) being compared (Marcus, 2000; von Cramon-Taubadel et al., 2007; Mitteroecker and Gunz, 2009). Outline-based methods describe shapes in more detail. This can be either using a more limited number of discrete points (homologous landmarks), possibly interspersed with semi-landmarks to specify the form further (Bookstein 1997; Perez et al., 2006), or using continuous functions (e.g. Fourier analysis) describing shape (Rohlf and Archie, 1984; Crompton, 1995). Where the forms being compared are more divergent (e.g. across higher taxa), it often becomes difficult to identify a sufficient number of homologous or functional landmarks to capture all but the most limited and conservative aspects of form variation (Bocxlaer and Schultheiß, 2010). Here, it is possible to use an array of discretely coded characters, each recognizing two or more alternative states, as descriptors of morphological variation (Wills et al., 1994; Wills, 1998). Such data are more flexible, but entail more assumptions and potential subjectivity concerning the selection and discretization of characters and states. The morphospaces that they define also have properties that differ from those derived from continuous character data (Gavrilets, 1999).

The first studies quantifying disparity explicitly with empirical data sets were published about 25 years ago (Foote, 1990, 1994; Briggs et al., 1992; Wills et al., 1994) (Fig. 6). The disparity profiles of numerous major animal clades were investigated over the next decade, before a wane in apparent interest. The last few years, however, have seen the resurgence of empirical studies, with a particular emphasis on the use of discrete character data sets (see Supplementary Data Table S1). As a general rule, metazoan clades tend to show an initial rapid increase in disparity, with early levels of disparity being at or close to the maximum levels observed throughout the group’s history.

### Biological homology and functional analogy

With all types of data, a distinction can be drawn between those approaches that attempt to capture variation in biologically homologous aspects of morphology (Rohlf, 2002; Klingenberg et al., 2004), and those that are more concerned with the functional parameters of shape (Nogueira et al., 2009; Figueirido et al., 2011; O’Higgins et al., 2011; Anderson et al., 2011, 2013). Morphological disparity can be used to refer to both aspects of variation in form, although the intention is sometimes unspecified (Love, 2007). The distinction can be illustrated with reference to the tails of derived sharks and ichthyosaurs, both of which have convergently evolved dorsal and ventral lobes with a relatively high aspect ratio for high-speed aquatic locomotion (Motani, 2002; Lingham-Soliar, 2005; Lingham-Soliar and Plodowski, 2007). In functional terms, the dorsal lobes of both groups are comparable, as are the ventral lobes. However, the vertebral column of sharks extends into the dorsal tail lobe, while that of ichthyosaurs deviates into the ventral lobe. The tip of both dorsal lobes might therefore constitute a valid functional landmark, but the tip of the dorsal lobe of sharks is biologically homologous to the tip of the ventral lobe in ichthyosaurs. Similar considerations apply to discrete, character data; much depends upon the manner in which characters and states are defined.

The exclusive use of putatively biologically homologous discrete variables restricts consideration to the same pool of characters used by cladists. In practice, and especially when dealing with fossil taxa, cladistic homology is established on operational grounds of detailed similarity and relationships to other structures (Pinna, 1991; Butler and Saidel, 2000). Such characters may also be functionally analogous, but are not necessarily so (Ruvinsky and Gibson-Brown, 2000; Shubin et al., 2009). Cladistic matrices therefore offer a rich resource for quantifying morphological variation across more conservative suites of putatively biologically homologous characters. Moreover, in the absence of homoplasy, we would expect the inter-OTU morphological distances assessed from such data to correlate closely with the evolutionary or patristic distances inferred on
most parsimonious or otherwise optimal phylogenetic trees. With the progressive introduction of more character conflict and homoplasy (Sanderson and Donoghue, 1989), this correlation will increasingly break down (Kelly et al., 2014), as will the inferred validity of many of the homology statements underpinning the data. Cladograms must account for the distribution of states across taxa by introducing hypotheses of convergence and reversal along branches. The metrics of morphological differences underpinning analyses of morphological disparity do not account for similarity due to the convergent acquisition or loss of traits, and are therefore intrinsically more phenetic in approach. Indeed, as levels of homoplasy increase (and more putative homologies are revealed to be analogies), patterns of morphological variety inferred from homologies and those inferred from statements of functional similarity become progressively more similar.

Morphospaces: theoretical and empirical

Once a set of morphological descriptors or variables has been established for a given group, it is possible to assess the morphological variety of constituent subgroups (e.g. clades) or of chronological sub-samples (e.g. taxa from successive geological periods). This can be done directly from the data, but it is more typical to visualize patterns of taxonomic distributions in some form of morphospace; an abstract, multidimensional space in which distances correlate with morphological differences. A distinction (although one not universally embraced; Mitteroecker and Huttegger, 2009) can be drawn between theoretical and empirical morphospaces. Theoretical morphospaces typically have dimensions that each capture a single quantifiable aspect of form, and (despite being parameterized with reference to real organisms) are defined a priori without the need for an empirical data set. The most frequently cited examples are those describing mollusc shells, which variously quantify form and growth using a very modest number of variables (Raup and Michelson, 1965; Skalak et al., 1997; Hammer and Bucher, 2005; Urdy et al., 2010). Real specimens can be located within theoretical morphospaces, but empirical data are not necessary in order to define them. Empirical morphospaces, in contrast, are constructed from a particular set of empirical morphological data. Their dimensionality tends to be high (Raup and Michelson, 1965; Foote, 1997; McGhee, 1999; Mitteroecker and Huttegger, 2009); much higher than that of their theoretical counterparts. For this reason, a number of data reduction techniques (usually multivariate ordination such as principal components or co-ordinates analyses) are used to condense the dimensionality of the space. This makes it possible to summarize morphological variation using a smaller number of abstracted variables, whilst minimizing distortion. These abstracted axes often cannot be described verbally, but may allow
the relative disparity of groups to be visualized and quantified more readily. Many of these approaches necessitate a distillation of the multivariate differences between taxa into a single measure of difference or distance for all possible taxon pairs (often realized as a triangular distance matrix analogous to that used to tabulate distances in a road atlas). The precise distance metric used depends upon the nature of the data and the desired properties of the resultant space and/or disparity indices. These complexities are discussed elsewhere at length (Wills, 1998, 2001; Hughes et al., 2013).

Two issues deserve emphasis. First, all morphospaces are abstractions, and necessarily based upon a sub-set of morphological variables. Variable choice inevitably determines the nature of the space. Many practitioners seek to sample variables as widely as possible from all aspects of morphology, thereby deriving spaces that reflect overall form. This is not always possible, however, as in many cases where only variation in particular organs or aspects of form can be codified across taxa (Pretorius and Scholtz, 2001; Lindbladh, 2002; Miller and Venable, 2003; Neige, 2006; Jones et al., 2009). Morphospaces derived from particular aspects of form or using data from particular organ systems or modules may be well suited to addressing particular evolutionary questions. However, ‘morphological disparity’ is usually conceived as referring holorically to overall form. Secondly, indices of disparity are necessarily relative, and comparisons are only possible within the parameters of a given morphospace or underlying data set. Hence, it is possible to make inferences regarding the relative disparity of a group at different times in its evolutionary history, or to compare the disparity of constituent subgroups within an analysis, it is not possible to make comparisons between groups from independently constructed morphospaces or data sets. This is also the reason why supermatrices uncritically assembled from multiple published data sets (and containing large blocks of inapplicable codes for large groups of taxa) may lose traction on some of the largest and deepest comparative questions.

A variety of disparity indices have been discussed in the literature (Foote, 1991, 1994, 1997; Wills et al., 1994; Wills, 2001; Hughes et al., 2013), but it is not our intention to rehearse the relative merits of these here. Among the most widely used approaches are those that distil the dispersion of taxa on multiple axes of the morphospace into a single value. The dispersion on a single axis can be quantified as either the range (defined by the outliers) or the variance of scores; the latter has the advantage of a relative insensitivity to sample size differences. Measures on multiple axes can be combined either as their product – effectively calculating the hypervolume of a hypercube – or as their sum. While hypervolumes are superficially more intuitive, they effectively give disproportionate weighting to smaller differences on later axes. Most ordination methods sequester progressively smaller fractions of total variance in later axes, but multiplying the univariate indices of dispersion means that halving the spread on any axis (whether the first or last) will halve the resultant hypervolume. Products also collapse to zero whenever the dispersion of taxa on a given axis is zero. Summing the univariate indices of dispersion (rather than multiplying them) avoids these problems. The sum of variances has particularly desirable properties, therefore, and has been used throughout the present study.

MATERIALS AND METHODS

Data collection

In general, we followed the protocols set out in Hughes et al. (2013). Morphological matrices for six major tracheophyte groups [Angiospermae (Doyle et al., 1994; Nandi et al., 1998; Doyle and Endress, 2014), Arecaceae (Baker et al., 2009), Nymphaeales (Borsch et al., 2008), Pinophyta (Hart, 1987), Pinaceae (Klymiuk and Stockey, 2012) and Polypodiidae (Pryer et al., 1995)] were selected from the literature. These represent the most diverse extant higher taxa of vascular plants, in addition to well-sampled sub-clades within both the angiosperms and conifers. A breakdown of the number and type (broad anatomical focus) of characters present in each data set is given in Supplementary Data Table S2. All of the matrices – with the exception of the cone-only pine data set – included a wide range of characters representing most aspects of morphology. The morphospaces derived from them might therefore reasonably be expected to capture overall form. Outgroup taxa were removed from these source matrices in several cases; either because there were missing taxa between the ingroup and outgroup, or because the outgroup OTUs were sampled at a higher taxonomic level than the ingroup. In some data sets, we also had to overcome inhomogeneity of sampling within the ingroup, which we achieved by amalgamating OTUs in such a way as to render them homogeneous at a higher taxonomic level. Character amalgamation utilized modal states. Some characters were rendered uninformative as a result of these condensations, and were therefore removed (specifically Pinaceae – 46, 47, 51; Arecales – 6, 10, 15, 21, 22, 48, 78, 91, 92, 10; Nymphaeales – 4, 5, 6, 12, 14, 22, 28, 39, 57). Stratigraphic ranges were assigned to stages using the International Stratigraphic Chart (Gradstein et al., 2004; Ogg et al., 2008). Stratigraphic range data were sourced from the Paleobiology Database (http://paleodb.org/), Sepkoski Online (Sepkoski, 2002) and The Fossil record 2 (Benton, 1993). Ranges were treated as continuous between first and last occurrences, with data being grouped into stage-level time bins. In cases where first and last occurrences were resolved only to intervals above the stage level, we coded for the stage corresponding to the midpoint of the interval. There were very few fossils within the Nymphaeales and we therefore estimated ranges using the time-calibrated molecular phylogeny of Yoo et al. (2005). Temporal bins with sample sizes of one were also amalgamated so that disparity could be calculated for these intervals. Comparisons between our sampled diversity curves and those compiled from more comprehensive sources are given in Supplementary Data Fig. S1.

Analyses

For each exemplary clade, intertaxon distance matrices were calculated using the Generalised Euclidean Distance (GED) metric of Wills (1998), and as implemented in Hughes et al. (2013). Distance matrices were ordinated in R (R Core Team, 2013) using principal co-ordinates analysis (Wills et al., 1994), and incorporating Cailliez’s correction for negative eigenvectors (Cailliez 1983). Disparity for each time bin was calculated as the sum of variances on all axes of the morphospace, yielding a
trajectory of disparity through time. The centre of gravity of each trajectory was used to distinguish between those clades whose temporal mean disparity was located early (bottom heavy), late (top heavy) or in the middle of their evolutionary history (symmetrical). A centre of gravity metric (Gould et al., 1987; Uhen, 1996) in absolute time (CG) was calculated for each clade as:

$$CG_m = \frac{\sum d_i t_i}{\sum d_i}$$

where \(d_i\) is the disparity at the \(i\)th stratigraphic interval and \(t_i\) the temporal midpoint in absolute time [million years (My)] of the \(i\)th stratigraphic interval. This was then scaled between the ages of the oldest \((t_{\text{oldest}})\) and youngest \((t_{\text{youngest}})\) intervals to yield an index of observed \(CG (CG_{\text{scaled}})\) between 0 and 1.

$$CG_{\text{scaled}} = \frac{t_{\text{oldest}} - CG_m}{t_{\text{oldest}} - t_{\text{youngest}}}$$

The expected \(CG_{\text{scaled}}\) for a clade of constant disparity through time is unlikely to be 0-50, but is determined by the durations of the time bins over which the profile is measured. The observed \(CG_{\text{scaled}}\) was therefore compared with the inherent \(CG (CG_i)\) for a clade of uniform disparity spanning the same intervals. A bootstrapping procedure was used to generate a distribution of 1000 resampled differences between \(CG_{\text{scaled}}\) and \(CG_i\), and clades for which >97.5% of bootstrapped replicates lay either above or below the centre of gravity inherent in the time scale \((P < 0.05)\) were deemed to be significantly top or bottom heavy, respectively (Foote, 1991). Observed \(CG_{\text{scaled}}\) was then expressed relative to \(CG\) as a baseline; hereafter simply \(CG\).

An ancillary test from Hughes et al. (2013) was used to determine whether the taxa observed in the first two stages had significantly less disparity than the maximum observed in any time bin. The disparity profile of the clade was bootstrapped 1000 times. For each replicate curve, the difference in disparity between the first two stages and the disparity maximum was calculated, yielding a distribution. If a difference of zero was within the 95% limits of this distribution, we were unable to reject the null hypothesis: namely that there was no difference between the initial disparity and the maximum (early high disparity). In such cases, maximal disparity was achieved in the earliest stages of the clade’s evolution. A similar test was applied to the end of each group’s history (late high disparity).

RESULTS AND DISCUSSION

Patterns of plant disparity through time

Our results are presented as preliminary explorations of the manner in which our selected clades have explored one form of morphospace through time. While more detailed work will certainly follow, our findings highlight several general patterns and permit certain conclusions.

For extinct clades with homogeneous birth/death dynamics and characters evolving under a Brownian model, the null expectation is that clade disparity profiles should be somewhat top heavy on average (a mean clade \(CG >0.5\)) (Foote, 1991). This is because the morphology of new lineages is contingent upon the morphology of those from which they have evolved; clades would therefore be expected to explore morphospaces in a progressive manner. The extinction of lineages, in contrast, can occur in any pattern with respect to the morphospace. Random extinction, in particular, will tend to maintain a relatively wide morphospatial distribution, introducing a fundamental asymmetry into clade evolution. This is an oversimplistic model for the clades studied here, because all are extant; the Recent effectively truncates their evolution. As demonstrated by Hughes et al. (2013), extant clades (as well as those becoming extinct coincident with one of the ‘Big Five’ mass events) have a much greater tendency towards top-heaviness merely by virtue of their persistence to the Recent. It is therefore unsurprising that most of our exemplar clades (with the exception of two of the three angiosperm data sets: Doyle et al., 1994; Nandi et al., 1998) show significantly top-heavy (\(CG >0.5\)) profiles (Table 1). More remarkably, several of these clades show initial disparity levels close to their maxima, or reach this level early in their history. Arecaceae (palms) (Baker et al., 2009) first appeared at their maximum disparity, while all three of our angiosperm data sets (Nandi et al., 1998; Doyle et al., 1994; Doyle and Endress, 2014) showed initial disparity levels >90% of the maxima in each case. Polypodiales (ferns) (Pryer et al., 1995) and Pinaceae (pines) (Klymiuk and Stockey, 2012) both reached their maxima within three time bins. The Nymphaeales (water lilies) (Borsch et al., 2008) are represented by a small data set (just 22 taxa) partitioned into larger time bins. Despite their apparently slow start, early disparity levels were not significantly different from the maximum (Table 1).

Conifers (Hart, 1987) have the most dynamic disparity trajectory, with initial Carboniferous and Permian levels significantly lower than at any subsequent times (Fig. 7). These modest levels persisted until after the end of the Permian, whereupon there were significant increases into the early Mesozoic. Although disparity appears to decline between the Middle and Late Triassic, it increases subsequently to reach maximum levels at the end of the Jurassic. Levels then decline gradually until the Recent, with extant disparity being significantly lower than the maximum levels observed at the end of the Jurassic. Conifers also show more intensive clustering of taxa in the morphospace at a variety of spatial scales than do the other clades in our study (Fig. 8). Disparity within the pine family (Fig. 9) shows broad similarities with conifers as a whole from their origins in the Jurassic; a reassuring finding given that pines represent a significant proportion of conifer diversity from this time. The initial increase in disparity for pines occurs slightly later than the corresponding increase in conifers as a whole, and is maintained until the present day.

Both angiosperms as a whole (Doyle et al., 1994) (Fig. 10) and the palm sub-clade (Baker et al., 2009) (Fig. 11) show approximately constant disparity through time. Palm disparity undergoes a slight decrease through the end of the Mesozoic and the early Palaeogene, such that the disparity of living taxa is lower than the realized maximum of the past. In contrast, our results suggest that the water lilies (Borsch et al., 2008) did not reach present levels of disparity until the Neogene (Fig. 12), with markedly lower levels for the first 10 My of their history. We note that this is our smallest data set (22 taxa), resulting in
### Table 1. Expected (or inherent) and observed centres of gravity (CG<sub>scaled</sub>) for clade disparity profiles, along with the results of bootstrapping tests (CG P-value) to determine if these differ.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Data set</th>
<th>Expected CG</th>
<th>Observed CG</th>
<th>Relative CG</th>
<th>CG P-value</th>
<th>Early high</th>
<th>Late high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiosperms</td>
<td>Doyle and Endress (2014)</td>
<td>0.757</td>
<td>0.759</td>
<td>0.502</td>
<td>0.001</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Angiosperms</td>
<td>Doyle et al. (1994)</td>
<td>0.718</td>
<td>0.722</td>
<td>0.504</td>
<td>0.228</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Angiosperms</td>
<td>Nandi et al. (1998)</td>
<td>0.714</td>
<td>0.718</td>
<td>0.504</td>
<td>0.846</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Conifers (Pinophyta)</td>
<td>Hart (1987)</td>
<td>0.556</td>
<td>0.712</td>
<td>0.655</td>
<td>0.001</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Leptosporangiate, Ferns (Polypodiidae)</td>
<td>Pryer et al. 1995</td>
<td>0.546</td>
<td>0.669</td>
<td>0.622</td>
<td>0.001</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Palms (Arecaceae)</td>
<td>Baker et al. (2009)</td>
<td>0.690</td>
<td>0.761</td>
<td>0.571</td>
<td>0.001</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Fines (Pinaceae)</td>
<td>Klymiuk and Stockey (2012)</td>
<td>0.604</td>
<td>0.753</td>
<td>0.649</td>
<td>0.001</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Water lilies (Nymphaeales)</td>
<td>Borsch et al. (2008)</td>
<td>0.626</td>
<td>0.794</td>
<td>0.668</td>
<td>0.001</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The expected CG is that determined for a clade with uniform disparity through time, and deviates from 0-5 because stratigraphic intervals and bins are of variable length. Relative CG is adjusted relative to the expected or inherent CG as a baseline. Clades that persist to the Recent typically have top-heavy profiles, since they are effectively truncated.

Early high and late high columns indicate the results of bootstrapping tests to determine if the disparity observed in the first and last intervals is distinguishable from the overall maximum for the clade (‘no’ indicates a difference with P < 0.05).

---

**Fig. 7.** (A) Disparity and diversity profiles for conifers using data from Hart (1987). Disparity (black circles) calculated as the sum of variances on all principal co-ordinate axes within several time bins. Values are the mean of 1000 bootstrap replicates ± s.e. Sampled generic diversity per stage is indicated by open, red circles. (B) Distribution of taxa on the first two principal co-ordinate axes of the empirical morphospace at four of the period time bins.
large estimates of error relative to observed fluctuations in disparity.

In polypod ferns, disparity increases through the Permian and Triassic, reaching or slightly exceeding modern levels by the Early Jurassic (Fig. 13). Disparity increased slightly thereafter to peak levels around the K/Pg but subsequently declined significantly in the last few million years.

An unexpected observation is that high levels of disparity were maintained for the past 80þ My in our largest clades (conifers, pines, ferns and angiosperms), despite successive radiations of sub-groups and catastrophic environmental and faunal upheavals over this time, including the K/Pg event (Ehleringer and Sage, 1991; Cerling et al., 1997; Zachos et al., 2001). Indeed, while there is evidence of significant local faunal turnover in plants (McElwain and Punyasena, 2007), recent work suggests that only two major extinction pulses are supported in the plant fossil record: one at the Carboniferous–Permian transition and another during the middle-late Permian (Cascales-Miñana and Cleal, 2014). Of the groups analysed, only conifers spanned this second event and actually show a significant increase in disparity during this time. It is therefore possible that conifers were evolving into areas of ecospace formerly occupied by other plant groups that declined at the end of the Permian (Retallack, 1995).

The high initial disparity of many of the plant groups investigated here results from the appearance of a small number of morphologically highly distinct taxa close to the base of each clade. In most of our groups, fossils quickly define the extremes of the empirical envelope as soon as they appear, with subsequent lineages gradually filling the intervening morphospace rather than colonizing more eccentric regions of it. Conifers exhibit a rather different pattern (Fig. 7), with the gradual appearance of sub-clades that each occupy distinct regions of the space (Fig. 8). Rather than rapid morphospace occupation followed by subsequent saturation, conifers appear to show several phases of morphospace colonization and subsequent diversity increase in tightly defined regions centred around pioneers with novel character combinations. This suggests that the evolution of conifers may have been characterized by the intermittent acquisition of novel morphologies or ‘key’ innovations, followed by subsequent diversification. Such events may include the radiation of the pines in the Jurassic and the cypresses in the

![Fig. 8. To what extent are taxa clustered within their respective morphospaces at different levels of granularity?](https://academic.oup.com/aob/article-abstract/117/5/859/1741305)
Cretaceous and early Palaeogene. The high degree of morphospace clustering may result from competition with other groups (such as angiosperms), constraining the available morphospace. However, it is more likely to be a function of greater structural or developmental constraints acting upon suites of characters within the conifer data set (moreover, conifers appear to show relatively tight clustering in the Triassic and Jurassic, prior to the inferred appearance of basal angiosperms). Pines show much weaker clustering than conifers as a whole. Characters within the pine data set (Klymiuk and Stockey, 2012) were derived from cone morphology, strongly implying that Pinacae were able to explore the majority of possible cone forms rapidly and early in their evolution in a relatively unconstrained manner.

Because most of the discrete character matrices analysed here included a broad sample of characters from many different anatomical regions, it is reasonable to assume that the gross morphology of the taxa in the sample was reasonably represented. Our three angiosperm matrices had marked differences in character and taxon composition (Fig. 14), but showed similar overall patterns of disparity through time.

**Why are there so few studies of plant disparity?**

There are a number of possible reasons why empirical morphospace approaches have been underutilized within the plant sciences, aside from the usual methodological considerations underpinning the choice of data and indices (Rohlf, 1998). Many morphometric approaches entail time-consuming data collection, which may limit tractable sample sizes. There are also difficulties in establishing variable or character sets that can be measured or coded across higher taxa. Most studies therefore focus upon smaller plant clades or else derive data...
from particular structures (Chartier et al., 2014) rather than investigating overall morphological disparity. Moreover, the often fragmentary nature of fossil material may mean that holistic treatments are impractical, or that many types of morphometric data cannot be obtained (Adams et al., 2004).

**Utilizing existing discrete morphological data matrices**

New morphological character matrices for plants are becoming increasingly rare (Gottlieb, 1988; Sytsma et al., 1991); mounting evidence from molecular phylogenetics implies that morphological convergence is obfuscating our understanding of plant relationships (Bowe and Coat, 2000; Donoghue and Doyle, 2000; Schneider et al., 2009). However, we believe that morphological character data have important uses beyond that of inferring phylogeny (Thorne et al., 2011); not least for quantifying patterns of disparity change throughout morphologically and taxonomically diverse clades with long evolutionary histories. In this context, the problems of homoplasy and convergence that bedevil phylogenetic inference are less marked, since morphospaces are conceived for a variety of purposes and can be intended to reflect a variety of aspects of evolution. Discrete character morphospaces offer a framework for quantifying patterns of morphological disparity within large clades, but also highlight questions that can be addressed in a more focused manner using other morphometric techniques (Goodman, 2002). More comprehensive analyses of existing plant character matrices would represent an efficient use of legacy data, allow some of the commonalities suggested in this paper to be properly tested and would powerfully complement existing and future morphometric studies.
Despite the abundance of discrete, morphological data in the literature, there are a number of considerations when using explicitly cladistic matrices for disparity studies. Morphological cladists usually seek to resolve phylogeny (Forey et al., 1998), but are not always concerned with representing accurate branch lengths and evolutionary distances. Even in the extreme approach adopted by pattern cladistics, which views the cladistic method as being divorced from evolutionary assumptions of descent through modification (Brady, 1982; Brower, 2000), there is still an imperative to recognize hierarchical groupings within sets of taxa (Hennig, 1966; Estabrook et al., 1975). There may therefore be a tendency to sub-divide morphological variety more finely within taxa that are morphologically conservative overall in order to resolve their relationships or structure. Conversely, taxa supported by long evolutionary branches may be morphologically very distinct from their nearest sampled relatives, but there may be no imperative to quantify all of these differences to the same degree of resolution as in highly diverse and morphologically similar groups. More generally, it is reasonable to expect character matrices to be biased towards distinctive features and/or those which have been demonstrated to be good at distinguishing groups in previous studies. An allied issue is the assumption that all characters should be treated equally. This may not always be desirable, particularly in cases where some groups are characterized by a limited number of highly distinctive and variable characters while others are defined by broader suites of gross morphological features that are nevertheless coded as a single character. For example, it is probably simplistic to treat the presence or absence of sclereids in the leaves on an equal footing with scandent vs. arborescent growth habits (Foster, 1956; Rury and Dickison, 1984). While there are a variety of objective approaches for the differential

FIG. 11. (A) Disparity and diversity profiles for palms using data from Baker et al. (2009). Disparity (black circles) calculated as the sum of variances on all principal co-ordinate axes within several time bins. Values are the mean of 1000 bootstrap replicates ± s.e. Sampled sub-familial diversity per stage is indicated by open, red circles. (B) Distribution of taxa on the first two principal co-ordinate axes of the empirical morphospace at four of the period time bins.
weighting of characters in phylogenetic studies, these are derived from predictions or empirical estimates of levels of homoplasy or the phylogenetic information content of characters (Farris, 1969; Sharkey, 1989; Goloboff, 1993, 2014). In disparity analyses, what may be required is rather some weighting derived from the ontogenetic priority, developmental (Riedl, 1977; Arthur, 1984, 1988; Wimsatt, 1986) or structural depth (Stebbins, 1969; Pettersson, 2009) of characters, although such weights are notoriously difficult to assign.

Some cladistic matrices are constructed in order to address particular questions, most commonly sequences of character acquisition across important evolutionary transitions; for example, tetrapods from fishes (Wagner and Chiu, 2001; Long and Gordon, 2004; Ruta et al., 2006; Wagner et al., 2006) and birds from dinosaurs (Garner et al., 1999; Xu, 2006; Heers et al., 2014; Brusatte et al., 2014). Such data intentionally focus on the taxa and characters bracketing these changes, with deliberately much sparser sampling outside of this. More generally, outgroup taxa – often included for rooting purposes – are more sparsely sampled than those of the ingroups (Graybeal, 1998; Heath et al., 2008). Morphological cladistic characters may therefore sample morphological variation unevenly across taxa and through time. Not all data sets are suitable for investigating temporal and taxonomic patterns of morphological variation therefore, and many require some form of moderation. Hughes et al. (2013), for example, standardized sampling according to higher taxonomy, and removed outgroups.

One final issue is the inclusion or otherwise of autapomorphic character states; those present in just a single taxon (Yeates, 1992; Bryant, 1995). Such states cannot influence inferred cladistic branching structure, but they do affect branch lengths (without introducing homoplasy) and indices of morphological difference. In two state characters, an autapomorphic state renders the entire character cladistically (but not phenetically) uninformative. This property is flagged by most phylogenetic software, which usually results in their removal from cladistic matrices. Autapomorphic states are more likely to be retained in multistate characters (those with three or more

---

**Fig. 12.** (A) Disparity and diversity profiles for water lilies using data from Borsch et al. (2008). Disparity (black circles) calculated as the sum of variances on all principal co-ordinate axes within several time bins. The sampled number of species per stage is indicated by open, red circles. Values are the mean of 1000 bootstrap replicates ± s.e. (B) Distribution of taxa on the first two principal co-ordinate axes of the empirical morphospace at four of the period time bins.
states), since the character remains informative overall. More generally, cladists do not actively seek to include autapomorphic states, such that cladistic matrices usually omit this aspect of morphological variation. Empirically, however, the inclusion/exclusion of autapomorphies makes relatively little difference to assessments of morphological variety (Wills, 2001).

The precise effect of autapomorphic states will depend upon the overall properties of the data set and the mode of analysis, but in general they merely cause the taxa possessing them to appear marginally more divergent from the overall mean morphology than they would otherwise be.

There is an increasing desire for large, complete phylogenies to underpin various forms of evolutionary and ecological analyses (Guyer and Slowinski, 1993; Phillimore and Freckleton, 2006; Tamura et al., 2012). Large matrices of molecular characters (supermatrices) are frequently assembled de novo using open data resources and automated algorithms (Liu et al., 2001; Davies et al., 2004; Bininda-Emonds, 2004; Davis and Page, 2008). There are no similar repositories or tools for morphological matrices. Assembling large matrices comprising hundreds or thousands of OTUs and characters from first principles would ensure greater consistency, but is hugely time-consuming. Hence, morphological supermatrices are often assembled by amalgamating the largest data sets or synthetic treatments available for constituent groups. However, this approach may entail its own set of problems. The first is alluded to above; the differential sampling of taxa and characters. Taxon sampling can be standardized more readily, but uniform character sampling requires more detailed knowledge and entails greater subjectivity. More problematically, it is often difficult or impossible to code many of the characters in the constituent matrices for the ‘out-group’ taxa (those represented in the other matrices), thereby

![Diagram](https://example.com/diagram1.png)

**FIG. 13.** (A) Disparity and diversity profiles for extant leptosporangiate ferns using data from Pryer et al. (1995). Disparity (black circles) calculated as the sum of variances on all principal co-ordinate axes within several time bins. Values are the mean of 1000 bootstrap replicates ± s.e. Generic diversity per stage (from the Paleobiology Database) is indicated by open, red circles. (B) Distribution of taxa on the first two principal co-ordinate axes of the empirical morphospace at four of the period time bins.
within several time bins. Values are the mean of 1000 bootstrap replicates ± s.e. Despite the inclusion of different taxa and characters, all three profiles show a rapid initial increase in disparity followed by relatively constant disparity over the rest of their history.

resulting in large blocks of inferred plesiomorphies (typically ‘0’ or absent) and inapplicable codings (‘?’). Depending upon the manner in which such inapplicabilities are treated, this phenomenon can result in artificially distinct clusters of taxa, strongly but spuriously demarcated by these discontinuities in knowledge and character sampling (Wilkinson et al., 2005; Cotton et al., 2006). For these reasons, large published cladistic matrices compiled from first-hand observations of specimens (or from careful treatments of the primary literature) have many potential advantages over those assembled by conjoining data from disparate published sources (de Dequeiroz and Gatesy, 2007).

CONCLUSIONS

1. The concept of morphological disparity is distinct from those of diversity and species richness (Wills, 2001). Indices of disparity attempt to codify the morphological variety of a sample of taxa, are calculated relative to some set of morphological variables or characters, and often utilize a plot of taxa in a multidimensional morphospace. Morphospaces are abstract spaces in which the geometric distances between taxa are proportional to some measure of the morphological differences between them. The nature of a morphospace is entirely contingent upon the underlying data, the manner in which differences between taxa are summarized as distances, and the methods used to project these distances into an n-dimensional space. The precise approach will depend upon the purpose for which the morphospace is intended. It follows that there is no objective morphospace (in the sense that there is an objective phylogeny), and that the dispersion of taxa in different spaces cannot be compared directly (comparisons between sub-groups within the space are possible, but these are necessarily only relative). Morphospaces derived from large samples of characters or variables encompassing most aspects of form are most likely to offer insights into overall morphological variety. Indices of disparity variously assess the relative dispersion of samples of taxa within a morphospace, or provide some distillation of the morphological differences between them.

2. Diversity and disparity appear to be fundamentally decoupled. A significant majority of the animal clades investigated show relatively high disparity early in their evolution (Hughes et al., 2013) at times when diversity is still comparatively low (i.e. there are modest numbers of taxa but these are morphologically highly distinct from each other). The subsequent evolution of such groups often sees an increase in diversity with little or no concomitant increase in disparity; there are increasing numbers of taxa within a restricted number of morphological ‘themes’. Disparity may even decline as diversity is rising, since some of the most speciose clades have particularly constrained bodyplans but are able to partition ecospace and morphospace particularly finely. A substantial minority of animal clades show other patterns, including high initial disparity at low diversity (Foote, 1990).

3. There have been relatively few studies of morphological disparity in plants, and no studies have attempted to assess patterns of overall disparity in major clades through time. Temporal patterns of diversity in plants and animals show significantly different patterns (Knoll et al., 1979), with plants counterintuitively being less affected at times of global mass extinction (Cascales-Miñana and Cleal, 2014). An assessment of patterns of disparity in major plant clades is therefore overdue, and may provide insights into plant macroevolution to complement those being obtained for animals.

4. There are numerous morphometric methods that allow shape and shape change to be quantified across taxa. However, as the morphological variety of the forms being compared increases (usually in tandem with the taxonomic scope of the study), the ability of such approaches to compare increasingly disparate forms becomes more limited. Discrete character data sets have certain advantages in this context. There are rich resources of discrete character matrices already available for numerous plant clades, and, although initially intended for inferring phylogeny, these data sets can be repurposed for disparity studies within certain strictures.
5. Our preliminary disparity analyses for six exemplary plant clades demonstrate that initial levels of disparity are usually high, if not indistinguishable from (or at) the maximum ultimately achieved by the group. Most regions of the morphospace are colonized early in the history of each plant clade, with subsequent evolution serving merely to increase diversity within these regions. The notable exception are the conifers, in which sub-clades appear intermittently, and progressively colonize distinct regions of the space. This results in conifer disparity increasing incrementally over the first half of the group’s history. All of our exemplary plant clades have disparity profile shapes with a centre of gravity higher than the intrinsic null (significantly so in all except for two angiosperm data sets). This is unsurprising, however, since all are extant groups, with profiles truncated by the Recent (Hughes et al., 2013). Combining detailed empirical morphometric studies of specific anatomical regions with the more holistic approach illustrated here will probably be reciprocally illuminating, and offer insights into plant macroevolution.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: list of published disparity studies. Table S2: distribution of characters in sampled plant data sets. Figure S1: comparison between actual and drafted one figure.

ACKNOWLEDGEMENTS

This paper evolved from a presentation by M.A.W. at Euro Evo Devo 2014 in Vienna, and we thank Rainer Melzer and Günter Theissen for their invitation to speak at their symposium. We are grateful to the authors who generated the cladistic matrices repurposed and reanalysed herein. This work was generously supported by grant no. 43915 from the John Simon Guggenheim Foundation (which funds J.W.O.’s PhD work), and by the Leverhulme Trust grant F/00 351/Z to M.A.W. (which underpinned our work on disparity). J.W.O. wrote the manuscript, collated plant morphological and stratigraphic data sets, drafted figures and analysed data. M.H. ran disparity analyses and statistics, summarized the disparity literature and drafted figures. S.G. produced clustering plots. M.A.W. conceived and designed the study, wrote the manuscript and drafted one figure.

LITERATURE CITED


Anderson PSL, Friedmann M, Ruta M. 2013. Late to the table: diversification of tetrapod mandibular biomechanics lagged behind the evolution of terrestriality. Integrative and Comparative Biology 53: 197–208.


