

Assessing the robustness of disparity estimates: the impact of morphometric scheme,
temporal scale, and taxonomic level in spatangoid echinoids

Loïc Villier and Gunther J. Eble

RRH: ROBUSTNESS OF DISPARITY IN SPATANGOIDS

LRH: LOÏC VILLIER AND GUNTHER J. EBLE

Abstract. – The joint quantification of disparity and diversity is an important aspect of recent macroevolutionary studies, and is usually motivated by theoretical considerations on the pace of innovation and the filling of morphospace. In practice, varying protocols of data collection and analysis have rendered comparisons among studies difficult. The basic question remains, how sensitive is any given disparity signal to different aspects of sampling and data analysis? Here we address this issue in the context of the radiation of the echinoid order Spatangoida during the Cretaceous. We compare patterns at the genus- and species-level, with time subdivision into subepochs and into stages, and with morphological sampling based on landmarks, traditional morphometrics, and discrete characters. In terms of temporal scale, similarity of disparity pattern accrues despite a change in temporal resolution. Different morphometric methods, however, produce somewhat different signals early in the radiation. Both the landmark analysis and the discrete character analysis suggest relatively high early disparity, whereas the analysis based on traditional morphometrics records a much lower value. This difference appears to reflect primarily the measurement of different aspects of overall morphology. Notwithstanding, a general deceleration in morphological diversification is apparent at both the genus and the species level. Moreover, inclusion or exclusion of the sister-order Holasteroida and stem-group Disasteroida in the reference morphospace did not affect proportional

changes in spatangoid disparity. The relative robustness of these patterns implies that the choice of temporal scale, morphometric scheme, and taxonomic level may not substantially affect the representation of large-scale morphospace structure and of broad trends in disparity. However, the choice need not be arbitrary, as particular questions may demand certain protocols.

Loïc Villier. UMR CNRS 6118 Géosciences, Université de Rennes 1, campus de Beaulieu, 163 avenue du Général Leclerc, 35043 Rennes, France and UMR CNRS 5561 Biogéosciences, Université de Bourgogne, Centre des Sciences de la Terre, 6 Bd. Gabriel, 21000 Dijon, France. E-mail: loic.villier@u-rennes1.fr

Gunther J. Eble. Interdisciplinary Centre for Bioinformatics, Universität Leipzig, Kreuzstrasse 7b, D-04103 Leipzig, Germany. E-mail: eble@bioinf.uni-leipzig.de

Introduction

In general evolutionary discourse, the notion of disparity is used to express morphological diversity and body plan variety, as opposed to taxonomic diversity. With refinements in quantitative methodology, comparisons between disparity and diversity are becoming a powerful tool to address evolutionary questions debated in theoretical circles (e.g., Gould 1991; Foote 1993, 1997; Wills et al. 1994; Roy and Foote 1997; Conway Morris 1998; Eble 1998a). Comparing disparity against diversity allows for fuller recognition and description of evolutionary radiations, assessing extinction selectivity, evaluating morphological responses to environmental or ecological factors, and testing macroevolutionary hypotheses (e.g. Foote 1993, 1996, 1997, 1999; Wills et al. 1994; Wagner 1995; Jernvall et al. 1996; Dommergues et al. 1996; Neige et al. 1997, 2001; Wills 1998a; Eble 1998a,b, 2000a; Lupia 1999; Smith and Lieberman 1999). However, perceived data limitations within studies have tended to encourage only general disparity patterns to be recognized. Varying protocols in data collection and analysis have usually discouraged comparisons among studies. Even though theoretical and conceptual work can suggest a number of research questions, their tractability may depend on how robust inferences about disparity may be relative to various potential sources of bias. Progress has been made in understanding the effect of taxonomic and character sampling (Foote 1995, 1999), as well as of choice of disparity metric (Foote 1991; Wills et al. 1994; Ciampaglio et al. 2001). Also, some cross-taxonomic

generalizations have proved possible in terms of patterns of proportional change (Foote 1996). Nevertheless, how different morphometric methods and temporal scales affect the estimation of a given disparity signal is an issue that is still poorly understood.

An additional source of debate relates whether species-level data are better suited for disparity studies than genus-level data. Many disparity studies have tended to focus on genera as units of sampling, as it is commonly done with taxonomic diversity data. A few studies suggest that sampling at the genus- and species-level provide equivalent estimates of global disparity signal (Foote 1995, 1999). Others suggest that the species-level is a priori preferable, at least for certain questions (Smith and Lieberman 1999). The issue is far from settled.

The aim of this paper is to propose an empirical account of how sensitive measured disparity may be relative to methods of analysis. We address this question in the context of the radiation of the echinoid order Spatangoida during the Cretaceous. With comparisons of patterns derived from separate datasets, we assess the influence of morphological descriptors, temporal scale, and taxonomic level. Previous work documented spatangoid disparity at the genus-level, using landmarks, and with time subdivision into subepochs (Eble 1998b, 2000a). A first interval of relatively high disparity associated with low initial diversity was suggested, followed by deceleration in morphological diversification and subdued response to major increase in diversity. Similar patterns have been found in other groups (see Foote 1997), and have been

interpreted as evidence of morphospace saturation. A new, independent study is presented here, focusing on the species-level, using traditional morphometrics and discrete characters, and with time subdivision into stages. It provides an opportunity to reassess spatangoid disparity patterns, and to evaluate whether and when particular methodological choices affect inferences about morphological diversification.

Material and Methods

Morphological descriptors. – Landmark data are taken from Eble (2000a), whose analysis employed 18 landmarks to describe global morphology, the architecture of the plastron, and the geometry of the ambulacra, peristome, periproct, and apical system (Fig. 1). Three-dimensional coordinates of landmarks were scaled to a common baseline, row-normalized, and standardized, thus minimizing the effect of size. After exclusion of redundant and uninformative coordinates, 38 variables were retained.

For the traditional morphometrics analysis, morphology was described by 16 distance measurements (Fig. 1). About 50% of the features correspond to aspects of morphology included in the landmark-based analysis. Distance measurements capture the width of the petals, but not the architecture of the plastron (16 coordinates in the landmark-based analysis). Measurements were used for calculation of 14 indices, corresponding to ratios easily understandable in term of anatomy or of functionality (appendix 1). These ratios do not constitute scaling by a baseline, but nonetheless

reduce the effect of size. Variables were treated using principal component analysis.

The first three axes were retained, and were considered as representative of spatangoid morphospace. They summarize only 51.5 % of the total variance, but underlie the most important features of dissimilarities, and are likely to reflect more interpretable aspects of disparity. Significant segregation occurs between taxa on the first three axes. On later axes, taxa appear to be randomly distributed and their contribution to disparity should be less determinant.

A third morphometric scheme is based on thirty-five discrete characters, which incorporate different aspects of morphology. Eighty percent of coded discrete characters have no equivalent in the morphometric schemes. Coding discrete states is an effective way to jointly take into account tuberculation, plate architecture (type of apical disc and of plastron), and the structure of the ambulacra (type of pores and of petals), which are difficult to integrate in a continuous morphometric scheme. Further, coded discrete characters encompass a set of features bearing potentially more phylogenetic information, while overall shape is suspected to be more sensitive to homoplastic changes due to environmental setting (Zaghib-Turki 1989; Néraudeau and Floquet 1991; Kanazawa 1992) or developmental biases (McNamara 1987; Eble 2002).

Measurement of disparity. – Different disparity estimates have been explored on theoretical and empirical grounds (e.g. Foote 1990, 1991, 1999; Wills et al. 1994;

Ciampaglio et al. 2001). They correspond to different aspects of morphological diversity (variance or range of morphology, average dissimilarity, nearest-neighbor distance). The variance, in particular, is a measure of disparity that is robust to sample size differences. The sum of univariate variances, or total variance, is equivalent to the mean squared Euclidean distance to the centroid (Van Valen 1974), and can be likened to measures of dissimilarity between individuals. We therefore selected the sum of variances as a measure of disparity for continuous morphometric variables, and the mean pairwise distance for discrete characters. Disparity in this study has therefore the same meaning, and comparisons can be directly made between analyses based on continuous and discrete data.

In traditional morphometrics, principal component axes can be considered as composite characters for estimation of disparity. Disparity is therefore calculated as the sum of variances of scores on the first three axes of the PCA. Before calculation of disparity, scores on each axis are scaled to the eigenvalue with multiplication by the square root of the eigenvalue. Such approach avoids artificial weighting of variables and emphasizes the main sources of variance.

We calculated error bars by bootstrap, following the procedure generally used in disparity studies (i.e. Foote 1992, 1994a, 1999; Wills et al. 1994, 1998a; Eble 2000a). For each interval, we used the mean value of the bootstrap distribution as the population equivalent of the disparity of the sample, and calculated error bars as one standard

deviation of the bootstrap distribution (500 replicates). Bootstrap values were calculated with a modified version of the freeware Rare 1.1 (Wills 1998b). In this study, error bars are given for illustration purposes only. While routinely used in disparity studies, error bar structure may vary from interval to interval. Resampled values are frequently unsettled in small samples. Error bars reflect not only the variability of statistical samples, sample size, and number of replications, but also the degree of homogeneity of the data, which may not always be high even for large samples. It is still unclear how disparity relates to these issues for different kinds of data. Statistical tests and general inferences based on error bars are possible, but here and elsewhere error bars are best treated as a guide to data structure.

Temporal scale. – Disparity signal was compared at two levels of temporal resolution (Fig. 2): at the stage level and at a larger scale previously used by Eble (1998b, 2000a), whereby amalgamated stages were used to define five Cretaceous intervals (K1 to K5). These time intervals were defined so as to minimize variation of their average duration (mean duration of 15.8 My, $sd=2$) and maximize the reliability of stratigraphic ranges, while preserving the potential for meaningful macroevolutionary patterns to be recorded. Late Cretaceous spatangoids have historically been extensively studied, and the stratigraphical range of most species and genera can be regarded as reliable at sub-stage level. Data are less precise for the early Cretaceous, and only stage-level

resolution is possible, as a result of the lack of comprehensive revisions. Consequently, stages seem to be the shortest operational time interval for Cretaceous spatangoids as a whole. Uncertainty and risks of erroneous stratigraphic attribution are likely to increase at finer resolution. Sampling at stage level improves stratigraphic resolution (mean duration of 6.6 My), but implies more heterogeneity in interval duration ($sd=3.34$).

Taxonomic sampling.— Eble (1998b, 2000a) analyzed disparity of the super-order Atelostomata (which comprises the orders Spatangoida, Holasteroida, and Disasteroida) with sampling of one to three species per genus. Sampled species are taken as representative of their genera. Following a range through method, a species may contribute to disparity estimates up to the entire stratigraphic range of a genus. Such approach increases the autocorrelation of disparity from interval to interval, which may result in some underestimation of disparity.

In another approach, measures were taken on one to four specimens of each species available in various paleontological collections (British Museum of Natural History, Muséum National d'Histoire Naturelle de Paris, Muséum d'Histoire Naturelle de Grenoble, Université de Bourgogne, Université de Toulouse, Université de Poitiers). The raw data matrix encompasses 206 specimens, representing 103 species spanning the time interval from Berriasian to Cenomanian. The data matrix was reduced to 138 lines in the matrix so as to homogenize representation, for each genus and each species, of

their different morphological types and of their entire stratigraphic range. This method of sampling avoids the range-through method and allows for potentially more morphological variability to be sampled.

For the discrete character analysis, the data matrix includes 34 early Cretaceous and Cenomanian species representing all nominal genera. The 35 characters correspond to a total of 91 character states. This data matrix was initially coded for a cladistic analysis (see Villier et al. in press). It does not include autapomorphies or invariant characters, and all multistate characters were treated as unordered.

Results

Influence of temporal resolution. – The genus-level signal constructed with amalgamated units shows an initially high disparity value in the Neocomian, followed by a slight drop in K2 and a general increase from K2 to K5 (Fig. 3). The dataset was reworked to describe the pattern at the stage-level. At this resolution, samples for the first three intervals are reduced to two genera each. Low sample size inflates error bars and renders disparity values at best suggestive. Even though quantitative comparison must be made with caution here, the qualitative pattern is strongly similar. There is conservation of the general trend observed with amalgamated stages. The similarity of disparity pattern despite a change in stratigraphic resolution supports the view that

broad trends in disparity may be relatively insensitive to temporal scale. Nevertheless, stage resolution allows for additional detail to emerge.

At stage resolution, disparity cannot be measured in the Berriasian because only one genus is present. A high value is expressed in the Valanginian but not in the Hauterivian, suggesting that high disparity in the Neocomian was more short-lived than could be inferred from coarser resolution. The increase in disparity from K2 to K5 appears at stage level as a two-stepped pattern. Indeed, disparity increases from the Barremian to the Albian, remains stable from Cenomanian to Coniacian, and afterwards increases again. In addition, breaking up the K5 unit into Campanian and Maastrichtian stages suggests a slight decrease of disparity before the end of the Cretaceous, which was not previously apparent.

The signal constructed from species-level sampling (using traditional morphometrics) (Fig. 4) yields slightly different interpretations relative to those stemming from genus-level sampling (which employed landmark-based morphometrics). With species-level sampling and amalgamated stages, initial disparity seems at its lowest level during K1. At stage-level resolution, this corresponds to the Valanginian and Hauterivian stages. Disparity regularly increases from the Hauterivian to the Albian and remains at an equivalent level during the Cenomanian. Switching from amalgamated stage to stage-level resolution does not affect the overall trend.

Sensitivity of the disparity signal to morphological descriptors. – The two species-level datasets analyzed with traditional morphometrics and with discrete characters, respectively, were originally designed to tackle the first phase of Cretaceous diversification, and to resolve phylogenetic relationships within the family Toxasteridae (Villier 2001). It includes only pre-Turonian species. Direct comparisons with the genus-level landmark-based study are thus possible only for the first four intervals (K1 to K4) (Fig. 5).

To different degrees, different morphometric methods reveal a tendency for disparity to increase in the long run. If one takes into account that the diversity of species and genera increases substantially over the same period, a pattern of eventual deceleration in morphological diversification is clear in all cases. Nonetheless, substantial differences exist. The relative timing and magnitude of disparity change is conspicuously different when derived from traditional morphometric data, and disparity is at its lowest in the Neocomian. More subdued change from interval to interval, and larger error bars characterize landmark-based and discrete character analyses. Importantly, the patterns derived from landmark data and discrete character data are also similar in that disparity is already relatively high early on and is not eventually exceeded, in contrast with the traditional morphometric analysis.

The difference in Neocomian disparity might reflect the measurement of different aspects of overall morphology, as well as the use of different data treatments

(character state coding, standardization, PCA). A sample size effect might also be possible. Disparity for the K1 interval is calculated from 3 genera for landmarks, 8 species for discrete characters, and 20 species for distance measurements. At face value, small sample size would here appear to be associated with high disparity and, conversely, large sample size with low disparity. In theory, given a common pool of variability, sampling twenty species instead of 3 should lead to a decrease in disparity. This reasoning breaks down, however, for at least two reasons. First, larger sample sizes after the Neocomian do not substantially alter disparity as inferred from both landmark data and discrete character data. Second, when Neocomian disparity is recalculated with the same taxa used in the landmark analysis, disparity does not change substantially (gray bars on Fig. 5). This suggests that sample size smallness per se is not invariably a source of artifact, the impact of sample size on disparity being a combination of statistical and evolutionary components, i.e., sample completeness and the nature of the variability pool. Here, the latter seems more important. Not only different aspects of morphology are likely to be captured by different methods, but also different potentials to vary, different degrees of redundancy, and different scales of change. The tendency of traditional morphometrics to emphasize global over local differentiation and to less effectively eliminate size differences might account for the pattern of initial diffusion and delayed deceleration observed, as well as for the low disparity level regardless of whether species or genera are sampled. In contrast, both landmark and discrete-

character data are more directly capturing, if partially, structural differences characteristic of genera.

Effect of taxonomic level. – Sampling at the genus-level is usually chosen to study macroevolutionary patterns. This choice is often grounded on the fact that fossil species tend to have shorter stratigraphic ranges, implying a higher sensitivity to sampling bias, and are less often taxonomically standardized in large databases, implying a higher sensitivity to monographic bias. Using genera as proxies for lineages is thus a priori preferable, even more so when a good correlation between number of genera and species accrues, as was the case here (Spearman rank correlation test highly significant, $p=0.003$).

Interestingly, sampling at genus level canonically makes two related assumptions about the hierarchical structure of morphological space: (1) that the differences among species of separate genera exceed differences among species within genera; (2) that, on average, the morphology of a measured species can be taken as representative of the morphology of its genus when the range-through method is applied. For echinoids, stasis of genera is common (Smith 1984; Donovan and Veale 1996), and in taxonomic and phylogenetic work synonymizing of species is far more frequent within genera than across genera. While this tends to support genera as appropriate operational units to study morphological diversification, as well as the

range-through approach, it does not necessarily follow that disparity patterns at the genus level and the species level should be congruent.

Yet in Cretaceous spatangoids at least, disparity signals have the same pattern at genus- and species-levels (compare Figs. 3b and 4b). The similar pattern between genera and species implies that the choice of taxonomic level may not substantially affect the representation of the structure of morphological space.

Influence of reference morphospace on disparity values. – In the landmark-based analysis of Eble (2000a), the reference morphospace encompassed the orders Spatangoida, Disasteroida and Holasteroida, i.e., the super-order Atelostomata. Standardization of variables was based on the overall range of variation in the more inclusive clade. As a consequence, one could argue that the distribution of spatangoids in morphospace might have been influenced by the other two orders. However, very similar patterns of disparity are found when the reference morphospace is constructed based on the Spatangoida alone (Fig. 6B). Such similarity points to macroevolutionary stability of morphospace structure despite major phylogenetic divergence. McGhee (1999) suggested that the sample-dependence of empirical morphospaces makes them inherently unstable, but the open issue is to what degree (Eble 2000b). The spatangoid case suggests that instability may be minor, and that inferences about disparity from

empirical morphospaces can be robust to the inclusion of taxa with varying degrees of phylogenetic propinquity.

To assess whether stability might also hold within spatangoids, we analyzed disparity of the genus Heteraster. The genus Heteraster comprises 33 species of spatangoids (Villier unpublished data). It appears first during the Hauterivian, reaches its acme in the Aptian and Albian, and disappears early in the Cenomanian. Disparity for the genus Heteraster has been calculated from traditional morphometrics following the procedure used in the analysis of spatangoids as a whole. The disparity signal is similar whether the reference space is based on species of the genus only or from a sample of Early and Middle Cretaceous spatangoids (Fig. 6). The disparity analysis of the genus Heteraster suggests morphospace stability within spatangoids as well. [as beyond.]

Discussion

Temporal scales and temporal data. – In large-scale studies, disparity is often estimated for long intervals of approximately similar duration, so as to minimize the effect of uncertainty of stratigraphic data and of small sample size, and to facilitate worldwide comparisons. In practice, using time units that are longer than the uncertainty associated with taxon first and last appearances reduces analytical error stemming from different time scales and from uncertain taxon ages.

However, stratigraphic resolution might affect disparity because of variation of sample size and sampling rate. As an example of this potential bias, sampling or not of outliers does contribute to variation in crinoid disparity (Foote 1994a). Analytical error related to stratigraphic resolution is likely to increase when short intervals entail samples that are too small to statistically justify removal of putative outliers, but this kind of problem seems to be generally avoided in most empirical studies.

In comparison, diversity is more uniformly affected by variation in interval length than disparity. Diversity tends to increase monotonically with the increase in duration of time intervals, but this may often not be the case with disparity. Consequently, contrasting disparity versus diversity always implies assumptions about the influence of stratigraphic resolution and about the quality of time scales. Some types of evolutionary interpretation may be influenced by such assumptions, but the degree to which this is the case is unclear.

Taxonomic level of sampling. – Operationally, genera may be represented by a sample of specimens, by a sample of species, by one species per stratigraphic interval, or by one species for the entire range of the genus (range through method). Here different protocols produce results that are largely in agreement. Foote (1996) obtained similar results for crinoids: the same broad disparity pattern accrues whether sampling is based on one or several species per genus.

Genera are considered as a better sampling unit than species for diversity studies at lower taxonomic levels (Raup and Boyajian 1988). Sampling at the species level is expected to magnify biases of the fossil record and of taxonomic practice. Genera have on average longer durations than species, minimizing the singleton effect. As genera usually include more than one species, they have a higher rate of preservation in the fossil record, and thus good approximation of the temporal range of genera is more likely than for species. On these grounds, the genus-level diversity signal is more relevant than that at the species-level. Extension of this reasoning to morphological data partly explains why authors favor and prioritize maximal representation of genera for disparity analysis.

However, sampling randomly one species per genus may not be the best way to sample morphological variety. In large taxonomic data sets, the definition of genera may be variously grounded on phenetic criteria, on phylogenetic criteria, or both. Therefore, genera can suggest different amounts of disparity depending on whether phylogenetic or phenetic criteria are used. In contrast, as minimally diagnosable taxonomic units or phenons (Smith 1994), species are arguably more stable in terms of morphology, favoring more accurate estimation of disparity. Our study, however, lends support to the view that sampling at the species- or genus-level is equivalent for disparity estimation when the focus is on broad patterns such as the evolution of higher taxa or large-scale trends. This may be because most paleontological species and genera

are identified on phenetic grounds, because genera are in geological and monographic time as stable as species, or because both species and genera are random samplers of the same underlying morphospace. The convergence of signals from species and genera can also justify mixed designs, whereby uncertainty in the estimation of genus-level disparity in intervals with small sample size is reduced with sampling of additional species, assuming the disparity metric is insensitive to data inflation.

Disparity and morphological descriptors. – Previous analyses of the robustness of disparity to different aspects of morphology have been made mostly a posteriori, through alteration or partition of a core signal. Partitioning of morphological data into datasets representing particular anatomical regions may produce signals similar to but also signals substantially different from that of the total signal. When differences are present, partitioned disparity patterns have proved interpretable in terms of differential evolutionary constraint, rather than sampling bias (Wagner 1995; Foote 1994b, 1999; Eble 2000a). Semi-independence in the evolution of body regions implies that disparity is not strictly equivalent when morphology is described by morphometric approaches that sample different aspects or scales of organization.

The differences among disparity profiles based on different morphometric methods are therefore probably a real reflection of differential sampling of morphology. For example, the plastron is a feature involved in important evolutionary changes during

the early spatangoid history and underlies a distinct pattern of disparity (Eble 2000a). Nearly 40% of landmark data code for the architecture of plastron plates. This latter aspect of disparity is much reduced in the discrete character analysis (3 states out of 91), and is not taken into account in traditional morphometrics. Despite the obvious impact such asymmetry in the description of morphology is expected to have, the fact that a global signal of deceleration is ubiquitous suggests that broad trends in disparity may exhibit little selectivity with respect to morphology.

Influence of reference morphospace on disparity values. – Morphospace structure may

affect estimates of disparity depending on which taxa are included. It has been suggested that empirical morphospaces are inherently unstable and that theoretical morphospaces provide a better basis for the study of disparity (McGhee 1999).

However, theoretical morphospaces are model dependent, and tend to focus on a small number of parameters in an attempt to capture broad aspects of organism construction.

This may be interesting for the study of certain aspects of disparity, but often multiple sources of variation, both global and local, are apparent, and high dimensionality needs to be reckoned with empirically. Empirical morphospaces, with or without ordination, are in this sense best suited for the estimation of disparity, the real issue becoming the relative instability of empirical morphospace and the robustness of inferences derived from it (Eble 2000b).

Results from our analyses show conservation of broad disparity patterns and presumably of morphospace structure under several conditions of sampling. Ordination via PCA, on the basis of which disparity was quantified in the analyses of traditional morphometric data, did not produce significantly different results when species were added or subtracted from the total sample. The relative stability of disparity profiles with different numbers of taxa and different levels of phylogenetic inclusiveness is not necessarily expected, and may be due to the morphogenetic constraints and morphospace saturation that a common bauplan induces across lineages and clades.

Conclusions

Our results suggest that the temporal geometry of disparity among Cretaceous spatangoid echinoids is robust to choice of taxonomic level, temporal scale, and rationale for morphospace construction. Such stability of the morphospace and of the patterns of disparity confirms previous results (Eble 1998b, 2000a), and suggests that the macroevolutionary signal captured by different disparity studies may be broadly comparable. Moreover, new data allows some refinement of evolutionary interpretations. Initially high disparity is confirmed and is dated more precisely as Valanginian. The hypothesis of high initial disparity should be later confirmed by examination of unpublished taxa from the Berriasian of the Tethyan realm, which have other associations of characters (A.B. Smith personal communication 2001, and other

unpublished data). The deceleration of morphological diversification, in turn, is pervasive and suggests that the saturation of morphospace is at least in part intrinsic to its structure, rather than a result of external controls.

That disparity patterns may be robust to methodological protocol, as this case-study suggests, emphasizes the reliability of disparity as a descriptor of evolutionary patterns. Obviously, rate and quality of sampling will always influence the estimation of disparity, the effect of sampling becoming more noticeable with small sample size. Based on simulations, Ciampaglio et al. (2001) suggest that sampling strategy has to take into account differential aggregation of taxa in morphospace, to avoid distortion of disparity estimates. Our empirical analyses show a reduced effect of sampling, probably because the distribution of taxa is relatively uniform in morphospace. This of course may not always be the case.

The main differences between separate analyses are largely due to the choice of morphological descriptors. This is not so much a problem or a limitation of disparity studies, but is rather linked to the open issue of whether particular aspects of morphology can document morphological evolution in general. While broad sampling of morphology is always advisable, the potential for pronounced morphological modularity means that more than one disparity signal and more than one pattern of morphological diversification may coexist in the same bodyplan.

In general, the extent to which different disparity studies are comparable and interpretable within the same theoretical framework is a function of the relative robustness of disparity estimates. As our understanding of disparity increases, so does our ability to refine generalizations about morphological diversification.

Acknowledgments

This paper is a contribution to the theme "Macroévolution et dynamique de la biodiversité" of the UMR CNRS 5561, Biogéosciences. We thank B. David, D. Néraudeau, N. Navarro, T. Saucède, and B. Viguié for discussion and/or comments.

Literature Cited

- Ciampaglio, C. N., M. Kemp and D. W. McShea. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* 27:695-715.
- Conway Morris, S. 1998. *The crucible of creation*. Oxford University Press, New York.
- Dommergues, J.-L., B. Laurin, and C. Meister 1996. Evolution of ammonoid morphospace during the Early Jurassic radiation. *Paleobiology* 22:219-240.

- Donovan, S. K. and C. Veale. 1996. The irregular echinoids Echinoneus Leske and Brissus Gray in the Cenozoic of the Antillean region. *Journal of Paleontology* 70:632-640.
- Eble, G. J. 1998a. The role of development in evolutionary radiations. Pp. 132-161 in M. L. McKinney and J. A. Drake, eds. *Biodiversity dynamics: turnover of populations, taxa, and communities*. Columbia University Press, New York.
- Eble, G. J. 1998b. Diversification of disasteroids, holasteroids and spatangoids in the Mesozoic. Pp. 629-638 in R. Mooi and M. Telford eds. *Echinoderms: San Francisco, Balkema, Rotterdam*.
- Eble, G. J. 2000a. Contrasting evolutionary flexibility in sister groups: disparity and diversity in Mesozoic atelostomate echinoids. *Paleobiology* 26:56-79.
- Eble, G. J. 2000b. Theoretical morphology: state of the art. *Paleobiology* 26:520-528.
- Eble, G. J. 2002. Developmental morphospaces and evolution. Pp. 35-65 in J. P. Crutchfield and P. Schuster, ed. *Evolutionary dynamics: exploring the interplay of selection, accident, neutrality, and function*. Oxford University Press, Oxford.
- Foote, M. 1990. Nearest-neighbor analysis of trilobite morphospace. *Systematic Zoology* 34:371-382.
- Foote, M. 1991. Morphological and taxonomic diversity in a clade's history: the blastoid record and stochastic simulations. *Contributions from the Museum of Paleontology, University of Michigan* 2:101-140.

- Foote, M. 1992. Rarefaction analysis of morphological and taxonomic diversity. *Paleobiology* 18:1-16.
- Foote, M. 1993. Discordance and concordance between morphological and taxonomic diversity. *Paleobiology* 19:185-204.
- Foote, M. 1994a. Morphology of Ordovician-Devonian crinoids. *Contributions from the Museum of Paleontology, University of Michigan* 29:1-39.
- Foote, M. 1994b. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. *Paleobiology* 20:320-344.
- Foote, M. 1995. Morphology of Carboniferous and Permian crinoids. *Contributions from the Museum of Paleontology, University of Michigan* 29:135-184.
- Foote, M. 1996. Ecological controls on the evolutionary recovery of post-Paleozoic crinoids. *Science* 274:1492-1495.
- Foote, M. 1997. The evolution of morphological diversity. *Annual Review of Ecology and Systematics* 28:129-152.
- Foote, M. 1999. Morphological diversity in the evolutionary radiation of the Paleozoic and post-Paleozoic crinoids. *Paleobiology Memoirs* 25:1-115.
- Gould, S. J. 1991. The disparity of the Burgess Shale arthropod fauna and the limits of the Cladistic analysis: why we must strive to quantify morphospace. *Paleobiology* 17:411-423.

- Gradstein, F. M., F. P. Agterberg, J. G. Ogg, J. Hardenbol, P. Van Veen, J. Thierry, and Z. Huang. 1995. A Triassic, Jurassic and Cretaceous time scale. Pp. 95-126 in *Geochronology time scales and global stratigraphic correlation*, SEPM Special Publication, 54.
- Jernvall, J., J. P. Hunter, and M. Fortelius. 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiation. *Science* 274:1489-1492.
- Kanazawa, K. 1992. Adaptation of test shape for burrowing and locomotion in spatangoid echinoids. *Palaeontology* 35:733-750.
- Lupia, R. 1999. Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. *Paleobiology* 25:1-28.
- McGhee, G. R. 1999. *Theoretical morphology: the concept and its applications*. Columbia University Press, New York.
- McNamara, K. J. 1987. Taxonomy, evolution and functional morphology of southern Australian tertiary hemiasterid echinoids. *Palaeontology* 30:44-48.
- Neige, P., S. Elmi, and L. Rulleau. 2001. Existe-t-il une crise au passage Lias-Dogger chez les ammonites? Approche morphométrique par quantification de la disparité morphologique. *Bulletin de la Société géologique de France* 172:257-264.
- Neige, P., D. Marchand, and A. Bonnot. 1997. Ammonoid morphological signal versus sea-level changes. *Geological Magazine* 134:261-264.

- Néraudeau, D., and M. Floquet. 1991. Les échinides Hemiasteridae: marqueurs écologiques de la plate-forme castillane et navarro-cantabre (Espagne) au Crétacé supérieur. *Palaeogeography, Palaeoclimatology, Palaeoecology* 88:265-281.
- Raup, D. M., and G. E. Boyajian. 1988. Patterns of generic extinction in the fossil record. *Paleobiology* 14:109-125.
- Roy, K., and Foote M. 1997. Morphological approaches to measuring biodiversity. *Trends in Ecology and Evolution* 12(7):277-281.
- Smith, A. B. 1984. *Echinoid paleobiology*. Allen and Unwin, London.
- Smith, A. B. 1994. *Systematics and the fossil record: documenting evolutionary patterns*. Blackwell Scientific Publications, Oxford and Malden.
- Smith, L. H., and B. S. Lieberman. 1999. Disparity and constraint in oleneloid trilobites and the Cambrian radiation. *Paleobiology* 25:459-470.
- Van Valen, L. 1974. Multivariate structural statistics in Natural History. *Journal of Theoretical Biology* 45:235-247.
- Villier, L. 2001. Evolution du genre Heteraster dans le contexte de la radiation de l'ordre des Spatangoida (Echinoidea, Echinodermata) au Crétacé inférieur. Thèse de l'Université de Bourgogne, 389 p. (unpublished)
- Villier L., D. Néraudeau, B. Clavel, C. Neumann, and B. David. in press. Phylogeny of Early Cretaceous Spatangoids (Echinoidea: Echinodermata) and taxonomic implications. *Palaeontology*.

Wagner, P. J. 1995. Testing evolutionary constraint hypotheses with early Paleozoic gastropods. *Paleobiology* 21:248-272.

Wills, M. A. 1998a. Cambrian and recent disparity: picture from priapulids. *Paleobiology*, 24:177-199.

Wills, M. A. 1998b. Rare 1.1. Rarefaction of morphological disparity indices. Basic code for Chimpunk Basic, University of Bristol.

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

Zaghib-Turki, D. 1989. Les échinides indicateurs des paléoenvironnements: un exemple dans le Cénomaniens de Tunisie. *Annales de Paléontologie* 75:63-81.

Figure Legends

FIGURE 1. Time scales investigated. Values indicate age of stage boundaries in My (from Gradstein et al. 1995). Units made of amalgamated stages (K1-K5) are defined in the upper box line and stages in the lower one.

FIGURE 2. Morphometric schemes used in this study. Landmarks are pointed onto the upper row of drawing (see Eble 1998, 2000a for a precise definition), and distances measured onto the lower row of drawing (see appendix 1 for a formal definition).

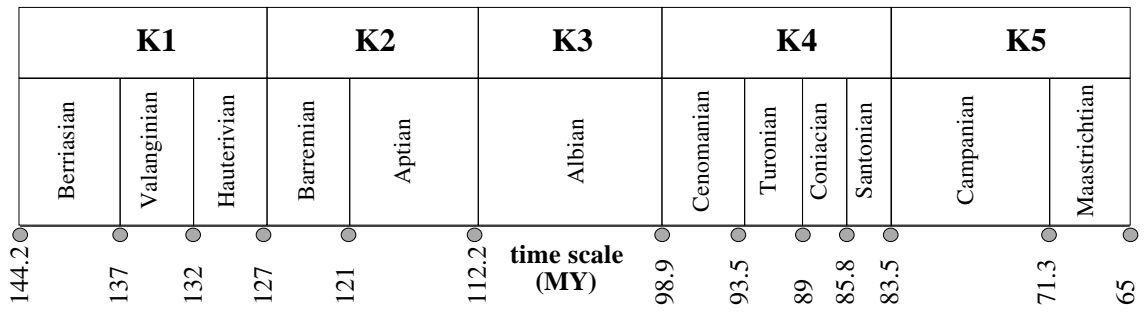
FIGURE 3. Disparity of the order Spatangoida during the Cretaceous: influence of time scale. A. Use of amalgamated stages. B. Stage-level analysis. Taxa are sampled at genus-level, one species representing each genus. Range through method is applied when time range of genus exceeds the duration of one interval. Values are plotted at the mean age of the interval. Disparity is measured as the sum of variance for 38 coordinates of landmarks. Error bars on disparity correspond to \pm one standard deviation calculated from 500 bootstrap replicates.

FIGURE 4. Disparity of the order Spatangoida during the Cretaceous: influence of time scale. A. Use of amalgamated stages. B. Stage-level analysis. Taxa are sampled at species-level, with sampling of nearly 50% of species known per genus, and one to four specimens representing each species. Range through method is avoided. Values are plotted at the mean age of the interval. Disparity is measured as the sum of variance for the three first axes of a PCA analysis calculated with 14 ratio of linear measurements (see appendix 1 for details). Error bars on disparity correspond to \pm one standard deviation calculated from 500 bootstrap replicates.

FIGURE 5. Disparity of the order Spatangoida during the Cretaceous: influence of the morphological descriptor and statistical treatment of raw morphometric data. A. Disparity from discrete characters is measured as the mean phenetic distance (number of character states) between taxa. Initial data matrix comprises 34 early Cretaceous and Cenomanian spatangoids sampled according to have a representation of the morphologic variety and the entire stratigraphic range of genera (see Villier et al. in press for character definition and details of taxon sampling). Several species may be sampled for one genus, but range through method was used when necessary. Gray bars indicate results for the K1 interval when same species are sampled as in the landmark-based analysis. B. Disparity from landmarks. Data and statistical treatment are equivalent as for Fig. 3A. Gray bar indicates value calculated for K4 with a taxonomic

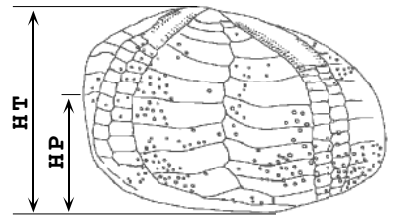
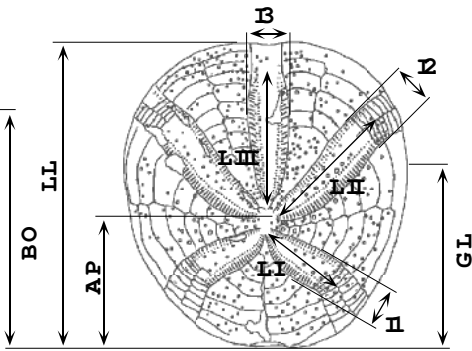
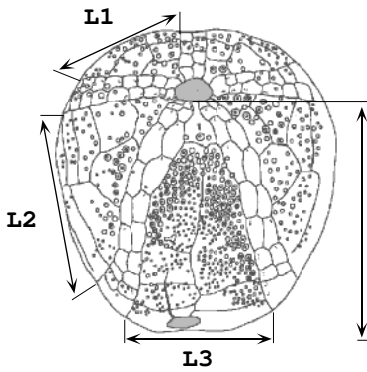
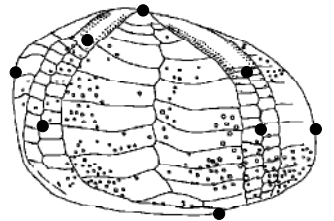
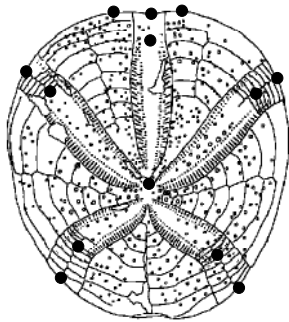
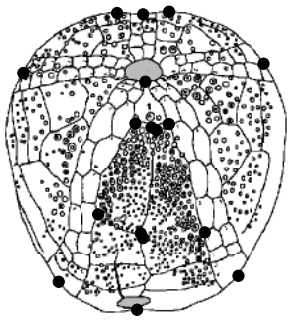
sampling equivalent to that of the analyses using discrete characters and traditional morphometrics. C. Disparity from traditional morphometrics. Data and statistical treatment are equivalent as for Fig. 4A. Gray bar corresponds to the value obtained when sampling only the species analyzed with landmarks.

FIGURE 6. Disparity of the order Spatangoida during the Cretaceous: structure of the morphospace. A. Pattern of spatangoid disparity calculated from a data set including the orders Disasteroidea, Holasteroidea and Spatangoida (data from Eble 2000a). B. Pattern of spatangoid disparity calculated from independent analysis using the same landmark data and the same procedure, but morphospace contains spatangoid only. C. Pattern of disparity of the spatangoid genus Heteraster in the morphospace constructed for analysis of disparity of the Spatangoida with traditional morphometrics and species-level sampling. Data include 18 species of the genus Heteraster. Disparity is calculated at stage-level such as in Fig. 4B. D. Pattern of disparity of the genus Heteraster in an independent morphospace based solely on 25 species of the genus (from 33 species retained after taxonomic revision, Villier unpublished data). Disparity is measured by the sum of variance on the first four axes of a PCA analysis of measurement ratios (see Fig. 2 and appendix 1 for details).

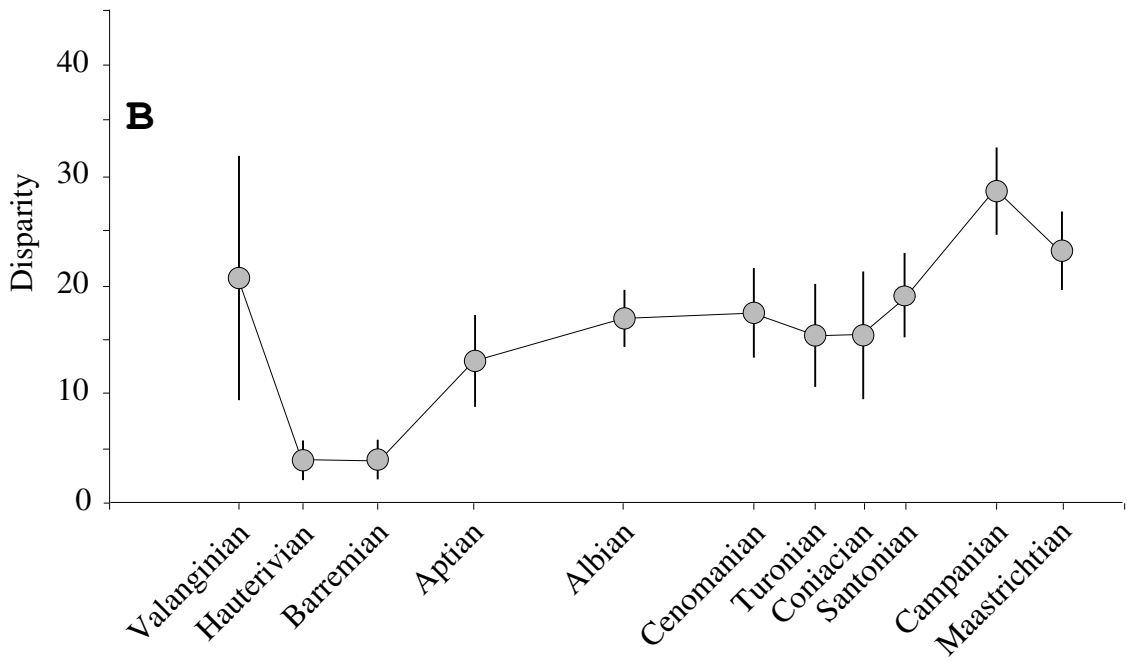
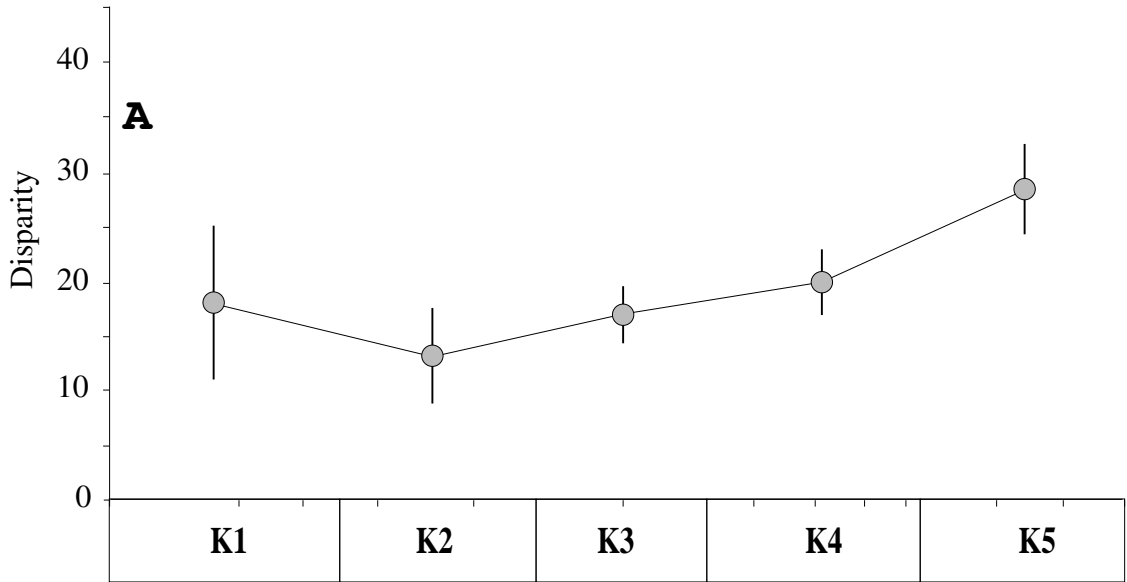


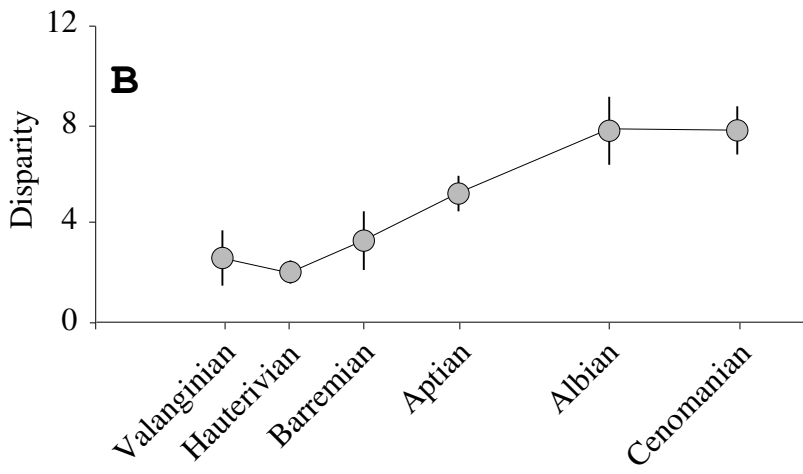
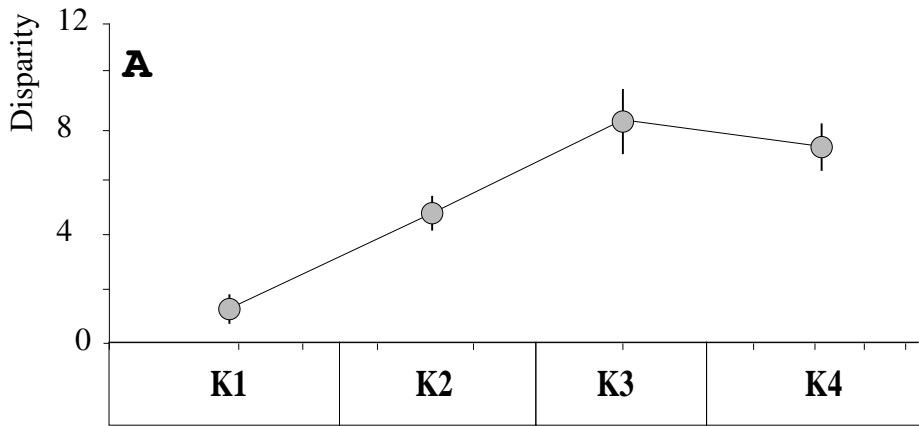
Villier & Eble, Fig. 1

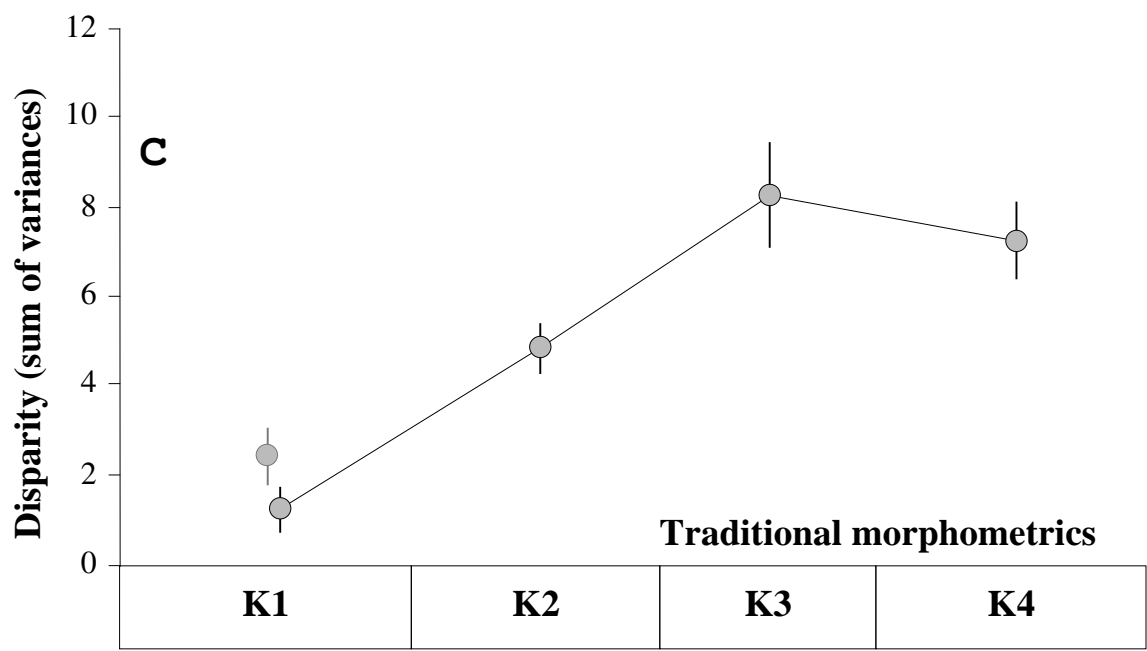
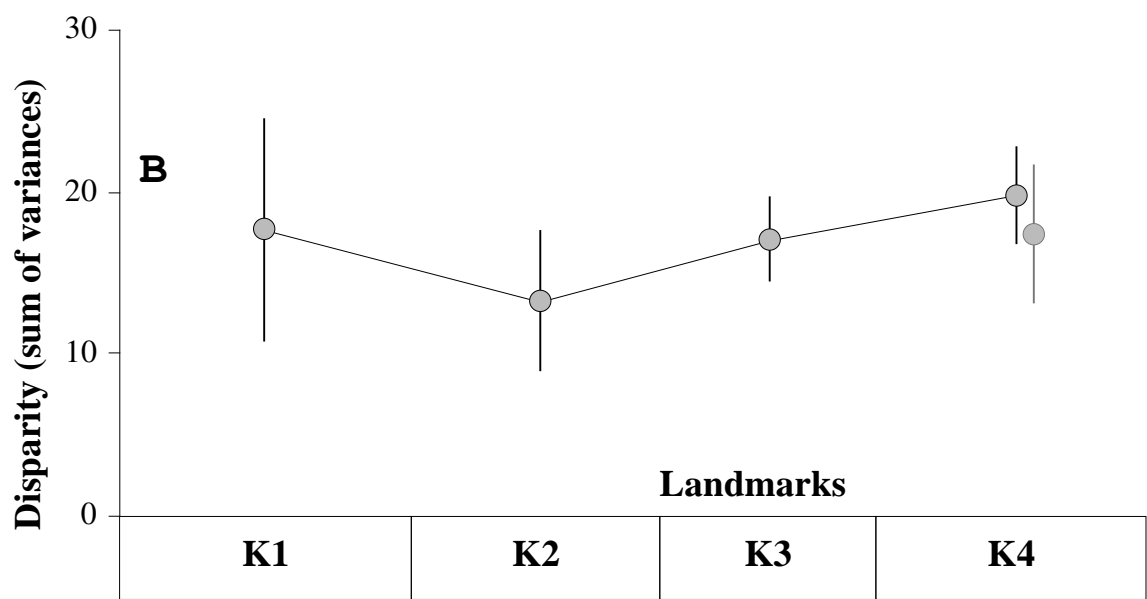
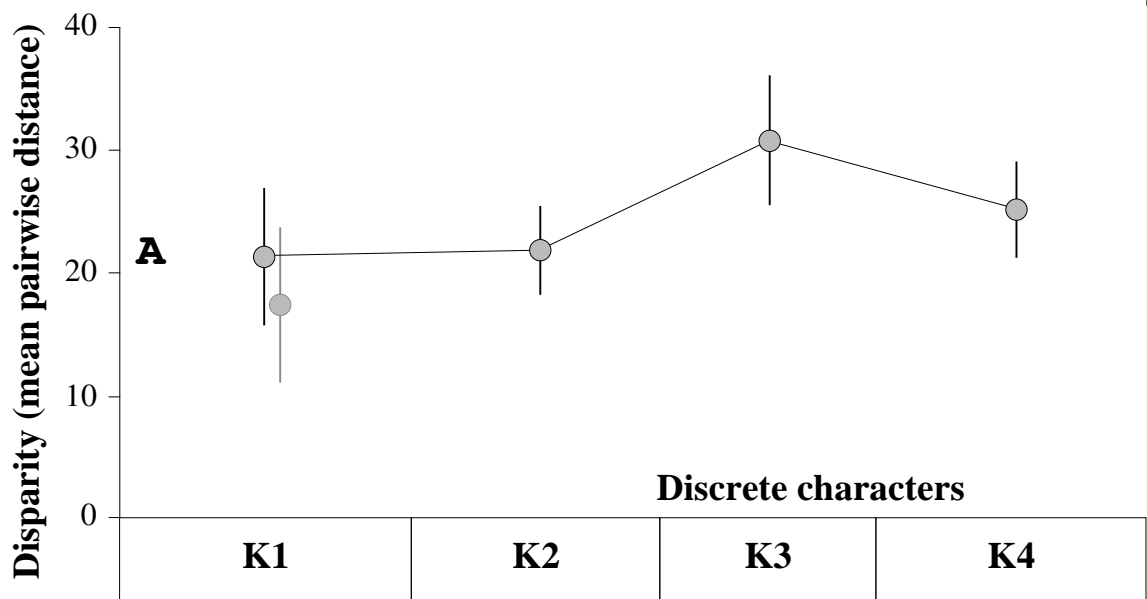
↑ Top



Villar & Eble, Fig. 2







↑ Top

