A morphospace of planktonic marine diatoms. II. Sampling standardization and spatial disparity partitioning

1 **A morphospace of planktonic marine diatoms, part II:**

2 **Sampling standardization and spatial disparity partitioning**

3 Benjamin Kotrc and Andrew H. Knoll

4

5 RRH: DIATOM MORPHOSPACE PART II: SUBSAMPLING

6 LRH: BENJAMIN KOTRC AND ANDREW H. KNOLL

7

Abstract.—Morphospace occupation through time provides a view of diversification distinct from the more familiar taxonomic tabulations. However, this view is subject to the same geological biases long recognized in studies of taxonomic diversification, where techniques for correcting secular bias in sampling have become standard practice. In this study, we apply sampling standardization techniques to a morphospace investigation to test whether observed stratigraphic trends in morphospace occupation are artifacts of trends in sampling. When sampling bias is corrected by randomized subsampling, all disparity metrics show stationary patterns, or at most directional changes of small magnitude. Metrics describing the average dispersion of taxa in morphospace are less subject to sampling bias than those describing the total extent of morphospace occupied. We also investigate a measure of disparity that is insensitive to sampling intensity, introducing a geographic component of morphological disparity. By analogy to α and β components of taxonomic diversity, we suggest the notions of α and β disparity, and find that α disparity remains roughly constant through time. Our analysis also allows us to present the first taxonomic diversity curve of diatoms under shareholder quorum subsampling (SQS), showing similar results to previously published subsampling methods: a roughly twofold rise over the Cenozoic, with peak diversity around the Eocene/Oligocene boundary. Tests for methodological bias from choices in ordination method and data culling during morphospace construction indicate that our results are relatively insensitive to both factors: Cenozoic occupation of planktonic diatom morphospace is largely unchanging. We find a similarly stationary pattern when we directly analyze the morphological data, seeing no change in the prevalence of taxa with different sets of morphological characters. More broadly, our results make clear that a complete view of morphological disparity must

consider sampling biases, which can be addressed with well-established, quantitative methods in morphospaces populated using occurrence-level data.

Benjamin Kotrc. Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138 U.S.A. Present address: *Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 U.S.A. E-mail: kotrc@mit.edu*

Andrew H. Knoll, Department of Organismic and Evolutionary Biology, Harvard University Cambridge, Massachusetts 02138 U.S.A.

Introduction

Studies of the fossil record make valuable contributions to our understanding of evolution, not least through documenting the diversification history of clades. By analyzing the occupation of morphospace through time, we can compare a morphological perspective on diversification (through metrics of disparity) to the more familiar taxonomic view of diversification history (through counts of species richness). Many of the groups in which this comparison has been made show "asymmetric diversification" where peak morphological disparity is reached early and then remains more or less stationary while taxonomic diversification continues (Gould 1989; Foote 1997; Erwin 2007). In marine planktonic diatoms, an ecologically important group of primary producers with siliceous cell walls called frustules, the history of taxonomic diversification has received more attention than morphological diversification. Their taxonomic diversification history has conventionally been read as a sharp Cenozoic rise to current levels of diversity, relatively late in an evolutionary history stretching back to at least the Early Cretaceous Period.

In the companion paper in this issue (Kotrc and Knoll submitted), we addressed the Cenozoic history of diatom morphological disparity through a study of their fossil record. We showed that as tabulated taxonomic richness increased, the range of morphospace occupied increased as well, while the distance between taxa in morphospace remained the same or even declined slightly. We stopped short of making strong biological inferences from these observations, however, because of the possibility that these results are subject to bias, particularly from temporally uneven sampling. Such biases have long been recognized in studies of taxonomic diversification, in which the impact of different methodological choices has been investigated (e.g., different taxon-counting methods, Bambach 1999); techniques that correct for

secular bias in sampling, like rarefaction, by-list subsampling, or shareholder quorum subsampling have become standard practice (Miller and Foote 1996; Alroy et al. 2001, 2010b). The impact of bias (particularly from sampling), however, is not as often considered in studies of morphospace, even though trends in morphological disparity are commonly compared to taxonomic diversity. This might be explained by the widespread recognition that morphological data represent a different window into evolutionary history than taxonomic data, perhaps distracting from the fact that both are derived from the fossil record, and are thus both subject to the well-known geological biases that have been the subject of research since the origins of the discipline (e.g., Darwin, 1859; Newell 1959; Raup 1972).

Nonetheless, Foote (1992) did recognize the importance of sample size in assessing morphological disparity and applied rarefaction to metrics of morphospace occupation in trilobites, blastoids, and ammonoids. However, Foote's definition of a "sample"—the unit being rarefied or subsampled to a common threshold—in that study is quite different from the definition in current studies seeking to correct for sampling bias in time series of taxonomic richness. In the morphospace study by Foote (1992), each taxon in a given time bin is considered a sample, while in diversity subsampling studies, each occurrence of a taxon (or assemblage of taxa) is considered a sample. Thus, rarefied time bins in Foote's morphospaces contain the same number of taxa, while in studies of taxonomic diversity, rarefied time bins contain the same number of occurrences (e.g., Miller and Foote 1996).

In a more recent morphospace study of the Ediacara biota, Shen et al. (2008) approached the problem of sampling bias by calculating a metric of morphospace occupation under rarefaction using the latter definition, treating taxon occurrences as samples. However, these authors report the results of rarefaction for just one of three time bins, and do not attempt to

correct comprehensively for sampling differences. We are not aware of any morphospace study to date in which sampling differences have been corrected by sampling standardization as has become common practice for studies of taxonomic diversity.

The need to correct for uneven paleontological sampling in studies of morphological diversification was recently highlighted in a study of pterosaur disparity (Butler et al. 2012). The authors demonstrated significant correlations between proxies of geological sampling and metrics of morphospace occupancy and concluded that disparity metrics based on the range of occupied morphospace, in particular, are strongly affected by uneven sampling of the fossil record. Although Butler et al. (2012) did apply rarefaction to standardize disparity metrics, it was subsampling of the sort performed by Foote (1992), to a standard number of taxa. Although occurrence-level data are available, due to the nature of the pterosaur fossil record—in which almost every occurrence is a singleton (i.e., the only occurrence of that taxon)—no meaningful sampling standardization is possible at the level of occurrences (in the sense of diversity studies like Alroy et al. 2008).

In the companion paper in this volume (Kotrc and Knoll submitted), we found that temporally uneven sampling was a possible confounding factor in interpreting the results of our diatom morphospace study. Because the diatom fossil record can yield many thousands of individuals in a spoonful of sediment, and since the *Neptune* database of microfossil occurrences (Lazarus 1994; Spencer-Cervato 1999) captures much of this information, we can directly address sampling biases.

Before beginning to address temporally uneven sampling, however, it is worth a brief aside to consider some of the factors contributing to bias in the diatom fossil record (for a more thorough treatment, see Lazarus 2011). We have already mentioned, in the companion paper, the

declining abundance of deep-sea sediments with age (due to subduction), as well as the necessity of drilling through younger sediments to access older ones. These factors, which affect all groups of planktonic marine microfossil groups alike, lead to an decrease in sampling with age. Siliceous microfossils additionally undergo a series of diagenetic mineral transitions with rising burial temperature and pressure (DeMaster 2003), beginning with the original amorphous silica and eventually leading to remobilization into chert layers (Moore 2008), with the eventual loss of recognizable morphological features along the way. This process also serves to bias preservation—including morphological diversity—against older diatom records, with very little preservation before the early Paleogene (Fenner 1985). The modern ocean is strongly undersaturated with respect to amorphous silica (Sarmiento and Gruber 2006), leading to rapid dissolution that limits preservation to areas with high rates of silica accumulation (Lazarus 2011). Much less than half of the extant species of marine planktonic diatoms are known from deep-sea sediments (Sournia et al. 1991). This bias may, however, have been less significant in the past, as a number of lines of evidence point toward a decline in the oceanic concentration of dissolved silica over the course of the Cenozoic (Siever 1991, Maldonado et al. 1999, Racki and Cordey 2000, Muttoni and Kent 2007, Lazarus et al. 2009). This decline would suggest a secular trend in the dissolution bias opposite to the others, namely, favoring preservation with increasing age.

In this study, we extend the techniques of sampling standardization developed for studies of taxonomic diversity history by applying them to a morphospace of diatoms in order to test whether the results presented in Kotrc and Knoll (submitted) are artifacts of secular trends in sampling. We use various subsampling methods, including the recently published shareholder quorum (SQS) method; in the process, we also report the first application of SQS to the diatom

record of taxonomic diversification. We further test for sampling bias by examining disparity metrics that ought to be insensitive to sampling differences. We also test for methodological bias in constructing the morphospace, from choices in ordination method and the choice of thresholds for data culling based on missing information. Finally, we look for biological signals in the data by examining the distribution of sets of characters expected to change under suggested drivers of macroevolutionary change over the Cenozoic Era.

Materials and Methods

Morphospace Construction and Disparity Metrics

We constructed an empirical morphospace (McGhee 1999) by coding the states of 123 discrete binary or unordered multistate morphological characters for 152 diatom genera. The chosen genera represent all valid genera in the *Neptune* database, less those identified as resting stages. This choice of taxa made it possible to use the *Neptune* database to populate the morphospace through time and apply sampling-standardization methods at the level of occurrences. We use principal coordinates analysis (PCO) to transform the data to continuous form, and binned occurrences into 2-Myr time intervals to calculate four disparity metrics describing the occupancy of this morphospace: the volume of the convex hull encompassing the taxa present, the volume of an alpha shape encompassing the taxa present, the alpha shape volume divided by the number of taxa, and the mean pairwise distance (measured as the number of character state mismatches divided by the number of possible matches). The first two measure the total amount of morphospace occupied, while the last two measure how close taxa are to one another in morphospace.

All analyses were carried out in the statistical programming language R (R Development Core Team 2011); the software written to carry out the analyses and produce the figures shown is provided in the online supplement. A detailed description of the method of morphospace construction, including the choice of morphological characters and the calculation of disparity metrics, is given in the companion paper in this issue (Kotrc and Knoll submitted).

Morphospace Subsampling

We carried out sampling standardization using four different subsampling methods: "classical" rarefaction (CR, Miller and Foote 1996), by-list unweighted (UW, Alroy et al. 2001), by-list weighted by occurrences (OW, Alroy 1996), and shareholder quorum subsampling (Alroy 2010b). These methods are reviewed in detail by Alroy (2010a) and are only briefly outlined here.

In each of these methods, occurrences are drawn from the full dataset until a given quota is reached. Morphospace metrics are calculated on this subsample and the process is repeated many times; the mean and confidence intervals of these iterations are reported. In CR, occurrences are drawn individually until a quota of a number of occurrences is reached. In UW, occurrences are drawn by taxonomic list (a list of taxa reported from one slide at one depth in one borehole) until a quota of a given number of lists is drawn. In OW, occurrences are also drawn by-list, but the quota is a given number of occurrences. These subsampling methods are the same as those carried out by Rabosky and Sorhannus (2009), although we do not apply O^2W subsampling (in which the quota is a sum of squared occurrences) due to the strong biases in that method when beta diversity is non-negligible, as demonstrated by Bush et al. (2004). Also, since we require a list of taxon names present in each subsample—rather than just the number of

178 taxa—in order to calculate metrics of morphospace occupancy, we do not apply a sampling
179 probability correction (the “three-timer” correction of Alroy et al. 2008).

180 These methods of sampling standardization seek to achieve uniform sampling through
181 time, but Alroy (2010a, 2010b) has argued that uniform sampling is not necessarily fair
182 sampling. He suggested that fair sampling should sample the same proportion of total diversity in
183 each interval—meaning that a more diverse interval will require more sampling than a less
184 diverse interval to accurately recover their relative diversities (if the shape of their rank-ordered
185 occurrence distributions is the same). He proposed a new sampling standardization method,
186 shareholder quorum subsampling (SQS), which hinges upon estimating the proportion of total
187 diversity represented by a sample. This is achieved using Good’s *u* (Good 1953), a metric from
188 ecology that uses the prevalence of singletons in a sample as an indication of coverage. Alroy
189 modified this for use in SQS by substituting taxa occurring in a single publication in place of
190 singletons. The *Neptune* database, however, does not include direct information about source
191 publications, and in any case, the source publications rarely contain singleton occurrences
192 because of the way micropaleontological data are collected (they report occurrences of a set of
193 taxa over a stratigraphic range). We thus apply a further modification to this estimate,
194 substituting for single-publication taxa the number of taxa occurring in only one DSDP/ODP
195 borehole. We also neither apply the largest collection correction nor do we discard the most
196 abundant taxon in each sample because we do not consider the related biases to apply to the
197 *Neptune* data. Finally, the current version of our software does not implement the “throwback"
198 refinement of Alroy (2010a, 2010b), meaning that each subsample will have a quorum level
199 slightly exceeding the target.

Because our morphospace is constructed at the genus level with some taxa excluded (see companion paper), we report both the genus richness recovered by the morphospace subsampling exercises as well as the species richness obtained in separate subsampling of the complete *Neptune* data.

Analysis

Distribution of Occurrences in Morphospace

Visualizations conventionally display the occupation of morphospace in a binary fashion: a point in morphospace is either occupied by a taxon—shown by a point plotted in the chosen ordination at the location representing the taxon's morphology—or it is not. When an occurrence-level database is used to populate the morphospace, however, an additional dimension of information can be shown by representing the number of occurrences of each taxon by the size of each plotted point.

Plotting Cenozoic diatom morphospace occupation in this way shows that some areas of morphospace are more sparsely occupied than others in terms of fossil occurrences (Fig. 1). In our companion paper, we pointed out the possibility that the Cenozoic rise in the number of *Neptune* occurrences might bias our metrics of disparity (Kotrc and Knoll, submitted). Figure 1 gives a more nuanced view of the need to consider sampling differences, making it clear that some regions of morphospace are occupied by few occurrences. Had the younger intervals been sampled less, at a level comparable to the Paleocene, those regions might have been seen as unoccupied.

Comment [BK1]: Fig. 1 here

The same observation could have been made by simply comparing rank-ordered abundance distributions for different *Neptune* time bins. But Figure 1 suggests something further: that these occurrences may not to be randomly distributed in morphospace, at least as

223 viewed through the first two PCO axes. In the Paleogene time bins, taxa defining the edges of
224 occupied morphospace appear to have relatively many occurrences. In contrast, in the Neogene
225 (and particularly the Plio-Pleistocene), the edges of morphospace are largely occupied by taxa
226 with few occurrences. This observation calls into question the interpretation of disparity metrics
227 based on the range or volume of morphospace occupied (Figs. 8B and C of the companion paper)
228 and suggests the possibility that, under sampling comparable to older time bins, the younger time
229 bins might not have shown the observed increase in the total extent of occupied morphospace.
230 We note that it ought to be possible (and may be interesting) to formulate a metric
231 describing the evenness of morphospace occupation, a morphological equivalent of the concept
232 of taxonomic evenness. This could be done, for example, by analogy to the ESS metric (Peters
233 2004), or by comparison to stochastic simulation of random partitioning of occurrences in
234 morphospace. Foote (1995: p. 283) adapted Shannon's information or evenness metric to
235 pairwise character-state combinations to describe morphospace; a similar application of this
236 metric might also be made to the occurrence-evenness in morphospace. In the following section,
237 we address the question of whether sampling differences might account for observed changes in
238 metrics of morphological disparity (Fig. 8 of the companion paper) by applying sampling
239 standardization methods to the diatom morphospace.
240
241 Subsampling of Morphospace
242 *Taxon Counting.*—Before delving into subsampling, it is worth considering how we
243 construct a list of the taxa present in a time bin from raw data of fossil occurrences. Curves of
244 diatom taxonomic diversity have conventionally been compiled using the range-through method
245 of taxon counting (Spencer-Cervato 1999; Rabosky and Sorhannus 2009), in which a taxon is

246 counted as present in any time bin between its first and last appearance, regardless of whether or
247 not it is actually observed in that time bin. Metrics of morphological disparity and taxonomic
248 diversity for the Cenozoic diatom morphospace under the range-through method are shown in
249 Figure 2.

250 Range-through (RT) taxon counting is intuitively appealing, because we know taxa must
251 have been extant between their first and last appearances. However, this method has fallen out of
252 favor because it has been shown to suffer from a number of significant biases (such as the
253 Signor-Lipps effect and other edge effects) that distort the form of the resulting diversity curve
254 (reviewed in Alroy 2010a). An alternative method counts only those taxa actually observed in a
255 time bin (sampled in-bin, SIB). Although SIB taxon counting underestimates standing diversity
256 in time bins with poor sampling, its immunity to many of the other biases affecting the RT
257 method has led Alroy (2010a) to champion it as the preferred method of taxon counting; this
258 view is not universally shared, however, and other methods are also favored for their particular
259 strengths. Disparity and diversity metrics for the Cenozoic diatom morphospace using SIB taxon
260 counting are shown in Figure 3.

261 Comparing the disparity metrics calculated under RT (Fig. 2) and SIB (Fig. 3) illustrates
262 that the method of taxon counting does not affect the first-order patterns observed. In both
263 figures, metrics of the separation between taxa in morphospace (A, D) are approximately
264 stationary through time, while metrics of the total volume of morphospace occupied (B, C) show
265 an increase with time. However, the curves drawn under SIB are noisier, while the RT curves are
266 smoother, reflecting bin-to-bin differences in sampling (with intervals of poor sampling, perhaps
267 due to differences in preservation, masked by the RT method). Besides the obvious sampling gap
268 in the earliest Eocene, for which no diatom data are present in the *Neptune* database, these “dips"

Comment [BK2]: Fig. 2 here

Comment [BK3]: Fig. 3 here

in the SIB curves (relative to the RT curve) also highlight the potential of sampling bias to influence the disparity metrics. The dips at 47 Ma and 39 Ma in the SIB diversity curve, for example, have corresponding dips in the convex hull and alpha shape volume curves, but these dips are absent in the RT curves. Since we thus know these dips are due to sampling (taxa not counted but known to have existed), this further reinforces the need to correct for sampling before interpreting disparity metrics, particularly those describing the volume of occupied morphospace.

Uniform Subsampling.—Under CR subsampling (Fig. 4), measures of taxonomic diversity and some measures of morphological disparity show different temporal trajectories from those under SIB (Fig. 3). Rabosky and Sorhannus (2009) described Cenozoic diatom diversity under various methods of subsampling in detail, so we go no further here than to confirm that our results (Fig. 4E) agree: we find a much-attenuated, roughly twofold rise in diversity, compared to the fourfold rise under SIB (Fig. 3E), over the course of the Cenozoic Era. Peak diversity under CR is reached in the early Oligocene Epoch (rather than in the Pleistocene under SIB), with a more pronounced Oligocene diversity crash and a subsequent recovery to early Oligocene-level diversity through the remainder of the Cenozoic Era.

Comment [BK4]: Fig. 4 here

The metrics of morphological disparity describing the distance separating taxa in morphospace show much the same trajectory under CR (Fig. 4A and D) as under SIB (Fig. 3A and D). The per-genus volume of morphospace occupied (Fig. 4D) shows a stationary pattern through time, much as under SIB (Fig. 3D). Similarly, mean pairwise distance (Fig. 4A) shows a broadly stationary pattern, much as under SIB (Fig. 3A), albeit with a less pronounced peak in the mid-Oligocene and a more accentuated Oligocene-Miocene trough.

In contrast, those metrics of disparity describing the total volume of morphospace occupied show results under CR (Fig. 4B and C) that are qualitatively different than under SIB (Fig. 3B and C). Both convex hull volume (Fig. 4B) and alpha shape volume (Fig. 4C) show a broadly stationary trajectory under CR, compared to the twofold increase under SIB. Although there is an increase in occupied volume from the Paleocene to the Eocene in both the CR and SIB results, the subsequent trajectory is flat under CR where there is an increase under SIB. The spikes in occupied volume at the 41, 29, and 12 Ma time bins are attenuated under CR, perhaps because taxa responsible for an expansion of occupied space, located at the extremes of morphospace, are sampled only in some of the subsampling iterations.

The results for UW and OW subsampling are very similar to those for CR (results for these analyses are thus provided in the online supplement).

In summary, all disparity metrics show broadly stationary patterns when based on *Neptune* occurrence data subsampled to a uniform sampling level. Those disparity metrics describing the separation among taxa in morphospace (mean pairwise distance and mean alpha shape volume occupied per list) do not change substantially compared to the raw (SIB) results, while those metrics describing the volume of morphospace occupied (by convex hull and alpha shape) lose the increasing trend seen under SIB when subsampled.

Subsampling by SQS.—Although SQS is conceptually distinct from the uniform item quota subsampling methods (CR, UW, and OW), the morphospace metrics calculated under our version of SQS (Fig. 5) are similar to those obtained through the other methods. Under SQS, mean pairwise distance (Fig. 5A) shows a generally stationary pattern (again with a very slight net decline representing at most a few percentage points in dissimilarity), much as in the other analyses.

Comment [BK5]: Fig. 5 here

Convex hull volume also shows a generally stationary pattern under SQS (Fig. 5B), albeit with slightly more variability than under CR. Alpha shape volume through time (Fig. 5C) also shows greater amplitude variability under SQS than CR, and although the net increase over the Cenozoic is still far less than under SIB, there is a clearer increase under SQS than under CR. However, this increase may be an artifact of the choice of α parameter; this was chosen at α=0.11 to optimally describe the arrangement of taxa in the raw dataset and may not adequately capture morphospace occupancy of smaller subsamples with a different arrangement of taxa. Indeed, volumes calculated with higher values of α (0.2 and 10, upper grey curves in Fig. 5D) show a more stationary pattern.

Finally, per-genus alpha shape volume (Fig. 5D) shows a stationary pattern over much of the Cenozoic under SQS, similar to the results under CR, though volumes in the Paleocene time bins and one Eocene time bin are lower under SQS than under CR, which suggests a slight increase over time.

The Cenozoic trajectory of taxonomic diversity (Fig. 5E) is greatly flattened, much as in the uniform subsampling method results. However, the Eocene-Oligocene peak in diversity under SQS greatly exceeds the diversity recovered subsequent to the Oligocene. In this regard, the SQS diversity curve resembles the O^2W curve presented by Rabosky and Sorhannus (2009).

It should be noted that the diatom diversity curves obtained by subsampling methods have not been universally accepted by micropaleontologists (Lazarus et al. 2012). A criticism of these methods, including SQS, is that they can perform poorly under changes in relative frequency distributions. In essence, if relative frequencies are evenly distributed to begin with and become very uneven through time, subsampling could significantly underestimate diversity in the more uneven intervals. A similar concern has been raised with regard to increases in

provinciality through time (changes in β diversity), and an alternative diversity curve more similar to the canonical view (Spencer-Cervato 1999) has been put forth by Lazarus et al. (2012), who added empirical correction factors to subsampled diversity curves to account for changes in evenness and provinciality.

Summary of Subsampling Results.—The results of morphospace analyses under different subsampling methods show the following:

1. When sampling bias is corrected by randomized subsampling, all disparity metrics show stationary patterns or, at most, directional changes of small magnitude (a small decrease in mean pairwise distance in all analyses and a small increase in occupied volume under SQS).
2. Morphological diversification in Cenozoic diatoms is described as stationary once sampling differences are taken into account. This is true for both measures of average morphological distances among taxa and the total range of morphologies explored, and is in agreement with the results of the comparison of morphological with molecular and phylogenetic distance in our companion paper.
3. Disparity metrics describing the average dispersion of taxa in morphospace (mean pairwise distance and per-genus alpha shape volume) are less sensitive to sampling bias than metrics describing the total extent of morphospace occupied (convex hull and alpha shape volume).

By using subsampling methods, we seek to discover something about the nature of morphological diversification by correcting for differences in sampling. In the following section, we pursue the same goal using a different approach, by examining aspects of the data that are independent of secular variations in sampling.

360

361 Occupied Morphospace Per List

362 An alternative means to overcome the problem of sampling bias is to look at measures of
363 morphological disparity calculated for individual lists (the sets of taxa reported from a particular
364 depth in a particular borehole). A helpful context for this approach is to consider the notions of α
365 and β taxonomic diversity.

366 We can consider global taxonomic diversity, S, to consist of a local component—
367 described by the average length of a taxonomic list at a particular location, □—and a component
368 describing how different any given list is from another. A useful definition relating these
369 components is that of Whittaker (1960), who defined this β component as β = S / □. Using this
370 definition, we can consider β diversity as the number of potentially unique communities (or the
371 number of nonoverlapping lists of average list length). By rearranging this expression as S = □
372 × β, it becomes clear that changes in observed global diversity can be due to either changes in the
373 per-locality diversity or changes in the taxonomic similarity among localities (or some
374 combination of the two). Such changes in the components of global taxonomic richness have
375 been explored, for example, in Paleozoic marine animals (Sepkoski 1988).

376 By analogy, we can think of geographic structure in morphological disparity, consisting
377 of a component describing the local morphological disparity ("α disparity") and a component
378 describing how morphologically different communities are from one another ("β disparity"). We
379 calculated the average α disparity for both mean pairwise distance and occupied convex hull
380 volume as per-list disparity metrics in each time bin (Fig. 6).

Comment [BK6]: Fig. 6 here

381 Both average per-list convex hull volume (Fig. 6A) and per-list mean pairwise distance
382 (Fig. 6B) show broadly stationary patterns, though the latter shows a slight decline through time

383 (as does mean pairwise distance at the global level, Fig. 3A). Although methodological bias
384 toward constant list length during data collection has been suggested for micropaleontological
385 data (Lazarus 2011), such a bias would simply imply that these results have been standardized
386 for secular changes in taxonomic diversity. These results are consistent with the largely
387 stationary patterns observed at the global scale under subsampling and support an overall picture
388 of Cenozoic diatom morphological evolution that is unchanging.

389 The per-list volume results (Fig. 6A) also suggest that the increase in occupied
390 morphospace volume seen at the global scale in the raw data (Figs. 3B and C) must have a
391 spatial component: if the increase in occupied morphospace volume is not due to an increase in
392 the volume occupied by individual lists (and, by extension, by local assemblages), it stands to
393 reason that the increase reflects the addition of more lists occupying similar-sized but non-
394 overlapping volumes of morphospace. As explained above, we describe this as a rise in β
395 disparity. One might imagine that the increasing latitudinal temperature gradients observed
396 through the Cenozoic Era (Zachos et al. 2001) might contribute to such an increase.

397 We note that we have introduced here a multiplicative notion of disparity partitioning by
398 analogy to Whittaker's (1960) scheme for taxonomic diversity. We can also conceive of the total
399 disparity as the sum of local disparities, leading to an additive disparity partitioning concept
400 analogous to that proposed by Lande (1996) for taxonomic diversity. In this approach, where α
401 and β are related by S = □ + β (Veech et al. 2002), beta diversity represents the portion of total
402 diversity absent from an average assemblage. By extension, beta disparity represents the portion
403 of total disparity absent from an average assemblage—for example, the difference between total
404 occupied morphospace volume and that occupied by one list. This approach might impart several
405 desirable properties on α and β championed for the additive scheme in the context of taxonomic

406 diversity, such as sharing the same units, allowing sampling at different hierarchical levels, and
407 enabling the computation in percentages of a taxon's contribution to each component (Holland
408 2010).

409 Though we can confidently infer this rise in β disparity in our data, we cannot determine
410 whether it represents a true geographic differentiation in diatom disparity or whether this is an
411 artifact of the secular increase in the number of lists sampled. Nonetheless, we can rule out
412 Cenozoic morphological diversification at the local scale, finding instead a stationary pattern in α
413 disparity consistent with that in our other results.
414
415
416 Sensitivity of Results to Methodological Choices

417 Constructing a morphospace and using it to measure secular changes in disparity involves
418 numerous methodological choices. We have already investigated the effect of one of these
419 choices, the taxon counting method. In the following, we test the sensitivity of our disparity
420 metrics to further important methodological choices that are commonly unexamined: how to find
421 a low-dimensional representation of the morphospace (the choice of ordination method) and how
422 much incomplete data to reject before constructing the morphospace (the choice of data culling
423 threshold).

424 *Ordination Method.*—In order to investigate the sensitivity of our results to the choice of
425 ordination method, we repeated the calculation of convex hull volumes and alpha shape volumes
426 through time using another ordination method commonly used in morphospace studies (e.g., by
427 Huntley et al. 2006; Shen et al. 2008): non-metric multidimensional scaling (NMDS). Unlike
428 PCO, NMDS is not an eigenvector method; rather, a fixed number of dimensions is chosen a

priori and the best representation of the data in those dimensions is found numerically. The method proceeds through successive iterations until an acceptable (but not necessarily unique or optimal) solution is found. We carried out this analysis using the *isoMDS()* function from the MASS package (Venables and Ripley 2002). An analysis using the *metaMDS()* function from the vegan package (Oksanen et al. 2013), which uses a variety of different starting configurations for the NMDS algorithm, gave very similar results.

Figures 7A and C show the resulting comparison of morphospace volume metrics calculated using NMDS with three dimensions specified (red points) and the first three PCO axes (blue points). The results are very similar, and when the timeseries resulting from one ordination procedure is plotted against those resulting from the other (Figs. 7B and D), the closeness of this correlation can be summarized with an R^2 value (0.90 and 0.93 for convex hull and alpha shape volumes, respectively).

Comment [BK7]: Fig. 7 here

These results suggest that metrics of occupied morphospace volume are not sensitive to the choice of ordination method.

Data Culling.—Virtually all paleontological datasets contain missing data, and this is particularly true of those used to construct morphospaces. The possible causes of missing entries in the morphospace matrix used here are discussed in more detail in the companion paper in this issue, but a crucial question at the outset of a morphospace study is: how much valid data should a genus or character have to be included in the analysis? The edge cases are trivial to decide: a genus with no valid character states or a character with no valid entries for any genus adds no information and obviously ought to be excluded. Likewise, genera and characters with entirely valid entries ought to be included. Where the line is drawn in between these extremes is to some extent an arbitrary decision; in this study, we chose a threshold of 80% completeness.

452 In order to investigate the sensitivity of our results to different choices of data quality
453 threshold, we repeated our analysis under the entire range of completenesses represented in our
454 data, ranging from including all the data collected at one extreme (a threshold of 57% or more of
455 observed states) to including only complete genera and characters at the other (a threshold of
456 100% observed states). The data culling algorithm was applied in the same manner as in the
457 companion paper, removing first characters and then genera until both reached the desired
458 threshold of data quality. As before, we only considered unobserved entries in calculating
459 completeness, since we consider the other types of missing data (the cases where states are either
460 inapplicable to a taxon or where multiple states apply) to constitute important information.
461 We compared the convex hull volume and mean pairwise distance results obtained under
462 each data quality threshold to the “reference results" under the 80% threshold presented in the
463 companion paper. Rather than plotting the timeseries for each of these comparisons (like in Figs.
464 7A and C), we summarized each comparison using the R^2 correlation coefficient (like in Figs. 7B
465 and D).
466 The R^2 values summarizing the comparison of analyses under each data quality threshold
467 with the reference results are shown in Figure 8. These results show that neither convex hull
468 volume (Fig. 8A) nor mean pairwise distance (Fig. 8B) are sensitive to the addition of more data
469 of lower quality. Even setting the most permissive threshold (including all the data collected)
470 yields time series that are highly correlated (R2>0.9) with the reference results. The results also
471 remain correlated above ~0.9 as data are removed until the data quality threshold exceeds about
472 90% completeness, beyond which correlations decline. Results under the most stringent data
473 quality threshold (100% complete characters and genera only) show relatively weak correlations

Comment [BK8]: Fig. 8 here

of only 0.5–0.6. Mean pairwise distance appears to be more sensitive than convex hull volume to changes in data quality threshold.

In order to clarify whether the results with R^2 values suggesting weak correlations with the reference results are in fact qualitatively different, we plotted a comparison between the results using the most stringent data quality threshold (100% completeness, with only 32 characters retained) and the reference results (80% completeness), shown in Figure 9. Despite the low R^2 values, the results are qualitatively similar. Convex hull volume increases in both cases (Fig. 9A) while mean pairwise distance remains roughly constant in both cases (Fig. 9C), although the absolute values of distance are lower under the more stringent threshold.

Comment [BK9]: Fig. 9

Testing Character Sets for Specific Evolutionary Hypotheses

We have thus far approached our goal, to make biological inferences about the morphological evolution of the diatom frustule, by summarizing morphological data and abstracting it through metrics of disparity and correcting those measures for sampling differences. These results all seem to point towards a stationary Cenozoic pattern.

The morphological data set underlying the morphospace analysis also permits an analysis of morphological evolution from a fundamentally different approach, if we momentarily set aside concerns about sampling. Rather than examining the data abstractly and in aggregate, we can analyze the morphological data directly to examine how the prevalence of taxa with different sets of morphological characters has changed through time. A similar approach has previously been used to categorize Phanerozoic animals by anatomical and ecological traits to document major shifts in the proportions of, for example, physiologically unbuffered to physiologically buffered taxa, or predator to non-predator taxa (Bambach et al. 2002). These categories were

497 associated with evolutionary hypotheses about mass-extinction kill mechanisms (Knoll et al.
498 1996) and ecological escalation (Vermeij 1987), respectively.

499 By analogy, we can parse our morphological data a priori by criteria related to
500 hypothesized drivers of diatom evolution. For example, predation has been suggested to play an
501 important role in diatom evolution (Smetacek 2001; Hamm and Smetacek 2007) and we can
502 identify characters that might relate to defense against predation, like spines and projections or
503 ribs and costae buttressing and strengthening the valve. Then, we can investigate whether the
504 prevalence of these characters changed through time, as would be expected under the
505 hypothesized selective pressure. If we were able to detect systematic changes in the proportion of
506 character states expected under a given scenario, we might question the stationary pattern
507 suggested by the subsampling exercises and the alpha disparity results above.

508 We assembled four lists of characters expected to change under changes in four factors
509 that have been identified as central to Cenozoic evolution in diatoms: predation, sinking (Raven
510 and Waite 2004), viral attack (Smetacek 1999), and silica availability (Finkel and Kotrc 2010).
511 For each chosen character, we sorted character states into one of two categories: either favorable
512 or unfavorable, with respect to the particular hypothesis (e.g., for predation, character states
513 indicating possession of spines were assigned to the favorable category, those states indicating
514 absence of spines to the unfavorable category). The complete listing of characters and assigned
515 states are tabulated in the online supplement.

516 The results (Fig. 10) show a remarkable absence of trends through time. The proportion
517 of morphological character states thought to be associated with specific hypothesized drivers of
518 evolution in diatoms are essentially constant through Cenozoic time. These results portray
519 untrended morphological evolution that is consistent with the other lines of evidence presented

Comment [BK10]: Fig. 10 here

here. However, we note that the absence of trends in these characters do not necessarily imply a lack of response to these selective pressures, since some responses may simply not be visible in our data. For example, our morphospace does not capture changes in cell size, although this may be an important factor in mechanical strength and thus predation resistance (Hamm et al. 2003), and (Finkel et al. 2005) documented a Cenozoic decrease in the size of diatom frustules that may point to just such a response.

Conclusions

The substantial Cenozoic rise in sampling through time calls into question the marine planktonic diatom disparity results presented in our companion paper, which show a rise in occupied morphospace volume, in contrast to the stationary pattern seen in all other metrics. Two further analyses presented here highlight the need to take sampling differences into account before interpreting disparity metrics. First, the differences between volume-based disparity metrics calculated under different methods of taxon sampling (SIB and RT) suggest that these metrics are affected by sampling. Second, illustrating the number of occurrences represented by each taxon in a morphospace plot shows that morphospace is occupied unevenly and raises the possibility that less-intensive sampling of more recent time bins may have led to lower reported volumes of morphospace occupation.

The plotting of morphospace occupation "density" permitted by the use of an occurrence-based database leads us to formulate a notion of "morphological evenness." Analogous to taxonomic evenness, which describes the distribution of individuals (or, in paleontological studies, occurrences) among taxa, morphological evenness would describe the distribution of individuals (or occurrences) in morphospace. Any given abundance distribution could be, at one

543 extreme, randomly distributed throughout morphospace; at the other extreme, occurrences could
544 be preferentially concentrated in one area. We suggest that quantifying this notion would be an
545 interesting target for future work.

546 In order to address the potential sampling bias identified in these ways, we recalculate the
547 disparity metrics presented in the companion paper in this issue under various methods of
548 subsampling. We find that, under subsampling, the increases in occupied morphospace volume
549 seen in the unsubsampled results largely disappear, and all disparity metrics show essentially
550 stationary results, (consistent with the known susceptibility of these metrics to sampling bias).
551 These results suggest a morphologically untrended Cenozoic when sampling differences are
552 corrected in this fashion.

553 Comparing the disparity metrics calculated under subsampling to those calculated from
554 the data directly suggests that the metrics describing the volume of occupied morphospace are
555 more sensitive to sampling differences than those describing the distances among taxa (or, put
556 another way, their dispersion in morphospace). These results agree with the findings of Butler et
557 al. (2012) and Ciampaglio et al. (2001), albeit using different metrics.

558 In seeking a direct measure of disparity insensitive to sampling intensity, we introduce
559 the concept of a geographic component to morphological disparity. By analogy to Whitaker's
560 (1960) α and β components of taxonomic diversity, we suggest the notions of α and β disparity.
561 We find that mean α disparity (as quantified by the mean of either convex hull volumes or the
562 mean pairwise distances across lists) remains roughly constant through time. These results
563 support untrended diatom morphological evolution through the Cenozoic Era.

564 Constant mean α disparity through time is compatible with the observations of roughly
565 constant total disparity under subsampling. If the subsampling results were to be rejected in favor

566 of the results in the companion paper, however (see caveats below), constant mean α disparity
567 would imply that the rise in total disparity resulted from an increase in β disparity.

568 As a by-product of applying subsampling methods to diatom morphospace, we present a
569 taxonomic diversity curve of diatoms under SQS based on the *Neptune* database. We find results
570 similar to other subsampling methods, with a flattened diversity curve showing peak diversity
571 near the Eocene/Oligocene boundary and a pronounced Oligocene decline in diversity. In the
572 SQS diversity curve, this Eocene/Oligocene peak far exceeds the species richness recovered
573 subsequently, and is thus most similar to the O^2W results reported by Rabosky and Sorhannus
574 (2009).

575 The diatom diversity curves obtained by subsampling methods, however, have not been
576 universally accepted by micropaleontologists (Lazarus et al. 2012), because they can perform
577 poorly under changes in relative frequency distributions. As we have presented a more detailed
578 discussion of this issue elsewhere (Kotrc and Knoll in review), it is sufficient here to point out
579 that the stationary results of the volume-based disparity metrics and the sampling-corrected
580 diversity curves are dependent on whether subsampling methods are believed to provide a more
581 accurate view than the raw data, or whether they simply trade on bias for another. The other
582 untrended results, however—the distance-based disparity metrics, the disparity metrics per-list,
583 and the comparison of morphospace to molecular phylogeny—do not depend on subsampling.

584 In a sensitivity test comparing our morphospace volume results using PCO to those using
585 NMDS, a substantively different, non-eigenvector ordination method, we find similar results in
586 both and conclude that our results are not sensitive to ordination method. In a similar sensitivity
587 test repeating our analyses after culling more or less of the data by completeness, we find that
588 our results are also robust to choices in data quality threshold.

589 In summary, when sampling biases are taken into account using subsampling methods as
590 well as sampling-independent metrics of disparity, our results point toward unchanging Cenozoic
591 occupation of planktonic diatom morphospace. This suggests diatoms had reached peak disparity
592 by the early Cenozoic Era, while taxonomic diversity continued to rise, albeit more gradually
593 than the canonical diversity curve would suggest. Though we have not analyzed diversity and
594 disparity from the origin of the clade, our results point to a decoupling of taxonomic and
595 morphological diversification akin to the “asymmetric diversification" reported for many other
596 groups.
597 More broadly, these results make clear that a complete view encompassing all aspects of
598 morphological disparity must consider sampling biases. The use of occurrence-based databases
599 to populate morphospaces allows these biases to be addressed using well-established,
600 quantitative methods.
601

602 **Acknowlegements**

603 Research leading to this paper was made possible by PlanktonTech, an integrative
604 research project sponsored by the Helmholtz Geminschaft and led by Christian Hamm at the
605 Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany. We are
606 grateful to D. Rabosky and J. Alroy for providing their code for subsampling analyses and thank
607 J. Dumais and C. Marshall for helpful discussions at an early stage. For helpful comments on
608 various stages of the manuscript, we thank B. Kock, J. Creveling, and W. A. Green, who also
609 provided much computational assistance. We thank M. Foote and G. Hunt for constructive and
610 helpful reviews.
611

612
613 **Literature Cited**

614 Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in
615 North American mammals. Palaeogeography, Palaeoclimatology, Palaeoecology
616 127:285–311.

617 ———. 2010a. Fair sampling of taxonomic richness and unbiased estimation of origination and
618 extinction rates. In J. Alroy and G. Hunt, eds., Quantitative Methods in Paleobiology,
619 Paleontological Society Papers, 16:55–80.

620 ———. 2010b. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine
621 diversification. Palaeontology 53:1211–1235.

622 Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezusko, M. Foote, F. T. Fürsich, T. A. Hansen, S.
623 M. Holland, L. C. Ivany, D. Jablonski, et al. 2001. Effects of sampling standardization on
624 estimates of Phanerozoic marine diversification. Proceedings of the National Academy of
625 Sciences 98:6261.

626 Alroy, J., M. Aberhan, D. Bottjer, M. Foote, F. Fürsich, P. Harries, A. Hendy, S. Holland, L.
627 Ivany, W. Kiessling, et al. 2008. Phanerozoic trends in the global diversity of marine
628 invertebrates. Science 321:97–100.

629 Bambach, R. 1999. Energetics in the global marine fauna: A connection between terrestrial
630 diversification and change in the marine biosphere. Geobios 32:131–144.

631 Bambach, R. K., A. H. Knoll, and J. J. Sepkoski. 2002. Anatomical and ecological constraints on
632 Phanerozoic animal diversity in the marine realm. Proceedings of the National Academy
633 of Sciences 99:6854–6859.

634 Bush, A., M. Markey, and C. Marshall. 2004. Removing bias from diversity curves: the effects
635 of spatially organized biodiversity on sampling-standardization. Paleobiology 30:666–
636 686.
637 Butler, R., S. Brusatte, B. Andres, and R. Benson. 2012. How do geological sampling biases
638 affect studies of morphological evolution in deep time? A case study of pterosaur
639 (Reptilia: Archosauria) disparity. Evolution 66:147–162.
640 Ciampaglio, C. N., M. Kemp, and D. W. McShea. 2001. Detecting changes in morphospace
641 occupation patterns in the fossil record: characterization and analysis of measures of
642 disparity. Paleobiology 27:695–715.
643 Darwin, C. 1859. On the Origin of the Species by Means of Natural Selection: Or, The
644 Preservation of Favoured Races in the Struggle for Life. John Murray, London.
645 DeMaster D. J. 2003. The Diagenesis of Biogenic Silica: Chemical Transformations Occurring
646 in the Water Column, Seabed, and Crust *in* Holland HD and Turekian KK (eds.). Treatise
647 on Geochemistry 7:87–98.
648 Erwin, D. 2007. Disparity: morphological pattern and developmental context. Palaeontology
649 50:57–73.
650 Fenner, J. 1985. Late Cretaceous to Oligocene planktic diatoms *in* Bolli, H. M., Saunders, J. B.
651 and Perch- Nielsen, K. (eds) Plankton Stratigraphy. Cambridge University Press,
652 Cambridge, 713–762.
653 Finkel, Z. and B. Kotrc. 2010. Silica use through time: macroevolutionary change in the
654 morphology of the diatom frustule. Geomicrobiology Journal 27:596–608.
655 Finkel, Z. V., M. E. Katz, J. D. Wright, O. M. Schofield, and P. G. Falkowski. 2005. Climatically
656 driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic.

Proceedings of the National Academy of Sciences of the United States of America 102:8927–8932.

Foote, M. 1992. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology 18:1–16.

———. 1995. Morphological diversification of Paleozoic crinoids. Paleobiology 21:273–299.

———. 1997. The evolution of morphological diversity. Annual Review of Ecology and Systematics 28:129–152.

Good, I. 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40:237–264.

Gould, S. J. 1989. Wonderful Life: The Burgess Shale and the Nature of History. W.W. Norton.

Hamm, C. and V. Smetacek. 2007. Armor: Why, When, and How. In P. G. Falkowski and A. H. Knoll, eds., Evolution of Primary Producers in the Sea. Elsevier, Burlington, MA.

Hamm, C., R. Merkel, O. Springer, P. Jurkojc, C. Maier, K. Prechtel, and V. Smetacek. 2003. Architecture and material properties of diatom shells provide effective mechanical protection. Nature 421:841–843.

Holland, S. M. 2010. Additive diversity partitioning in palaeobiology: revisiting Sepkoski's question. Palaeontology, 53(6):1237-1254.

Huntley, J. W., S. Xiao, and M. Kowalewski. 2006. 1.3 billion years of acritarch history: An empirical morphospace approach. Precambrian Research 144:52–68.

Knoll, A. H., R. Bambach, D. Canfield, J. Grotzinger, et al. 1996. Comparative Earth history and Late Permian mass extinction. Science 273(5274):452–457.

Lande, R. 1996. Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. Oikos 76(1):5–13.

680 Lazarus, D. 2011. The deep-sea microfossil record of macroevolutionary change in plankton and
681 its study. Geological Society, London, Special Publications 358:141–166.
682 Lazarus, D. B., Kotrc, B., Wulf, G., & Schmidt, D. N. 2009. Radiolarians decreased silicification
683 as an evolutionary response to reduced Cenozoic ocean silica availability. Proceedings of
684 the National Academy of Sciences. 106(23): 9333-9338.
685 Lazarus, D., J. A. Barron, A. Türke, P. Diver, and J. Renaudie. 2012. Diversity history of
686 cenozoic planktic marine diatoms. In The Micropalaeontological Society AGM and
687 Warm World Symposium, British Geological Survey, Nottingham, UK, Nov. 11th-13th.
688 The Micropalaeontological Society.
689 Lazarus, D. B. 1994. Neptune: a marine micropaleontology database. Mathematical Geology
690 26:817–832.
691 Lazarus, D. B. 2011. The deep-sea microfossil record of macroevolutionary change in plankton
692 and its study. In McGowan, A. J. and Smith, A. B. (eds) Comparing the Geological and
693 Fossil Records: Implications for Biodiversity Studies. Geological Society, London,
694 Special Publications, 358, 141–166.
695 Maldonado, M., Carmona, M. C., Uriz, M. J., and Cruzado, A. 1999. Decline in Mesozoic reef-
696 building sponges explained by silicon limitation. Nature 401(6755):785-788.
697 McGhee, G. R. 1999. Theoretical morphology: the concept and its applications. Columbia
698 University Press.
699 Miller, A. and M. Foote. 1996. Calibrating the Ordovician radiation of marine life: implications
700 for Phanerozoic diversity trends. Paleobiology 304–309.
701 Moore, J. T. C. 2008. Chert in the Pacific: biogenic silica and hydrothermal circulation.
702 Palaeogeography, Palaeoclimatology, Palaeoecology 261, 87–99.

703 Muttoni, G. and Kent, D. V. 2007. Widespread formation of cherts during the early Eocene
704 climatic optimum. Palaeogeography, Palaeoclimatology, Palaeoecology. 253:348–362.
705 Newell, N. 1959. Adequacy of the fossil record. Journal of Paleontology 488–499.
706 Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin P. R., O'Hara, R. B., Simpson,
707 G.L., Solymos, P., Stevens, M. H. H. and Wagner, H. 2013. vegan: Community Ecology
708 Package. R package version 2.0-10. http://CRAN.R-project.org/package=vegan
709 Pearson, E. S. 1926. A Further Note on the Distribution of Range in Samples Taken from a
710 Normal Population. Biometrika 18(1/2):173–194.
711 Peters, S. E. 2004. Relative abundance of Sepkoski's evolutionary faunas in Cambrian–
712 Ordovician deep subtidal environments in North America. Paleobiology 30:543–600.
713 R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R
714 Foundation for Statistical Computing, Vienna, Austria.
715 Rabosky, D. L. and U. Sorhannus. 2009. Diversity dynamics of marine planktonic diatoms
716 across the Cenozoic. Nature 457:183–186.
717 Racki, G., and Cordey, F. 2000. Radiolarian palaeoecology and radiolarites: is the present the
718 key to the past? Earth-Science Reviews, 52(1):83–120.
719 Raup, D. M. 1972. Taxonomic diversity during the Phanerozoic. Science 177:1065–1071.
720 Raven, J. A. and A. M. Waite. 2004. The evolution of silicification in diatoms: inescapable
721 sinking and sinking as escape? New Phytologist 162:45–61.
722 Sarmiento, J. L. and Gruber N. 2006. Ocean Biogeochemical Dynamics. Princeton University
723 Press, Princeton, NJ.
724 Sepkoski Jr, J. J. 1988. Alpha, beta, or gamma: where does all the diversity go? Paleobiology
725 221–234.

726 Shen, B., L. Dong, S. Xiao, and M. Kowalewski. 2008. The Avalon explosion: evolution of
727 Ediacara morphospace. Science 319:81.
728 Siever, R. 1991. Silica in the Oceans: Biological-Geochemical Interplay. In S. Schneider and P.
729 Boston (eds.), Scientists on Gaia, pp. 287–295. Massachusetts Institute of Technology,
730 Cambridge, MA.
731 Smetacek, V. 1999. Diatoms and the ocean carbon cycle. Protist 150:25–32.
732 ———. 2001. A watery arms race. Nature 411:745–745.
733 Spencer-Cervato, C. 1999. The Cenozoic deep sea microfossil record: explorations of the
734 DSDP/ODP sample set using the Neptune database. Palaeontologia Electronica 2.
735 Veech, J. A., Summerville, K. S., Crist, T. O., and Gering, J. C. 2002. The additive partitioning
736 of species diversity: recent revival of an old idea. Oikos 99(1):3–9.
737 Venables, W. N. and B. D. Ripley. 2002. Modern Applied Statistics with S. Springer, New York,
738 fourth ed. ISBN 0-387-95457-0.
739 Vermeij, G. 1987. Evolution and Escalation: An Ecological History of Life. Princeton University
740 Press.
741 Whittaker, R. 1960. Vegetation of the Siskiyou mountains, Oregon and California. Ecological
742 Monographs 30:279–338.
743 Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and
744 aberrations in global climate 65 Ma to present. Science 292:686–693.
745
746
747

748
749 **Figures**

750 Figure 1: [one- or two-column, print color] Morphospace plot of the first two PCO axes through
751 time, with the size of each plot point representing the number of occurrences of that taxon in the
752 *Neptune* database. Plot points are sized relative to the mean number of occurrences in each time
753 bin, shown (rounded to the nearest whole number) in the legend to the right of each time slice.
754 The colored polygons at the bottom of the plot are convex hulls enclosing the taxa present at
755 each time bin, labeled in the corresponding colors.

756

757

758 Figure 2: [full-page, print B&W, color online] Metrics of morphological disparity (A-D) and
759 taxonomic diversity (E) for the Cenozoic morphospace of marine planktonic diatoms, populated
760 using range-through (RT) taxon counting of *Neptune* database occurrences. A, Mean pairwise
761 distance between genera, (character state mismatches over possible matches). B, Convex hull
762 volume containing genera, normalized to largest value; black line is volume calculated over the
763 first three PCO axes, grey lines are volume over the first four, five, etc. up to ten PCO axes. C,
764 Alpha shape volume containing genera; black line is volume for α-value chosen by inspection to
765 best capture occupied volume across time bins, grey lines are other α-values. α=10 recovers the
766 convex hull solution. D, Alpha shape volume (as in C) divided by number of genera. E, Species-
767 level diversity from the *Neptune* database (includes taxa omitted from morphospace analysis) in
768 black; genus-level diversity in morphospace analysis in grey.

769

770

Figure 3: [full-page, print B&W, color online] Metrics of morphological disparity (A-D) and taxonomic diversity (E) for the Cenozoic morphospace of marine planktonic diatoms, populated using sampled-in bin (SIB) taxon counting of *Neptune* database occurrences. Metrics as explained in Fig. 2.

Figure 4: [full-page, print B&W, color online] Metrics of morphological disparity (A-D) and taxonomic diversity (E) for the Cenozoic morphospace of marine planktonic diatoms, populated using *Neptune* database occurrences subsampled to a quota of 100 occurrences by classical rarefaction with 10,000 iterations. Metrics as explained in Fig. 2; error bars show 95% confidence intervals of subsampling. Error bars omitted from genus diversity curve for clarity.

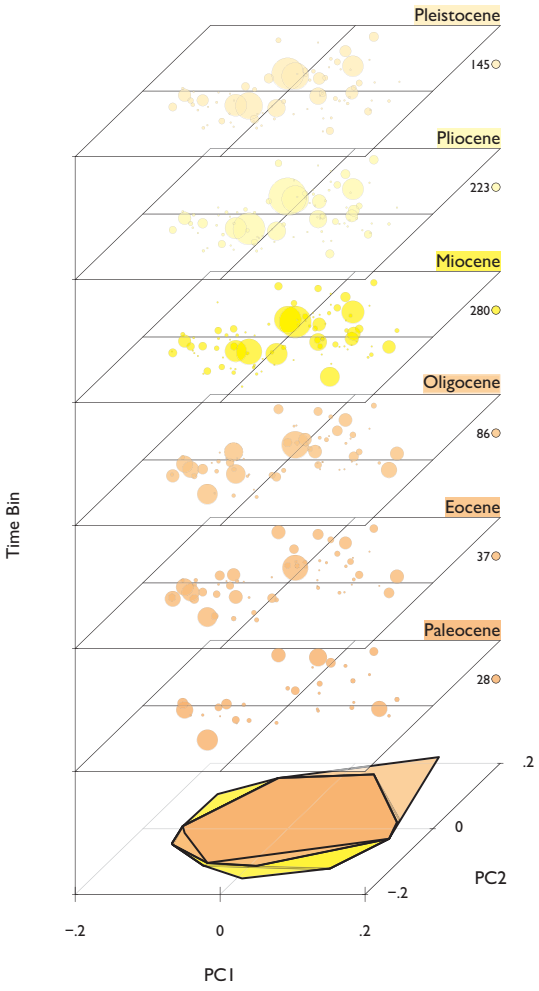
Figure 5: [full-page, print B&W, color online] Metrics of morphological disparity (A-D) and taxonomic diversity (E) for the Cenozoic morphospace of marine planktonic diatoms, populated using *Neptune* database occurrences subsampled by to a uniform coverage of 0.5 by shareholder quorum subsampling with 1,000 iterations. Metrics as explained in Fig. 2; error bars show 95% confidence intervals of subsampling.

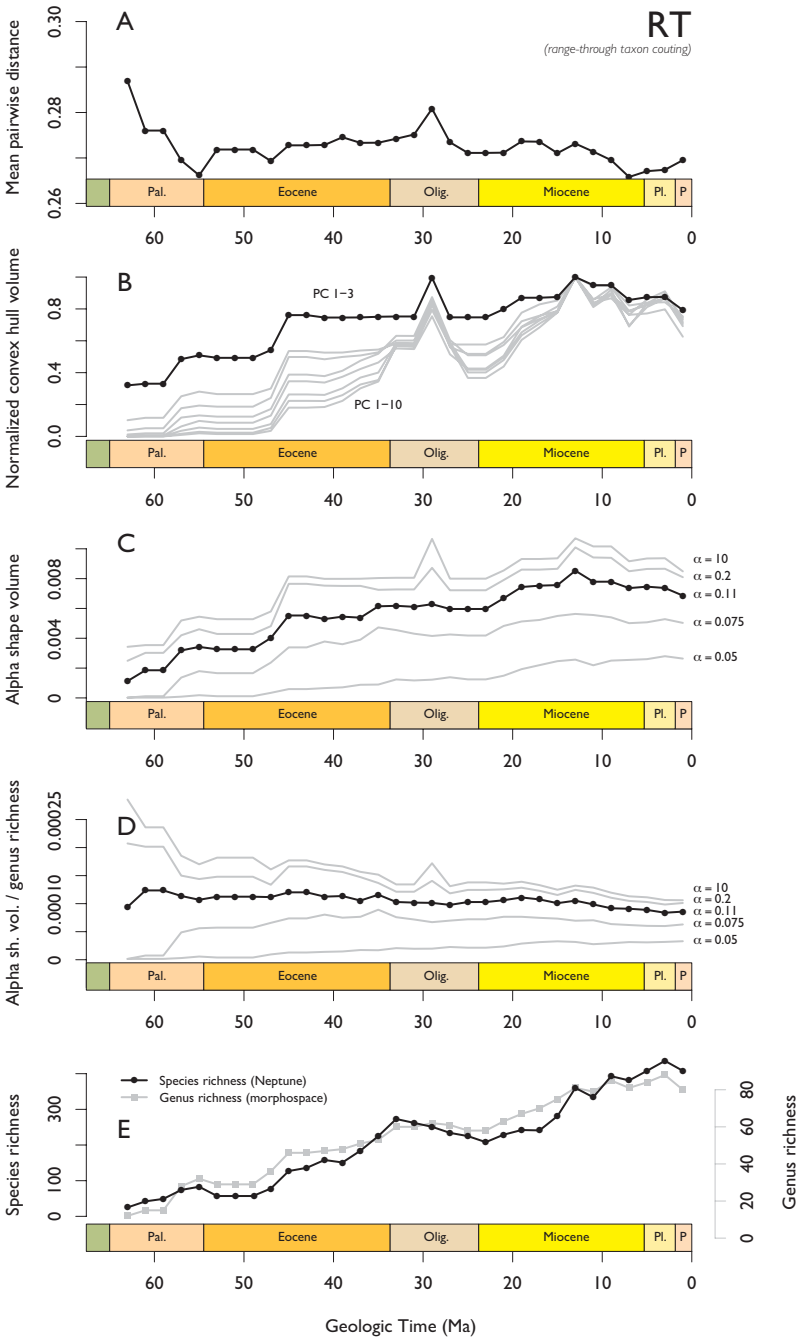
Figure 6: [two-column, print B&W, color online] Metrics of "α disparity", the average morphological disparity represented by a taxonomic list, measured in (A) convex hull volume (in three dimensions) and (B) mean pairwise distance. Error bars show the middle 50% of values,

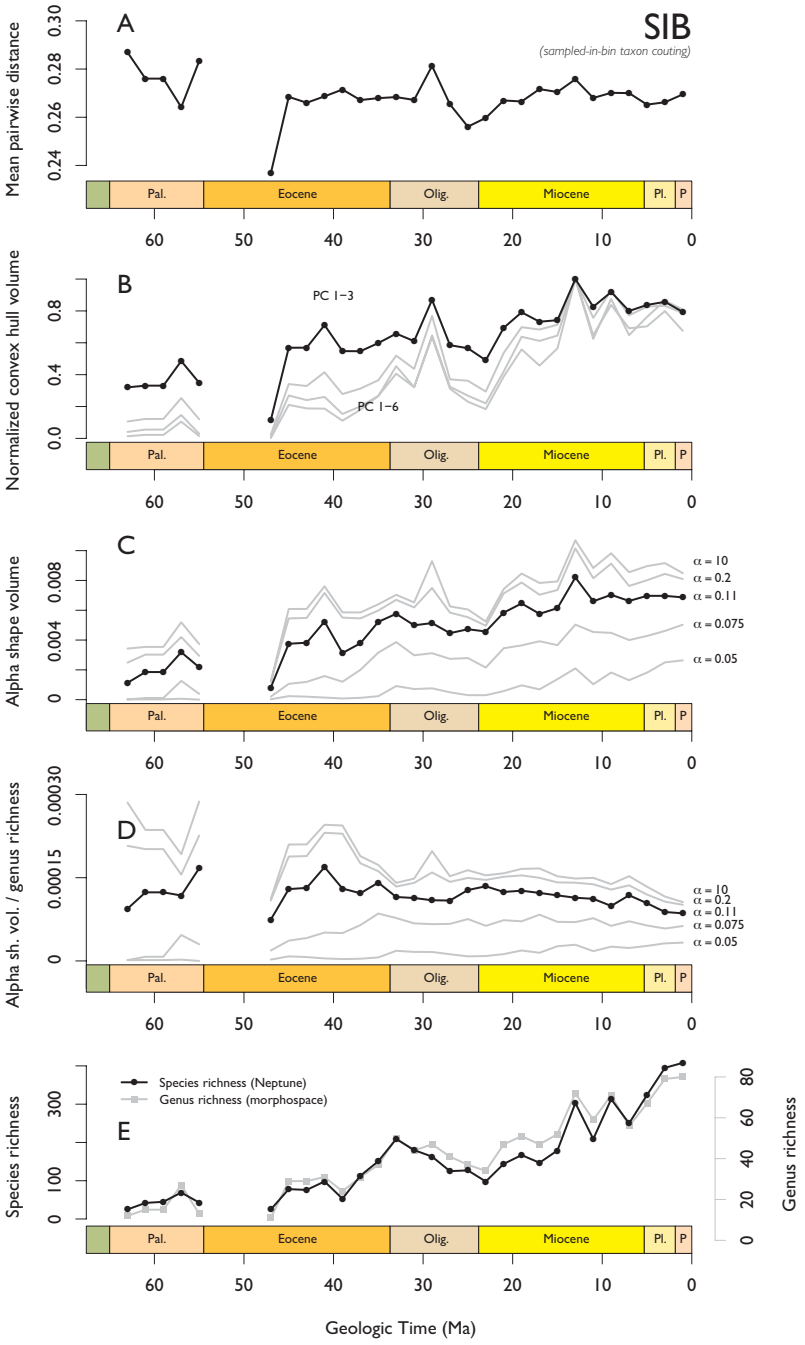
794 i.e., the 25th and 75th percentiles. Note that α disparity is unrelated to the concept of alpha
795 shapes used to quantify occupied morphospace volume.
796
797
798 Figure 7: [two-column, print B&W, color online] Plots illustrating the (in-)sensitivity of the
799 volume-based disparity metrics to the choice of ordination method. A, normalized convex hull
800 volume through time, calculated for three dimensions only, using PCO ordination (blue plot
801 points) and NMDS (red plot points). B, crossplot of the PCO and NMDS results in (A), with
802 linear model and squared correlation shown. C, alpha shape volume through time for both
803 ordination methods. D, crossplot and squared correlation of results in (C).
804
805
806 Figure 8: [one-column, print B&W] Plots showing the sensitivity of disparity metrics to the
807 quality threshold required for data included in the analysis. In both plots, each plot point
808 represents a comparison between the results reported in the companion paper (the “reference
809 results") and the results of an analysis with data satisfying a certain level of completeness,
810 expressed as a correlation coefficient (R^2) between the two sets of results. The plot above shows
811 results for a metric of the total extent of occupied morphospace (convex hull volume), the plot
812 below shows results for the dispersion metric (mean pairwise distance). Because the reference
813 analysis used an 80% completeness threshold, the correlation is perfect at that threshold (the
814 method of taxon counting in all cases was SIB).
815
816

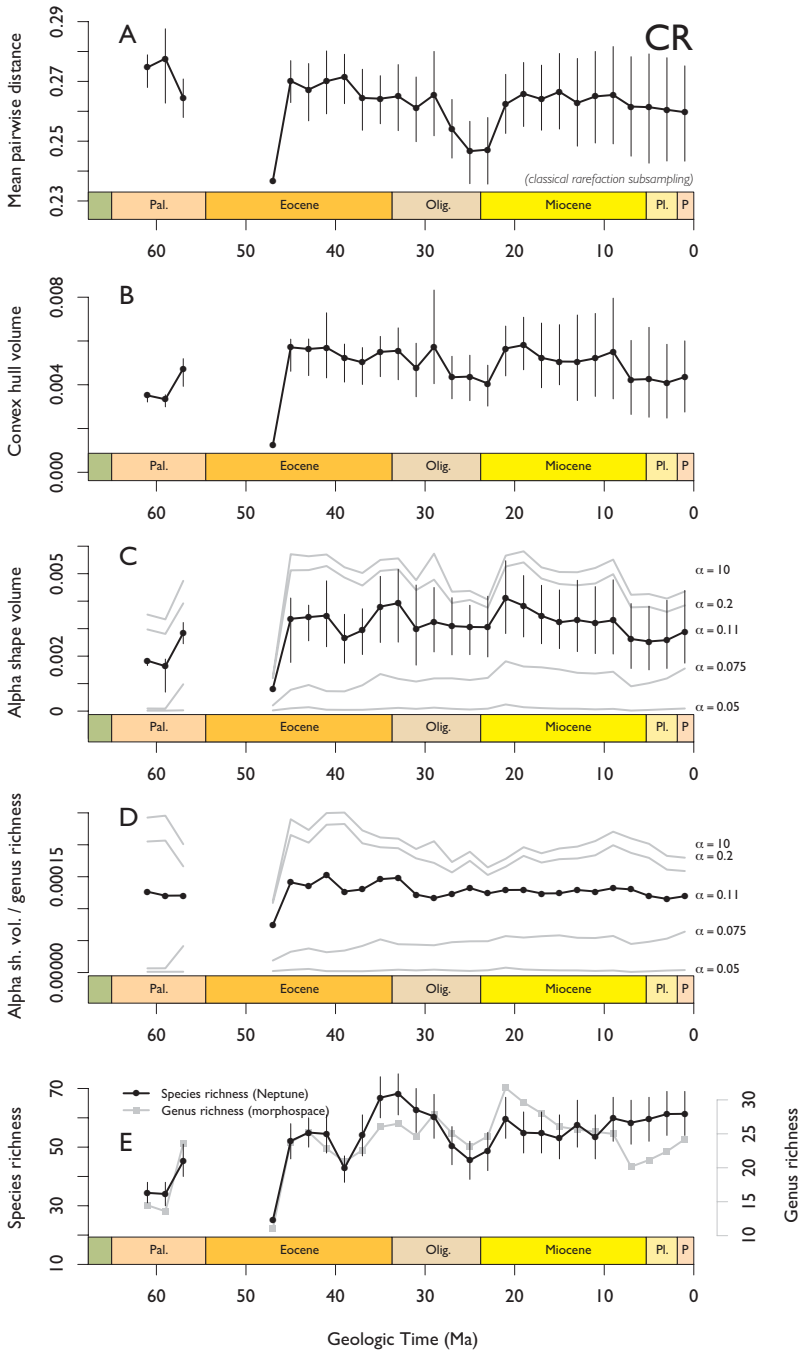
Figure 9: [two-column, print B&W, color online] Comparison of results under two different thresholds of data quality, 80% (as used in the results above and in the companion paper, shown in blue) and 100% observed character states (no “? " entries in the morphospace matrix, shown in red). A, normalized convex hull volume through time, calculated for three dimensions only, using PCO ordination. B, crossplot of the results in (A), with linear model and squared correlation shown. C, alpha shape volume through time under both data quality thresholds. D, crossplot and squared correlation of results in (C). MPWD stands for mean pairwise distance.

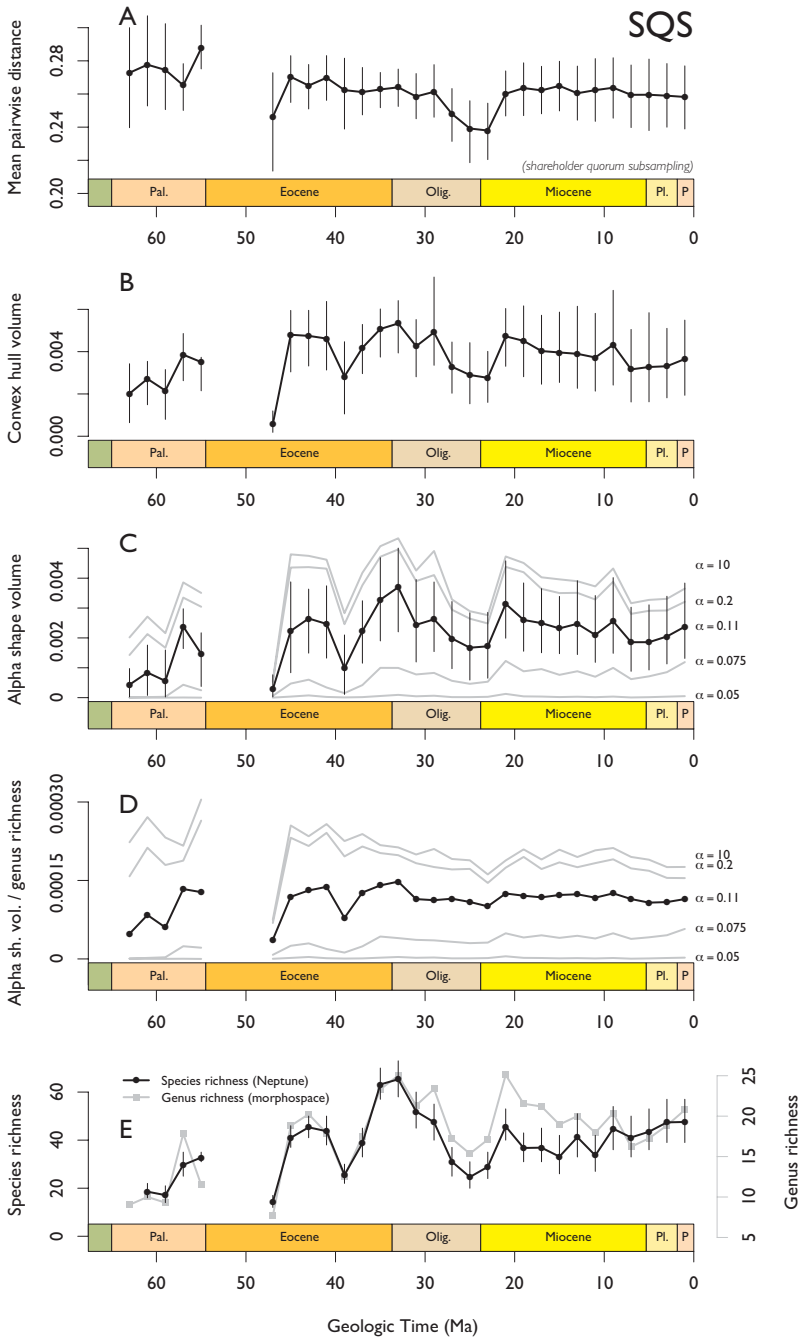
Figure 10: [two-column, print B&W, color online] Prevalence through time of sets of characters expected to change under different hypothetical Cenozoic drivers of diatom evolution. A, characters related to predation resistance; B, characters indicating cell-cell linkage, thought to impact sinking rates; C, characters thought to confer resistance against viral attack; D, characters impacting silica use.

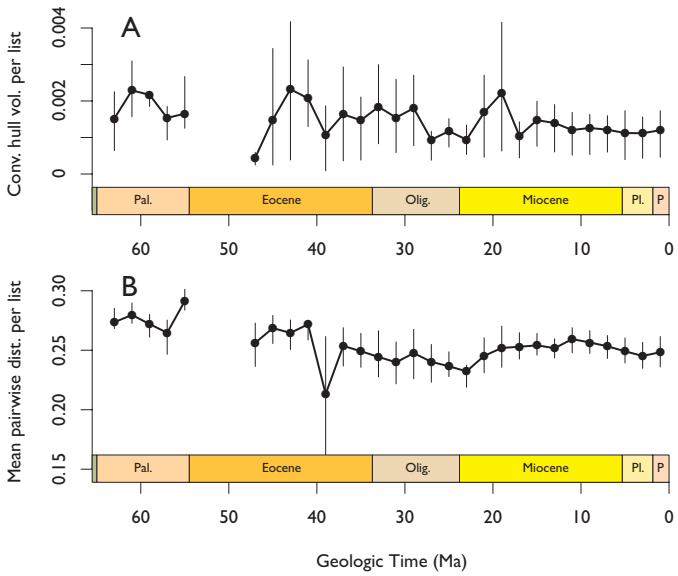

Pleistocene
145
Pliocene
223
Miocene
280
Oligocene
86
Eocene
37
Paleocene
28
Time Bin
−.2
0
.2
PC1
.2
0
−.2
PC2

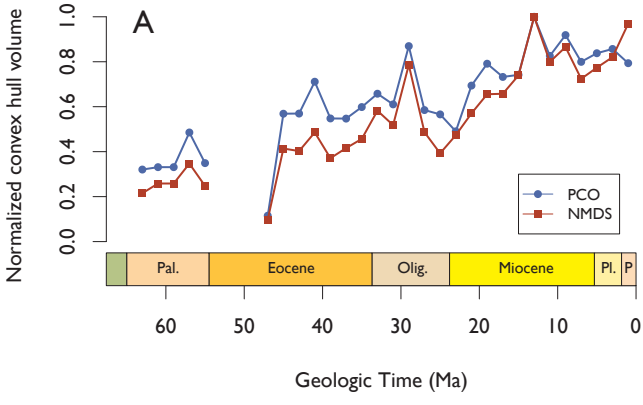

A
RT
(range-through taxon couting)
Mean pairwise distance
0.26
0.28
0.30
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
B
PC 1–3
PC 1–10
Normalized convex hull volume
0.0
0.4
0.8
C
α = 10
α = 0.2
α = 0.11
α = 0.075
α = 0.05
Alpha shape volume
0
0.004
0.008
D
Alpha sh. vol. / genus richness
0
0.00010
0.00025
E
Species richness (Neptune)
Genus richness (morphospace)
Species richness
0
100
300
Genus richness
0
20
40
60
80
Geologic Time (Ma)

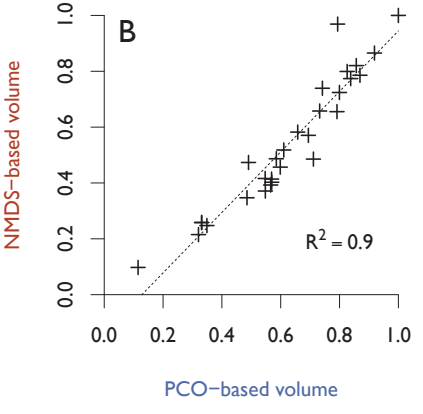

A
SIB
(sampled-in-bin taxon couting)
Mean pairwise distance
0.24
0.26
0.28
0.30
B
PC 1–3
PC 1–6
Normalized convex hull volume
0.0
0.4
0.8
C
α = 10
α = 0.2
α = 0.11
α = 0.075
α = 0.05
Alpha shape volume
0
0.004
0.008
D
Alpha sh. vol. / genus richness
0
0.00015
0.00030
E
Species richness (Neptune)
Genus richness (morphospace)
Species richness
0
100
300
Genus richness
0
20
40
60
80
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)

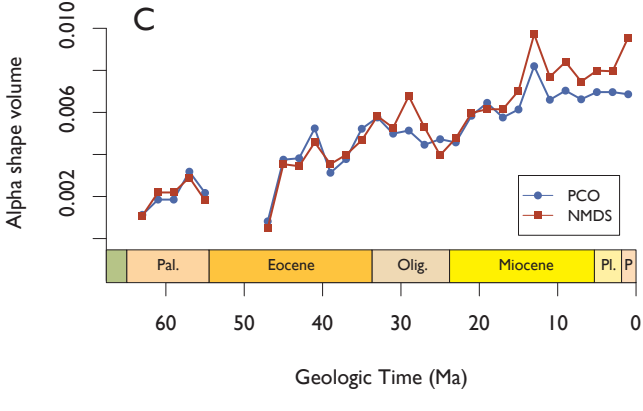

A
CR
Mean pairwise distance
0.23
0.25
0.27
0.29
(classical rarefaction subsampling)
B
Convex hull volume
0.000
0.004
0.008
C
Alpha shape volume
0
0.002
0.005
α = 10
α = 0.2
α = 0.11
α = 0.075
α = 0.05
D
Alpha sh. vol. / genus richness
0.00000
0.00015
α = 10
α = 0.2
α = 0.11
α = 0.075
α = 0.05
E
Species richness
10
30
50
70
Species richness (Neptune)
Genus richness (morphospace)
Genus richness
10
15
20
25
30
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)

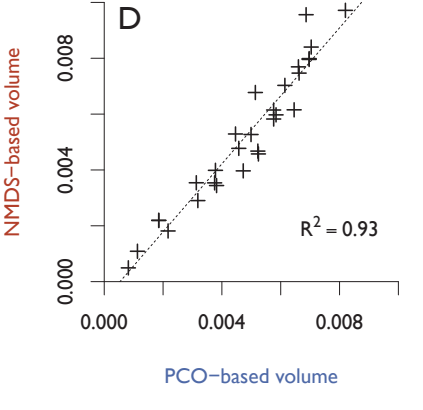

A
SQS
Mean pairwise distance
0.20
0.24
0.28
(shareholder quorum subsampling)
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
B
Convex hull volume
0.000
0.004
C
Alpha shape volume
0
0.002
0.004
α = 10
α = 0.2
α = 0.11
α = 0.075
α = 0.05
D
Alpha sh. vol. / genus richness
0
0.00015
0.00030
E
Species richness (Neptune)
Genus richness (morphospace)
Species richness
0
20
40
60
Genus richness
5
10
15
20
25
Geologic Time (Ma)

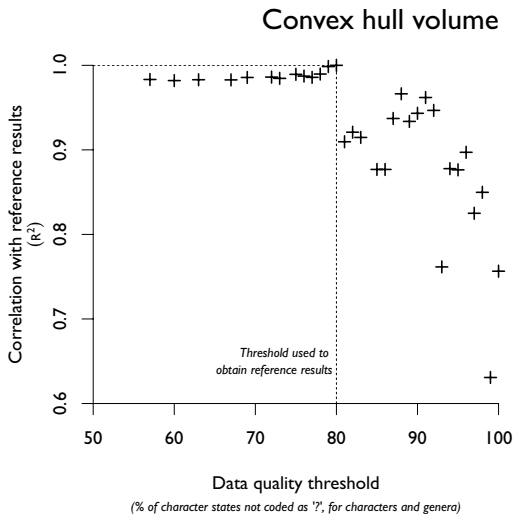

A
Conv. hull vol. per list
0
0.002
0.004
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
B
Mean pairwise dist. per list
0.15
0.20
0.25
0.30
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)

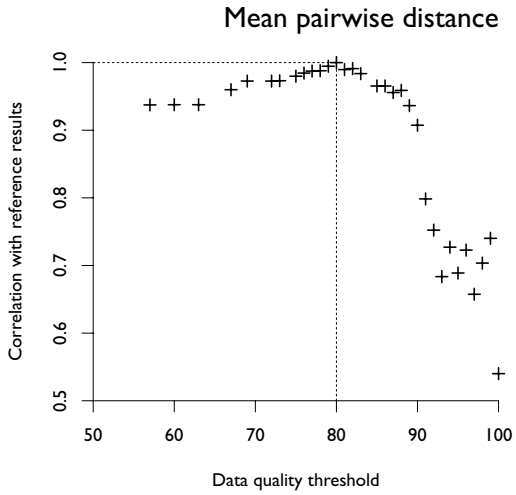

A
Normalized convex hull volume
0.0
0.2
0.4
0.6
0.8
1.0
PCO
NMDS
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)

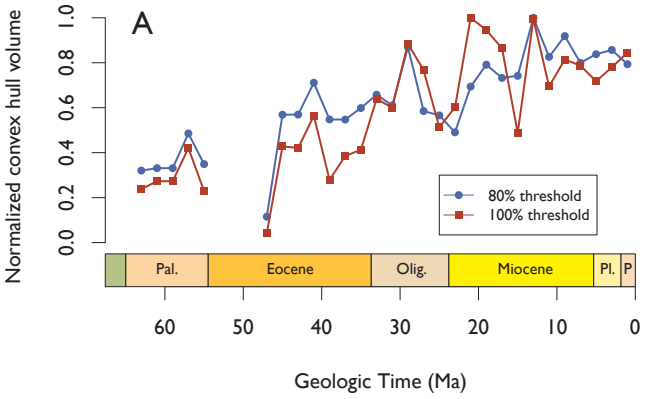

B
NMDS−based volume
PCO−based volume
$R^2 = 0.9$
0.0
0.2
0.4
0.6
0.8
1.0

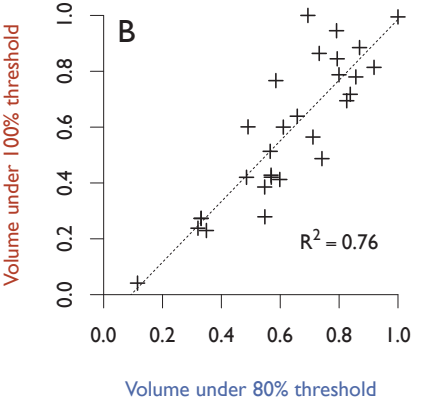

C
Alpha shape volume
0.002
0.006
0.010
PCO
NMDS
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)

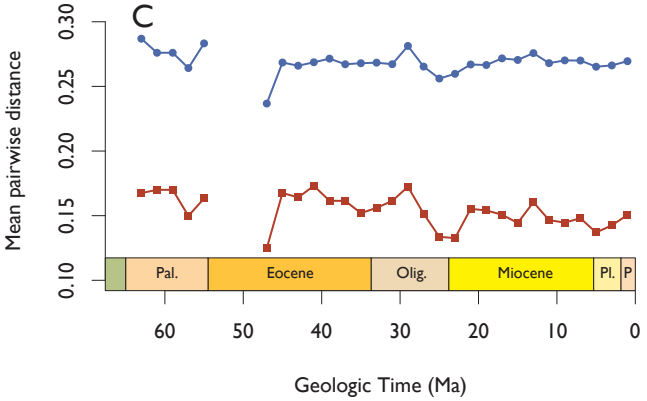

D
NMDS−based volume
PCO−based volume
$R^2 = 0.93$
0.000
0.004
0.008

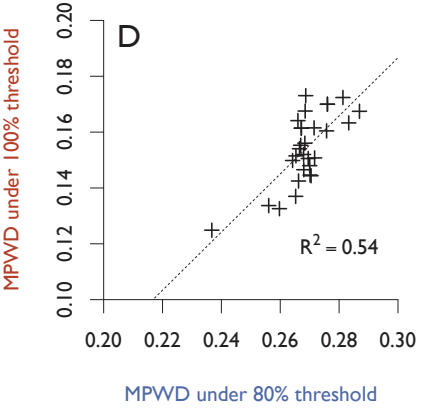

Convex hull volume
Correlation with reference results
(R^2)
1.0
0.9
0.8
0.7
0.6
Threshold used to
obtain reference results
50
60
70
80
90
100
Data quality threshold
(% of character states not coded as '?', for characters and genera)

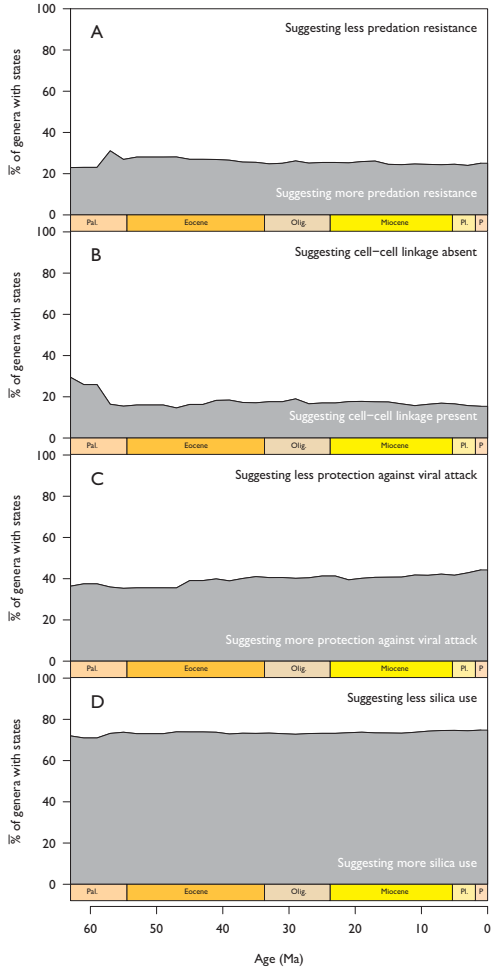

Mean pairwise distance
Correlation with reference results
1.0
0.9
0.8
0.7
0.6
0.5
50
60
70
80
90
100
Data quality threshold


A
Normalized convex hull volume
0.0
0.2
0.4
0.6
0.8
1.0
80% threshold
100% threshold
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)


B
Volume under 100% threshold
0.0
0.2
0.4
0.6
0.8
1.0
$R^2 = 0.76$
Volume under 80% threshold


C
Mean pairwise distance
0.10
0.15
0.20
0.25
0.30
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Geologic Time (Ma)


D
MPWD under 100% threshold
0.10
0.12
0.14
0.16
0.18
0.20
$R^2 = 0.54$
0.20
0.22
0.24
0.26
0.28
0.30
MPWD under 80% threshold


A
Suggesting less predation resistance
Suggesting more predation resistance
B
Suggesting cell−cell linkage absent
Suggesting cell−cell linkage present
C
Suggesting less protection against viral attack
Suggesting more protection against viral attack
D
Suggesting less silica use
Suggesting more silica use
% of genera with states
0
20
40
60
80
100
Pal.
Eocene
Olig.
Miocene
Pl.
P
60
50
40
30
20
10
0
Age (Ma)