Chapter 2

Morphospaces and Databases: Diatom Diversification through Time

Benjamin Kotrc and Andrew H. Knoll
Department of Earth and Planetary Sciences
Harvard University
Cambridge MA 02138, USA

Abstract
The diversity of diatom form inspired Art Nouveau designers, an interest renewed by recent advances in biomimetic design. The fossil record provides two windows on the diversification history of diatoms: taxonomic diversity and morphological disparity. Marine planktonic diatom diversity is conventionally interpreted to describe a steep, almost monotonic rise through Cenozoic time. Subsampling methods used to address the associated rise in sampling reveal a more stationary pattern, with peak diversity in the mid-Cenozoic, whether by established methods or a new method (shareholder quorum subsampling, SQS). However, these methods may underestimate diversification if evenness decreases. In order to measure morphological disparity, we constructed an empirical morphospace based on discrete characters. Mean pairwise distance, a disparity metric describing the density of taxa in morphospace, shows little secular change, while convex hull volume, a measure of the extent of occupied morphospace, increases through time. Since we populated the morphospace with occurrence-based data, we can apply subsampling algorithms to these disparity metrics. Mean pairwise distance is largely unaffected, while the increase in occupied volume largely disappears under subsampling. Depending on the metric used, characterizing diatom diversification thus depends upon whether a literal reading of the fossil record or the use of subsampling algorithms is preferred. While this may prompt a reexamination of evolutionary narratives prominently featuring diatom diversification, changes in abundance and
silicification may also affect the diatom’s biogeochemical importance. For biologically inspired design, an early exploration of diatom morphospace suggests that fossil forms should be considered alongside extant diatoms.

1.1 Introduction

The diversity of diatom form has been a source of fascination and inspiration since diatom frustules were first described by the 19th Century pioneers of micropaleontology (Ehrenberg 1838, Haeckel 1904) and their shapes applied to Art Nouveau architecture and design, like René Binet’s design for the Printemps department store (Proctor 2006) or Hendrik Petrus Berlage’s jewelry imitating chain-forming diatoms (Netherlands Architecture Institute 2012). Many thousands of extant diatom species have been described (Mann and Droop 1996), their shapes representing a wide range of variations on a basic pill-box Bauplan—from circles to triangles, needles, and curves—with staggering variety in the geometrically arranged, hierarchical pore structure (Round et al. 1990), lending an aesthetic that evidently appealed to turn-of-the-century designers. With biomimetic design advancing from superficial aesthetic inspiration to an application of underlying structural and evolutionary principles, renewed interest in diatoms warrants efforts toward a deeper understanding of their diversification, a cardinal feature of any clade’s evolutionary history.

The fossil record provides two windows on clade diversification history: taxonomic diversity and morphological disparity. The former, often referred to as taxonomic richness or simply as diversity, is the familiar measure that tallies numbers of taxa (commonly species). The latter, disparity for short, describes the variety of shapes or the “within-group variance of form” (Erwin 2007) by directly quantifying organismal morphology. In a sense, diversity and disparity both measure variety of form, because fossil taxonomy is, of course, itself based on morphology. It is also intuitive, however, that the two approaches measure that variety in very different ways. As an extreme example, a collection containing one species of fish, one species of elephant, and one species of insect has the same taxonomic diversity as a collection of three fish species, though the former clearly represents much greater morphological disparity.
We have two tie points on both the taxonomic and the morphological diversification of diatoms—their origin and their present diversity and disparity—from which we can trivially infer a net increase through time. But the more interesting questions about what happened in between are less trivial. What were the trajectories of diatom diversity and disparity through time? Has there been a monotonic increase, or was an early rise followed by stasis or even decline? Did diversity and disparity vary in lockstep or independently?

While the fossil record of diatoms extends back to at least the early Cretaceous Period (Harwood and Gersonde 1990) and includes many occurrences from nonmarine environments, the most robust and abundant data come from deep-sea sediments of the Cenozoic Era. Although it does not represent the entire clade’s evolutionary history, we focus on this record here because it allows us to consider the biases that uneven sampling through time may impart to our view of evolutionary history and process.

1.2 Reconstructing Taxonomic Diversity

Conventionally, the Cenozoic history of marine planktonic diatom diversity describes a steep, almost monotonic rise of about an order of magnitude (Spencer-Cervato 1999). This view plays a central role in a number of evolutionary narratives involving the diatoms, including their coevolution with grasses (Falkowski et al 2004) and whales (Marx and Uhen 2010), their role in reshaping the silica cycle, and its effect on radiolarians (Lazarus et al 2009). Although widely accepted, this view has recently been challenged (Rabosky and Sorhannus 2009). The conventional diversity curve is generated from Neptune, a large database of marine microfossil occurrences reported from the Deep Sea Drilling Program and Ocean Drilling Program, representing several decades of micropaleontological research (Lazarus 1994, Spencer-Cervato 1999). The diversity history derived from these occurrences is not, however, a unique result, since different methodological choices can be made in taxon counting, dealing with data imperfections, and accommodating secular variations in sampling intensity. Each of these can change substantially the diversity curve generated.

Taxon counting. In order to get from database to diversity curve, occurrence data need to be divided into time bins (in our examples below, of two million year duration) and the number of
taxa in each bin counted. Traditionally, this has been done by counting taxa as present in all time bins between their earliest and latest occurrences, then tallying taxa known to have existed in each time bin regardless of whether observed or inferred. This “range-through” method (RT, see Fig. 1) has an advantage over simply counting taxa observed, or sampled in-bin (SIB, see Fig. 1), because it takes into account variations in preservation and sampling from one bin to another. For example, in Fig. 1, the plot of diatom diversity from Neptune using RT taxon counting makes up for missing data between 54–48 Ma; diversity during that interval was clearly not zero, as a literal reading of the SIB curve might imply.

Despite its advantage in accounting for ‘Lazarus taxa’ (taxa that appear to go extinct, only to reappear later in the record; Flessa and Jablonski 1983), however, the RT method has fallen out of favor among many paleobiologists because it imparts a number of undesirable and potentially severe biases: the Signor-Lipps effect, the Pull of the Recent, and other edge effects reviewed by Alroy (2010). The SIB sampling method is preferred over RT and other methods such as tallying only those taxa known to cross the boundary between adjacent time bins because it is immune to these biases, and the bin-to-bin sampling differences that remain can be counteracted with corrections like the part-timer sampling probability, which effectively performs a temporally localized range-through among adjacent time bins (Alroy 2008). While the diversity curve for the Neptune diatom data obtained by SIB taxon counting does differ from the conventional curve obtained using RT in the details, the curves are rather similar in shape to first order.

**Data quality.** Generally, paleontologists worry that the fossil record underestimates the true ranges of taxa (e.g. Marshall 1990), but the record of marine microfossils is so unusually rich that the opposite has been suggested for the *Neptune* database. Marine microfossils can appear outside of their true range due to "RATs", that is, because of the physical reworking of sediments (erosion and redeposition in a stratigraphically younger position), errors in the age model assigning a fossil occurrence to the wrong time bin, or taxonomic error (Lazarus 2011). For the curve that has become the canonical depiction of diatom diversity, these problems were addressed by manually excluding occurrences in *Neptune* considered unreliable, including occurrences near depositional hiatuses (Spencer-Cervato 1999). The effect of all but the most severe instances of reworking can be obviated by setting sufficiently wide time bins, and misplaced occurrences far from the true range of a taxon are much less of an issue for SIB than RT taxon counting. Nonetheless, outliers could also be identified for removal by applying hat-shaped models of the
rise and fall in occurrences through a taxon’s range (Liow and Stenseth 2007, Liow et al. 2010), but a much simpler method recently proposed just trims a certain calibrated percentage of occurrences from the beginning and end of a taxon’s range—aptly named Pacman profiling (Lazarus et al. 2012).

**Sampling biases.** An arguably graver concern for the accurate reconstruction of diatom paleodiversity is that the amount of data in the *Neptune* database greatly increases with time, as shown in Figure 2. This is worrisome because it is easy to imagine a situation where true diatom diversity in fact remained constant, but a steadily increasing number of samples through time captures more species from younger intervals, giving a false impression of rising diversity. Although this concern was noted in the first explorations of the *Neptune* dataset (Spencer-Cervato 1999), it was only recently addressed in detail (Rabosky and Sorhannus 2009).

Such temporal sampling biases are common in paleontological datasets, for a number of reasons. In general, older sediments are less abundant than younger ones. For microfossils from deep-sea drilling cores more specifically, sediments are progressively destroyed as ocean crust becomes subducted by plate tectonic processes, making older sediments less common. Perhaps more importantly, the deep drilling commonly required to reach older sediments is expensive, and requires drilling through younger sediments for which samples are usually also collected. Finally, diatoms undergo a series of diagenetic mineral transitions as burial temperatures and pressures increase, making the preservation of recognizable morphological features less likely with age (DeMaster 2003). In recent decades, paleobiologists have directed much research effort towards developing numerical methods to correct for these sampling biases.

**Subsampling methods.** The idea at the heart of these methods is that we can obtain a more accurate reconstruction of diversity history if we measure diversity using a standardized, comparable sample size in each time bin. It is important to note that the goal of such subsampling methods is not to get closer to the true absolute values of diversity (as range-through taxon counting does, for example), but rather to reconstruct the relative shape of the diversity curve as accurately as possible. The basic procedure underlying these methods involves randomly drawing items from the full data set of the time bin in question until some quota is reached. The number of taxa in that subsample are then tallied and the process is repeated a large number of times to obtain an average diversity and an associated confidence interval for that time bin. This whole process, in turn, is repeated for all time bins, using the same quota for each.
The well-established subsampling methods all use a uniform quota of items, but differ in how items are drawn and how the quota is set. For example, in classical rarefaction (CR, Miller and Foote 1996), occurrences are drawn from the full dataset until a quota of some fixed number of occurrences is reached. The items drawn can also be taxonomic lists, which in the case of the *Neptune* data means the list of taxa found in a particular borehole at a particular depth. In this case, the quota can be simply set as a number of lists, in which case the method is referred to as by-lists, unweighted subsampling (UW, Alroy 2000), or as a number of occurrences, (weighted by occurrence subsampling, or OW for short, Alroy 1996). In practice (at least for the Neptune diatom data set), these methods give broadly similar results.

Subsampling methods were only recently applied to diatoms (Rabosky and Sorhannus 2009), painting a picture of diatom diversification very different from curves generated from RT and SIB tabulations of *Neptune* data (Fig. 1). A representative diversity curve generated by OW subsampling (Fig. 1 OW), similar to those produced by CR and UW subsampling, still shows a net increase over the Cenozoic Era, but the total increase is only about twofold. Perhaps more importantly, maximum diversity is sited in the late Eocene Epoch, with tabulated diversity falling substantially through the Oligocene, before recovering through the Neogene Period.

Another method of subsampling, which uses a quota measured by the sum of squared occurrences (O2W), has also been applied to diatoms and results in an even flatter diversity curve with a still more pronounced Eocene-Oligocene diversity peak (Rabosky and Sorhannus 2009). The reasoning that underpins this method (Alroy 2000b) is a little complicated, but can be understood by considering how taxon occurrences are recorded. A paleontologist describing the species in a sample will examine specimens, i.e. individual fossils, recording new species as they are encountered. Because not all taxa are equally abundant, the common taxa will be found quickly, while rare taxa are only likely to be counted after many specimens have been examined. This results in an asymptotic collector curve (or accumulation curve), in which taxa discovered are plotted against specimens examined (or sampling effort expended). The O2W method attempts to account for this non-linear relationship between specimens and species. Unfortunately, diversity estimates under this method have been found to be strongly biased when the geographic structure of diversity changes through time (Bush et al. 2004), so we do not show O2W results here.
A major shortcoming of all of the fixed-quota subsampling methods is that they can systematically undersample intervals with high diversity, because uniform sampling is not necessarily fair sampling (Alroy 2010). Consider an ideal case where there has been no change in the true diversity through time, but an increase in sampling results in an apparent increase in diversity: in this situation, fixed-quota methods will theoretically perform well. But if there has been a true increase in diversity in addition to an increase in sampling, these methods may artificially flatten the resulting diversity curve, because more diverse intervals will require more sampling to capture the same proportion of the total (true) diversity. As an extreme example, consider that a population of five species can be sampled completely with 100 occurrences, while the same size sample will underestimate radically the diversity of a population with 500 species. When we allow for the possibility that true diversity can vary widely, it makes intuitive sense to allow the quota of items drawn in subsampling to vary also—and from this perspective, we can also see that the level to which we ought to standardize samples is not a fixed amount of sampling, but some amount of sampling that aims to return a fixed proportion of the total diversity, or coverage.

A recently published subsampling method, shareholder quorum subsampling (SQS), is based on this principle of taxonomic coverage (Alroy 2010). The method is named by analogy to a corporate shareholder's meeting, where a quorum of shareholders needs to be present such that the sum of their shares meets a threshold proportion of the total shares in the company. In SQS, we can think of taxa as shareholders and their share of the frequency distribution (proportion of total occurrences) as shares. Occurrences are drawn much as before, but rather than stopping at a certain number of occurrences or lists, samples are drawn until a 'shareholder quorum' is reached—that is, until the sum of the frequencies of sampled taxa exceeds some threshold. The number of lists or occurrences it takes to reach this quota is free to vary, making this method philosophically quite distinct from the uniform sampling methods like CR, UW or OW.

In order to ensure that the proportion of frequencies drawn represents a certain proportion of taxonomic coverage, however, we need to know how taxonomically complete each sample is—that is, we need an estimate of coverage. Another way to think of this problem is to consider the observed frequency distribution to overestimate the frequency of those taxa observed in favor of those taxa not observed, whose frequencies are rounded down to zero. We need an estimate of how much of the underlying, true frequency distribution has been muted by such rounding down.
One such estimate, used commonly in ecology, uses the proportion of observations that are singletons, i.e. taxa only seen once in that sample, calculated as Good's $u$ (Good 1953). In ecology these observations are individuals, but the approach can be extended to compiled paleontological data where these observations are occurrences (i.e. the presence of a taxon at a particular location and stratigraphic position, irrespective of its abundance; Alroy 2010). While for the purposes of ecological studies, singletons are those taxa that occur only once in a sample area such as a quadrat, Alroy (2010) argued that for paleontological data the best analogical equivalent was to count singletons as taxa occurring only in a single publication.

We applied a modified version of this basic SQS algorithm to the diatom occurrence data from Neptune. Because of the way micropaleontological data are collected (the occurrence of a set of taxa reported over a stratigraphic range), taxa will almost always have more than one occurrence in any given publication. Instead, we used taxa occurring only in a single borehole in place of singletons.

The results for SQ subsampling of the Neptune diatom data (Fig. 1, SQS) are broadly similar to those of the fixed-quota subsampling methods (Fig. 1, OW), suggesting that true diversity increased only slightly over the Cenozoic Era. Much as under the fixed-quota methods, peak diversity is reached in the latest Eocene/earliest Oligocene, but under SQS this peak is exaggerated, suggesting that diversity then was significantly higher than today—similar to the results of the O2W method (Rabosky and Sorhannus 2009).

In spite of the obvious sampling bias in the Neptune data, the largely stationary view of Cenozoic diatom diversity suggested by subsampling methods has not been universally accepted by micropaleontologists. A potential vulnerability of subsampling methods, including SQS, is that they may not give accurate results if there are large changes in relative abundance structure (or evenness) through time (Lazarus et al 2012b). We can understand this problem by considering the relative frequency distribution of a time interval—a rank-ordered plot of the proportion of occurrences of each taxon (such that the extent along the x-axis represents total diversity). Fixed-quota subsampling methods can be thought of as sampling all taxa falling above a threshold ‘veil line’ of some relative frequency (Alroy 2010). The failing of these methods, addressed by SQS, can be visualized by considering what happens if the diversity increases, but the shape of this distribution stays the same: because each taxon now has a smaller relative frequency, a greater proportion of the taxa falls under the veil line, underestimating diversity under subsampling.
Under SQS, a constant area under this curve is sampled, so even if the total diversity increases, the same proportion of the frequency distribution will be recovered—and, if the shape of the distributions stays the same, the same proportion of total diversity. If the shape of the frequency distribution were to change drastically, however, SQS might not work as well.

Empirically, the diatom occurrence data in *Neptune* do show a change in frequency distribution from more even to more uneven, and it has been argued that these changes may cause subsampling methods (including SQS) to mask a true rise in diversity (Lazarus et al. 2012b). If we imagine SQS subsampling to recover a fixed area under a rank-ordered relative frequency distribution (see supplement to Alroy 2010b), the area under a flat curve (an equitable frequency distribution) will sample a greater proportion of the total diversity than the same area under a hollow curve (an uneven frequency distribution). Lazarus et al. (2012b) apply an empirical correction factor to account for the changes in frequency distribution and recover a rise in diatom diversity more similar to the canonical view. A similar correction factor with even greater leverage is used to account for an increase in provinciality through time, particularly regarding the development on an endemic polar fauna (Lazarus et al. 2012b).

Lazarus et al. (2012b) marshal further support for the conventional view of diatom diversification from a catalogue of about 500 diatom species’ ranges compiled from both marine and land-based sections under expert curation against taxonomic and stratigraphic error. The curve generated is similar in form to the canonical diatom diversity curve (Spencer-Cervato 1999), albeit showing a net increase that is slightly less steep. This data set has certainly been better flushed of "RATs" (the sorts of errors described in the section on data quality above) than *Neptune*, but the question of sampling bias arguably remains: while there is no strong correlation in this compilation between diversity in a time bin and the number of publications from which this diversity is derived (Lazarus et al. 2012), the relationship between a taxonomic or biostratigraphic publication and the amount of sampling it represents is not clear and not necessarily fixed.

To summarize, the taxonomic window on diatom diversification provides an uncertain picture of Cenozoic diatom evolution. Interpreted at face value, the fossil record suggests a steep Cenozoic rise in species richness, whether from deep-sea occurrences in the *Neptune* database (Spencer-Cervato 1999) or from a biostratigraphic catalogue of first and last appearances (Lazarus et al. 2012). When the stark secular rise in the amount of available data is taken into account
using item quota (Rabosky and Sorhannus 2009) or SQS subsampling methods, however, a more stationary pattern emerges, showing at most a modest overall increase in species richness and peak diversity around the Eocene/Oligocene boundary. With changes in relative abundance potentially biasing the results of these subsampling methods, we are left with a level of uncertainty about the true diversification history of the diatoms. Recalling that there is another window on diversification, however, we turn to the history of diatom morphological disparity to gain another perspective on this question.

1.4 Reconstructing Evolution in Shape Space

In common paleobiological usage, disparity describes a quantification of morphological differences among organisms (Wills 2001, p. 56). Unlike species richness for diversity, there is no singular metric for disparity; commonly used measures can be more easily understood in the conceptual framework of morphospace—a mathematical construct used to quantify and describe organismal morphology.

**Morphospaces.** Morphospaces are n-dimensional mathematical spaces describing the form of a group of organisms. As such, morphospaces are an example of what, in the context of ecology, Lewontin (1969, p. 13) called “the concept of the vector field in n-dimensional space,” which he described as “the most fundamental [concept] we have for dealing with the transformations of complicated dynamical systems in time.” Familiar, conceptually related notions include adaptive landscapes (Wright 1932) and niche space (Hutchinson 1978, p. 158), but rather than gene alleles or ecological variables, the axes of morphospaces represent morphological characters or parameters. Each point in morphospace represents a particular, unique morphology, and it can either be occupied (i.e. represent a morphology actually realized by an organism) or not.

With this framework in mind, we can consider the morphological disparity of a group as a description of how the group is distributed in morphospace—are the taxa spread out widely (signifying large morphological differences) or clustered together (signifying morphological similarity)? As discussed in more detail below, this spatial distribution of taxa can be quantified
in a number of ways, leading to multiple metrics of disparity. Before considering how to measure morphospace occupation, however, it is worth briefly examining the different ways in which morphospaces can be constructed.

Morphospaces are often divided into two kinds, those whose axes are parameters of a shape-generating function, called generative or theoretical morphospaces, and those whose axes are measurements of organisms, called empirical morphospaces (McGhee 1999). Theoretical morphospaces generally have only a few axes and thus a small number of dimensions that is easy to visualize; the first and best-known example is Raup’s (1966) classic morphospace of coiled shells. Empirical morphospaces, in contrast, often have a very large number of axes (representing a large number of measured morphological characters) and generally require an ordination procedure such as principal components analysis (PCA) or principal coordinates analysis (PCO) for visualization and analysis, an approach pioneered by Foote (1989). Because of this, empirical morphospaces have been described as having axes that are data-dependent or unstable, since different measurements of the same morphology will result in different ordinated axes (McGhee 1999). Seen from a more general perspective, however, the distinction between theoretical and empirical morphospaces can become conceptually and mathematically blurred if the latter are considered in their full, unordinated dimensionality (sometimes called “raw morphospaces,” Eble 2000): the number of axes could then be seen as the most significant difference between the two. From this perspective, both sorts of morphospace can be used to investigate the realms of unrealized as well as realized morphologies—although theoretical morphospaces can undoubtedly generate a wider range of unrealized form than empirical morphospaces can.

**Limitations of theoretical morphospaces.** While there is broad consensus that theoretical morphospaces are preferable because their use of explicit, measurement-independent growth models that allow one to explore a wider range of unexplored as well as impermissible forms (e.g. Erwin 2007), their application is unfortunately not always possible (McGhee 1999, p. 26). Growth models for theoretical morphospaces are more readily devised for organisms with accretionary or branching growth (e.g. Raup 1966, Niklas 1999), but mathematical shape models with a reasonable number of parameters can only reproduce so many aspects of form. The applicability of generative morphospaces with a small number of parameters is thus limited in a two ways that are well illustrated by the case of the diatoms: the range of overall forms that can be generated, and the difficulty of including complex and higher-order morphological features.
**Previous diatom morphospaces.** The diversity of fundamental forms that can be generated by a mathematical model with a few parameters is limited. In diatoms, for example, capturing the great variety of different symmetries of the valve in plan view alone (circular-elliptical, triangular, rectangular, curved, isopolar or heteropolar, and so on) in a generative model would require many parameters, and even then the plan-view outline shape says nothing about the obviously important three-dimensional shape of the valve. Generative shape models that have been developed for the diatoms are thus by necessity limited both in terms of covering only a subset of the full taxonomic and morphological diversity, and in terms of describing a subset of the overall frustule morphology. Examples include models for a particular species (Stoermer and Ladewski 1982) or genus (Mou and Stoermer 1992), a more widely applicable model describing only valve outlines (Arita and Ohtsuka 2004), and a model based on 3D parametric equations limited to a group of asymmetrical pennate diatoms (Pappas 2005). While generative morphospaces of this nature have been profitably applied to questions of taxonomic distinction or morphological evolution within particular groups, they capture neither the total diversity of overall diatom form, nor the higher-order features of diatom morphology such as pore arrangement, spines, processes, or the raphe—even though these may well be of biological and evolutionary significance.

We are thus led to an empirical morphospace approach in trying to understand the Cenozoic evolution in morphospace of the marine diatoms in as a whole. This approach has been successfully applied to many other groups with complex morphologies, highlighting important features that are hard to model with simple geometric models (e.g. Foote 1991, Lupia 1999, Boyce and Knoll 2002). Under the auspices of the PlanktonTech initiative, we conducted an empirical morphospace study of Cenozoic marine planktonic diatoms, the detailed methodology and results of which are published elsewhere (Kotrc and Knoll in review [a] and [b]). Summarized briefly, this approach involved quantifying the morphology of Cenozoic diatom taxa using a large number of morphological characters to construct a morphospace, applying an ordination procedure to visualize this morphospace, and then populating the morphospace through time based on the fossil record. Since the *Neptune* database is a readily available compilation of fossil diatom occurrences, we used it to populate the morphospace. Because an analysis at the species level would be intractable (there are over 1,000 diatom species in *Neptune*), we chose to
work at the genus level; our final analysis included 140 genera (for complete methods and data see Kotrc 2013, and digital supplement at http://dash.harvard.edu/handle/1/11051218).

**Choice and coding of characters.** We used descriptions of frustule morphology and taxonomic descriptions of the genera in Neptune to compile a list of morphological characters. In formulating these characters, we were careful to strictly describe morphology independent of taxonomy or phylogeny—meaning that structures were quantified by their similarity in form regardless of whether they are equivalent in development or evolutionary origin. In some cases, this meant bridging substantial gaps in the nomenclature used to describe structures in different groups within the diatoms, for example the terms applied to the arrangement of pores in centric diatoms (“areolation”) versus in pennate diatoms (“striation”). This strictly morphological approach distinguishes morphospace analysis from morphometric approaches in which the importance of choosing homologous characters is often strongly emphasized (e.g. Rohlf and Bookstein 1990). It has the advantage of allowing the evolutionary exploration of form on its own merit, independent of how this form is achieved phylogenetically or developmentally. Our final analysis included 100 morphological characters, coded as unordered, discrete character states (including many binary characters), allowing us to account for both intrageneric variation and the categorical nature of many characters (such as the presence or absence of a raphe, a slit along the valve face of some diatoms that enables locomotion).

**Ordination.** The resulting morphological character codings for each genus define a 100-dimensional, categorical morphospace, a “raw morphospace” in the sense alluded to above. Visualizing this morphospace requires that we find a lower-dimensional projection of the relative locations of our genera, something that can be accomplished with PCO, a method analogous to the more familiar PCA that works with unordered discrete characters. This transformation results in a representation of the morphospace along continuous axes, the first two of which (the two capturing the greatest amount of the information in the full-dimensional morphospace) are shown on the left in Figure 3.

**Visualization.** Empirical morphospace plots resulting from ordination procedures like this one can be hard to interpret if the only information presented is the location of genera on ordinated axes of mysterious meaning. Providing representative images of morphologies at selected locations is helpful, but because we have morphological data for each of the points, it can make sense to modify the shape of plot points themselves to reflect morphological character
states. In the left panel of Figure 3, we have represented the states of three characters relating to
the plan (valve) view shape of each genus: the overall shape (elliptical, triangular, square, or
ovate.), the aspect ratio, and the presence or absence of a raphe. The plot shows equant forms
toward the top left, and elongate forms with and without raphes toward the bottom right. For
examples of these morphologies, we refer the reader to the excellent SEM images in Round et al.
(1990). This division reflects the largest-scale taxonomic division within the diatoms—centrics
versus pennates—and raises the question to what extent phylogenetic structure is evident in
morphospace.

**Morphospace and phylogeny.** Comparing the proximity of genera in morphospace to
their proximity on a molecular phylogenetic tree shows only a very weak correlation;
phylogenetic proximity at finer levels of resolution beyond the centric-pennate divide is not a
good predictor of proximity in morphospace (Kotrc and Knoll in review. [a]). For example, the
two major subdivisions within the pennate diatoms (raphids and araphids) do not appear to
occupy distinct regions of morphospace. This may seem surprising, considering diatom
phylogenies from before the molecular era broadly agree with molecular ones. It helps to
remember, however, that morphological phylogenies identify key features with defined polarities,
called synapomorphies, signifying inclusion in groups; this morphospace, in contrast, consists of
equally weighted characters.

These results seem to suggest that the major groups of diatoms iteratively recolonized
already-occupied regions of morphospace. In the case of the raphe, for example, this might make
sense if we consider that—in providing for locomotion—it represents a key innovation in the
radiation of diatoms in benthic and terrestrial environments. The taxa in our analysis come from
the marine plankton, however, where—in the absence of substrates upon which to locomote—this
innovation may have been of relatively little consequence. As a result, raphid diatoms in the
plankton may have radiated to fill ecological niches indistinct from those occupied by their
araphid cousins, resulting in overlapping occupation of morphospace.

While on the topic of diatom phylogenies, we divert briefly to consider that a clade’s
diversification history can also been addressed from a phylogenetic perspective directly,
something commonly achieved using molecular phylogenies of their extant members and “lineage
through time” plots. While such an approach could be taken for the diatoms (perhaps even using a
phylogeny derived from morphological data, such as the data matrix underlying the morphospace
analysis here), we point out the limited inferences that can be drawn from such analyses. Diversification rates result from a balance between origination and extinction, and as Quental and Marshall (2010) showed, molecular phylogenies are sensitive to the former, but blind to the latter. As an example, they compare a diversification history for whales derived from a molecular phylogeny, suggesting a history of unchecked diversification since the Oligocene (Steeman et al. 2009), to one based on the group’s rather complete fossil record, which reveals a sharp decline in diversity from the Late Miocene to the present, a worrying discrepancy due to extinctions not revealed by the molecular phylogeny.

Returning to morphospace, molecular clocks suggest that the four major groups of diatoms (the raphid and araphid pennates and two groups of centric diatoms, the radial centrics and bi- and multipolar centrics) had diverged by the late Cretaceous Period (e.g. Kooistra et al. 2007). Given the apparent overlap of these groups in morphospace, we might expect to see broad morphological stasis across the Cenozoic Era. Fortunately, we can further explore this question by extending our morphospace back through time using the fossil data from Neptune.

Morphospace through time. The right hand side of Figure 3 shows the occupation of the two-dimensional morphospace plot on the left through geologic time, tilted to an oblique view, for several Cenozoic time bins (with the youngest at the top). The polygons at the bottom of the plot are convex hulls enclosing the points in each of the time bins above in the corresponding colors; convex hulls being the shape that would be made by stretching a rubber band around the points in a time bin and thus describe the extent of occupied morphospace. While the Cenozoic assemblages cover a larger area of the plot than the Cretaceous ones, the area does not increase very much through the Cenozoic Era. One interesting observation is that the quadrant at the center of pennate diatom morphology (the bottom right) is sparsely occupied up until the Miocene. This general impression of how morphospace has been occupied through time can be quantified much more rigorously by calculating metrics of disparity for each time bin in the Neptune data set.

Metrics of disparity. Many different ways of measuring disparity have been devised. A thorough review of these different metrics is beyond the scope of this chapter (for more detailed treatments see Wills et al. 1994, Ciampaglio et al. 2001, and Erwin 2007), but a key point is that these metrics do not all describe the same aspects of morphospace occupation. To illustrate this notion, we present two disparity metrics here: convex hull volume and mean pairwise distance. The former is a higher-dimensional extension of the rubber band method introduced in Figure 3;
instead of measuring the area enclosed by the polygon, the volume or hypervolume enclosed in three or more dimensions is calculated. While convex hull volume is a measure of the total extent of morphospace occupied, the latter metric—as implied by its name—measures the average distance between pairs of taxa, measured as the proportion of character state mismatches out of possible matches.

It is not hard to see that because these metrics describe different aspects of morphospace occupation—the total extent of occupied morphospace and the degree of dispersion of taxa in morphospace, respectively—they can give different answers regarding disparity trends through time. In the case of diatom morphospace, for example, mean pairwise distance (Fig. 4, top left panel) stays roughly constant through time, while convex hull volume (Fig. 4, top right) shows a substantial increase (regardless of how may dimensions are considered in the volume calculation). While these results would appear to be at odds if we were to consider disparity as a monolithic concept, inspection of the stacked morphospace plot in Figure 3 makes clear which aspects of morphospace occupation they describe: on the one hand, the total extent of occupied morphospace is increasing as new genera expand into previously unoccupied morphologies, while on the other hand, the number of genera is increasing, leading to a roughly constant (or slightly increasing) “packing” of genera into the occupied region. We can thus think of mean pairwise distance as representing something of a density measure and convex hull volume as the total extent of occupied morphospace. As is the case in Figure 4, the density (mean pairwise distance) can remain roughly constant as occupied morphospace (convex hull volume) expands, but only if there is a commensurate rise in diversity.

**Sampling.** This increase in the number of genera in morphospace raises an interesting question about sampling. Since we spent much of the first half of this chapter voicing concern about the possibility that the evident rise in taxonomic diversity may simply be an artifact of sampling, it seems natural to ask: could sampling also affect metrics of morphological disparity? Although we introduced disparity as a different window onto the diversification history of a clade, it is based on the same fossil record and is thus subject to the same geological sources of bias. Studies on the effect of sampling on disparity metrics have suggested the use of mean pairwise distance, which has been observed to be more robust to variations in sampling than other disparity measures (Foote 1992, Ciampaglio et al. 2001, Butler et al. 2012). As we have just seen, however, considering one disparity metric alone limits analysis to just one aspect of morphospace
occupation. While it is always possible to consider the effect of increasing numbers of taxa in a morphospace indirectly, for example, by rarefaction to a standardized number of taxa (Foote 1992), we can do better in the unique situation where a morphospace has been populated using a database of fossil occurrences. In this case, we can apply those subsampling metrics (discussed above) developed for addressing sampling bias in studies of taxonomic diversity to our morphospace, and calculate metrics of disparity under subsampling.

The results of applying SQS to the diatom morphospace and calculating our two chosen disparity metrics are shown in the bottom half of Figure 4. Mean pairwise distance does not appear to be significantly impacted by the subsampling exercise, lending further support to the notion that this metric is robust to sampling variations. For convex hull volume, in contrast, the increasing trend over the Cenozoic seems to all but disappear, revealing a much more stationary pattern through time. When sampling is taken into account, then, disparity metrics paint a picture of relative Cenozoic stasis in diatom morphological evolution that stands in agreement with both the comparison between morphospace and phylogeny and the results of subsampled taxonomic diversity.

1.5 Synthesis

The fossil record provides us with two windows onto the Cenozoic diversification history of diatoms, through taxonomic diversity and morphological disparity. When read at face value, the record suggests steep Cenozoic diversification from both a view through diversity and through convex hull volume, a disparity measure describing the total extent of occupied morphospace; mean pairwise distance, in contrast, suggests a stationary pattern through Cenozoic time. When secular trends in sampling intensity are taken into account using both well-established and new subsampling methods, however, the records through both windows broadly suggest stasis, a pattern also predicted from a comparison of morphospace and molecular phylogeny. While disparity as measured by mean pairwise distance seems to be robust to sampling, the other results—disparity as well as diversity—hinge upon whether we believe that subsampling algorithms do a better job at uncovering true diversity history than a literal reading of the fossil
record, or whether they simply trade sampling bias for another bias resulting from changing relative abundance distributions.

A similar pattern has been discovered in another group of open ocean microfossils, the planktonic foraminifera: a literal reading of their deep-sea record also suggests a steep and roughly monotonic rise in species richness through Cretaceous and Cenozoic time, but, when sampling is accounted for by either subsampling or modeling, a much more gradual rise is recovered, with peak diversity in the Cretaceous (Lloyd et al. 2012a). And the fossil record of cocclithophorids also contrasts a literal reading of the Cretaceous-Cenozoic record with results obtained when sampling is taken into account (Lloyd et al. 2012b).

Without the compelling empirical suggestion of a Cenozoic rise in diatom diversity, it is worth considering whether an unfettered diversification would be expected *a priori*. The answer is not immediately obvious, but surely requires consideration of the relationship between phytoplankton diversity and both physical and chemical oceanography—as highlighted, for example, by global marine ecosystem models that implicate the role of resource availability and dispersal in controlling phytoplankton diversity (Follows et al. 2007). While these oceanographic factors are undoubtedly linked to climate, how exactly changes in climate would be expected to affect diatom diversity is a question deserving of further attention. We do note the correspondence between the high rates of diversity change around the Eocene-Oligocene boundary and the major shift in the Earth’s climate system observed around that time. The decline in diatom diversity, particularly pronounced in the subsampled results, coincides broadly with a sharp drop in ocean temperature (about 5°C cooling in the high latitudes, Liu et al 2009), thought to be driven by the establishment of circum-Antarctic circulation (Kennett 1982) and/or decreases in atmospheric CO2 (Pagani et al. 2005).

If the pattern of relative Cenozoic stasis in diversity and disparity to which the results presented here point is accurate, most of the marine planktonic diatoms’ diversification was a Mesozoic to earliest Cenozoic event, perhaps prompting a reexamination of evolutionary narratives in which a Cenozoic rise in diatom diversity features prominently. In many of these narratives, however, diatom diversity merely stands as a proxy for diatom participation in the silica cycle, yet the number of taxa is only one factor in their importance to silica cycling. It is also conceivable that there were changes through time in diatom abundance or the rate of diatomaceous sediment deposition. Such changes could, at least theoretically, be independent of
diversity; consider, for example, that in the Southern Ocean diatom ooze belt, perhaps the most important area of diatomaceous silica deposition today, sediments are dominated by just one species, *Fragilariopsis kerguelensis*, constituting up to 60 to 90% of total diatom abundance (Zielinski and Gersonde 1997). In addition to its abundance, *F. kerguelensis* is rather heavily silicified, illustrating also the potential role of changes in silicification to the diatoms’ biogeochemical impact. The origination of even a small number of such numerically dominant or robustly silicified taxa could potentially expand the diatoms’ role in the silica cycle to an extent much greater than the concomitant taxonomic diversification. Indeed, as described in the discussion of subsampling methods above, frequency distributions in the *Neptune* database indicate an increase in dominance through time compatible with such a scenario.

From the perspective of diatoms serving as starting points for biologically inspired design, one implication of an early exploration of morphospace might be that fossil morphologies ought to be considered alongside those of extant diatoms. Particularly if the maximum range of diatom form was achieved early in the Cenozoic, fossils from that time period may provide a range of biological constructions of engineering value not available in recent forms.

There are also important limitations of this study that must be considered. We have only considered the taxonomic and morphological diversity of marine planktonic diatoms here, although these only account for as estimated one-tenth the total diversity of diatoms (Kooistra et al 2007). While we are not aware of systematic compilations of global diversity history of either benthic or freshwater diatoms, we expect the trajectories may be different than in the marine plankton—freshwater diatoms, for instance, being dominated by the pennate clade, whereas centric diatoms predominate in the marine planktonic realm (Round et al 1990). Constructing a morphospace capable of representing the full taxonomic and temporal sweep of a clade as large and diverse as the marine planktonic diatoms also requires trade-offs in the level of morphological detail that can be recorded. For example, changes in the degree of silicification of diatom frustules are not well captured by the morphological characters in this study, since these are not necessarily visible in those characters that can be coded cohesively at the genus level. If predictions from the fossil record of radiolarians (Lazarus et al. 2009) and semi-quantitative observations of the diatom fossil record (Finkel and Kotrc 2010) hold true, diatoms ought to show a reduction in silicification over the Cenozoic Era, a pattern of interest to engineers seeking structures that maximize strength with minimal use of constructional material. Such trends may
be best investigated by looking at morphological changes within long-ranging genera (such as *Stephanopyxis*), where insight might be gained to how nature does more with less. For figured examples of such morphologies, we refer the reader to Gombos (1980), Olshtynskaya (1990 and 2002), Sims (1986, 1988 and 1990), and Round (1990, particularly pp. 144, 166, 172, 182-189, 202, 216, 226-229, 268-279, 316, 330, 416 and 564).
Bibliography


Kotrc B and Knoll AH (in review, [a]) A morphospace of planktonic marine diatoms, part I: Two views of disparity through time. Submitted to Paleobiology
Kotrc B and Knoll AH (in review, [b]) A morphospace of planktonic marine diatoms, part II: Sampling standardization reveals stasis. Submitted to *Paleobiology*


Lazarus DB, Kotrc B, Wulf G and Schmidt DN (2009) Radiolarians decreased silicification as an evolutionary response to reduced Cenozoic ocean silica. PNAS 106(23): 9333—9338


Figure Captions

Fig. 1. Diatom diversity through time over the Cenozoic era, based on analysis of the *Neptune* database by range-through taxon counting (RT), sampled-in-bin taxon counting (SIB), subsampling by-lists, weighted by occurrences (OW), and shareholder quorum subsampling (SQS).

Fig. 2. Number of lists (of taxa found at a particular depth in a particular borehole) in the *Neptune* database through the Cenozoic era.

Fig. 3. Left: first two principal coordinate (PCO) axes of marine planktonic diatom morphospace; plot symbols reflect states morphological characters describing valve shape in plan view and presence/absence of raphe. Right: the same plot populated through time using occurrence data from the *Neptune* database; colored polygons below show convex hull areas enclosing points in time bins above of the corresponding color.

Fig. 4. Metrics of diatom disparity through time under in-bin sampling of the *Neptune* occurrence data taken at face value (top row) and shareholder quorum subsampling of the same dataset (bottom row). Plots on the left show the average dissimilarity (or morphological distance) between taxa measured as the proportion of character state mismatches to possible matches, plots on the right show the volume of convex hulls enclosing genera present in each time bin (in the top plot, 3–10 dimensions, in the bottom plot, 3 dimensions only).
Subsampled species richness

Sampled species richness

Geologic Time (Ma)

RT
(Sample through)

SIB
(Sampled in-bin)

OW
(Subsampled by list, weighted by occurrences)

SQS
(Shareholder quorum subsampling)

SIB
(Sampled in-bin)

RT
(Sample through)

Geologic Time (Ma)
Mean pairwise distance
(mean pairwise character mismatches)

Convex hull volume
(standardized to highest value)

Geologic Time (Ma)

Mean pairwise distance

Convex hull volume

Geologic Time (Ma)

Mean pairwise distance

Convex hull volume

Geologic Time (Ma)