

## GRAPH MATCHING A SKELETONIZED THEORETICAL MORPHOSPACE WITH A CLADOGRAM FOR GOMPHONEMOID-CYMBELLOID DIATOMS (BACILLARIOPHYTA)

JANICE L. PAPPAS

*Museum of Paleontology, University of Michigan  
1109 Geddes Ave., Ann Arbor, MI 48109-1079, USA  
[jlppapas@umich.edu](mailto:jlppapas@umich.edu)*

Received 2 March 2010

Accepted 5 April 2010

A three-dimensional (3D) theoretical morphospace of gomphonemoid and cymbelloid diatoms was skeletonized using concepts from extended Reeb graph analysis and Morse theory. The resultant skeleton tree was matched to a cladogram of the same group of related taxa using adjacency matrices of the trees and ordinated with multidimensional scaling (MDS) of leaf nodes. From this, an unweighted path matrix based on the number of branches between leaf nodes was ordinated to determine degree of matched tree structures. A constrained MDS of the path matrix, weighted by ranked MDS leaf node groups as facets, was used to interpret taxon environmental tolerances and habitat preferences with respect to adaptive value. The methods developed herein provided a way to combine results from morphological and phylogenetic analyses and interpret those results with respect to an aspect of evolutionary process, namely, adaptation.

*Keywords:* Adjacency Matrix; Cymbella; Diatoms; Facet Diagram; Gomphonema; Morse Theory; Multidimensional Scaling; Path Matrix; Reeb Graph; Skeleton Tree.

### 1. Introduction

One of the most important concepts in biology is the concept of morphological change over time. This concept is measured in many ways, but in all instances, capturing this concept as a quantity is a useful assessment and indicator for the direction and timing of evolution.

The multitude of organisms and how they are related to one another is encompassed in the concept of morphological change over time. We identify the extent of variety in many ways, but our first inclination is to look at morphological forms, and we want to know how the forms are related, so we look at organisms' common ancestor. In this regard, the study of taxa morphological variation and phylogenetic history is instrumental. More specifically, total morphological variation may be used as a kind of boundary to phylogenetic history, and phylogenetic history may be used as a direction in which only particular forms of morphological variation exist.

What is the relation between all the possible morphological variants and shared common ancestry for a group of related taxa? There have been many studies done separately on morphological variation and cladistic analyses of taxon relatedness, but rarely has there been an attempt to combine results to produce a single outcome. To determine how morphological change and shared common ancestry are related, deriving a commonality between these disparate types of biological assessment is methodologically necessary.

Morphospace is an important concept in studying organism form, for example, by shape,<sup>1</sup> surface,<sup>2,3</sup> size,<sup>4</sup> growth,<sup>5</sup> character spaces,<sup>6</sup> or genome.<sup>7</sup> A theoretical morphospace is the extent of all possible morphologies for a group of organisms.<sup>4</sup> In general, a theoretical morphospace will be broadest at the base or in valleys indicating less adaptive morphology (depicting most similarity)<sup>8</sup> and smallest at local peaks indicating more adaptive morphology (depicting most dissimilarity).<sup>9</sup> Functions, relations, statistics, or other mathematics may be the basis of theoretical morphospaces.

Cladograms are topological trees<sup>10</sup> as nested hierarchies<sup>11</sup> that represent the descendants of a shared common ancestor for a group of related taxa<sup>11</sup> as directed, connected acyclic graphs.<sup>12</sup> Such graphs are topological spaces since twisting tree branches at their vertices does not change the relation among terminal endpoints or leaf nodes. The process of producing a “best” phylogenetic tree is an NP-complete optimization problem.<sup>13</sup> As such, cladograms are not representative of a fixed tree of shared common ancestry. Rather, cladograms represent a best possible hypothesis at a given time.

Theoretical morphospace is defined by morphological transitions and the extent that they are representations of adaptive fitness as an adaptive landscape.<sup>9,14</sup> That is, those characteristics of an organism’s existence that contribute to fitness also define adaptive value. Using the geometric characteristics of a theoretical morphospace, different types of fitness may be identified. From terminology in geomorphic studies, “peaks” are maximum fitness, “valleys” are minimum fitness, and “passes” or saddles (= saddle points) between maxima and minima represent intermediate fitness that implies morphological transition.

The cladogram is also a representation of adaptive radiation<sup>15</sup> and is composed of nodes and branches whereby internodes, outgroups, and terminal or leaf nodes may be characterized as saddles, minima, and maxima, respectively. Both theoretical morphospaces and cladograms contain information about adaptive changes over time. That is, an adaptive walk<sup>14</sup> may be used to map the changes in morphology for a group of taxa, or cladograms may show changes in morphology or other characters that are successful at a given time of adaptive radiation. A possible inference made from a shared common ancestry is that the taxa depicted in a cladogram exhibit a combination of characters that are successful in perpetuating the existence of those taxa at a given time. That success may be related to fitness as an adaptive value.

A cladogram may be embedded in, mapped to, or matched within a theoretical morphospace to determine the relation that shared common ancestry has to morphological variants and may be a generalization of the placement of descendants, and if known, their immediate ancestor. The degree of matching from node to node is a measure of placement of the cladogram within a theoretical morphospace. Theoretical morphospace may act as the boundary extension of variation within a clade or for an entire phylogeny, or a phylogeny may act as a constraint on all the possible morphologies allowed in theoretical morphospace.

Creating a framework relating theoretical morphospace and cladogram involves characterizing the “essential backbone” or “skeleton” of a morphospace as a topological structure and matching this to a cladogram. If maxima, minima and saddles on the “essential backbone” of morphospace and nodes in a cladogram have meaning in terms of adaptive fitness with shared common ancestry, then the degree of matching between the two topological structures will exhibit varying degrees of an evolutionary process, such as adaptive fitness.

Using a general, rough framework of a fitness or adaptive landscape, theoretical morphospaces and cladograms may be related to each other using graph-matching techniques. To illustrate the method, the 3D theoretical morphospace for gomphonemoid and cymbelloid diatoms<sup>3</sup> will be skeletonized and matched to the cladogram<sup>16</sup> for these same diatoms. The results from graph matching may be used in an assessment of taxon adaptation of gomphonemoid and cymbelloid diatoms to their environment.

To this end, we will develop the characterization of a 3D theoretical morphospace as a topological space and how a skeleton tree is extracted. Then, the skeleton tree will be matched to the cladogram and evaluated by mapping information about taxon environmental preferences to the graph matched structures indicating taxon adaptation.

## 2. Methods

The methodology for constructing a skeleton tree from the topological properties of the theoretical morphospace, and for graph matching the skeleton tree to a cladogram of gomphonemoid and cymbelloid diatoms, is fully detailed in the Appendix.

## 3. Results

The 3D theoretical morphospace was rotated approximately  $180^\circ$  (Fig. 1), and its outline curve is shown in Fig. 2. Contours are drawn in the  $y - z$  plane based on peaks, valleys, and at points of inflection in its outline curve (Fig. 3). Contours of the rotated 3D theoretical morphospace in the  $x - y$  plane are illustrated in Fig. 4. Within the contours (Fig. 3), triangulations are drawn (Fig. 5), and critical areas and their influence zones are marked by critical points as maxima, minima and saddles (Fig. 6). Critical points are thinned and illustrated in Fig. 7. Connections

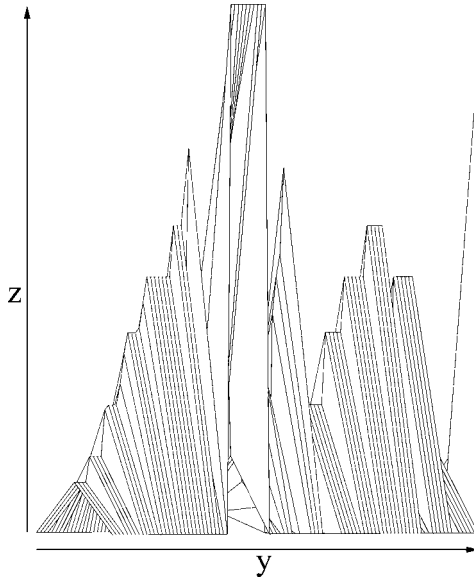


Fig. 1. 3D theoretical morphospace  $z - y$  projection.

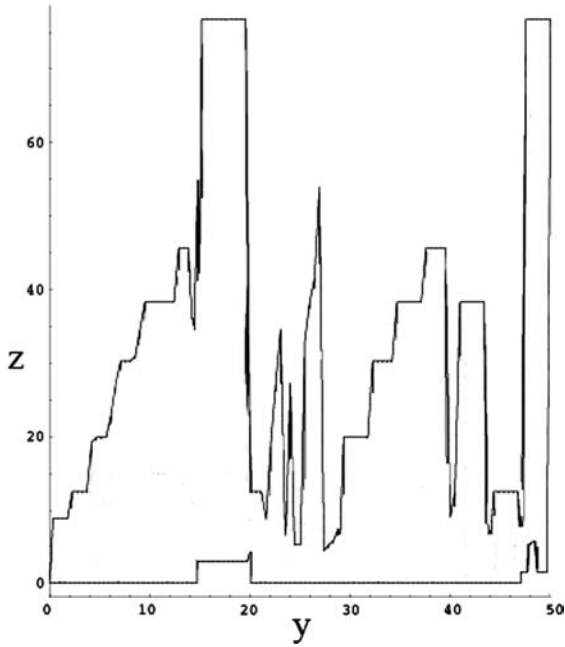


Fig. 2. 3D theoretical morphospace  $z - y$  projection outline.

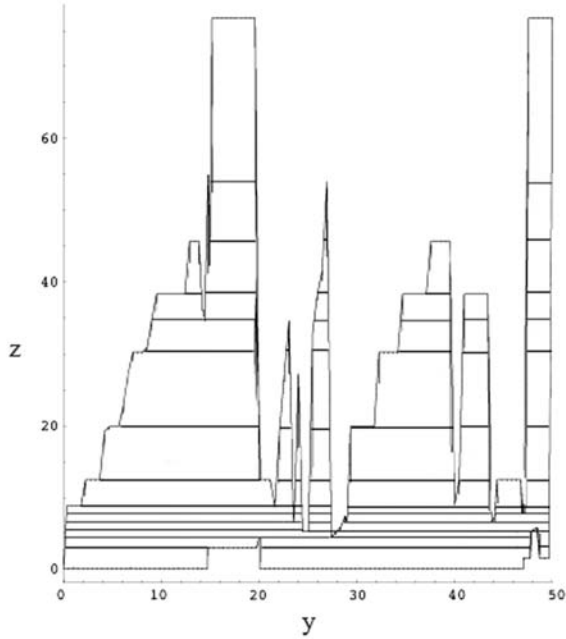


Fig. 3. 3D theoretical morphospace  $z - y$  projection of contour lines on outline.

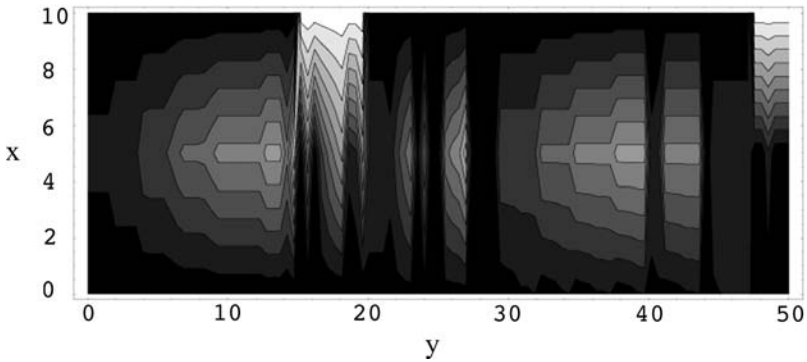


Fig. 4. 3D theoretical morphospace  $x - y$  projection of contour lines. Darkest areas are minima, lightest areas are maxima, and shades of gray are local maxima, minima or saddles.

between critical points are overlaid on the contour diagram as a skeleton tree via an extended Reeb Graph (Fig. 8).

From Kocielek and Stoermer<sup>16</sup>, the naviculoid-gomphonemoid-cymbelloid cladogram is depicted and modified to show critical points (Fig. 9). The 3D theoretical morphospace<sup>3</sup> with taxon labels Didymosphenia, Reimeria, Encyonema, and Cymbella for cymbelloid forms and Gomphonema, Gomphoneis, and

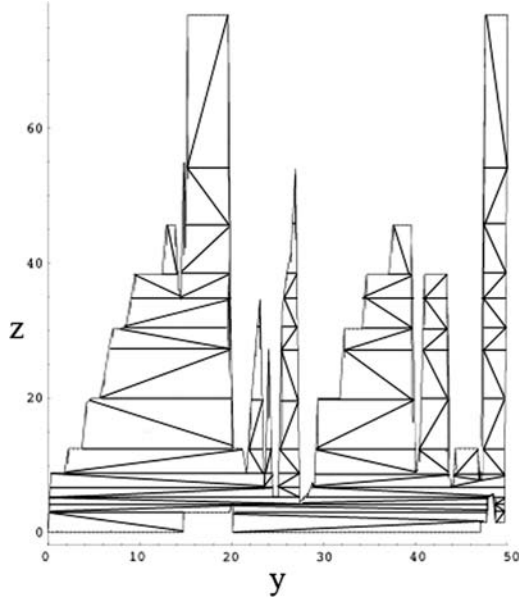


Fig. 5. 3D theoretical morphospace  $z - y$  projection of triangulated contour lines.

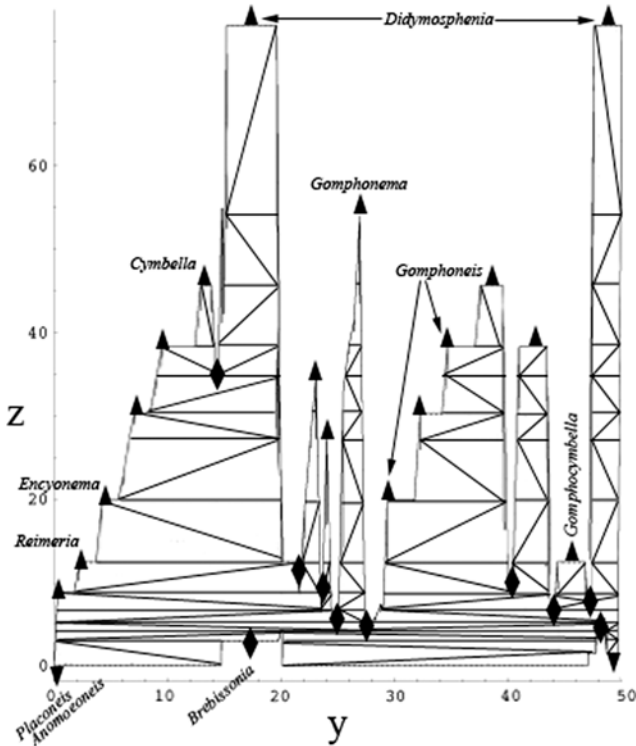


Fig. 6. 3D theoretical morphospace  $z - y$  projection of all maxima (▲), minima (▼) and saddles (◆).

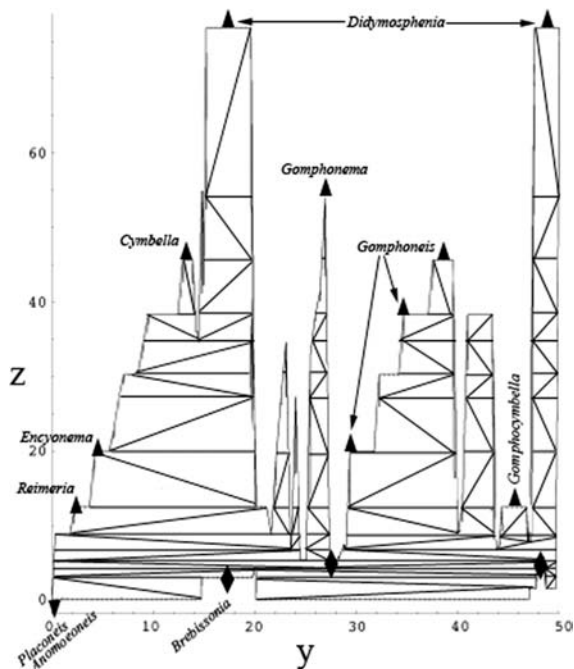


Fig. 7. 3D theoretical morphospace  $z - y$  projection of thinned critical points for taxon labeled critical areas and influence zones.

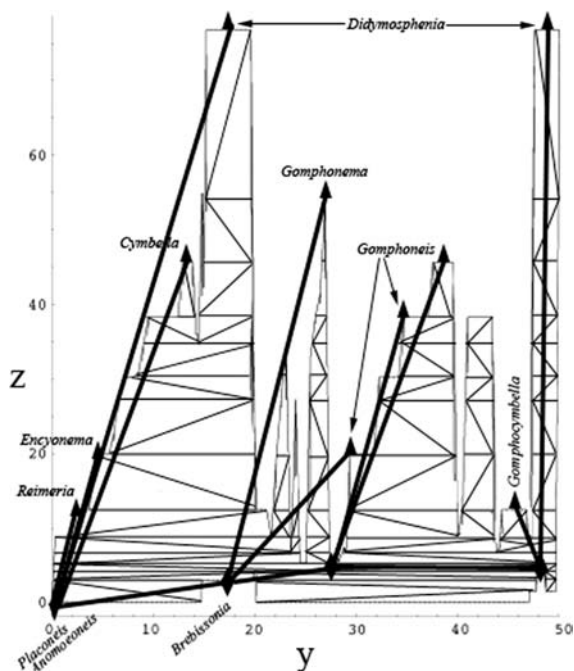


Fig. 8. 3D theoretical morphospace  $z - y$  projection of connections between critical points as a skeleton tree.

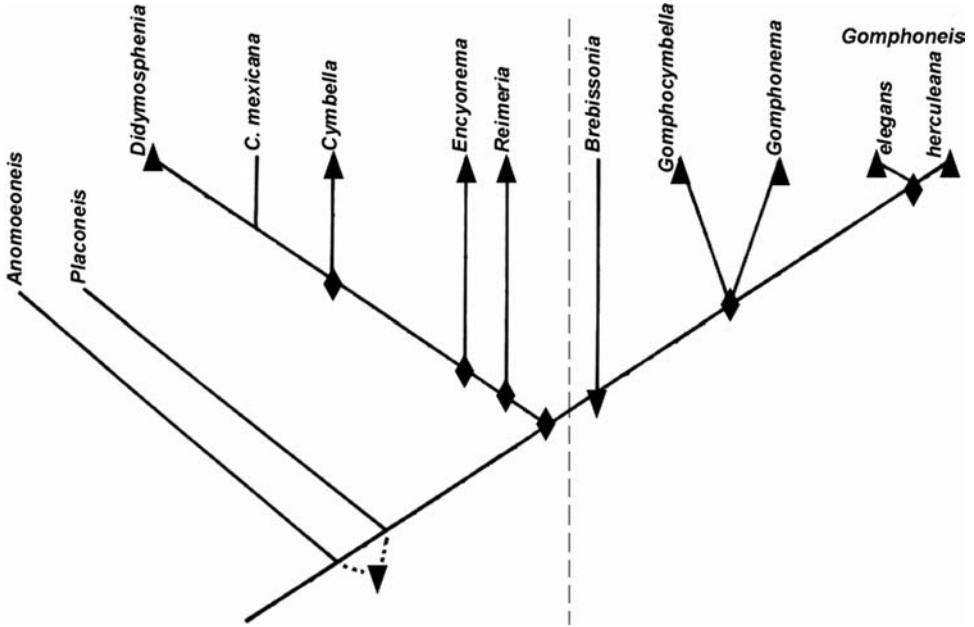


Fig. 9. Cladogram of naviculoid-cymbelloid-gomphonemoid taxa<sup>16</sup> with maxima (▲), minima (▼) and saddles (◆).

Gomphocymbella for gomphonemoid forms is also depicted with skeleton tree overlain on the morphospace (Fig. 10). The naviculoid forms of Anomoeoneis and Placoneis are represented as one basal node and are at the origin of the 3D theoretical morphospace and cladogram as the root (Fig. 9). The methods used to produce Figs. 8 and 10, depicted in a stepwise fashion in Figs. 3, 5, 6, and 7, are fully elucidated in the Appendix A, Sections A.1–A.5.

Initially, cladogram and the skeleton tree of 3D theoretical morphospace are depicted as Gini trees<sup>17</sup> in Fig. 11. Note that, at each clade level, the graphs are matched for naviculoid forms, Gomphonema, Gomphoneis, and Didymosphenia (Fig. 11). In addition, note that there are mismatches for Reimeria, Encyonema, Cymbella, and Gomphocymbella (Fig. 11). However, because 3D theoretical morphospace includes morphological variation for each taxon, some of the taxa occur more than once in the skeleton tree. This depiction of variation in the skeleton tree is helpful in matching it to a static graph such as a cladogram.

Adjacency matrices were determined for both the skeleton tree and the cladogram, and the matrices were analyzed by MDS. *S*-stress was 0.005. The resultant ordination depicts the structures in eigenspace (Fig. 12). The coordinates in a large ring on the periphery are internodes, and coordinates grouped in the center are leaf nodes (Fig. 12). It is the leaf nodes that we turn our attention to, since they are of interest in further analyses.



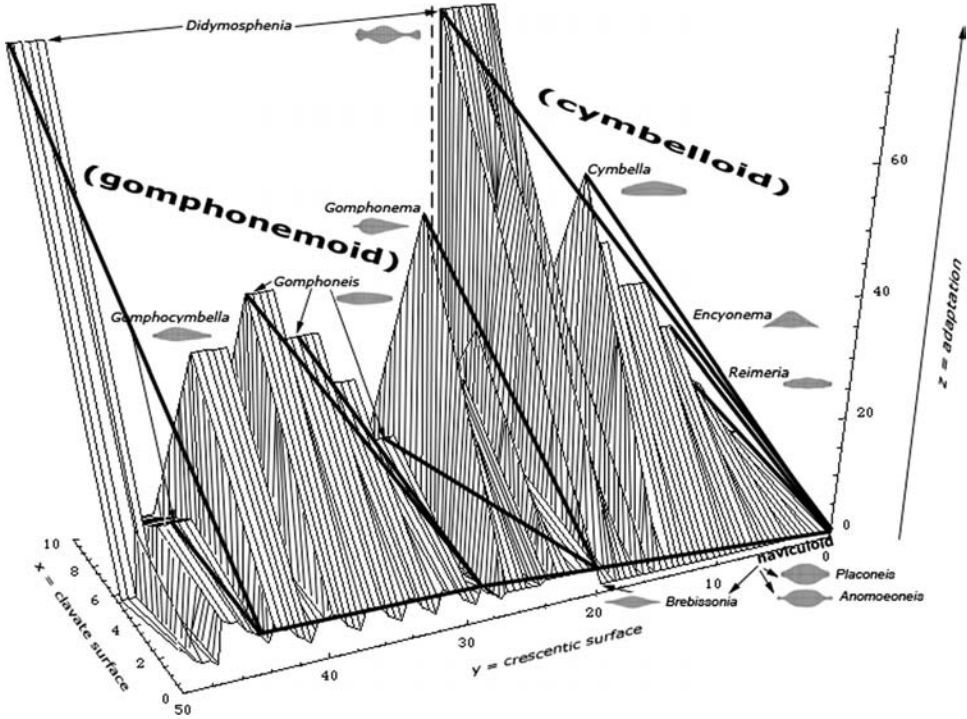


Fig. 10. Labeled 3D theoretical morphospace<sup>3</sup> with superimposed skeleton tree.

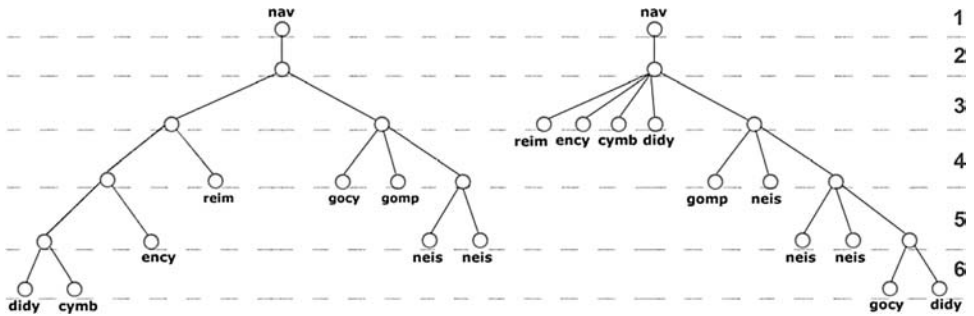


Fig. 11. Gini trees of cladogram on the left and skeleton tree on the right. Taxon names: nav = naviculoid, reim = Reimeria, ency = Encyonema, cymb = cymbelloid (Cymbella), didy = Didymosphenia, gomp = gomphonemoid (Gomphonema), gocy = Gomphocymbella, and neis = Gomphoneis.

MDS of only the leaf nodes from both trees is shown in Fig. 13. Clustered groups are circumscribed as facets in a facet diagram. Axial partitions<sup>18</sup> were drawn to separate the larger general taxon affiliations. From left to right, taxa progress from most recent common ancestor, naviculoid, to mostly crescentic forms, then to



clavate forms for both the skeleton tree and cladogram, and this agrees with both tree structures independently.

Each leaf node group (Fig. 13) is composed of particular taxa. On the left, naviculoid forms from both trees and Reimeria from the skeleton tree comprise the first group. In the center and at the top of the leaf node diagram, Encyonema and Gomphoneis from the cladogram form another group. Around the origin, a third group has Didymosphenia, Cymbella, Encyonema, and Gomphocymbella from the skeleton tree with Reimeria and Gomphoneis from the cladogram. In the center and at the bottom of the leaf node diagram, Gomphoneis from the skeleton tree with Cymbella and Didymosphenia from the cladogram form the next group. Finally, Gomphoneis and gomphonemoid forms from the skeleton tree with Gomphocymbella and gomphonemoid forms from the cladogram are the taxa in the group on the right (Fig. 13).

### 3.1. Graph matched structures and interpretation

An unweighted path matrix was devised from the number of branches in each tree from taxon to taxon and ordinated using MDS. The path matrix is a comparison of each taxon from one kind of tree to the other (Fig. 14). For taxa with multiple leaf nodes, only one representation per taxon was used. In this regard, the leaf node that was first maximum encountered on the skeleton tree was used in the unweighted path matrix.

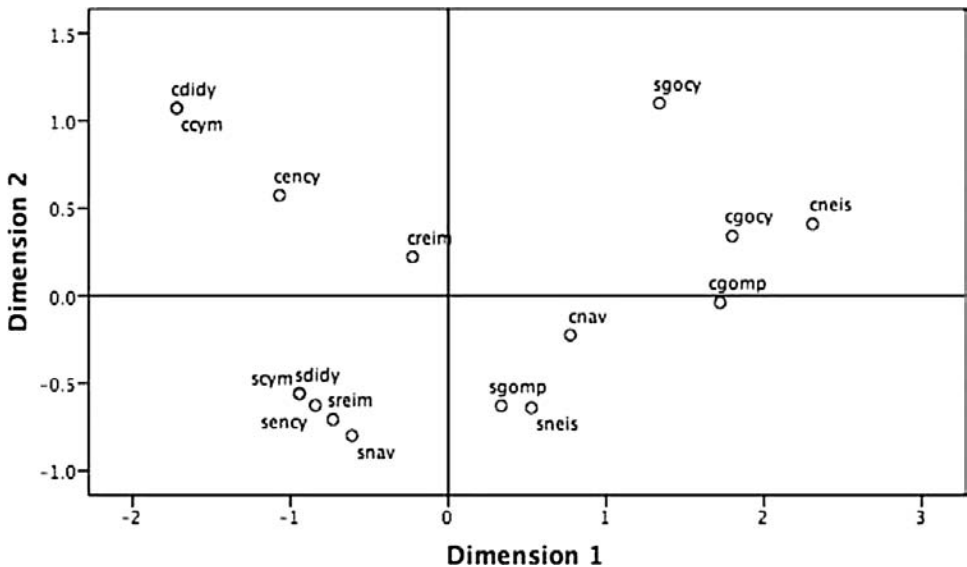


Fig. 14. MDS of path matrices for skeleton tree and cladogram, showing divergent paths of branching patterns for each tree. For taxon name identification, see Fig. 12.

Divergent paths occurred between the skeleton tree and cladogram branching patterns (Fig. 14). For the skeleton tree, the trajectory of taxa is a cluster in the lower left of the MDS diagram that includes Cymbella, Encyonema, Reimeria, Didymosphenia, and naviculoid, followed by gomphonemoid and Gomphoneis in the lower center, and finally in the upper right, Gomphocymbella is depicted (Fig. 14). For the cladogram, the trajectory of taxa starts in the upper left of the MDS diagram with Cymbella and Didymosphenia, followed by Encyonema, Reimeria, and naviculoid descending into the lower right. The taxa that occur on the right side are gomphonemoid, Gomphocymbella, and Gomphoneis (Fig. 9). When considering tree paths, graph matching for all taxa does not occur, and this is expected given the comparison of trees in the Gini tree diagram (Fig. 11).

Branching and leaf node patterns were combined and used in weighted path matrix analysis. Using MDS of leaf nodes (Fig. 13) as a facet diagram, taxon groups as cluster constraints were used as weights in the path matrix. The weights were set at 5 = within-group taxa, 4 = adjacent between-group taxa, and 3 = non-adjacent between group taxa. Weights are arbitrary values but must be set in a descending order to identify relative taxon position as approximate change in clade age or skeleton tree level with respect to both trees. Each group weights taxon paths by constraining each path, thereby equilibrating paths with respect to clade age or skeleton tree level, so skeleton tree and cladogram taxa are best matched. For further details on graph matching using path and weighted path matrices, see the methodological explanation given in Appendix A, Sections A.6–A.8.

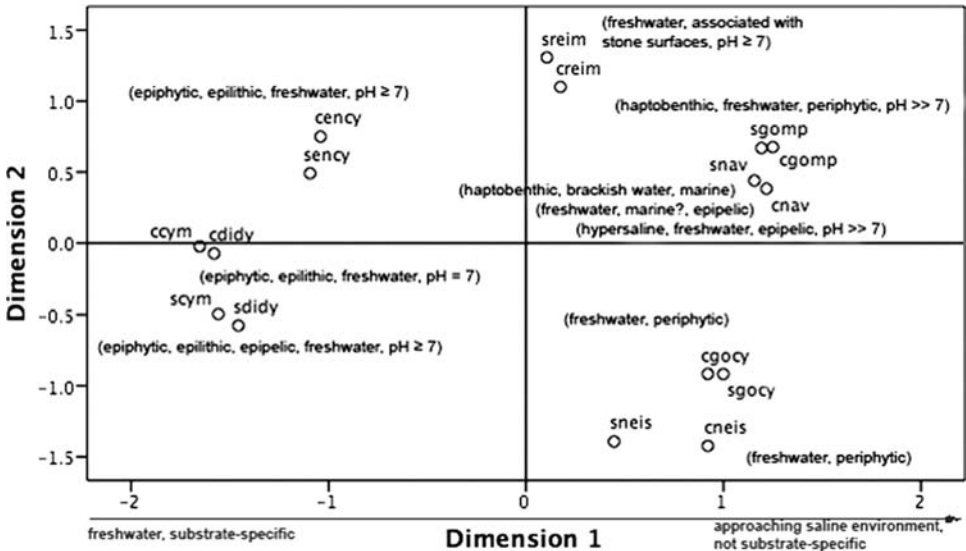


Fig. 15. Constrained MDS of skeleton tree and cladogram weighted path matrix, where weights are based on clusters from Fig. 13. Constrained MDS shows proximity of taxa from each tree and habitat preferences and environmental tolerances<sup>19–22</sup> in an approximate adaptive walk. For taxon name identification, see Fig. 12. *S*-stress = 0.005.

Constrained MDS eigenspace depicts degree of graph matching between cladogram and skeleton tree (Fig. 15). In addition, habitat preferences<sup>19,20,21</sup> and environmental tolerances<sup>19,22</sup> are super imposed on constrained MDS path space to depict an approximate adaptive walk with regard to graph matched cladogram and skeleton tree. This adaptive walk is depicted on the first dimension of constrained MDS path space (Fig. 15). From left to right, there is a progression from freshwater and specific substrates to almost saline and non-specific substrates for habitat preferences of taxa (Fig. 15). In addition, taxa on the left are tolerant of pH 7 or slightly above this, while taxa on the right are tolerant of pH much greater than 7 (Fig. 15). Cymbelloid taxa generally occur on the left, while gomphonemoid and naviculoid taxa occur on the right (Fig. 15). Constrained MDS path space agrees with previous analyses of naviculoid, gomphonemoid, and cymbelloid habitat preferences and environmental tolerances.<sup>3</sup>

#### 4. Discussion

Construction of theoretical morphospaces and cladograms is a well-established practice in biological research. Constructing a skeleton tree from a theoretical morphospace is a novel way to evaluate the matching of morphological with phylogenetic change. The branching in the skeleton tree of gomphonemoid and cymbelloid morphologies depicts adaptive environmental characters. The branching in the cladogram depicts distribution of those adaptive characters with respect to how gomphonemoid and cymbelloid morphologies are related.

It has long been established that there is no single method for quantitatively assessing the evolutionary history of a group of related taxa.<sup>23</sup> Both morphological and phylogenetic studies and combination studies thereof are valuable in advancing our knowledge about the dynamics of morphological trends in evolving lineages.<sup>24</sup> However, care must be taken when combining morphospaces and cladograms. For instance, in one study, properties of right triangles were used in a geometric algorithm to linearly map a cladogram to morphospace.<sup>25</sup> This method may produce erroneous conclusions if results are based on linear transitions in large clades or a sparse record of morphological diversity.<sup>26</sup>

In our work, we matched skeletonized theoretical morphospace with a cladogram in a topological framework as a way to circumvent geometric assumptions. Optimization methods may be applicable in determining the best possible skeleton tree, just as such methods are used to determine the best cladogram. For multiple taxon labels identified in a skeletonized theoretical morphospace, choices other than first maximum encountered may be used. The key is consistency throughout the graph matching process. For larger numbers of taxa, optimization concerns about the skeletonization and graph matching processes need to be addressed.

Using an extended Reeb graph was a simple, efficient way to skeletonize a 3D theoretical morphospace. Reeb graphs are well known to be reliable in terms of reconstructing 3D shapes from the extracted skeleton.<sup>27,28</sup> Medial axis models,<sup>29</sup>

including shock graphs,<sup>30</sup> are also used to skeletonize 3D shapes. However, these methods may entail high computational cost and are sensitive to noise.<sup>29</sup>

Sometimes, minor critical points, especially multiple saddles, are indicated in the extended Reeb graph because of the presence of noise or an insufficient sampling of morphologies that constitutes the theoretical morphospace. In this case, simplification of the skeleton may be necessary to remove such obscuring factors, thereby preserving the global structure of the skeleton.<sup>31</sup> To thin<sup>32,33</sup> minor critical points, averaging multiple saddles<sup>32</sup> may be used to preserve the original 3D theoretical morphospace shape. Theoretical morphospaces are usually simple figures, and dealing with multiple saddles should be a relatively straightforward exercise. If need be, optimization algorithms can be used.

There are many ways to graph match structures. The methods that we used are simple, efficient, and computationally inexpensive. Other methods involving metrics, such as the geodesic<sup>34</sup> or Fréchet distances,<sup>35</sup> may be used, but are not necessarily applicable in the case of theoretical morphospaces. Other methods are computationally expensive.<sup>36</sup> Such methods may be warranted when considering more complicated structures, but this situation is unwarranted with simple structures such as theoretical morphospaces.

In the most difficult cases, inexact graph matching or fuzzy graph matching between the skeleton tree and cladogram may still be achieved. This entails comparing the structures in an overall sense and the compatibility of their graph labels (i.e., names of taxa) by using relaxation-labeling methods.<sup>37,38</sup> In this way, an approximate matching will occur.

In general, graph matching tree structures and evaluation of them may induce additional research. Such research may include matters concerning existence of organism form, directionality in clade branching, and diversification of taxa. Such matters are important in furthering our understanding of evolution.

Overall, skeletonizing a 3D theoretical morphospace and matching it to a cladogram for a group of related taxa produces meaningful, novel ways of interpreting adaptive fitness. Such interpretations have implications not only for understanding the evolution of the taxa used in our study, gomphonemoid and cymbelloid diatoms, but also for the interpretation of particular adaptive measures such as environmental tolerances and habitat preferences.

## Acknowledgment

We would like to thank an anonymous reviewer for suggestions that improved this manuscript.

## References

1. Kendall DG, Barden D, Carne TK, Le H, *Shape and Shape Theory*, p. 306, John Wiley & Sons, Ltd., Chichester, UK, 1999.

2. Pappas JL, Theoretical morphospace and its relation to freshwater gomphonemoid-cymbelloid diatom (Bacillariophyta) lineages, *J Biol Systems* **13**:385–398, 2005.
3. Pappas JL, More on theoretical morphospace and its relation to freshwater gomphonemoid-cymbelloid diatom (Bacillariophyta) lineages, *J Biol Systems* **16**:119–137, 2008.
4. McGhee, Jr. GR, *Theoretical Morphology: The Concept and Its Applications*, p. 316, Columbia University Press, New York, 1999.
5. Raup DM, Geometric analysis of shell coiling: General problems, *J Paleontol* **40**:1178–1190, 1966.
6. Thomas RDK, Shearman RM, Stewart GW, Evolutionary exploitation of design options by the first animals with hard skeletons, *Science* **288**:1239–1242, 2000.
7. Christensen K, DiCollobiano SA, Hall M, Jensen HJ, Tangled nature: a model of evolutionary ecology, *J Theor Biol* **216**:73–84, 2002.
8. Costello JH, Colin SP, Dabiri JO, Medusan morphospace: phylogenetic constraints, biomechanical solutions, and ecological consequences, *Invert Biol* **127**:265–290, 2008.
9. McGhee, Jr. GR, *The Geometry of Evolution: Adaptive Landscapes and Theoretical Morphospaces*, p. 200, Cambridge University Press, Cambridge, UK, 2007.
10. Hendy MD, Penny D, Cladograms should be called trees, *Syst Zool* **33**:245–247, 1984.
11. Eldredge N, Cracraft J, *Phylogenetic Patterns and the Evolutionary Process*, p. 349, Columbia University Press, New York, 1980.
12. Bollobás B, *Graph Theory: An Introductory Course*, p. 180, Springer-Verlag, New York, 1979.
13. Graham RL, Foulds LR, Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time, *Math Biosci* **60**:133–142, 1982.
14. Niklas KJ, *The Evolutionary Biology of Plants*, p. 449, University of Chicago Press, Chicago, 1997.
15. Felsenstein J, Phylogenies and the comparative method, *Am Nat* **125**:1–15, 1985.
16. Kocielek JP, Stoermer EF, A preliminary investigation of the phylogenetic relationships among the freshwater, apical pore field-bearing cymbelloid and gomphonemoid diatoms (Bacillariophyceae), *J Phycol* **24**:377–385, 1988.
17. Billera LJ, Holmes SP, Vogtmann K, Geometry of the space of phylogenetic trees, *Adv Appl Math* **27**:733–767, 2001.
18. Borg I, Groenen P, *Modern Multidimensional Scaling: Theory and Applications*, p. 471, Springer, New York, 1997.
19. Patrick R, Reimer CW, *The Diatoms of the United States, Exclusive of Alaska and Hawaii, Vol. 2, Part 1* p. 213, The Academy of Natural Sciences of Philadelphia, Pennsylvania, 1975.
20. Round FE, *The Ecology of Algae*, Cambridge University Press, Cambridge, 1981.
21. Round FE, Crawford R, Mann DG, *The Diatoms: Biology and Morphology of the Genera*. Cambridge University Press, Cambridge, 1990.
22. Lowe R, *Environmental Requirements and Pollution Tolerance of Freshwater Diatoms*, EPA-670/4-74-005, National Environmental Research Center, Office of Research and Development, United States Environmental Protection Agency, Cincinnati, Ohio, 1974.
23. Duncan T, Philips RB, Wagner, Jr. WH, A comparison of branching diagrams derived by various phenetic and cladistic method, *Syst Bot* **5**:264–293, 1980.
24. Alroy J, Understanding the dynamics of trends within evolving lineages, *Paleobiology* **26**:319–329, 2000.
25. Stone JR, Mapping cladograms into morphospace, *Acta Zool* **84**:63–68, 2003.



26. Erwin DH, Disparity: morphological pattern and developmental context, *Palaeontology* **50**:57–73, 2007.
27. Shinagawa Y, Kunii TL, Kergosien YL, Surface coding based on Morse theory, *IEEE Comp Graphics Appl* **11**:66–78, 1991.
28. Biasotti S, Marini S, Mortara M, Patanè G, Spagnuolo M, Falcidieno B, 3D shape matching through topological structures, in *Proceedings of 12th Discrete Geometry for Computer Imagery Conference, Lecture Notes in Computer Science (LNCS) 2886*, Nyström I, (ed.), pp. 194–203, 2003.
29. Hilaga M, Shinagawa Y, Kohmura T, Kunii TL, Topology matching for fully automatic similarity estimation of 3D shapes, in *ACM SIGGRAPH*, New York, pp. 203–212, 2001.
30. Siddiqi K, Shokoufandeh A, Dickinson S, Zucker SW, Shock graphs and shape matching, *Inter J Comp Vis* **35**:13–32, 1999.
31. Takahashi S, Yakeshima Y, Fujishiro I, Topological volume skeletonization and its application to transfer function design, *Graphical Models* **66**:24–49, 2004.
32. Zhu X, Shape recognition based on skeleton and support vector machines, in *ICIC, Vol. 2*, Huang D-S, Heutte L, Loog M (eds.), Springer-Verlag, Berlin, pp. 1035–1043, 2007.
33. Sundar H, Silver D, Gagvani N, Dickinson S, Skeleton based shape matching and retrieval, in *Proceedings of the Shape Modeling International*, IEEE Computer Society, pp. 280–290, 2003.
34. Toponogov VA, *Differential Geometry of Curves and Surfaces — A Concise Guide*, p. 206, Birkhäuser, Boston, Basel, Berlin, 2006.
35. Veltkamp RC, Hagedoorn M, *State-of-the-Art in Shape Matching, Technical Report UU-CS-1999-27*, Utrecht University, The Netherlands, 1999.
36. Mailund T, Pedersen CNS, QDist—quartet distance between evolutionary trees, *Bioinformatics Appl Note* **20**:1636–1637, 2004.
37. Medasani S, Krishnapuram R, Choi Y, Graph matching by relaxation of fuzzy assignments, *IEEE Trans on Fuzzy Systems* **9**:173–182, 2001.
38. Torsello A, Hancock ER, Computing approximate tree edit distance using relaxation labeling, *Pattern Recogn Lett* **24**:1089–1097, 2003.
39. Kahn DW, *Topology: An Introduction to the Point-Set and Algebraic Areas*, Dover Publications, Inc., New York, 1995.
40. Milnor J, *Morse Theory, Annals of Mathematics Studies, Vol. 51*, p. 153, Princeton University Press, Princeton, New Jersey, 1963.
41. Matsumoto Y, An Introduction to Morse Theory, in Hudson K, Saito M (translators), *Translations of Mathematical Monographs, Vol. 208*, p. 219, American Mathematical Society, Providence, Rhode Island, 2002.
42. Reeb G, Sur les points singuliers d’une forme de pfaff complètement intergrable ou d’une fonction numérique [on the singular points of a complete integral pfaff form or of a numerical function], *Comptes Rendus Acad Science Paris* **222**:847–849, 1946.
43. Biasotti S, Mortara M, Spagnuolo M, Surface compression and reconstruction using Reeb graphs and shape analysis, in *Proceedings of Spring Conference on Computer Graphics*, Falcidieno B (ed.), Comenius University, Bratislava, Slovakia, pp. 175–184, 2000a.
44. Christiansen HN, Sederberg TW, Conversion of complex contour line definitions into polygonal element mosaics, *Computer Graphics* **12**:187–192, 1978.
45. Shinagawa Y, Kunii TL, Constructing a Reeb graph automatically from cross sections, *IEEE Comp Graphics Appl* **11**:44–51, 1991.
46. Lazarus F, Verroust A, Level set diagrams of polyhedral objects, in *Proc Symp Solid Modeling*, pp. 130–140, 1999.



47. Attene M, Spagnuolo M, Automatic surface reconstruction from point sets in space, in *Eurographics 2000 Vol. 19*, Gross M, Hopgood FRA (eds.), The Eurographics Association and Blackwell Publishers, UK, pp. 1–9, 2000.
48. Biasotti S, Topological techniques for shape understanding, in *Proceedings of the 5th Central European Seminar on Computer Graphics*, Budmerice, Slovakia, pp. 1–10, 2001.
49. Pascucci V, Scorzelli G, Bremer P-T, Mascarenhas A, Robust on-line computation of Reeb graphs: simplicity and speed, *ACM Trans Graphics* **26**:58-1–58-9, 2007.
50. Biasotti S, Falcidieno B, Spagnuolo M, Extended Reeb graphs for shape analysis and model compression, *Lecture Notes in Computer Science (LNCS)*, pp. 1–18, 2000b.
51. Biasotti S, Falcidieno B, Spagnuolo M, Extended Reeb graphs for surface understanding and description, in *Proceedings of 9th Discrete Geometry for Computer Imagery Conference, Lecture Notes in Computer Science (LNCS ) 1953*, Borgefors G, Nyström I, Sanniti di Baja G (eds.), Springer-Verlag, Berlin, pp. 185–197, 2000c.
52. Attene M, Biasotti S, Spagnuolo M, Shape understanding by contour-driven retiling, *The Visual Computer* **19**:127–138, 2003.
53. Xiao Y, Wergli, Siebert P, A discrete Reeb graph approach for the segmentation of human body scan, in *Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling*, IEEE Computer Society, Alberta, Canada, pp. 1–8, 2003.
54. Shokoufandeh A, Dickinson S, A unified framework for indexing and matching hierarchical shape structures, in *IWVF4, LNCS 2059*, Arcelli C, et al. (eds.), Springer-Verlag, Berlin, pp. 67–84, 2001.
55. Sá LB de, Mesquita A, Evolutionary synthesis of low-sensitivity equalizers using adjacency matrix representation, in *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation, GECCO '08, Atlanta, Georgia*, Association for Computing Machinery, New York pp. 1283–1290, 2008.
56. Cox TF, Cox MAA, *Multidimensional Scaling*, p. 213, Chapman & Hall, London, 1994.
57. Torsello A, Hancock ER, A skeletal measure of 2D shape similarity, *Comp Vis Image Underst* **95**:1–29, 2004.
58. Canter D (ed.), *Facet Theory: Approaches to Social Research*, Springer-Verlag, New York, 1985.
59. Takane Y, Young FW, DeLeeuw J, Nonmetric individual differences multidimensional scaling: an alternating least-squares method with optimal scaling features, *Psychometrika* **42**:7–67, 1977.
60. Young FW, Lewyckyj R, *ALSCAL-4 User's Guide*, Psychometric Laboratory, University of North Carolina, Chapel Hill, 1979.
61. Borg I, Shye S, *Facet Theory: Form and Content*, p. 208, Sage Publications, Inc., Newbury Park, California, 1995.
62. Heiser WJ, Groenen PJF, Cluster differences scaling with a within-cluster loss component and a fuzzy successive approximation strategy to avoid local minima, *Psychometrika* **62**:63–83, 1997.

## Appendix A

The following is a compendium of mathematical steps necessary in order to construct a skeleton tree from a theoretical morphospace, and match the tree to a cladogram for a group of related taxa.

### A.1. Theoretical morphospace and topological characteristics

A 3D theoretical morphospace is a topological space,  $X$ , that has topological properties<sup>9</sup> and is a collection of subsets,  $A \subseteq X$ . Then,  $X$  is a closed set if  $X - A = \{x \mid x \in X, x \notin A\}$  is an open set.<sup>39</sup> That is, for topological space  $X$ , and  $x \in X < A$ , the set is open, and for  $x \in X \leq A$ , the set is closed if the boundary of  $A$  is equal to  $A$ . The subsets,  $A$ , are the taxon spaces that occupy 3D theoretical morphospace,  $X$ . Each taxon space, when removed from the entire 3D theoretical morphospace, is an open set in 3D theoretical morphospace,  $X - A$ , when each  $A$  does not include its own boundary. That is, each  $A$  is an open subset if a point and those points in its neighborhood are in  $A$ . In 3D theoretical morphospace, if each taxon group is identified as a non-overlapping open set of pairwise comparisons,<sup>2,3</sup> and these open sets occupy distinct areas of the topological space that comprise the 3D theoretical morphospace, then the topological space,  $X$ , is also a Hausdorff space.<sup>39</sup>

Consider that the total 3D theoretical morphospace is bounded in  $x - y$ ,  $y - z$ , and  $x - z$  planes. Within this bounded space, we can define the 3D theoretical morphospace itself as a closed surface. Moreover, this closed surface can be shown to be compact since on the surface of a closed 3D theoretical morphospace, a finite number of polygons, such as triangles, can be drawn. The 3D theoretical morphological space is a compact closed topological space.<sup>39</sup>

Even more specifically, we can define this 3D theoretical morphospace surface as a two-dimensional (2D) surface that is a 2-manifold,  $M$ . In the neighborhood of a point on the 2-manifold, there is a planar surface or chart, and the collection of charts is an atlas on the surface of a 3D theoretical morphospace that resembles Euclidean space.<sup>39</sup> If there is a differentiable function,  $f$ , on  $M$  with respect to two critical points, then  $M$  is homeomorphic to a sphere.<sup>40</sup> The critical points may be degenerate or non-degenerate, and homeomorphism still holds true. An open subset,  $A$ , of  $M$  is also a manifold<sup>18</sup> that is homeomorphic to an  $n$ -dimensional space,  $\mathfrak{R}^n$ .

Globally, we also define the 2-manifold as embedded in a metric space,  $S$ , that is a generalized 3D Euclidean space,  $\mathfrak{R}^3$ , and for every pair of points  $x, y \in S$ , there is a distance function,<sup>39</sup>  $d$ , having properties of  $d(x, y) \begin{cases} \geq 0, & \text{if } x \neq y \\ = 0, & \text{if } x = y \end{cases}$ ,  $d(x, y) = d(y, x)$ , and  $d(x, y) \leq d(x, z) + d(z, y)$  for any  $z \in S$ .

For  $M$  as a topological space  $X$  embedded in a metric space,  $S$ , for subsets  $A \subseteq X$ , there is a function,  $d$ , for every pair of points in  $A \subseteq X$ , and every metric space is Hausdorff.<sup>39</sup> A 3D theoretical morphospace is a compact 2-manifold,<sup>39,41</sup> and as such, 3D theoretical morphospace has the same local properties as 3D Euclidean space and local coordinates and functions that are smooth.<sup>41</sup> That is, 3D theoretical morphospace as a compact 2-manifold has a differentiable structure between smooth functions on the 2-manifold.

### A.2. Theoretical morphospace and Reeb graphs

Within the compact 2-manifold that is the 3D theoretical morphospace, morphological transitions may be modeled by using a Reeb graph<sup>42</sup> as a proxy for an adaptive

walk that is finite. Reeb graphs are topological graphs of distinct points, including critical points such as maxima, minima or saddles that may occur on the boundary of a 3D theoretical morphospace.

Reeb graphs are quotient spaces of manifolds.<sup>43</sup> A quotient space is the topological gluing together of distinct boundary points of a space that are identified by an equivalence relation. That is, a quotient space is the set of equivalence relations where  $(X_1, f(X_1)) \sim (X_2, f(X_2))$  if and only if  $f(X_1) = f(X_2)$  and  $X_1$  and  $X_2$  are in the same connected space of  $f^{-1}(f(X_1))$ . For the Reeb graph of manifold  $M$  with real-valued function,  $f : M \rightarrow \mathfrak{R}$ , the quotient space,  $M \times \mathfrak{R}^n$ , is defined as  $(x_1, f(x_1)) \sim (x_2, f(x_2)) \Leftrightarrow f(x_1) = f(x_2)$ . That is, the equivalence relation of a Reeb graph is a quotient space, and it follows since the 3D theoretical morphospace that is a compact 2-manifold is a quotient space.<sup>39</sup>

### A.3. Theoretical morphospace and level sets

A contour diagram of 3D theoretical morphospace may be defined as a topologically discrete surface on which contour lines may be drawn. A contour line is the intersection of an arbitrary surface and a plane.<sup>44</sup> The plane is parallel to the base of the theoretical morphospace, and evenly spaced planes parallel to the base will intersect the theoretical morphospace forming cross-sectional spaces<sup>45</sup> or contours.

That is, to define a contour line, let  $\{(x_1, \dots, x_n) | f(x_1, \dots, x_n) = c\}$  be a level set of points for a real scalar-valued function,  $f$ . The contour line is a level curve and is a cross-section of  $f$  when  $f(x, y) = c$ . For each of  $m$ -level curves, the function  $f$ , has a different constant value. For  $c = 2$ , a level curve is produced for a function in two variables,  $f(x, y)$ . For  $c = 3$ , level surfaces are produced, and the contour plot is a function of three variables,  $f(x, y, z)$ . Level curves on the 3D theoretical morphospace are contours that are parallel to the base of the morphospace and can be defined by a height function,  $h$ . Then,  $f : M \rightarrow \mathfrak{R}$  is a function on  $M$  where each point on  $M$  is  $p = (x, y, z)$ , and from the  $z$ -coordinate,  $f$  is a height function.<sup>41</sup>

To further characterize the contours, those that are adjacent to one another can be subjected to triangulation, if connections made between contours proceeds in the same direction, and the initial nodes are designated to be proximal. A triangulated mesh of the contours will define the surface in terms of polygons.<sup>46</sup> Triangularization of the contours of the 2-manifold surface<sup>47</sup> serves as a way to interpolate the surface.<sup>31</sup> The triangulated portions of a contour are localized flat areas.<sup>48</sup>

### A.4. Theoretical morphospace and determining critical points

Reeb graphs represent contours of a surface as points,<sup>45</sup> are a generalization of contour trees,<sup>49</sup> and are the fundamental structures defining the topology of shape.<sup>49</sup> Reeb graphs are dependent on distinct boundary points for their construction. Such points are identified as critical points,  $x_0$ , since the tangent space to  $M$  at  $x_0$  is in the tangent space at  $x_0$  of the level set at  $h$ . At each point for several values on  $M$ , a Jacobian matrix of partial derivatives can be calculated, and the minimum rank

is the highest value of regular points, while singular values are critical points.<sup>27</sup> The type of critical point — maximum, minimum or saddle — is determined by eigenvalue sign in a Hessian matrix of second partial derivatives.<sup>40,41,43</sup>

For a function  $f : M \rightarrow \mathfrak{R}$  defined on the surface of the 2-manifold,  $M$ , there is a point  $p_0$ , of  $M$  that is a critical point since  $\frac{\partial f}{\partial x}(p_0) = 0$  and  $\frac{\partial f}{\partial y}(p_0) = 0$  with respect to local coordinates in the neighborhood of  $p_0$ . If a function has all non-degenerate critical points (i.e.,  $f''(p_0) \neq 0$ ), then it is a Morse function.<sup>40</sup> Morse theory can be used to construct Reeb graphs,<sup>43</sup> and is concerned with how the shape of space (manifold) controls distribution of the critical points of a function locally, and how critical points of a function affect the shape (topologically) of space globally.

However, in 3D theoretical morphospace, degenerate points (i.e.,  $f''(p_0) = 0$ ) may be very important in characterizing the space. With maxima, minima and saddles as well as using functions other than Morse, an extended Reeb graph (ERG)<sup>50,51</sup> may be constructed and is based on finding critical areas of the 3D theoretical morphospace from its triangulated contour surface, influence zones of the critical areas, and using the relations among critical areas to devise a new graph structure.<sup>43,52</sup> Taxon labels are used to find inclusive important critical areas as influence zones, and therefore critical points in the ERG. Another important point is the origin in a 3D theoretical morphospace since it is a constraining factor for the start of ERG construction.

ERGs may be applied to a 3D theoretical morphospace discrete surface  $M$ , to determine the height function  $h$ , where  $h$  is not required to be Morse.<sup>43</sup> For discrete surfaces, critical areas of the smallest size are considered, and finding a critical point is topologically equivalent to the critical area as an influence zone.<sup>48</sup> Alternatively, if  $M$  is a continuous surface with  $f : M \rightarrow \mathfrak{R}^3$ , then in two variables,  $h : M \rightarrow \mathfrak{R}$  is the height function, and  $M = \{x, y \in M | h = f(x, y)\}$  in the ERG.<sup>48</sup>

Intersections of the 2-manifold surface with planes orthogonal to the  $x - z$ ,  $y - z$  or  $x - y$  plane and the contours of the surface also define level sets of the height function  $h$ . That is, the 3D theoretical morphospace as a 2-manifold can be depicted as a contour diagram where contours (level curves) are drawn at critical points of the critical areas, and the 2-manifold of  $f$  in one variable,  $M_k = \{x \in M | h = f(x)\}$ , is a change in level sets of a smooth function  $f$ , as  $h$  varies. For an ERG as a quotient space, the height function  $h$ , is defined with respect to the contours and used with respect to the critical points in critical areas with influence zones found on the 2-manifold<sup>50</sup> in 3D Euclidean space. If the height function is Morse, then the local behavior around a critical point is defined by the cross-sectional contours.<sup>43,50,51,52</sup>

#### A.5. Connecting critical points and constructing a skeleton tree

The ERG of critical points of critical areas and their influence zones represents how changes in level sets of the 2-manifold surface are connected by splitting and merging as a contour tree. That is, for an ERG  $G$ , of  $M$  under  $h$ ,  $G_h(M) = P_h(M)$ ,  $Arc_h(M)$ , where  $P$  is a node (=critical point) set and  $Arc$  is an arc (=connection

or edge) set. For  $P_h(M) = \{P_i \in M\}$ ,  $P_i$  is a critical point of  $M$  in that  $P_i \in h(M)$ . The topological space between points and arcs does not change.<sup>41,43</sup>

Change in height from lowest to highest elevation provides the general direction of critical point connections. To connect critical points, a maximum is a peak, and sequentially adjacent critical points are maxima, if there is an increase in the height function, minima if there is a decrease, and saddles if there is a decrease then an increase. A minimum is a valley, and sequentially adjacent critical points are minima, if there is a decrease in the height function, maxima if there is an increase, and saddles if there is an increase then a decrease. A saddle is a transition between maxima and minima with respect to the height function. To connect saddle areas, all points closest to each other that occur as maxima are connected to a saddle until the first point encountered beyond this is less than the maxima connected to the saddle. At this stage, the saddle is connected to an adjacent saddle determined in the same way, and the process continues until the ERG is complete.

Let the critical points of the ERG be  $p$ , and the first critical point is  $p_1$ . Of the critical areas,  $CA$ , the first critical area is  $CA_1$ . By definition, if the area is a maximum, it is a part of the Reeb graph.<sup>43</sup> If the area is a minimum, then its influence zone is the area adjacent to it. If the nearest adjacent area is a saddle, then from the critical point and arc sets, the arc,  $Arc_{p_1-saddle_1} = (p_1, CA_1)$  is the connection between critical areas defined by their critical points. If the nearest adjacent area is not a saddle, then  $CA_2$  is determined by recursion until a maximum is found. In this way, ERG branch terminations are determined from connection to  $p_1$ . Subsequently, the next critical area is found, and the process is repeated until all branches of the ERG are formed.

Change in elevation is the representation of the hierarchical features of the critical point extraction process that are translated into the edge connections of critical points (=vertices). That is, from triangulated contours of the 2-manifold, points adjacent to one another from one contour to the next may be compared in terms of the elevation that each point represents. The higher the point, it then becomes the next connecting point in the hierarchical sequence of skeleton connectors in an ascending order.<sup>50</sup> Across a contour boundary, multiple points may be evaluated to determine whether there is an upward or downward change in direction, and this will affect the hierarchy of the skeleton connectors. The skeleton tree reflects a hierarchical relation among the points<sup>31,32</sup> where a minimum point (origin) is the root node, a maximum point is a leaf node, and branched nodes are the saddle points.<sup>53</sup>

Taxon names as labels within 3D theoretical morphospace constrain branching with respect to hierarchical meaning from point to point of the ERG. Moreover, contours of 3D theoretical morphospace have a hierarchical structure that may be nested from the base to the apex of peaks, and this structure influences the hierarchical nature of the tree.<sup>27</sup> The resultant skeleton tree is a directed acyclic graph.

### A.6. Numerical representation of skeleton trees and cladograms

Simple, less time-consuming graph matching methods include clustering algorithms in conjunction with adjacency matrices that are representations of tree decomposition.<sup>54</sup> Adjacency matrices include elements as 0 and +1 to represent no connection between adjacent nodes and connection between adjacent nodes, respectively.<sup>32</sup> The skeleton tree and cladogram are oriented graphs without parallel edges. For either of the graphs, the set of vertices,  $V$ , where there are  $n$  vertices,  $v_1, \dots, v_n$ , in the set, and  $E$  is the set of edges, an  $n \times n$  adjacency matrix of the oriented graph,  $Adj_{i,j} = [a_{i,j}]$ , is defined as  $Adj_{i,j} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \\ 0, & \text{otherwise} \end{cases}$ ,<sup>55</sup> where the  $i$ th and  $j$ th elements of  $v$  in  $V$  are in  $E$  when taxa are adjacent to each other. In the skeleton tree and cladogram, there are no self-looping or 1s on the diagonal of the adjacency matrices since each graph is directed and acyclic.

### A.7. Graph matching of tree nodes

Since proximity between tree structures is an important result of graph matching, the multivariate distance method, multidimensional scaling (MDS),<sup>18,56</sup> is ideally suited to evaluate graph-matched structures that have been decomposed into adjacency matrices. Moreover, MDS facilitates understanding the structure of the data that represent skeletonized theoretical morphospace and cladogram and interpretation of results with respect to adaptive characters of taxa in terms of environmental preferences information.

MDS is used to compare multiple eigenspaces within one ordination and evaluate the graph matching on a node-to-node basis.<sup>57</sup> MDS is used to depict points in space so that geometric distances reflect empirical relationships.<sup>58</sup> One of the best computer programs in use to perform MDS is alternating least squares scaling (ALSCAL).<sup>59,60</sup> ALSCAL can handle any type of data, including data where some of the values are missing.<sup>56</sup>

ALSCAL is an optimization program where, initially, a dissimilarity matrix is calculated from adjacency matrices of the skeleton tree and cladogram. From this, disparities are calculated so that the dissimilarities are mapped to the disparities as  $\varphi(\delta_{rs,i}^2) = \hat{d}_{rs,i}^2$ , where  $\{\hat{d}_{rs,i}^2\}$  are least squares estimates of  $\{d_{rs,i}^2\}$ .<sup>56</sup> These estimates depend on minimization of the loss function,  $S$ -stress,<sup>59</sup> which is defined as  $S$ -stress =  $\sum_r \sum_s \sum_i (d_{rs,i}^2 - \hat{d}_{rs,i}^2)^2$ , where the squared distance between pairs of taxa is calculated for all possible combinations of taxa. Minimization of  $S$ -stress means that the ordination depicts actual distances<sup>56</sup> between taxa.

### A.8. Graph matching of tree paths

Adjacency matrices are representative of node identity in trees. Branches are paths between nodes, and counting the number of paths between nodes from one taxon

to another can be used as the content of a path matrix of trees for graph matching purposes. MDS of path matrices can be used as another indication of the degree to which a skeleton tree and cladogram are matched structures. By counting the number of steps from leaf node to leaf node, paths between taxa for the skeleton tree and cladogram can be determined as the sum of the number of steps. For the skeleton tree and cladogram, taxa are numbered consecutively with regard to their placement in a Gini tree.<sup>17</sup> From this, the sums for each taxon-to-taxon set of branching connections are the elements of the unweighted path matrix, and MDS is used to determine the degree of graph matching between skeleton tree and cladogram paths. The unweighted path matrix for either the cladogram or skeleton tree is square, symmetric. Elements on the diagonal for the path between a given taxon and itself are 0s, and the form of the unweighted path matrix is

$$\begin{bmatrix} tree_{(path)11} & \cdots & tree_{(path)1n} \\ \vdots & \ddots & \vdots \\ tree_{(path)n1} & \cdots & tree_{(path)nn} \end{bmatrix} \quad (\text{A.1})$$

where *tree* represents either the cladogram or skeleton tree, and there are  $n$  elements representing taxa for each tree. For MDS, taxon names are the rows, and taxon names for the cladogram and skeleton tree are the columns. From this, the paths of taxa in the cladogram and skeleton tree can be depicted in MDS path space to determine degree of graph matching.

Graph-matched structure evaluation is also done using path matrices in a constrained MDS. In this case, weights in the matrices<sup>18</sup> indicate the relation among nodes across both tree structures as an ordering of branches from root to leaf nodes. Using facet theory,<sup>18,58,61</sup> MDS leaf node clusters serve as facets of matched taxa that are cluster constraints.<sup>62</sup>

From MDS of leaf node clusters in a facet diagram, weights are based on the closeness of leaf nodes using each taxon label only once. These weights are numbered with respect to closeness that taxa in a facet of leaf nodes have to each other and to other taxa in other facets. The highest ranking goes to within-cluster group adjacency, next highest ranking goes to between-cluster group adjacency, and lowest ranking goes to taxa in non-adjacent clusters. Rankings are used as weights and constrain the order relation<sup>18</sup> between branching patterns in skeleton tree and cladogram.

To construct weighted path submatrices, each submatrix is constructed by only the branching between leaf nodes, and these values are reported in the lower half. In the upper half of the weighted path submatrix, weights as rankings between taxa are given. Each weighted path submatrix is square, asymmetric and has integers as its elements. However, because the facet diagram may contain more than one representation of a given taxon in more than one facet, weights from taxon to taxon are not necessarily the same for each weighted submatrix. For either the cladogram

or skeleton tree, the weighted path submatrix is of the general form

$$\begin{bmatrix} tree_{11} & \cdots & tree_{(weight)1n} \\ \vdots & \ddots & \vdots \\ tree_{(path)n1} & \cdots & tree_{nn} \end{bmatrix} \quad (\text{A.2})$$

where there are  $n$  elements for the *path* lower matrix or *weight* upper matrix for the cladogram or skeleton tree or a combination of both. For paths and weights from a given taxon to itself, the diagonal elements of each submatrix are 0s.

Four weighted path submatrices result in cladogram paths to cladogram weights, skeleton tree paths to cladogram weights, cladogram paths to skeleton tree weights, and skeleton tree paths to skeleton tree weights taxa comparisons. Each of the submatrices is arranged as one of four blocks in a single weighted path matrix and is partitioned as

$$\left( \begin{array}{c} \left[ \begin{array}{ccc} clad_{11} & \cdots & clad_{(weight)1q} \\ \vdots & \ddots & \vdots \\ clad_{(path)j1} & \cdots & clad_{jq} \end{array} \right] \\ \left[ \begin{array}{ccc} clad \leftrightarrow skel_{11} & \cdots & clad_{(weight)1s} \\ \vdots & \ddots & \vdots \\ skel_{(path)l1} & \cdots & clad \leftrightarrow skel_{ls} \end{array} \right] \end{array} \right) \left( \begin{array}{c} \left[ \begin{array}{ccc} skel \leftrightarrow clad_{11} & \cdots & skel_{(weight)1r} \\ \vdots & \ddots & \vdots \\ clad_{(path)k1} & \cdots & skel \leftrightarrow clad_{kr} \end{array} \right] \\ \left[ \begin{array}{ccc} skel_{11} & \cdots & skel_{(weight)1t} \\ \vdots & \ddots & \vdots \\ skel_{(path)m1} & \cdots & skel_{mt} \end{array} \right] \end{array} \right) \quad (\text{A.3})$$

where *clad* and *skel* elements are the  $q$ th,  $r$ th,  $s$ th, and  $t$ th entries for *weights* as well as the  $j$ th,  $k$ th,  $l$ th, and  $m$ th entries for *paths* for the cladogram or skelton tree or a combination of both.

The four-block square, asymmetric matrix is analyzed using MDS, and the constrained results are ordinated in path space. Published data on environmental and habitat preferences for each taxon are mapped to the weighted MDS ordination in path space, and taxon adaptation or adaptive ability is inferred.