Applications in Plant Sciences

APPLICATION ARTICLE

Shape outline extraction software (DiaOutline) for elliptic Fourier analysis application in morphometric studies

Asher Wishkerman[1] and Paul B. Hamilton[2,3]

Manuscript received 10 July 2018; revision accepted 21 October 2018.

[1] School of Marine Sciences, Ruppin Academic Center, Michmoret, Israel

[2] Research and Collections, Canadian Museum of Nature, Ottawa, Ontario K1P 6P4, Canada

[3] Author for correspondence: phamilton@nature.ca

Citation: Wishkerman, A., and P. B. Hamilton. 2018. Shape outline extraction software (DiaOutline) for elliptic Fourier analysis application in morphometric studies. *Applications in Plant Sciences* 6(12): e1204.

doi:10.1002/aps3.1204

PREMISE OF THE STUDY: Studies of plant cell and organ outline using shape analysis for taxonomic and morphological research have increased in the past decade. However, there are a limited number of available modern, intuitive, and easy software tools to conduct this work.

METHODS: We developed a tool for shape outline extraction using MATLAB accompanied with R scripts to perform elliptic Fourier analysis. To demonstrate the shape tool, we applied the software and scripts for genera and species shape determinations of diatom (single cell) species with x-, y-, and x- + y-shape symmetries.

RESULTS: Using the shape analysis tool, we were able to identify and distinguish different diatom taxa based on forms representing size diminutions associated with population changes.

DISCUSSION: Independent of symmetry, species were successfully distinguished using supervised and unsupervised analyses. We hope that these shape analysis tools will be used to add another metric to plant science studies.

KEY WORDS diatom; elliptic Fourier analysis; linear discriminant analysis (LDA); principal component analysis (PCA); shape; species identifications.

Shape tools for visual identification and classification based on qualitative and quantitative criteria have interested scientists (positively and negatively) for decades (Jensen, 2003). With advancements in image processing and data analysis (both in cost and speed), shape analysis tools have become more accessible in many fields including biology, geography, medicine, and archaeology. Studies of higher plant features such as leaves, petals, and seeds have been extensively performed using morphometric techniques; however, the cryptogams and other single-celled organisms have received less exposure to shape analysis (Neustupa, 2013; Pappas et al., 2014; Stanton and Reeb, 2016). For example, Stoermer and Ladewski (1982) used Legendre polynomial shape descriptors to examine type and modern populations of the single-celled *Gomphonema herculeana*; however, the approach has not been extensively adopted in diatom taxonomic and population studies. Overall, Fourier transformations (harmonics) are useful in modeling shape for any diatom outline without the requirement of identifying specific morphological features. See Pappas et al. (2014) for a complete review of morphometrics in diatom research.

Shape analysis tools offer precise and accurate descriptions, enable rigorous statistical analysis, and allow visualization interpretation and communication of the results. The ability to link plant organ shapes, architectures, and dynamic changes in phenotype expression with underlying environmental, genetic, and molecular drivers can produce interesting results for complex systems (Chitwood and Topp, 2015). Furthermore, morphometric techniques can greatly assist the limited number of well-trained taxonomists in linking phylogenetic and molecular studies with classical species identification (Tomaszewski and Górzkowska, 2016). Shape morphology tools are highly diversified, and in the past decade there has been an increasing interest in the use of modern geometric morphometrics (GM). Several approaches in GM have developed, such as landmarks (LMs; Potapova and Hamilton, 2007; Ros et al., 2014), semi-landmarks (Vieira et al., 2014; Ros et al., 2014; Glennon and Cron, 2015), or the use of elliptic Fourier analysis (EFA; Viscosi and Fortini, 2011; Adebowale et al., 2012; Kloster et al., 2014).

GM analysis using LMs is a popular and common method in which a landmark is a point of correspondence on each object that matches a shape position between and within a population (Zelditch et al., 2012; Brombin and Salmaso, 2013). The LMs coordinates mark morphological and anatomically definable points. Superimposing landmark configurations to a common coordinate system is used to generate a set of shape variables, also known as generalized Procrustes analysis (GPA; Zelditch et al., 2012). After superimposition, the aligned Procrustes shape coordinates describe the location of each specimen in a curved space related to Kendall's shape space. These are typically projected orthogonally into a linear tangent space yielding Kendall's tangent space coordinates on

which multivariate analyses of shape can be performed (Elewa, 2012). Biological hypotheses can then be examined using statistical methods like multivariate analysis of variance (MANOVA), partial least squares (PLS), principal component analysis (PCA), and linear discriminant analysis (LDA) (e.g., Stoermer and Ladewski, 1982; Silva et al., 2012; Adams and Otárola-Castillo, 2013; Boglino et al., 2013). Although GM analysis performs extremely well and is broadly used, it has limitations where no clear observable and repetitive LMs exist. One way to solve the problem is to apply a series of LMs on curves or perimeters that are known as sliding landmarks (SLMs). SLMs are defined in relation to other landmarks (e.g., arc between LMs) and, although they lack anatomical identifiers, SLMs can improve study results (Brombin and Salmaso, 2013).

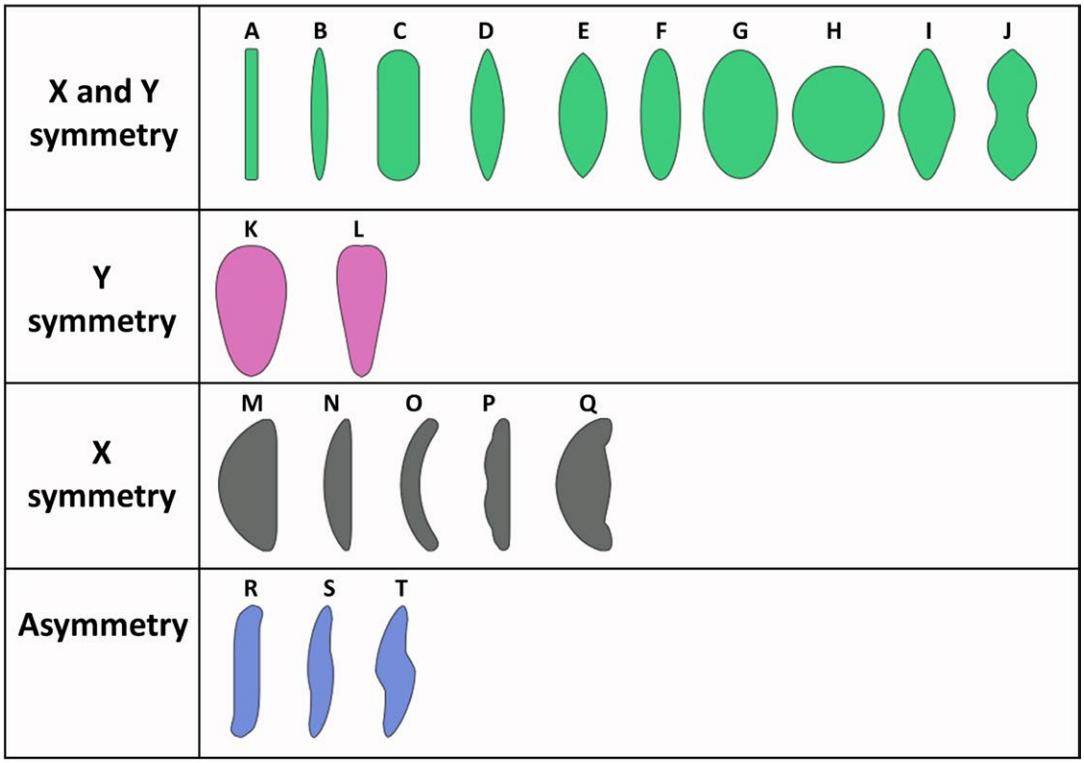
The LM approach does not represent the true shape outline, and when it is necessary EFA is often employed (e.g., Chitwood and Otoni, 2017). The use of EFA on objects allows a comprehensive depiction and quantification of shapes (outlines). Furthermore, this approach does not require the biological knowledge (recognition of discrete morphological characters) that is sometimes needed to identify and mark LMs (Carlo et al., 2011). EFA was first described by Giardina and Kuhl (1977) and Kuhl and Giardina (1982). Their development of a general method for fitting separately the x and y coordinates of an outline projected on a plane was an improvement compared to other Fourier-based approaches as no equally spaced points are needed. The outline does not require the prior definition of a biologically homologous centroid or geometric center, and the elliptic Fourier coefficients are independent of outline position on the digitization grid (Rohlf and Archie, 1984; Crampton, 1995). EFA decomposes the outline of an object into a series of closed curves (harmonics) that are generated by a known mathematical function (Haines and Crampton, 2000). Each harmonic is described by four Fourier coefficients (elliptic Fourier descriptors [EFDs]); two each for the x- and y-axes, generating a total of four n coefficients labeled a_n, b_n, c_n, and d_n, where n is the number of harmonics. Further details can be found in Claude (2008). Programs, source code files, and scripts have been developed to extract outlines and/or perform two-dimensional (2D) EFA analyses, for example, PAST, SHAPE, VisioBioShapeR, ShapeR, SHERPA, and Momocs (Hammer et al., 2001; Iwata and Ukai, 2002; Kloster et al., 2014; Bonhomme et al., 2014; Libungan and Pálsson, 2015; Stela and Monleón-Getino, 2016). Additional image cleaning and shape recognition tools are also available in free programs like ImageJ and Fiji (https://fiji.sc/). Some of the above-mentioned applications like PAST and several R packages lack the capability to perform image analysis operations. The software SHAPE has not been updated since 2006 and lacks compatibility with new operating systems. SHERPA is a detailed and advanced extraction tool for diatom studies, with a complex user interface that requires an experienced user, understanding of diatom shapes, and advanced and specialized slide scanning equipment (Kloster et al., 2014, 2017). DiaOutline was developed in order to provide a modern, intuitive, fast, and easy-to-use interface tool for general applications in order to extract outlines (e.g., leaves, seeds, bones). DiaOutline ensures compatibility with modern operating systems, supports various image file formats, and provides an outline x-y coordinate vector metric for each shape that can be further analyzed. We used MATLAB to construct the DiaOutline user interface, apply the necessary image analysis procedures (e.g., image threshold, extract shape/biological entity coordinates), and handle data (e.g., save coordinates files). R was used to perform the EFA analysis and visualization (extended information can be found in the Methods section).

EFA is a powerful tool for outline morphometric studies, but some drawbacks are described in Haines and Crampton (2000). The digitized outlines may provide "pixel noise" and smoothing the raw data may be performed. The smoothness level must consider parameters of varying scales, like the quality of the image, equipment used, and object complexity. It should be noted that there are no rules regarding smoothing level. EFA yields a relatively large number of Fourier coefficients (that are not computationally independent of each other and are, in part, redundant). This redundancy should be considered when choosing the number of coefficients. EFA increasingly downweights higher-order harmonics, which may result in reducing or suppressing the discriminatory power of outline details. The starting point of the outline trace can also influence the results, and therefore a normalization should be performed. DiaOutline solves this problem as long as the orientations of the objects are similar because the tracing is performed from left to right. The potential problems discussed above can be solved by several programs after outline extraction, which include HANGLE, HMATCH, and HCURVE (Haines and Crampton, 2000). In addition, the eFourier argument functions (smooth.it and norm) in the Momocs packages also deal with some of these limitations (Bonhomme et al., 2014).

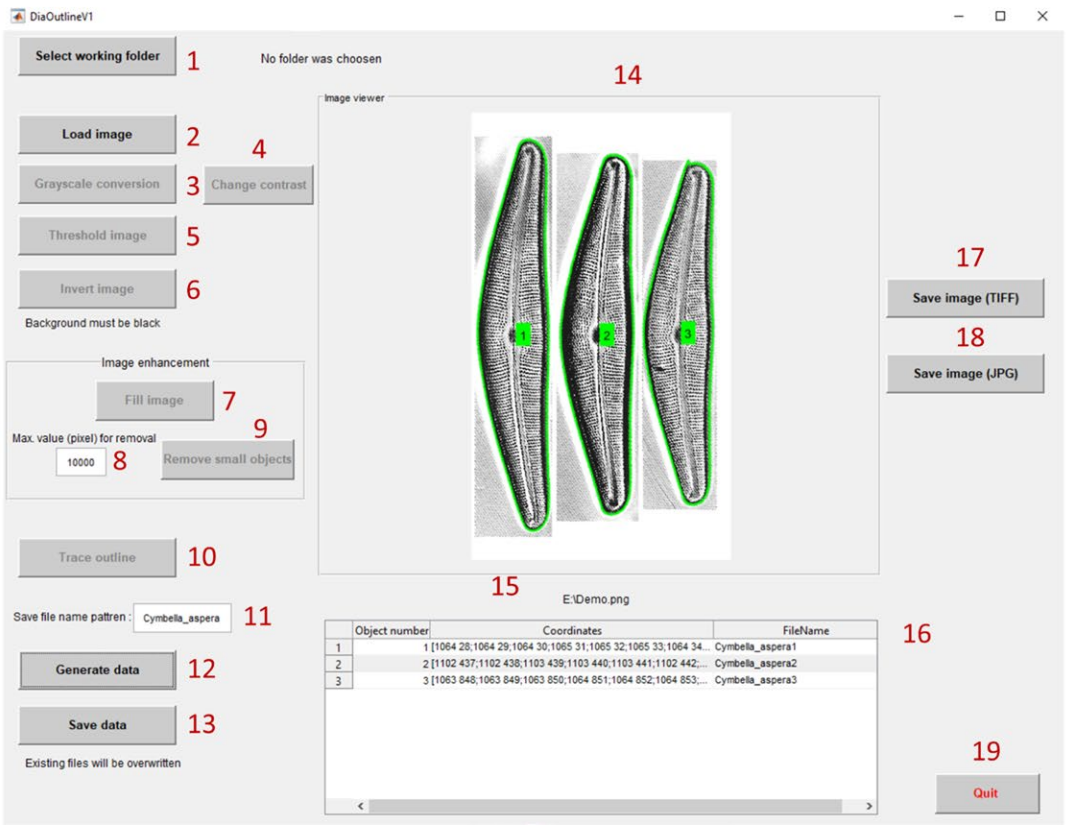
Light microscopy images of single cells (e.g., microalgae) present limited options for shape variations, both in 2D and 3D space. Spherical shapes (in two dimensions round), along with rods and spirals, represent the most biologically efficient living forms for cells and are the most common shapes observed in protists. Thus, simple cell differences in shape are hard to distinguish in classical taxonomy when phenotypic and genotypic expression within a species are also considered. However, species with single cells of distinct definable shapes are available for morphometric shape analysis. Diatoms have evolved to control their shape through the use of a silica shell, which is composed of two components (epitheca and hypotheca). These shells (valves) are held together by extracellular polymeric substances (EPS) and silicate bands (copulae/girdle bands). The primary morphological parameters (valve length, valve width, and stria density) can be used in identifications, but may not be sufficient for species identifications (e.g., Blanco et al., 2017). The shape of the diatom typically forms a definable face and a rectangular-like shape in side view. Therefore, diatoms as single cells represent good candidates for 2D shape analysis based on the valve's primary face. Surface morphology is often compared and considered an important metric in diatom species identifications (Round et al., 1990). Perfect symmetry

FIGURE 1. Diatom valve shapes and shape extraction graphical user interface (GUI). Part A: (A) linear, (B) linear-lanceolate, (C) linear-oblong, (D) linear-rhombic, (E) lanceolate, (F) oblong, (G) elliptical, (H) circular, (I) rhombic, (J) panduriform, (K) ovate, (L) clavate, (M) semi-circular, dorsiventral, (N) sigmoid-rhombic, (O) lunate-arcuate, dorsiventral, (P) semi-lanceolate-undulate, dorsiventral, (Q) semi-circular with tumid apices, (R) sigmoid-cylindrical, (S) sigmoid, lanceolate, (T) sigmoid-rhombic (based on John, 2015). Part B: eFourier shape extraction software DiaOutline GUI. Program workflow buttons and elements from 1–19, including image extraction (3–9), trace outline (10), and outline data generation (12). The images in Fig. 1B are reprinted with permission from Koeltz Botanical Books.

A



B



(half shape reflection) along the *x*- and *y*-axes in diatoms (isovalvar) typically make identifications between species difficult due to the number of shapes available (Fig. 1A). Other shapes with asymmetry in the *x*-, *y*-, or both the *x*- and *y*-planes give more shape definition, but are still limiting when subjectively evaluating 2D shape space. Therefore, the use of shape tools can enhance the differentiation of simple forms (with the exception of a sphere), representing the basic metric for a single-celled species identification.

The objective of this paper is (1) to provide easy, straightforward, and user-friendly software to extract 2D shape outline coordinates (DiaOutline), and (2) to demonstrate the utility of shape analysis using single cells with limited shape options. The software can be used for various protist shapes (microalgae) as well as for more complex morphologies like leaves and petals. Complementary R scripts are presented to evaluate EFA results using several multivariate analysis techniques. The DiaOutline software and associated R script were employed to examine diatom genera and species with different functional valve shapes; the software and related documentation are available for free download at GitHub (https://github.com/wishkerman/DiaOutline). Shape analysis has the utility to address and investigate a wide range of biological, evolutionary, taxonomic, and ecological questions, ranging from single-celled organisms to complex multicellular shapes.

METHODS

Biological material/samples

In total, 331 diatom specimen outlines were extracted from 23 species using DiaOutline (Table 1; Appendices S1–S3). Light microscopy images of diatoms were selected to evaluate the use of shape analysis in single-celled 2D biological shapes. The common use of the light microscope for taxa identifications in diatoms, coupled with a relative optical resolution of approximately 1 μm, was the reason for selecting images from this mode of viewing. Light microscopy images, depending on the optics used, can sometimes have a poor resolution of edges and artifacts, which was also a reason for using light microscopy images to challenge the utility of shape analysis (see Haines and Crampton [2000] for limitations). Taxa and genera were selected to evaluate EFA based on simple variations in reflectional *x*- and *y*-axis shapes (Fig. 1A). Diatom specimens from the genus *Luticola* were selected to represent the nearly perfect form (symmetric), while specimens from the genera *Gomphonema* (*y*-symmetric) and *Cymbella* (*x*-symmetric) were used to evaluate symmetric shape differences along the *x*-axis and *y*-axis, respectively. Finally, specimens from the genus *Gyrosigma* (reflectional asymmetric) were selected to represent complete asymmetry along the *x*- and *y*-axes. Images from six species within the genus *Luticola* representing a size diminution series for each species were taken from Levkov et al. (2013). Likewise, published light microscope images for five *Cymbella* (Krammer, 2002) and eight *Gomphonema* (Levkov et al., 2016) species were selected for study, along with images of *Gyrosigma* taxa accrued by the authors from studies across North America. The taxa selected for study included both distinct and subtle shape differences across a population diminution series. Methodologies for sample preparation and image processing follow a standard protocol of sample oxidation to remove organic matter followed by the removal of oxidant and mounting portions of the cleaned samples in Naphrax (Brunel Microscopes Ltd., Chippenham, United Kingdom) with a refractive index of 1.65. Differential interference contrast optics and bright-field optics were used to image the specimens for this study.

TABLE 1. Taxa used to evaluate four basic reflective shape groups (also see Fig. 1A).

Taxon	Authority	Shape group
Cymbella aspera	(Ehrenb.) H. Perag.	*x*-symmetry
Cymbella cymbiformis	C. Agardh	*x*-symmetry
Cymbella excisa	Kützing	*x*-symmetry
Cymbella neogena	(Grunow) Krammer	*x*-symmetry
Cymbella parva	(W. Sm.) Kirchner	*x*-symmetry
Gomphonema acuminatum	Ehrenb.	*y*-symmetry
Gomphonema brebissonii	Kützing	*y*-symmetry
Gomphonema gautieriforme	Mitić-Kopanja Wetzel, Ector & Levkov	*y*-symmetry
Gomphonema metzeltinii	Levkov, Mitić-Kopanja & E. Reichardt	*y*-symmetry
Gomphonema micropus	Kützing	*y*-symmetry
Gomphonema naviculoides	W. Sm.	*y*-symmetry
Gomphonema parvulum	(Kützing) Kützing	*y*-symmetry
Gomphonema truncatum	Ehrenb.	*y*-symmetry
Gyrosigma acuminatum	(Kützing) Rabh.	Asymmetry
Gyrosigma attenuatum	(Kützing) Rabh.	Asymmetry
Gyrosigma obscurum	(W. Sm.) Griff. & Henfr.	Asymmetry
Gyrosigma spenceri	(Quek.) Griff. & Henfr.	Asymmetry
Luticola crozetensis	Van de Vijver, Kopalová, Zidarova & Levkov	*x*- and *y*-symmetry
Luticola groeppertiana	(Bleisch) D. G. Mann	*x*- and *y*-symmetry
Luticola katkae	Van de Vijver & Zidarova	*x*- and *y*-symmetry
Luticola murrayi	(West & G. S. West) D. G. Mann	*x*- and *y*-symmetry
Luticola saprophila	Levkov, Metzeltin & A. Pavlov	*x*- and *y*-symmetry
Luticola yellowstonensis	Levkov, Metzeltin & A. Pavlov	*x*- and *y*-symmetry

Software interface

The DiaOutline software was written in MATLAB 2017b and requires installation of MATLAB Runtime (instructions and more details are available in the Readme.txt); its goal is to provide a single, simple, efficient, and integrated environment for shape outline extractions. The software graphical user interface (GUI) allows data collection and analysis workflow in a linear sequence (operate highlighted button/element selection in order from 1–19; Fig. 1B). A selection of default folder is available in case of multiple images stored in a specific location (Button 1). Several image formats are supported (e.g., JPEG, GIF, PNG, TIFF, and BMP) and can be uploaded to the image viewer panel. It is also possible to load a single image or an image with multiple shapes (Button 2). Grayscale conversion of the image is required (Button 3) and, when necessary, it is possible to

change the contrast (Button 4) in order to get better threshold results (Button 5) based on Otsu's method (Otsu, 1979; Kloster et al., 2014). Microscope images often contain artifacts on or around the specimen that make identification of shape outline problematic. If shape recovery is poor (trace outline, Button 10), correct problems (using an outside graphics tool, e.g., ImageJ or Fiji) and reload the corrected image. Image inversion is possible in order to achieve the mandatory black background (Button 6). Several image enhancements are included (Buttons 7–9) and are used to fill the object in order to get the complete shape and for the removal of small objects/artifacts. Tracing the outline is achieved by Button 10; the trace always starts from the left side of the image, and each object is assigned a number for easy future identification. It is possible to use a defined sequence of characters as a code for each object name (Button 11); DiaOutline will automatically enumerate objects in relation to the shape number as seen in the image viewer (Element 14) as well as the selected file (Element 15). Object numbers, coordinates, and file names will be added to the table (Element 16) upon selection of "Generate Data" (Button 12). Each object coordinates will be saved in a separate file based on the table's given file name (Button 13). The output image in the viewer can be saved (in any step) using the "Save Current Image (JPG)" button as a low-resolution JPG file, or using the "Save Current Image (TIFF)" button as a 500-dpi tiff file (Buttons 17 and 18, respectively). Termination and exiting the application can be done by selecting the Quit button (Button 19).

Statistical analyses

All statistical analyses were performed using R (version 3.3.1; http://www.R-project.org/) with the following packages installed: MASS, ggplot2, GGally, doBy, data.table, plyr, grid, gridExtra, and Momocs (Bonhomme et al., 2014). Multivariate statistical techniques can be applied; these included PCA and LDA. Both PCA and LDA are linear transformation techniques, and whereas LDA is supervised, PCA is unsupervised. The imported outline files can be further analyzed using R to perform other statistical tests (e.g., Manly and Navarro Alberto, 2016). MANOVA analysis of shapes was also used to evaluate the significance of the determined shape groups. The R scripts for all the analyses are available for free download at GitHub (https://github.com/wishkerman/DiaOutline).

RESULTS

Four reflective shapes represented by the genera *Luticola*, *Cymbella*, *Gomphonema*, and *Gyrosigma* were differentiated using both LDA (Fig. 2A, Table 1) and PCA analyses (Appendix S2A; MANOVA $P < 0.0001$). The combined explained variations across the first two components were 85.1% and 76.3% for LDA and PCA, respectively. In LDA, shape distinction over the first component (starting at the left upper quadrant) was from symmetric to *x*-symmetric, then to *y*-symmetric, and finally to non-symmetric biplot space. The first two PCA components distinguished shape groups, ranging from symmetric (upper left quadrant) to *x*-symmetry, then *y*-symmetry, and finally to asymmetry (lower right quadrant). The PCA biplot compared to the LDA plot showed a gradient along axis PC1 with less variability along PC2 for *Gomphonema* and *Gyrosigma*. The unsupervised approach of PCA did not separate the groups (genera) as well as the supervised LDA approach, although both analyses were significant in distinguishing genera.

Shape comparisons within genera and species provided clear differentiation among taxa with distinct shapes, but were less differentiated for taxa with similar shapes (Fig. 3A–D, Appendix S3A–D). Six *Gomphonema* taxa were distinguished in the LDA analysis, explaining 90.4% of the variance across two axes (Fig. 3A). Two taxa (*G. brebissonii*, *G. truncatum*) with similar crown-like apex morphologies over the diminution size range were not separated (MANOVA $P = 0.193$). In less distinguished morphologies with cuneate apices, one specimen (Appendix S1: *Gomphonema micropus*, specimen 30) of *G. micropus* was clearly not distinguished from *G. parvulum*. PCA analysis distinguished six taxa, with clear examples of outliers for four of the taxa (Appendix S3A). The combined explained variance across axes 1 and 2 was 86.5%. *Gomphonema parvulum* and *G. micropus* alone were not separated, whereas *G. truncatum* and *G. brebissonii* showed overlap. *Gomphonema gautieriforme* (Appendix S1: specimen 14) and *G. brebissonii* (Appendix S1: specimen 2) each had one specimen clearly separated from the population, whereas *G. acuminatum* (Appendix S1: specimens 1, 3) and *G. truncatum* (Appendix S1: specimens 1, 11) had two.

Taxa within the *x*- and *y*-symmetric genus *Luticola* were also distinguished using LDA. *Luticola murrayi*, *L. katkae*, and *L. yellowstonensis* were identified, whereas *L. groepertiana*, *L. saprophila*, and *L. crozetensis* could not be distinguished (Fig. 3B). The explained variance across two axes was 91%. PCA results were similar to LDA, with more variations within taxa (Appendix S3B). The largest shape variances in the PCA results were within *L. murrayi* and *L. yellowstonensis* (Appendix S3B).

The *y*-asymmetric *Cymbella* was identified in LDA with three easily separated taxa: *C. excisa*, *C. cymbiformis*, and *C. parva* (Fig. 3C). Two taxa, *C. aspera* and *C. neoacuta*, were not separated. The combined explained variance for the biplot was 90.6%. PCA results showed one clear taxonomic separation and four taxa with some level of similarity. *Cymbella aspera* and *C. neoacuta* were similar, whereas *C. cymbiformi*s and *C. parva* were not separated (Appendix S3C). The largest variations in shape were noted for *C. parva* and *C. aspera* (Appendix S3C).

The LDA results for the *x*- and *y*-asymmetric *Gyrosigma* showed three clear shape forms: one each for the taxa *G. acuminatum* and *G. obscurum*, and a third for two overlapping species (*G. spenceri*, *G. attenuatum*) (Fig. 3D). The two axes combined explained variance was 97.95%. The PCA results were similar to the LDA results (Appendix S3D). The larger shape variances were observed for *G. obscurum* (Appendix S3D).

When multiple genera with different shape forms were compared using LDA, the *x*- and *y*-symmetric forms along with the *x*- and *y*-asymmetric taxa showed the clearest separation in biplot space (Fig. 2B). The *x*-asymmetric *Gomphonema* and *y*-asymmetric *Cymbella* taxa displayed extensive overlap, with the *Cymbella* taxa more central in the biplot. The PCA results were similar to the LDA results. However, extensive overlap among the genera *Gomphonema*, *Cymbella*, and *Luticola* was evident (Appendix S2B). Large shape form variances were observed for *Cymbella* and *Luticola*.

DISCUSSION

Microscopic form (shape) of single cells is limited by geometric constraints and the biological directive to keep cells functionally simple (e.g., spheres, rods, spirals). In three-dimensional space, spheres (when flattened round) are a dominant cell form

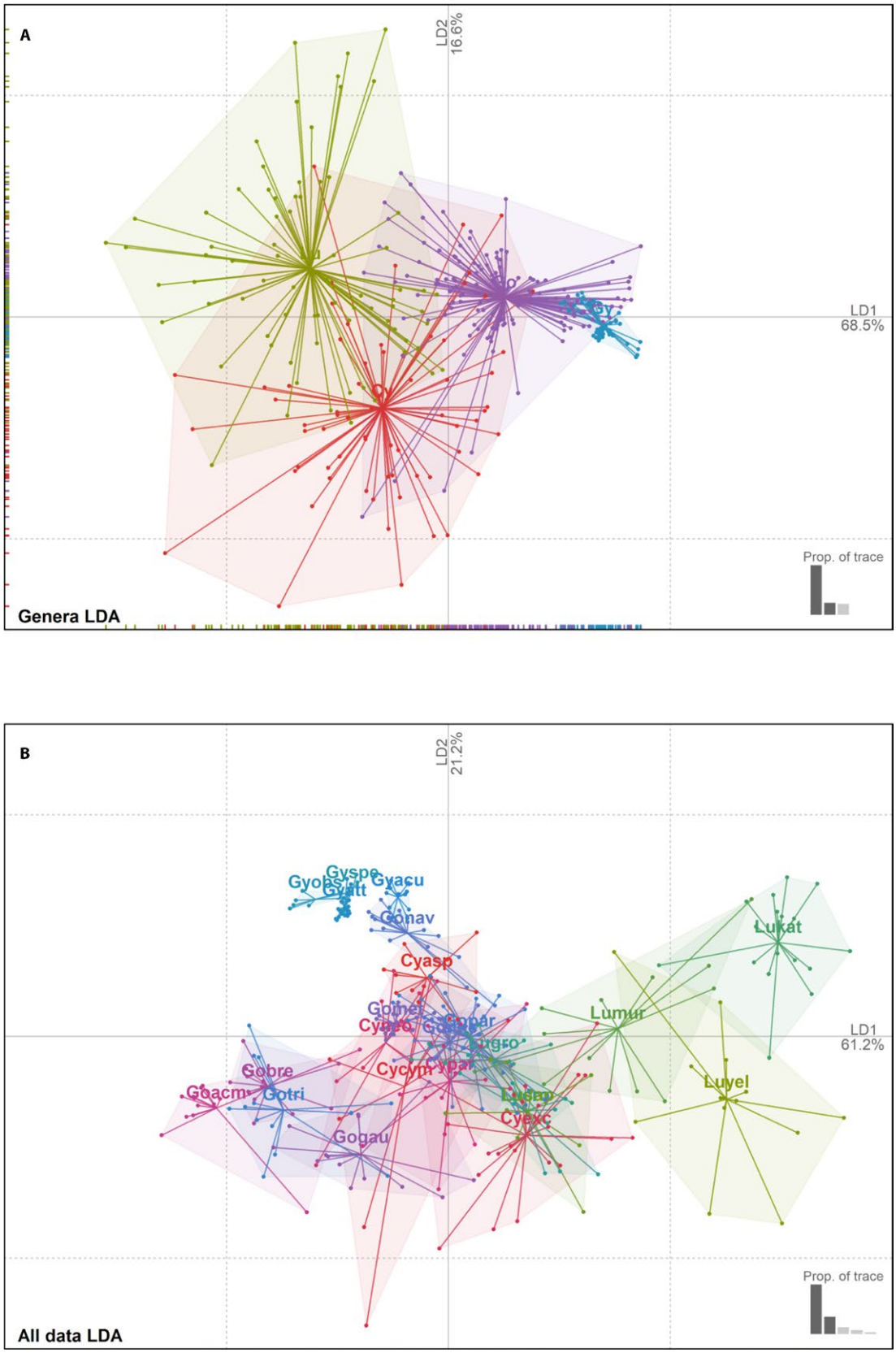


FIGURE 2. Linear discriminant analysis (LDA) plots distinguishing the genera *Cymbella* (red), *Gomphonema* (purple), *Gyrosigma* (turquoise), and *Luticola* (lime green) (A), and an LDA plot of all examined species (B).

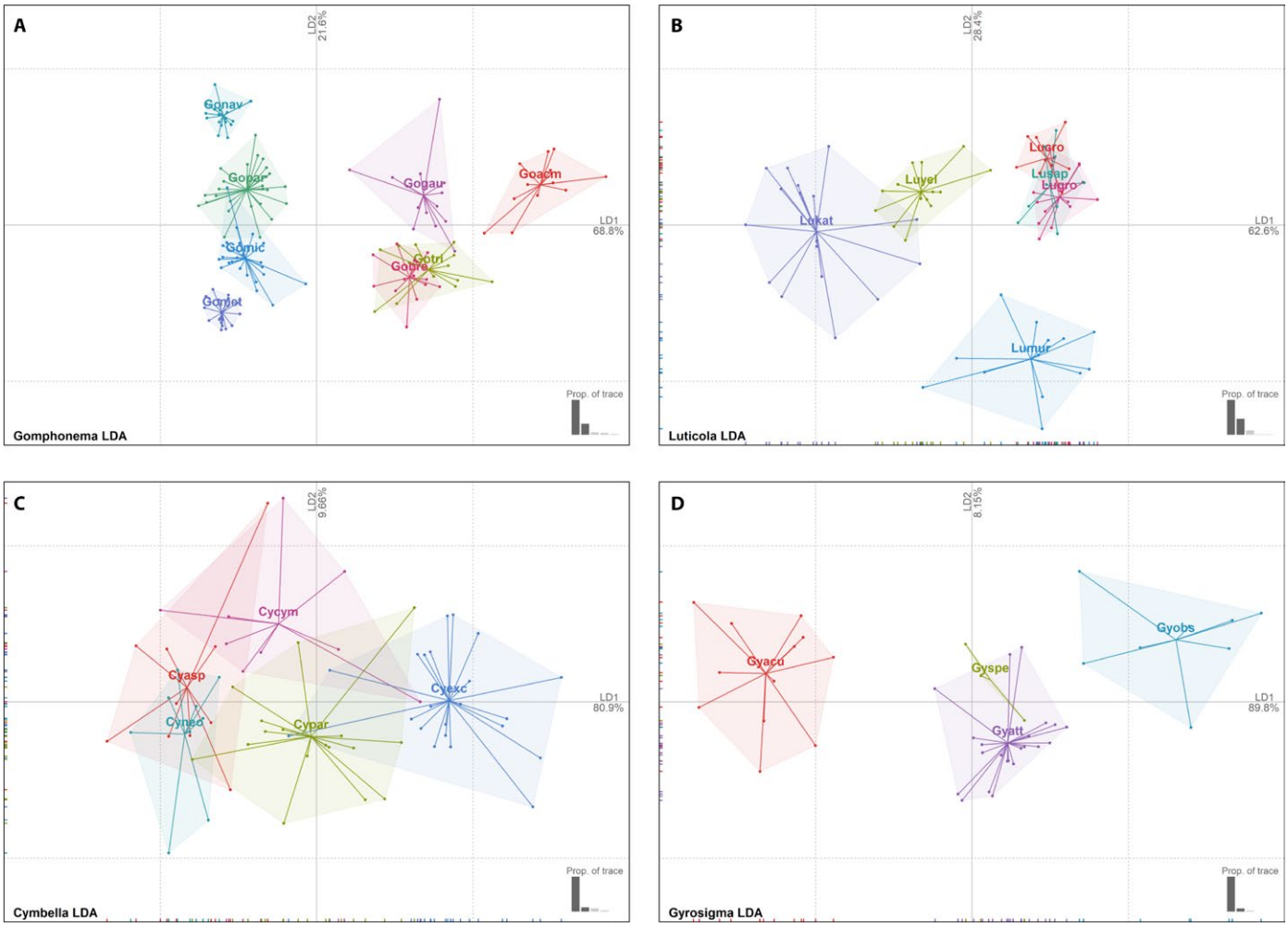


FIGURE 3. Linear discriminant analysis (LDA) plots representing four genera. (A) *Gomphonema acuminatum* (Goacm, *n* = 13), *G. brebissonii* (Gobre, *n* = 14), *G. gautieriforme* (Gogau, *n* = 14), *G. metzeltinii* (Gomet, *n* = 20), *G. micropus* (Gomic, *n* = 24), *G. naviculoides* (Gonav, *n* = 18), *G. parvulum* (Gopar, *n* = 27), and *G. truncatum* (Gotri, *n* = 13). (B) *Luticola crozetensis* (Lucro, *n* = 10), *L. groeppertiana* (Lugro, *n* = 13), *L. katkae* (Lukat, *n* = 17), *L. murrayi* (Lumur, *n* = 13), *L. saprophila* (Lusap, *n* = 9), and *L. yellowstonensis* (Luyel, *n* = 14). (C) *Cymbella aspera* (Cyaps, *n* = 10), *C. cymbiformis* (Cycym, *n* = 9), *C. excisa* (Cyexc, *n* = 22), *C. neogena* (Cyneo, *n* = 9), and *C. parva* (Cypar, *n* = 17). (D) *Gyrosigma acuminatum* (Gyacu, *n* = 12), *G. attenuatum* (Gyatt, *n* = 22), *G. obscurum* (Gyobs, *n* = 8), and *G. spenceri* (Gyspe, *n* = 3).

and are not applicable for EFA analysis. When single-celled organisms have non-spherical shapes, shape analysis tools can be used for taxonomic and physiological evaluation. For centuries, researchers have subjectively used morphological shape to distinguish single-celled life forms. Although statistical tools for the evaluation of cell outline have been available for more than 40 years, their use has been limited (Pappas et al., 2014). Current taxonomic publications in diatom research typically do not use mathematical evaluations for species shape. Subjective evaluations of diatom valve shape typically include Type I (splitting the same species into two species) and Type II (merging different species into one) errors. In diatoms, the problem of shape recognition is further complicated by the natural diminution of shapes within a species across a natural vegetative reproduction cycle. As the population ages, individual diatom valves may become less distinctive in shape and more problematic in identification. As individual diatoms in a growing population get smaller, shapes become less defined, more symmetric, and often become more similar across species (e.g., Theriot and Ladewski, 1986). In this study, the biplot convergence of the genera represents the diminution (size and shape reduction) of individuals in the populations. For example, *Cymbella excisa* (specimen 9), an older specimen in the life cycle, shows some symmetry and a similarity to other taxa (Appendix S3C).

Species concepts through time have changed with advancing analytical and genetic sequencing technologies. The limitations of light microscopy observations have in the past influenced the metrics used for identifications. Poor understanding of diatom valve structure in past research was caused in part by the inability to see fine structural details in the valve. Therefore, subjective evaluations of valve shape were a major identification variable, and Type II errors were historically more prevalent. Recent fine-grain taxonomy practices using light and scanning electron microscopy have increased the number of available morphometric variables, including more resolution of the valve shape. When subjectively identifying taxa using shape, the predominance of Type I errors is now more evident.

At the generic level, DiaOutline extracted defined shape symmetries that were easily identified under LDA analysis. The association between symmetric and asymmetric shapes illustrates a biplot convergence from symmetry to asymmetry. With diatoms, total asymmetry converged with *y*-symmetry, representing the importance of the longest measurement in the valve shape associations. Shape analysis using DiaOutline extractions was able to distinguish shape asymmetries in a predictable model, illustrating the utility in this analysis for biological phylogenetic associations at the genus level using shape. Studies on other genera lacking parallelogram or elliptic shapes, like *Tabellaria* (Mou and Stoermer, 1992) and *Asterionella* (Pappas et al., 2014), will also be well characterized with Fourier shape analysis.

Species and population distinctions within the four shape groups were identified by EFA, which is in line with other studies using diatoms (e.g., Kloster et al., 2017) and different biological organisms (e.g., Rohlf and Archie, 1984; Tracey et al., 2006; González-Wevar et al., 2011). Distinct shapes were separated in shape space, irrespective of asymmetric or symmetric valve forms (Fig. 3). The complexity of the shape did not necessarily improve shape recognition. *Gomphonema brebissonii* and *G. truncatum* showed similarities over the diminution size range (MANOVA, $P = 0.193$). In contrast, the simple shape forms of *Gyrosigma accuminatum* and *Gyrosigma* cf. *spenceri* were better separated (MANOVA, $P < 0.003$), illustrating that simple shape differences can be separated (Appendix S3). Similar shapes clustered together (e.g., *Luticola crozetensis*, *L. saprophila*), and degree of overlap was easily evaluated. In this study, we selected some taxa with almost identical shape forms (e.g., *Gomphonema micropus* and *G. parvulum*) in order to examine how well shape was able to distinguish species. Taxa of similar shape forms were generally not distinguished, which is useful for taxonomists in evaluating the significance (or not) of shape differences. Shape analysis is not useful when studying cryptic species.

Shape analysis can identify clusters of diatom taxa with similar form (e.g., Stoermer et al., 1986). The groups could be a biological cluster of shape-related taxa, represent the variable forms of phenotypic expression, or even reveal phylogenetic relationships. Woodard and Neustupa (2016) used geometric morphometrics with valves from *Luticola polickovae* to demonstrate that total asymmetry and fluctuating asymmetry were stable within a strain, whereas changes in directional asymmetry were more evident. Even seasonal changes in unicellular shape can be studied (Steinman and Ladewski, 1987). Taxa separation in shape space can represent meaningful separations in sub-taxa groups and phylogeny. This utility in shape evaluation can be illustrated with the examination of plant leaves from a single individual (Appendix S2C). The serrate leaf edges are extracted for each leaf shape. In the examination of leaf shape, differences in phenotypic expression of shape can be evaluated (Chitwood and Otoni, 2017). Shape analysis is able to access simple and complex shapes with an unbiased scoring of complexity. Furthermore, additional geometric morphometrics using landmarks on leaf and insect wing venations can allow for neural network analyses to further distinguish shape forms (Lorenz et al., 2015; Badi et al., 2017). In combination with other modalities like genetic analyses, shape (or other tools like curvature [Wishkerman and Hamilton, 2017]) and geometric morphometric analyses (Potapova and Hamilton, 2007; Chitwood et al., 2014) can improve taxonomic identifications by adding another statistical layer to microbial assessments.

The distinction of 2D cell shape is limited by the 3D structure of the cell or by the quality of tools used for examination (Haines and Crampton, 2000). When transferred into 2D space, a 3D curving shape will not always give a clear shape edge. For example, diatoms with curving valve margins could present a problem in defining edges (e.g., *Amphora* spp., *Neidium* spp., *Cocconeis* spp.). Although tools for 3D shape analyses are being developed to distinguish more complex biological shapes, diatoms with limited 3D structure still require good 2D analytical tools (e.g., Kloster et al., 2014). The optical alterations of specimens can also create false representations of the shape. Differential interference contrast and phase-contrast optics in light microscopy studies have an effect on the visual shape (distortion) of a specimen. High-resolution imaging techniques using electron microscopy (scanning electron microscopy and transmission electron microscopy) present higher-quality images for analysis with less chance of image distortion. The presence of cellular projections like spines can be informative features in cell shape analysis; however, the potential presence of projections into three-dimensional space can be problematic when flattened into two dimensions for analysis. Broad ranges in shape within the analyzed data set may also influence interrelation results. Large shape ranges can constrain finer shape differences or change biplot cluster associations (Fig. 2B vs. Fig. 3). LDA results on the larger data set showed a closer association of the asymmetric in *y*-symmetry than in *x*-symmetry, whereas PCA results showed a similar trend to our genera shape interpretations (Figs. 2, 3). Likewise, shape descriptors using low harmonic numbers are subject to typical statistical anomalies. The DiaOutline software presents a modern GUI workflow, which is easy to use and includes image-enhancing tools for improved outline extraction. DiaOutline imports the outline coordinates in a text file, which is compatible with other available data analysis software. We encourage the use of shape analysis tools in taxonomy, environmental monitoring, micropalaeontology, biostratigraphy, and forensic research as this analytical tool facilitates additional accurate and reproducible numerical metrics and adds scientific rigor to classical identifications.

ACKNOWLEDGMENTS

The authors thank Koeltz Botanical Books and Prof. Zlatko Levkov for approving the use of published images (*Gomphonema*, *Luticola*, and *Cymbella*) from the book series *Diatoms of Europe*.

AUTHOR CONTRIBUTIONS

A.W. developed the DiaOutline software and the R scripts, conducted the statistical analyses, and contributed equally to the writing of this manuscript. P.B.H. produced diatom images for this study, produced the outlines data set, and contributed equally to the writing of this manuscript.

DATA ACCESSIBILITY

The data used in this study are available as Supporting Information associated with the manuscript (Appendices S1–S3). The DiaOutline software and related documentation are available for free download at GitHub (https://github.com/wishkerman/DiaOutline).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

APPENDIX S1. A subset of specimens selected for the shape analysis. Specimens within each species represents a size diminution series. Reprinted with permission from Koeltz Botanical Books.

APPENDIX S2. Principal component analysis (PCA) plots for the diatoms and leaf shape extraction example. (A) PCA plot for the identification of diatom genera (*Luticola*, *Gomphonema*, *Cymbella*, and *Gyrosigma*). (B) PCA plot for all the specimens examined in the shape study. (C) Example of shape extractions for leaves.

APPENDIX S3. Principal component analysis (PCA) plots for specimens and species within the four genera: *Gomphonema* (A), *Luticola* (B), *Cymbella* (C), and *Gyrosigma* (D).

LITERATURE CITED

Adams, D. C., and E. Otárola-Castillo. 2013. geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4: 393–399.

Adebowale, A., A. Nicholas, J. Lamb, and Y. Naidoo. 2012. Elliptic Fourier analysis of leaf shape in southern African *Strychnos* section *Densiflorae* (Loganiaceae). *Botanical Journal of the Linnean Society* 170: 542–553.

Badi, A., C. Pandolfi, S. Mancuso, and A. Lenzi. 2017. A leaf-based back propagation neural network for oleander (*Nerium oleander* L.) cultivar identification. *Computers and Electronics in Agriculture* 142: 515–520.

Blanco, S., M. Borrego-Ramos, and A. Olenici. 2017. Disentangling diatom species complexes: Does morphometry suffice? *PeerJ* 5: e4159.

Boglino, A., A. Wishkerman, M. J. Darias, K. B. Andree, P. De La Iglesia, A. Estevez, and E. Gisbert. 2013. High dietary arachidonic acid levels affect the process of eye migration and head shape in pseudoalbino Senegalese sole *Solea senegalensis* early juveniles. *Journal of Fish Biology* 83: 1302–1320.

Bonhomme, V., S. Picq, C. Gaucherel, and J. Claude. 2014. Momocs: Outline analysis using R. *Journal of Statistical Software* 56: 1–24.

Brombin, C., and L. Salmaso. 2013. Theoretical aspects on permutation tests and shape analysis. *In* C. Brombin and L. Salmaso [eds.], Permutation tests in shape analysis, 17–36. Springer, New York, New York, USA.

Carlo, J. M., M. S. Barbeitos, and H. R. Lasker. 2011. Quantifying complex shapes: Elliptical Fourier analysis of octocoral sclerites. *Biological Bulletin* 220: 224–237.

Chitwood, D. H., and W. C. Otoni. 2017. Morphometric analysis of *Passiflora* leaves: The relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade. *GigaScience* 6: 1–13.

Chitwood, D. H., and C. N. Topp. 2015. Revealing plant cryptotypes: Defining meaningful phenotypes among infinite traits. *Current Opinion in Plant Biology* 24: 54–60.

Chitwood, D. H., A. Ranjan, C. C. Martinez, L. R. Headland, T. Thiem, R. Kumar, M. F. Covington, et al. 2014. A modern ampelography: A genetic basis for leaf shape and venation patterning in grape. *Plant Physiology* 164: 259–272.

Claude, J. 2008. Morphometrics with R. Springer, New York, New York, USA.

Crampton, J. S. 1995. Elliptic Fourier shape analysis of fossil bivalves: Some practical considerations. *Lethaia* 28: 179–186.

Elewa, A. M. T. 2012. Morphometrics for nonmorphometricians. Springer, New York, New York, USA.

Giardina, C. R., and F. P. Kuhl. 1977. Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Computer Graphics and Image Processing* 6: 277–285.

Glennon, K. L., and G. V. Cron. 2015. Climate and leaf shape relationships in four *Helichrysum* species from the Eastern Mountain Region of South Africa. *Evolutionary Ecology* 29: 657–678.

González-Wevar, C. A., T. Nakano, J. I. Canete, and E. Poulin. 2011. Concerted genetic, morphological and ecological diversification in *Nacella* limpets in the Magellanic Province. *Molecular Ecology* 20: 1936–1951.

Haines, J., and S. Crampton. 2000. Improvements to the method of fourier shape analyses as applied in morphometric studies. *Palaeontology* 43: 765–783.

Hammer, O., D. Harper, and P. Ryan. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* 4: 1–9.

Iwata, H., and Y. Ukai. 2002. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. *Journal of Heredity* 93: 384–385.

Jensen, R. J. 2003. The conundrum of morphometrics. *Taxon* 52: 663–671.

John, J. 2015. A beginner's guide to diatoms. Koeltz Scientific Books, Oberreifenberg, Germany.

Kloster, M., G. Kauer, and B. Beszteri. 2014. SHERPA: An image segmentation and outline feature extraction tool for diatoms and other objects. *BMC Bioinformatics* 15: 218.

Kloster, M., O. Esper, G. Kauer, and B. Beszteri. 2017. Large-scale permanent slide imaging and image analysis for diatom morphometrics. *Applied Sciences* 7: 330.

Krammer, K. 2002. *Cymbella*. *In* H. Lange-Bertalot [ed.], Diatoms of Europe: Diatoms of European waters and comparable habitats, Vol. 3. A. G. R. Ganter, Ruggell, Liechtenstein.

Kuhl, F., and C. Giardina. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18: 236–258.

Levkov, Z., D. Metzeltin, and A. Pavlov. 2013. *Luticola* and *Luticolopsis*. *In* H. Lange-Bertalot [ed.], Diatoms of Europe: Diatoms of European waters and comparable habitats, Vol. 7. A. G. R. Ganter, Ruggell, Liechtenstein.

Levkov, Z., D. Mitic-Kopanjz, and E. Reichardt. 2016. The diatom genus *Gomphonema* from the Republic of Macedonia. *In* H. Lange-Bertalot [ed.], Diatoms of Europe: Diatoms of European waters and comparable habitats, Vol. 8. A. G. R. Ganter, Ruggell, Liechtenstein.

Libungan, L. A., and S. Pálsson. 2015. ShapeR: An R Package to study otolith shape variation among fish populations. *PLoS ONE* 10: e0121102.

Lorenz, C., A. S. Ferraudo, and L. Suesdek. 2015. Artificial neural network applied as a methodology of mosquito species identification. *Acta Tropica* 152: 165–169.

Manly, B. F., and J. A. Navarro Alberto. 2016. Multivariate statistical methods: A primer, 4th ed. CRC Press, Boca Raton, Florida, USA.

Mou, D., and E. F. Stoermer. 1992. Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors. *Journal of Phycology* 28: 386–395.

Neustupa, J. 2013. Patterns of symmetric and asymmetric morphological variation in unicellular green microalgae of the genus *Micrasterias* (Desmidiales, Viridiplantae). *Fottea* 13: 53–63.

Otsu, N. 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9: 62–66.

Pappas, J. L., J. P. Kociolek, and E. F. Stoermer. 2014. Quantitative morphometric methods in diatom research. *Nova Hedwigia Beihefte* 143: 281–306.

Potapova, M., and P. B. Hamilton. 2007. Morphological and ecological variation within the *Achnanthidium minutissimum* (Bacillariophyceae) species complex. *Journal of Phycology* 43: 561–575.

Rohlf, F. J., and J. W. Archie. 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Biology* 33: 302–317.

Ros, J., A. Evin, L. Bouby, and M. P. Ruas. 2014. Geometric morphometric analysis of grain shape and the identification of two-rowed barley (*Hordeum vulgare* subsp. *distichum* L.) in southern France. *Journal of Archaeological Science* 41: 568–575.

Round, F. E., R. M. Crawford, and D. G. Mann. 1990. The diatoms: Biology and morphology of the genera. Cambridge University Press, Cambridge, United Kingdom.

Silva, M. F. S., I. M. de Andrade, and S. J. Mayo. 2012. Geometric morphometrics of leaf blade shape in *Montrichardia linifera* (Araceae) populations from the Rio Parnaiba Delta, north-east Brazil. *Botanical Journal of the Linnean Society* 170: 554–572.

Stanton, D. E., and C. Reeb. 2016. Morphogeometric approaches to non-vascular plants. *Frontiers in Plant Science* 7: https://doi.org/10.3389/fpls.2016.00916.

Steinman, A. D., and T. B. Ladewski. 1987. Quantitative shape analysis of *Eunotia pectinalis* (Bacillariophyceae) and its application to seasonal distribution patterns. *Phycologia* 26: 467–477.

Stela, B., and A. Monleón-Getino. 2016. Facilitating the automatic characterisation, classification and description of biological images with the VisionBioShape Package for R. *Open Access Library Journal* 3: 1.

Stoermer, E., and T. Ladewski. 1982. Quantitative analysis of shape variation in type and modern populations of *Gomphoneis herculeana*. *Nova Hedwigia Beihefte* 73: 347–386.

Stoermer, E. F., Q. Yu-Zao, and T. B. Ladewski. 1986. A quantitative investigation of shape variation in *Didymosphenia* (Lyngbye) M. Schmidt (Bacillariophyta). *Phycologia* 25: 494–502.

Theriot, E., and T. B. Ladewski. 1986. Morphometric analysis of shape of specimens from the neotype of *Tabellaria flocculosa* (Bacillariophyceae). *American Journal of Botany* 73: 224–229.

Tomaszewski, D., and A. Górzkowska. 2016. Is shape of a fresh and dried leaf the same? *PLoS ONE* 11: e0153071.

Tracey, S. R., J. M. Lyle, and G. Duhamel. 2006. Application of elliptical Fourier analysis of otolith form as a tool for stock identification. *Fisheries Research* 77: 138–147.

Vieira, M., S. J. Mayo, and I. M. de Andrade. 2014. Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piaui, Brazil. *Brazilian Journal of Botany* 37: 315–327.

Viscosi, V., and P. Fortini. 2011. Leaf shape variation and differentiation in three sympatric white oak species revealed by elliptic Fourier analysis. *Nordic Journal of Botany* 29: 632–640.

Wishkerman, A., and P. B. Hamilton. 2017. DiaCurv: A value-based curvature analysis application in diatom taxonomy. *Diatom Research* 32: 351–358.

Woodard, K., and J. Neustupa. 2016. Morphometric asymmetry of frustule outlines in the pennate diatom *Luticola poulickovae* (Bacillariophyceae). *Symmetry* 8: 150.

Zelditch, M. L., D. L. Swiderski, and D. H. Sheets. 2012. Geometric morphometrics for biologists: A primer, 2nd ed. Academic Press, Cambridge, Massachusetts, USA.

