A Protocol for the Creation of Useful Geometric Shape Metrics Illustrated with a Newly Derived Geometric Measure of Leaf Circularity

Author: Krieger, Jonathan D.

A PROTOCOL FOR THE CREATION OF USEFUL GEOMETRIC SHAPE METRICS ILLUSTRATED WITH A NEWLY DERIVED GEOMETRIC MEASURE OF LEAF CIRCULARITY[1]

JONATHAN D. KRIEGER[2,3]

[2]Herbarium, Library, Art and Archives Directorate, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom

- *Premise of the study:* I present a protocol for creating geometric leaf shape metrics to facilitate widespread application of geometric morphometric methods to leaf shape measurement.
- *Methods and Results:* To quantify circularity, I created a novel shape metric in the form of the vector between a circle and a line, termed geometric circularity. Using leaves from 17 fern taxa, I performed a coordinate-point eigenshape analysis to empirically identify patterns of shape covariation. I then compared the geometric circularity metric to the empirically derived shape space and the standard metric, circularity shape factor.
- *Conclusions:* The geometric circularity metric was consistent with empirical patterns of shape covariation and appeared more biologically meaningful than the standard approach, the circularity shape factor. The protocol described here has the potential to make geometric morphometrics more accessible to plant biologists by generalizing the approach to developing synthetic shape metrics based on classic, qualitative shape descriptors.

Key words: eigenshape analysis; geometric morphometrics; leaf physiognomy; leaf shape.

Morphometric tools have been tested many times on leaf form and proven to be useful (e.g., White et al., 1988; Premoli, 1996; McLellan and Endler, 1998; Jensen et al., 2002; Krieger et al., 2007; Bensmihen et al., 2008; and many others reviewed in Krieger, 2010). Generally, differences in shape (e.g., among species or genotypes) or patterns of shape variation (e.g., whether shape varies continuously between exemplar morphologies) that are detectable subjectively by the investigator become very apparent under morphometric analysis. However, these methods have not been explored with the specific goal of developing a useful, universal toolkit for plant biologists. To perform an eigenshape analysis or elliptic Fourier combined with principal components analysis (EFA-PCA), two of the most popular morphometric methods, it is first necessary to build up a set of specimens, on the order of several hundred to thousands, perform an eigenanalysis to generate a morphospace, and then assess the utility of the shape metrics that define this space (in this context, each axis x, y, z, etc. in the multidimensional morphospace is a shape metric, with movement along each axis describing some pattern of shape variation). It is often the case that the shape metrics generated this way are very similar between studies, and the differences are trivial;

nevertheless, these empirically derived shape metrics, created through an analysis of variance in the data set, will necessarily reflect the samples used, so although this approach succeeds at generating useful, novel shape metrics, it fails to generate a universal toolkit of shape metrics.

One possible approach to having shape metrics that are directly comparable between studies would be to request the matrices underlying the morphospaces generated in a published morphometric study. If it is possible to match a given data set to the published data set—e.g., by correctly interpolating and downsampling specimen outlines, then converting to phi functions, submitting to Fourier analysis, performing a relative warps analysis, etc., as needed—it would then be possible to project the specimens into the published morphospace, allowing a researcher to make the same shape measurements on his or her specimens as in a published study. However, even if the technical challenge of projecting specimens into a morphospace is not too much of a hurdle, convincing an investigator to share the requisite matrices may be an insurmountable challenge. When the alternative is a proscribed measurement, such as "length multiplied by width at a point 2/3rds down the midvein," the latter becomes very attractive by virtue of ease of use but disappoints in failing to describe the morphology for which there is such a rich set of qualitative terminology.

Here I present a general protocol for generating broadly useful measures of outline geometry, which can be computed without the need of a large-scale morphometric analysis. The proposed protocol is very straightforward: select two end-point geometries (specifically, mathematical geometries such as a circle or ellipse, not exemplar specimens) and submit them to a coordinate-point eigenshape (CPES) analysis, which defines the vector between them, generating a single, geometric shape

metric. The example explored here is circularity as measured using geometric morphometrics. Unlike existing metrics of circularity, such as the circularity shape factor (area multiplied by 4π, divided by perimeter squared) or Feret diameter ratio (diameter of the circle with same area as the specimen, divided by major axis length), this metric does not approximate circularity using perimeter (or length) and area; rather, it is measured directly from the geometry of the leaf. This geometric circularity vector represents a one-dimensional morphospace, what McGhee (1999) would term a "theoretical morphospace." The creation of theoretical plant morphospaces has so far been a very mathematical endeavor (e.g., Niklas, 1978; Prusinkiewicz and Lindenmayer, 1990; Prusinkiewicz et al., 2007); McGhee (1999) encouraged the further development of theoretical morphospaces by plant biologists, but provided only a series of ad hoc, mathematical examples, with no protocol to be followed or guidance for those who are not mathematically inclined.

The protocol described here is the next logical step from the work of MacLeod (2002a), where a variety of exemplar shapes were subjected to eigenshape analysis to inform the delineation between the different character states they represented. The most relevant example in that paper was an eigenshape analysis of three leaf shape exemplars: simple, palmately lobed, and pinnately lobed. Use of actual leaves as exemplars meant that there was both taxon- and specimen-specific information incorporated into the analysis (a shortcoming acknowledged and discussed at length in MacLeod, 2002a). While three leaf exemplars were used (and the two other examples in the same paper used eight and 12 exemplars), it was certainly inferred that two exemplars could be used, and moving from exemplars to pure, mathematical geometries (e.g., using a circle and a line, as opposed to selecting two leaves, one reasonably circular and the other reasonably linear) is also only a slight change to the methodology described there. Nevertheless, using two, pure geometries as a way to generate a useful shape metric was never explicitly suggested, and it does not appear to have been subsequently attempted by the botanical community as a means to generate shape measures or theoretical morphospaces. Therefore, it seems useful to publish this protocol as a very specific application of the more generalized technique described in MacLeod (2002a), and to explicitly describe how this type of analysis of pure geometries can be of use to the plant biologist, building on their knowledge of the natural variation in plant geometry. The current protocol also benefits from the incorporation of CPES analysis, vs. the extended eigenshape analysis used in that study.

I created a geometric circularity vector, independent of any biological specimens, as a sample application of the protocol. This can be thought of as a line connecting the two shapes, with movement along the line changing shape from a line to a circle. Having to derive the mathematical description of this transformation, much less the transformation between two more complicated geometries, would be challenging; the approach described here automates this step. I calculated scores along the line-circle vector for a sample of fern leaves showing a range of morphologies between linear and circular. These "scores" can be thought of as any other measurement of leaf shape: it is simple enough to imagine a score of "3.2," when the measurement is length in centimeters (i.e., 3.2 cm long); in the case of scores along the line-circle vector, the geometric meaning of scores can be visualized easily through modeling variation along the vector, which will show what shape corresponds to a particular score along the gradient from line to circle. To assess the line-circle vector against natural patterns of shape covariation, I compared scores along this vector to scores along a circularity axis from an empirical morphospace, generated through CPES analysis (MacLeod, 1995) of the fern leaf sample. The first principal component (or, eigenshape axis) in the CPES analysis, representing 85% of the variance in the sample, has previously been characterized as circularity (Krieger, 2007). This pattern of variation has appeared in diverse samples of leaves, usually explaining the majority of the sample variance, except where size has been left in the analysis (compare with models for a broad sample of dicot leaves in Krieger et al. [2007]; *Antirrhinum* L. in Weight et al. [2007]; *Arabidopsis* (DC.) Heynh. and *Antirrhinum* in Bensmihen et al. [2008]; and *Hoya* R. Br. in Torres et al. [2008]; in the second and third studies, size was the first principal component, and circularity was the second). Therefore, a metric that captures this pattern is likely to be both biologically meaningful and broadly useful. The line-circle vector was also generated using CPES analysis, a geometric morphometric technique, which is why I refer to it as the geometric circularity vector (abbreviated as "circularity vector" here). The initial empirical morphospace was centered on the overall sample mean shape (standard practice in this type of analysis). I made a second comparison, instead centering the sample on the mean of the circularity vector, comparing scores in this space to scores of the same specimens along the circularity vector (see Fig. 1). This ensured that the circularity vector intersected the center of the empirical morphospace, potentially bringing the two into closer correspondence. However, this would not be standard practice for a CPES analysis; if specimens are far from the center of the morphospace, the curvature of the space may distort the relationships among specimens. Finally, I standardized scores of leaves along the circularity vector to range from 0 to 1 and compared to calculated values of the circularity shape factor (CSF), the most commonly used metric of circularity, for each specimen. CSF ($4\pi \cdot$ area / perimeter^2) is mathematically constrained to range from 0 for a line (where area = 0) to 1 for a circle (where perimeter$^2 = 4\pi \cdot$ area).

METHODS AND RESULTS

I generated the geometric circularity metric as a line-circle vector, achieved very simply by performing CPES analysis on two shapes, a line and a circle, resulting in a one-dimensional morphospace. This is the essence of the proposed protocol: select two end-point geometries and submit them to a CPES analysis, which defines the vector between them. The two shapes must be centered at the origin, have unit centroid size, and have the correct number of points for the number used in the interpolation of our sample outlines. The "line" is not technically a line, but a linear outline with a width of zero (a description of how to compute these two shapes is given in Appendix 1). "Scores along the line-circle vector" and "geometric circularity" will be used interchangeably.

For comparison, I performed a CPES analysis on 938 leaves from 339 specimens, selected from 17 taxa (see Appendix S1) in the genus *Pleopeltis* Humb. & Bonpl. ex Willd. (Polypodiaceae), or closely related genera, depending on the circumscription (for additional information on specimen selection and preparation of leaf outlines, see Krieger, 2007), with each leaf represented by a 498-point outline. The CPES analysis generated a series of variance-optimized axes, like a principal components analysis, which together defined a morphospace (see Macleod, 1995, 1999). This space was centered at the overall sample mean shape, to which each shape was aligned using Procrustes superposition, part of the CPES analysis. The first axis in this space, ES1, appeared to correspond to circularity (Fig. 2A). Scores along the line-circle vector were strongly correlated to scores along ES1 in the original CPES analysis morphospace; there was no apparent structure to the residuals from a major axis regression, and there were no notable outliers (Fig. 2D). This showed that the geometric
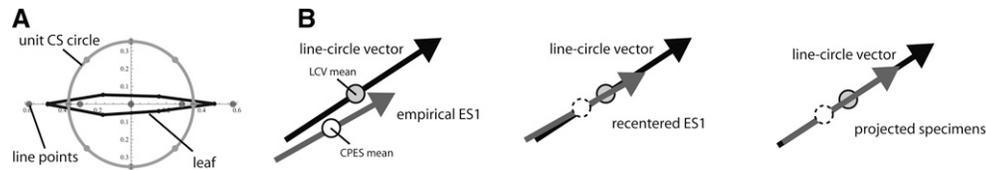
Fig. 1.   Concepts underlying construction of the geometric circularity vector. (A) Constructing the line-circle space. The one-dimensional space is constructed using a circle and a line, both centered at the origin with unit centroid size (the square root of the sum of squared distances of the points in the outline to the outline centroid). The size of the two shapes will depend on the number of points in the specimens, which the circle and line must match (see Appendix 1 for the relevant computations). (B) Comparing the line-circle vector (LCV) to other sample ordinations. The line-circle vector is the geometric metric of circularity. The other vector is the first axis in the coordinate-point eigenshape (CPES) analysis of fern leaves, empirical ES1. These two vectors may be highly correlated or not, and the line-circle region along the LCV may or may not intersect with the region of the ES1 vector occupied by specimens. To bring the two vectors together, it is possible to rerun the CPES analysis, instead centering the morphospace on the mean along the line-circle vector (the original CPES analysis was centered on the sample mean shape). The two vectors then pass through the same mean shape, and the angle between them is the correlation between the two metrics. It is also possible to project specimens onto the line-circle vector, to calculate values of geometric circularity for each specimen. The fern leaf sample only occupied a subset of the vector, because there were no perfectly circular or linear leaves.

circularity metric closely matched natural patterns of shape covariation in this sample. It also showed, definitively, that this axis in the CPES analysis corresponded to linear-circular variation, which previously had been inferred (Krieger, 2007).

I generated a second morphospace by instead aligning specimens on the mean shape along the line-circle vector (if a series of lines are drawn between corresponding landmarks on the two shapes, the mean shape is the shape formed of the midpoints of these lines). Scores along the new ES1 (see models in Fig. 2B) were strongly correlated to scores along the line-circle vector, with no apparent structure to the residuals or outliers (Fig. 2E). Despite recentering the morphospace, the results were largely the same as with the original CPES analysis. Points scoring high on the x-axis fell slightly below the regression line. This may reflect curvature of the recentered morphospace, due to centering it on a new mean shape.

I rescaled scores along the line-circle vector (see models in Fig. 2C) so they ranged from 0 (for a line) to 1 (for a circle), to match the range of values for the CSF. The two linear operations to achieve this—translation of the score at the "line" end of the line-circle vector to 0 and rescaling so the score at the "circle" end is at 1—were necessary so that every study measuring geometric circularity is using the same scale. To model variation at specific points along the rescaled vector, it is necessary to back-transform from scores along (0,1) to the original range, which is easily achieved. There was a strong correlation between the CSF (using only area and perimeter) and the geometric circularity metric. The slope of the major axis regression between these two metrics was not unity (Fig. 2E), as would be expected if they matched precisely. This could not be explained by the slight difference in the area:perimeter ratio for the circle as interpolated to 498 points. More likely, it reflected the fact that the actual leaves have rough margins, which will inflate perimeter with respect to area, and that the more circular leaves tended to have rougher margins and a greater prevalence of incisions along the margin. The sample was composed largely of simple leaves with entire margins; however, there were some noticeable outliers in Fig. 2F, specimens where the two metrics noticeably differed. One such leaf is shown in Fig. 3A. Because the margin of this leaf is incised, which is typical of this fern genus, the perimeter is inflated relative to area. To the degree that a plant biologist would assess circularity *as if that incision were not there*, it is clear that the geometric circularity metric is superior, because leaves with similar geometric circularity values have a similar overall shape, ignoring the incision.

An application of the line-circle vector is shown in Fig. 3B, with three of the taxa used in this study, showing three levels of dimorphy between sterile and fertile leaves: strong, weaker, and very weak (but still significant). Note that all of the distributions are along an axis scaled 0 (line) to 1 (circle). Using this protocol and the approach described in Appendix 1, another set of dimorphic leaves could be measured by another researcher, and directly compared to these values. This could be used as the basis for refining the terminology as applied to varying degrees of dimorphy (e.g., dimorphic, slightly dimorphic, slightly subdimorphic), as described by MacLeod (2002a) for degree and type of leaf lobing.

CONCLUSIONS

Based on the very strong correlations between geometric circularity scores and CPES analysis scores, it is clear that geometric circularity is an example of a synthetic shape metric that can capture natural patterns of shape covariation without the need of a morphometric analysis of a large data set. This alone means an enormous amount of time can be saved, digitizing a few specimens instead of the hundreds to thousands needed to generate a morphospace. The other pressing question is whether this synthetic metric is measuring what a plant biologist would consider to be "circularity." The example shown in Fig. 3A would suggest that geometric circularity is closer to a qualitative sense of circularity than the classic metric, CSF. Because the position along the line-circle vector is determined by Procrustes superposition, a leaf with a toothed or wavy margin will have the same score on this vector as one with a smooth margin of the same overall shape, whereas its increased perimeter will give it a much lower (more linear) value for CSF. In both of these cases, geometric circularity is relatively insensitive to features that would likely be ignored by a plant biologist making a qualitative assessment of circularity, whereas they would both have a significant impact on the measurement of CSF. There may be cases where shape factor is more biologically meaningful, e.g., when quantifying a physical process like heat exchange; however, using the term "circularity" for this seems misleading, as illustrated in Fig. 3A. The other significant issue with the classic CSF is that there is no mapping from a value to a shape, other than at the ends of the spectrum, whereas this modeling ability is inherent in the geometric approach. That is not to say there will not be multiple leaf shapes with the same geometric circularity value, just that this value is grounded in the distance between a sample leaf and the models shown in Fig. 2C. Unlike CSF, it is possible to have specimens that are past the ends (0,1) of the line-circle portion of the geometric circularity vector. The sample used here does not incorporate the vast variability found in leaves; there are certainly cases where the geometric circularity metric could break down, such as highly dissected leaves with little lamina, or leaves with lamina that overlaps itself. Whether real leaves will map outside of the range (0,1) remains to be seen, but it is a possibility. However, it is not necessarily a problem if they do, and such leaves are likely to also have misleading values of CSF, so it may be that neither metric is useful in such cases.

The intent of this protocol is to provide a generalized approach to generating shape metrics useful in the comparison and analysis of leaves. Ideally, we could understand geometric circularity as one component in overall form, along with patterns of variation like ovate-obovate and elliptic-oblong, which both have a grounding in classic, qualitative terminology and
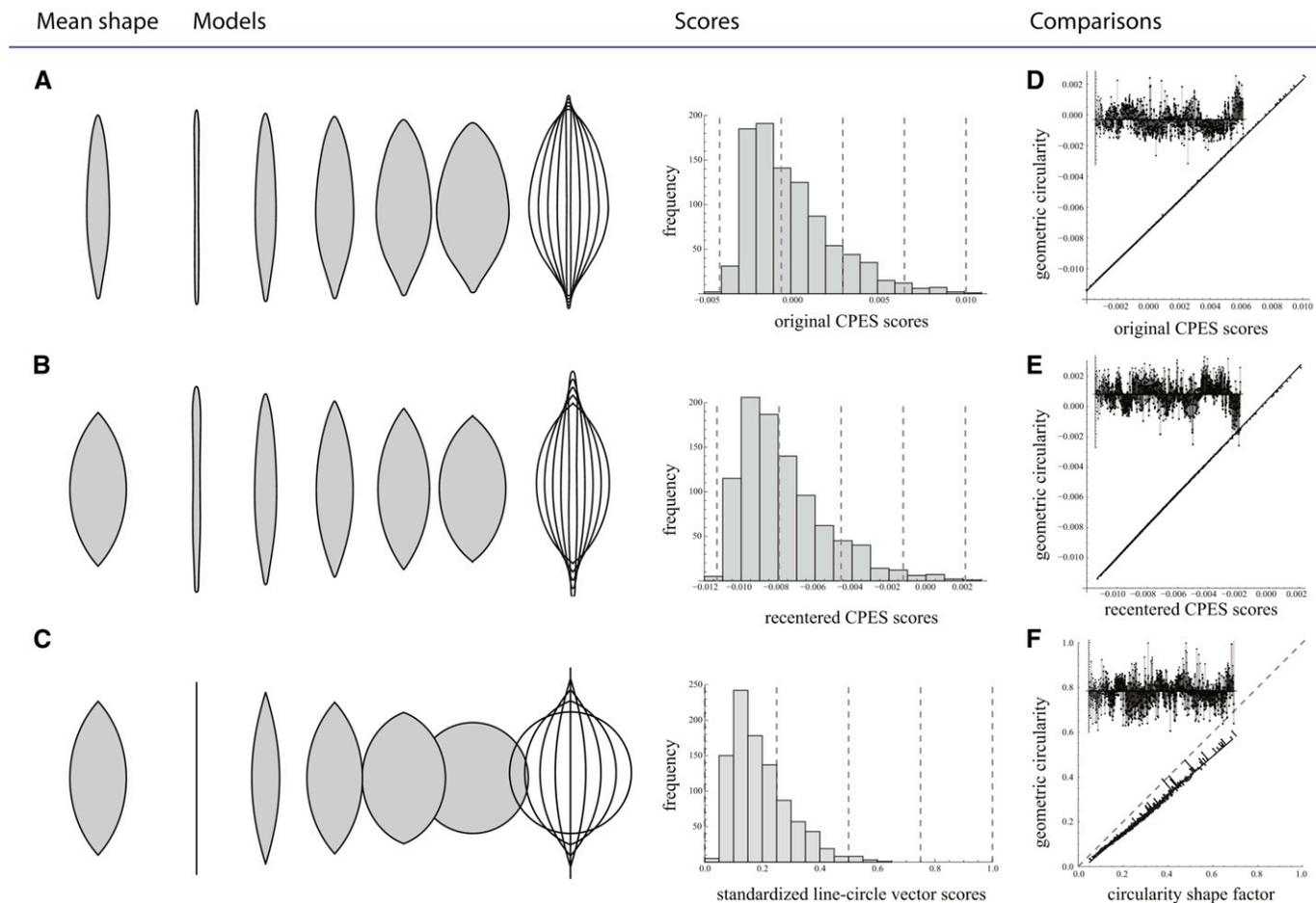
Fig. 2. Models and comparisons. For each analysis, five models were generated, at equal steps between the lowest and highest scoring specimens. An overlay of all five models is shown for each analysis. (A) Models and mean shape for the coordinate-point eigenshape (CPES) analysis of fern leaves. Specimens scoring low on this axis were nearly linear, and those scoring higher were more circular. There were no circular leaves in the analysis, and the shape models were limited to the range of scores of actual specimens. As models are generated higher along this axis, they become even more circular (not shown). (B) Models and mean shape for the CPES analysis recentered on the line-circle vector mean shape. The clearest difference between this and the original analysis is that the models look more geometric, i.e., less like real leaves. The models at the high end of the axis in (A) have a slightly more organic, asymmetric appearance, different in shape at the base and tip. (C) Models and mean shape for the geometric circularity vector. These models extend right up to a circle, a shape that is not present in the fern sample, although it would be possible to extend the models in (A) and (B) farther up the axis to better align them with those shown in (C). Real leaves also have nonzero widths, although there were some highly linear leaves in the sample. The distribution for the scores along the circularity vector looks very similar to that in (A) and (B), except that it is plotted along the full vector (from 0 to 1). (D) Major axis regression of CPES analysis circularity scores (along ES1 in the original analysis, as shown in [A]) to geometric circularity. There is a strong relationship between these two variables ($r^2 = 0.9999$, $F = 2.44 \times 10^7$, $P \ll 0.001$), and there is no apparent structure to the residuals (inset). (E) Major axis regression of the recentered CPES analysis circularity scores (as shown in [B]) to geometric circularity. There is a strong relationship between these two variables ($r^2 = 0.9999$, $F = 2.02 \times 10^7$, $P \ll 0.001$), and there is no apparent structure to the residuals. (F) Major axis regression of geometric circularity and circularity shape factor. The dashed line indicates a slope of 1. While there is a strong relationship between the two variables ($r^2 = 0.9975$, $F = 3.71 \times 10^5$, $P \ll 0.001$), there are clear outliers, which are informative as to the differences between these two metrics (see text and Fig. 3A; extreme outliers have been cropped out of the residuals plot).

appear as individual axes in morphometric analyses of diverse samples of leaves (both in the set of fern leaves used here, and in the diverse set of dicot leaves used in Krieger et al. [2007]; there are hints that they are present in other studies, including EFA-PCA analyses, but shape models are not published commonly enough to be sure). The challenge to generalizing the protocol is that the selection of end point geometries is not trivial. Unlike the case for circles and lines, it is not as clear as to what corresponds to the ideal "ovate" or "obovate" geometry. One solution is to fall back to the approach advocated by MacLeod (2002a), and use exemplar shapes provided by plant

biologists (e.g., Hickey, 1973; Leaf Architecture Working Group, 1999), which have been developed through the observation of large numbers of leaves by many botanists, an organic form of shape decomposition. The downside of exemplars is that they may contain taxon- or specimen-specific shape information that distracts from the pure geometries of interest. Another solution is to assemble a large, diverse set of leaves and see what geometries appear as orthogonal axes. The use of a morphospace in this way, for exploring patterns of shape covariation and identifying useful characters, is well established (e.g., MacLeod, 2002b). This is not inherently part of the protocol,
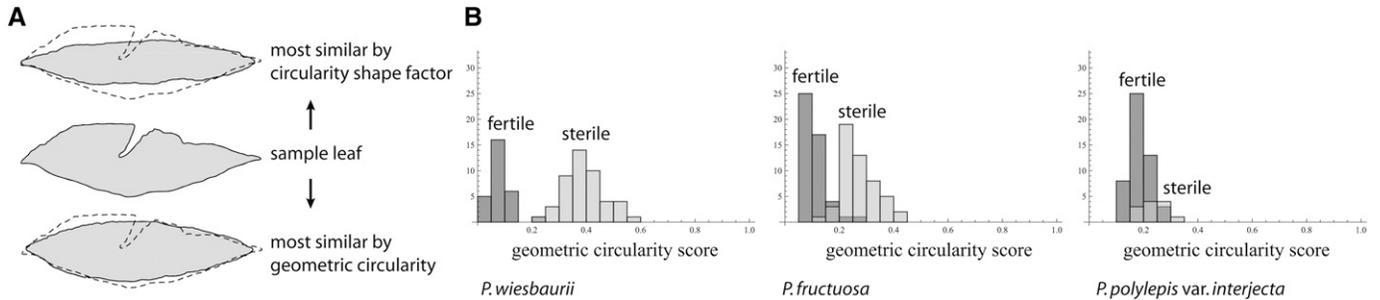
Fig. 3.   Circularity as a biologically useful metric. (A) Geometric circularity vs. shape factor. The sample leaf shown is one of the outliers in Fig. 2F, with different values for circularity shape factor (CSF) and geometric circularity (GC). The reason this leaf is an outlier is because it has an incised margin, a regular feature in these taxa. This inflates the perimeter relative to area, making the CSF more linear. The two closest-scoring leaves on the two circularity metrics are shown. Geometric circularity is clearly better at capturing the overall geometry. The leaf at the top is certainly, qualitatively, much less circular than the sample leaf, whereas the leaf at the bottom is a good match. (B) Frequencies of fertile and sterile leaves along the GC vector for three taxa (*Pleopeltis* species) with varying degrees of dimorphy. Given the strong correlation between GC and ES1 in the eigenshape analysis, the *x*-axis could instead be ES1, giving an identical result. This is current practice. However, the use of GC as the *x*-axis means that another researcher could calculate the same values for another set of dimorphic fern taxa (or, really, any closed outline) and directly compare the results to those shown here.

but it is useful for identifying uncorrelated characters, because the axes in a CPES analysis are orthogonal, as well as for assessing the proportion of variance explained by a particular character. As sample size and morphological diversity increase, the individual axis geometries tend to become more geometrically pure and less taxon- or specimen-specific. These axes can be used to inform the selection of pure forms of, e.g., "obovate," "ovate," "elliptic," or "oblong" (even for "elliptic," which should be easy to define, it is not clear what ratio of major and minor axes would be appropriate). It is certainly possible to use this approach to generate nonsensical metrics, if two unrelated geometric forms are selected. Therefore, it is likely to remain valuable to use empirical morphospaces to identify naturally occurring patterns of shape covariation. I suspect that an elliptic-oblong axis will be the next easiest to develop, perhaps by building an ellipse to match the line-circle vector mean shape, and using the vector between the mean and that ellipse as a starting point. The ultimate goal is to develop a standardized geometric toolkit that replicates the qualitative metrics already in use, such as those of Hickey (1973), in a quantitative, repeatable form.

LITERATURE CITED

Bensmihen, S., A. I. Hanna, N. B. Langlade, J. L. Micol, A. Bangham, and E. S. Coen. 2008.   Mutational spaces for leaf shape and size. *Human Frontier Science Program Journal* 2: 110–120.

Hickey, L. J. 1973.   Classification of the architecture of dicotyledonous leaves. *American Journal of Botany* 60: 17–33.

Jensen, R. J., K. M. Ciofani, and L. C. Miramontes. 2002.   Lines, outlines, and landmarks: Morphometric analysis of leaves of *Acer rubrum*, *Acer saccharinum* (Aceraceae) and their hybrid. *Taxon* 51: 475–492.

Krieger, J. D. 2007.   Quantification and analysis of variation in leaf shape: Climate signals in dicot leaves and evolution of leaf form in *Pleopeltis* (Polypodiaceae). Doctoral dissertation, University of Colorado at Boulder, Boulder, Colorado, USA.

Krieger, J. D. 2010.   Controlling for curvature in the quantification of leaf form. *In* A. M. T. Elewa [ed.], Morphometrics for non-morphometricians, 27–72. Lecture Notes in Earth Sciences, vol. 124. Springer, Berlin, Germany.

Krieger, J. D., R. P. Guralnick, and D. S. Smith. 2007.   Generating empirically determined, continuous measures of leaf shape for paleoclimate reconstruction. *PALAOIS* 22: 212–219.

Leaf Architecture Working Group. 1999.   Manual of leaf architecture: Morphological description and categorization of dicotyledonous and net-veined monocotyledonous angiosperms. Smithsonian Institution Press, Washington, D.C., USA.

MacLeod, N. 1995.   Testing morphometric data for phylogenetic and functional covariance. *Journal of Vertebrate Paleontology* 11(Suppl. 3): 41A–42A.

MacLeod, N. 1999.   Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* 5: 107–138.

MacLeod, N. 2002a.   Geometric morphometrics and geological shape-classification systems. *Earth-Science Reviews* 59: 27–47.

MacLeod, N. 2002b.   Phylogenetic signals in morphometric data. *In* N. MacLeod and P. L. Forey [eds.], Morphology, shape, and phylogeny, 100–138. CRC Press, Boca Raton, Florida, USA.

McGhee, G. R. 1999.   Theoretical morphology. Columbia University Press, New York, New York, USA.

McLellan, T., and J. A. Endler. 1998.   The relative success of some methods for measuring and describing the shape of complex objects. *Systematic Biology* 47: 264–281.

Niklas, K. J. 1978.   Morphometric relationships and rates of evolution among Paleozoic vascular plants. *In* M. K. Hecht and W. C. Steere [eds.], Evolutionary biology, vol. 11. Plenum Publishing Co., New York, New York, USA.

Premoli, A. C. 1996.   Leaf architecture of South American *Nothofagus* (Nothofagaceae) using traditional and new methods in morphometrics. *Botanical Journal of the Linnean Society* 121: 25–40.

Prusinkiewicz, P., Y. Erasmus, B. Lane, L. D. Harder, and E. Coen. 2007.   Evolution and development of inflorescence architecture. *Science* 316: 1452–1456.

Prusinkiewicz, P., and A. Lindenmayer. 1990.   The algorithmic beauty of plants. Springer, New York, New York, USA.

Torres, M. A. J., C. G. Demayo, and S. V. Siar. 2008.   Elliptic Fourier analysis of leaf outline differences between and among sixteen species of *Hoya. Philippine Agricultural Scientist* 91: 18–28.

Weight, C., D. Parnham, and R. Waites. 2007.   LeafAnalyser: A computational method for rapid and large-scale analyses of leaf shape. *Plant Journal* 53: 578–586.

White, R. J., H. C. Prentice, and T. Verwijst. 1988.   Automated image acquisition and morphometric description. *Canadian Journal of Botany* 66: 450–459.

APPENDIX 1. Computing geometric circularity.

The general protocol is to take two geometries and submit them to a coordinate-point eigenshape (CPES) analysis, to generate a vector spanning the two shapes. There is no set approach to selecting the two geometries (but see recommendations in the text), so the procedure described here will be specific to using a line and circle as end points.

First, make two synthetic shapes, a line and a circle. Then, perform a mean-centered CPES analysis on them and examine ES1, the single vector through the line-circle meanshape that describes variation between the two shapes.

The parameter n is the number of points. This may be slightly less than the interpolation values used with the original data analysis, which correspond to numbers of points in the outline before redundant points are discarded. For example, if two segments in a closed outline each have 100 points, with overlapping ends, the total number of points is actually 198 ($n = 198$), even though the interpolation values were (100, 100). The number of points n is constrained to be even, because the line is composed of two end points, with an equal number of points (zero or more) between the end points on each side (so all of the points between the ends are duplicated).

There are different options for scaling the objects. For the CPES analysis, the shapes need to be Procrustes aligned, thus, centered at the origin and set to a centroid size of 1.0. A circle with n points and a radius = $1/\sqrt{n}$ has a centroid size of 1, regardless of how the n points are spaced along the perimeter. The circle starts at $(-1,0)$, running counterclockwise.

The coordinates for the circle are computed as $(\cos(2\pi t/n), \sin(2\pi t/n))$, with t running from n/2 to n − 1, then 0 to n/2 − 1.

For a line of length 1 centered at the origin, centroid size is entirely squared x-distances. The formula is a bit more complicated than for the circle:

$$\text{line centroid size} = 2\sqrt{\frac{1}{n}\left(\frac{2}{3} + \sum_{k=1}^{n/2}\frac{2k-1}{3}\right)}$$

The line itself is a list of points, computed in one direction as $((4(t-1)/n) - 1, 0)$, with t running from 1 to (n + 2)/2, and in the other direction, filling in the middle of the line, as $((4(t-1)/n) - 1, 0)$, with t running (backward) from n/2 to 2.

The rest of the protocol will be the same, even if shapes other than a line and circle are used.

The mean shape is computed as the mean of these two shapes, which can be calculated as the average between each of the n corresponding coordinates for the line and circle. The mean shape should then be subtracted from each of the shapes, after which they are subjected to a CPES analysis (or, relative warps analysis, which is mathematically identical) to generate the sole eigenshape. This can be performed in most morphometric software that can perform a relative warps or eigenshape analysis.

To project specimens onto this vector, they need to be interpolated to the same number of points, rescaled to unit centroid size, then aligned to the mean shape using Procrustes superposition. This, too, can be achieved in a variety of commonly used morphometric programs. Finally, the line-circle mean shape is subtracted from specimens. Then, the scores for each specimen are computed as the covariance of each mean-centered specimen on the line-circle vector eigenshape.