APPLICATION ARTICLE

Applications in Plant Sciences

Digital morphometrics: Application of MorphoLeaf in shape visualization and species delimitation, using Cucurbitaceae leaves as a model

Oluwatobi A. Oso | Adeniyi A. Jayeola

Plant Anatomy Laboratory, Department of Botany, University of Ibadan, Oyo State, Nigeria

Correspondence
Oluwatobi A. Oso, Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511, USA.
Email: oluwatobi.oso@yale.edu

Present address
Oluwatobi A. Oso, Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA.

Abstract

Premise: Plant leaves are one of the most important organs for plant identification due to their variability across different taxonomic groups. While traditional morphometrics has contributed tremendously to reducing the problems accompanying plant identification and morphology-based species delimitation, image-analysis digital solutions have made it easy to detect more characters to complement existing leaf data sets.

Methods: Here, we apply MorphoLeaf to generate a morphometric data set from 140 leaf specimens of seven Cucurbitaceae species via landmark extraction, the reparameterization of leaf contours, and data quantification and normalization. A statistical analysis was performed on the resulting data set.

Results: A principal component analysis revealed that leaf blade area, blade perimeter, tooth area, tooth perimeter, the measure of the distance from tooth position to the tip, and the measure of the distance from tooth position to the base are important and informative landmarks that contribute to the variation within the species studied.

Discussion: MorphoLeaf can be applied to quantitatively track leaf diversity, thereby functionally integrating morphometrics and shape visualization into the digital identification of plants. The success of digital morphometrics in leaf outline analyses presents researchers with opportunities to carry out more accurate image-based research in areas such as plant development, evolution, and phenotyping.

KEYWORDS
botanical digitization, geomorphometrics, landmarks, leaf outline analysis, MorphoLeaf, shape visualization

Many quantitative analytical tools have been used to analyze the diversity of plant structures, with one of the most popular being morphometrics (Manacorda and Asurmendi, 2018). Morphometrics is used for the study of shape variations, particularly those in biologically relevant structures, and has been applied to multiple fields, including plant systematics, plant development and evolution, zoology, geology, geography, and other fields that depend on comparisons of structures, outlines, and contours (Itgen et al., 2019; Hernández-Esquivel et al., 2020; Pérez-Miranda et al., 2020; Terhune et al., 2020).

For most plants, leaves are one of the most important identifying features (Wäldchen et al., 2018), and present an opportunity to study diversity and the pattern of evolution in plants via morphometry. The diversity of leaf forms is a result of multiple factors, including genetic sequence, the regulated expression of specific molecular pathways, developmental patterns, and the environment (Nicotra et al., 2011; Dkhar and Pareek, 2014; Ichihashi et al., 2014; Chitwood and Sinha, 2016; Edwards et al., 2016). To fully understand the variation in leaf forms, it is important to perform an accurate analysis of the different leaf landmarks (Page et al., 2015; Soltis, 2017; Willis et al., 2017).

Morphometrics analyses are performed using one of three methods, depending on the data set used. Traditional

morphometrics have largely focused on basic measurements, including length, breadth, and angles. Outline-based morphometry focuses on the summary of the shape outlines, whereas landmark-based morphometry is the summary of the shape in relation to the homologous points. Geometric morphometrics (GMM) combines both landmark and outline analyses in the study of diversity and shape variation within species (Cope et al., 2012; Punyasena and Smith, 2014). GMM allows for the reconstruction of an average leaf shape and a visualization of the morphospace in which the leaf shape of each species belongs (Fu et al., 2017; Klein and Svoboda, 2017; Li et al., 2018; Borges et al., 2020).

Multiple programs have been used to analyze the landmark data of leaves, including geomorph (Adams and Otárola-Castillo, 2013), tpsUtil and tpsDig2 (Rohlf, 2015), MorphoJ (Klingenberg, 2011), ImageJ (Abramoff et al., 2004), LEAF-PROCESSOR (Backhaus et al., 2010), MorphoLeaf (Biot et al., 2016), and MASS (Chuanromanee et al., 2019). The present study uses MorphoLeaf (Biot et al., 2016) to integrate the basic steps of a GMM analysis with an all-in-one method of extracting the vital details of the leaf for a multiscale analysis, while still preserving the leaf outline and integrity. This method identifies biologically relevant and homologous landmarks along the leaf outlines with the aim of computing mean shapes. MorphoLeaf is available as a plug-in for Free-D software (Andrey and Maurin, 2005).

This work is an investigation of the common exploratory and confirmatory techniques in landmark-based geometric morphometrics. Due to the diversity reflected in the shape of leaves within the Cucurbitaceae, members of this family were used in this study. We asked two questions: (1) What structural differences and similarities are revealed from the foliar designs in the Cucurbitaceae family by landmark-based morphometrics? (2) Do naturally homologous points in Cucurbitaceae leaves allow trait diversity to be quantitatively tracked and used for the digital identification of plants?

MATERIALS AND METHODS

Sampling

Seven species of the Cucurbitaceae family were included in this study on the basis of their common leaf shape and patterns. Ordered from the most simple to the most complex leaf shape, these are: *Lagenaria siceraria* (Molina) Standl., *Coccinia grandis* (L.) Voigt, *Cucurbita pepo* L., *Benincasa hispida* (Thunb.) Cogn., *Trichosanthes cucumerina* L., *Momordica charantia* L., and *Citrullus colocynthis* (L.) Schrad.

The samples were collected from multiple populations in different locations to avoid bias arising from domestication. The sample sites were Ibadan, Lagos, Jebba, Taraba, and Jos in Nigeria, and Lucknow in India.

Approximately 50 mature leaves were collected per species, but only 20 leaves per species ($n = 20$) were scanned after careful selection so that each species was represented equally in the sampling size. Although there might be differences in leaf

sizes among the species within the family, we focused on leaf shapes that have identifiable homologous landmarks and noticeable differences in outlines (Figure 1A–G).

Imaging

An HP LaserJet Pro M1136 Mono Multi-Function Laser Printer (HP, Palo Alto, California, USA) was used to scan the samples. A 20-cm metallic ruler was positioned at the side of each scanned sheet as a size marker. The leaves were placed directly onto the scanner, adaxial face down, and scanned at a resolution of 300 ppi. Fresh leaves were used for specimens of all of the species except for *Coccinia grandis* and *Momordica charantia*, for which one-week-old pressed specimens were used.

Landmark data extraction

We followed the landmark data extraction guidelines in the MorphoLeaf manual (Biot et al., 2016), which involved several steps including contour extraction, peaks (lobes or teeth) and sinuses (indentations between peaks) extraction, data extraction, data quantification, data normalization, and data representation. Previous studies used 17 (Chitwood et al., 2016a) and 21 (Chitwood et al., 2016b) landmarks; for this study, we selected 17 biologically relevant landmarks, and these are listed in the last two pages of the MorphoLeaf manual.

Extraction of the leaf contour (or outline)

The watershed method employed by the MorphoLeaf application for automatic segmentation separates the leaf (the foreground) from the background and removes biologically nonrelevant details along the contour in the process. Because of the high quality of the images, the number of descriptors was sufficient to automatically retain the fidelity of the contour after extraction. The leaf contour was automatically extracted and manually corrected where necessary, particularly for leaves with unclear borders, which were mostly caused by deep sinuses. Using good-quality images allows for the easy detection of landmarks and reduces the need for manual correction. During the next step, two landmarks corresponding to the petiole were set manually, which allowed the automatic identification of the blade and the petiole. The leaf tip was then automatically determined as the point of the blade contour farthest away from the midpoint between petiole landmarks. This also defined the base–tip (or longitudinal) axis, which separates the blade into two halves.

Identification of sinuses and tips of teeth

In the next step, we automatically identified the teeth, which are defined as portions of the blade contour between two
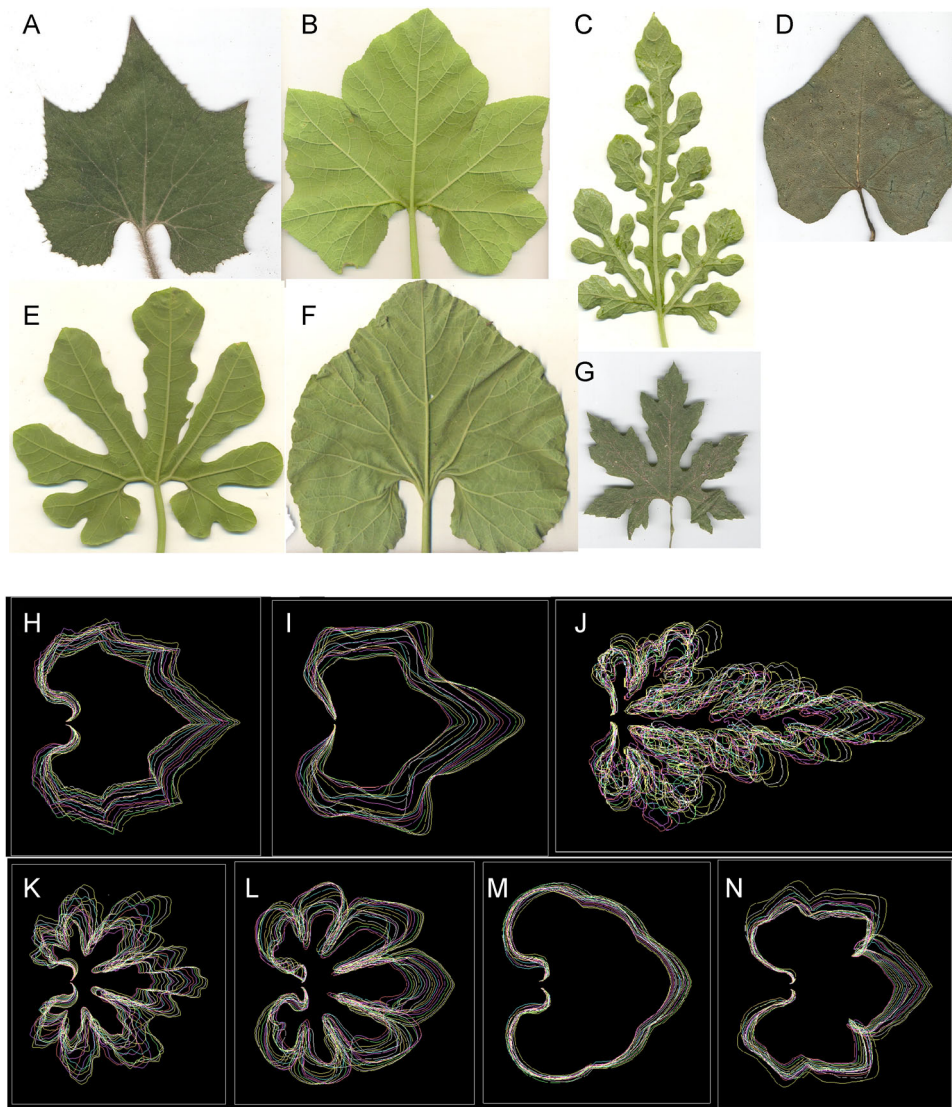
FIGURE 1 (A–G) Seven species of Cucurbitaceae representing seven genera. Each image corresponds to a leaf specimen, showing the leaf outline and teeth after scanning, before processing and analysis. (A) *Benincasa hispida*. (B) *Cucurbita pepo*. (C) *Citrullus colocynthis*. (D) *Coccinia grandis*. (E) *Trichosanthes cucumerina*. (F) *Lagenaria siceraria*. (G) *Momordica charantia*. (H–N) Mean shape visualization after landmarking and the reparameterization of biologically relevant points. The mean shapes are visualized to view the stack of each species' contours. (H) *Benincasa hispida*. (I) *Coccinia grandis*. (J) *Citrullus colocynthis*. (K) *Momordica charantia*. (L) *Trichosanthes cucumerina*. (M) *Lagenaria siceraria*. (N) *Cucurbita pepo*

sinuses. They are sometimes also known as lobes in the Cucurbitaceae. Sinuses, which correspond to contour points with a high concave curvature, were identified in a two-step procedure, known as the maximum local curvature method. In the first step, candidate intervals of the contour were determined as continuous domains where the curvature remains concave and above a user-defined threshold, which in this study was set on a leaf-by-leaf basis. Second, within each candidate interval, the point with the maximal curvature was selected as a sinus. After this automatic detection of sinuses, any errors were manually corrected to ensure that the tooth limits were correctly positioned and to avoid biases in subsequent analyses. After the sinus identification, MorphoLeaf determined the position of the tooth tip between consecutive sinuses. Another strategy available for

sinus identification, known as the local symmetry maximum method, is used for rather round structures. The user can choose either of these two strategies, depending on tooth shape. In this study, we chose the maximum local curvature method because it identifies the sharpest angle of the landmark.

The identification of teeth hierarchy is crucial for proper characterizations when there are serrations on the leaves; however, this was not performed here because the hierarchy of palmately lobed leaves with several levels of dissection cannot be established with MorphoLeaf. The appropriate setting of the parameters was sufficient to detect the sinuses on the main lobes and thus enabled proper quantitative analyses and mean leaf shape reconstruction.

Shape analysis

We extracted the quantitative parameters of all the sampled leaves for each species. Before proceeding with the shape analysis, the measures of the leaf contours, petiole/blade junctions, tooth sinuses, tooth peaks, and leaf apex landmarks were validated, i.e., it was confirmed that homologous landmarks were selected in all leaf specimens considered for each stack. If the landmarks were not homologous in all leaves in each stack, this confirmatory step would not be processed. The leaves were therefore processed to generate two files containing the quantifications of 17 landmark parameters, consisting of the blade length, two blade width parameters (BB, defined as the measurement from the farthest peak of one side of the symmetry to the other, and IS, defined as the measurement from the farthest sinus of one side of the symmetry to the other), blade area, blade perimeter, petiole width, number of upper teeth (the number of teeth on the left side of the symmetry), number of lower teeth (the number of teeth on the right side of the symmetry), total number of teeth, tooth position, tooth width, tooth height (latitude), tooth length (median), tooth area, tooth perimeter, tooth position from the leaf tip, and tooth position from the leaf base.

Normalization

We reconstructed the average leaf shapes using the moving average method. First, we resampled the leaf contours using the parameters of only two landmarks, the leaf apex and the petiole junction. We did not include the parameters of the other primary sinuses and tips because these do not change the contour resamples, and also because this causes difficulty in species with an extremely high number of vertices. Reducing the number of vertices thus meant that the shape outline might not be accurate. We then computed the moving average for leaf shapes. We defined the neighbor rank as 20, corresponding to the number of leaves that contributed to each average shape, which results in a smoother overall form of normalized shape (Figure 1H–N).

Morphometric analyses

All the morphometric data and codes used for this study (used to create Figures 1–8 and Appendix S1) are available from the Zenodo repository (Oso, 2020). Initial analyses used only the overall leaf variables (i.e., the width [BB and IS], length, area, and perimeter); subsequent analyses used a combined data set of the leaf and teeth variables. The trait means and their distribution were formally compared. All statistical functions were performed in R version 4.0.3 (R Core Team, 2020); the principal component analysis (PCA) was performed using prcomp with scale = TRUE, the Pearson's correlation coefficient analysis was performed using ggcorrplot, and the plots were produced using the package ggplot2 version 3.6.6 (Wickham, 2016).

RESULTS

Landmarks and extraction of leaf data

The primary reason for using MorphoLeaf is to automatically extract homologous landmark data for shape analysis. This includes all the processes of landmarking, sinus and peak selection, landmark editing and correction, and data averaging and normalization. The 140 leaves analyzed in this study represent seven different species and genera of the Cucurbitaceae family, with a focus on species with primary teeth or lobes (Figure 1A–G), and excluding secondary leaf types with leaflets. MorphoLeaf automatically extracted all the leaf landmarks; however, editing and manual correction were performed on a few leaves because of overlap in some of the leaves' lobes during scanning. The extracted data set was then used to generate mean leaf shapes after normalization (reparameterization); the reconstructed leaf shapes can be visualized using Sviewer, a standalone version of Free-D's 3D/2D rendering module for viewing curvature (".cv") files (Andrey and Maurin, 2005). The visualization of the mean shape outlines after reparameterization is presented in Figure 1H–N.

Leaf traits analysis

The coefficient matrix of the leaf traits data set, including the width (BB and IS), length, area, and perimeter (Figure 2A), showed a strong positive correlation between these five variables across all seven species, but the distributions of each variable differed between species, as qualitatively described in Figure 2A. The length and width correlated to each other in all species in the following order: *C. grandis* and *M. charantia* were similar in length and are the shortest (i.e., the distance from leaf apex to leaf base), followed by *B. hispida*, *T. cucumerina*, *C. pepo*, *C. colocynthis*, and *L. siceraria*. Leaf width differed among species in a similar manner to leaf length, with a significant level of variation. A PCA analysis was then performed on the leaf data set, excluding the teeth-specific variables, to investigate the quantitative differences in the width (BB and IS), length, area, and perimeter. The species clustered into seven groups according to the outline variables (Figure 2B). The scree plot of the PCA for the leaf variables is shown in Figure 2C, revealing that most of the variations are observed in the first and second principal components (PC1 and PC2), accounting for 76.2% and 19.8% of the variation, respectively. Although the PCA (Figure 2B) included only the leaf variables, it effectively separated the species without an overlap between species. When considered individually, some of the variables overlapped between some of the species. PC1 and PC2 were further plotted against the leaf length (Figures 3A, 3B), leaf width (Figures 3C, 3D), and leaf perimeter (Figures 3E, 3F) to determine the variations between the species. These variables were found to be negatively correlated with PC1 and positively correlated with PC2.
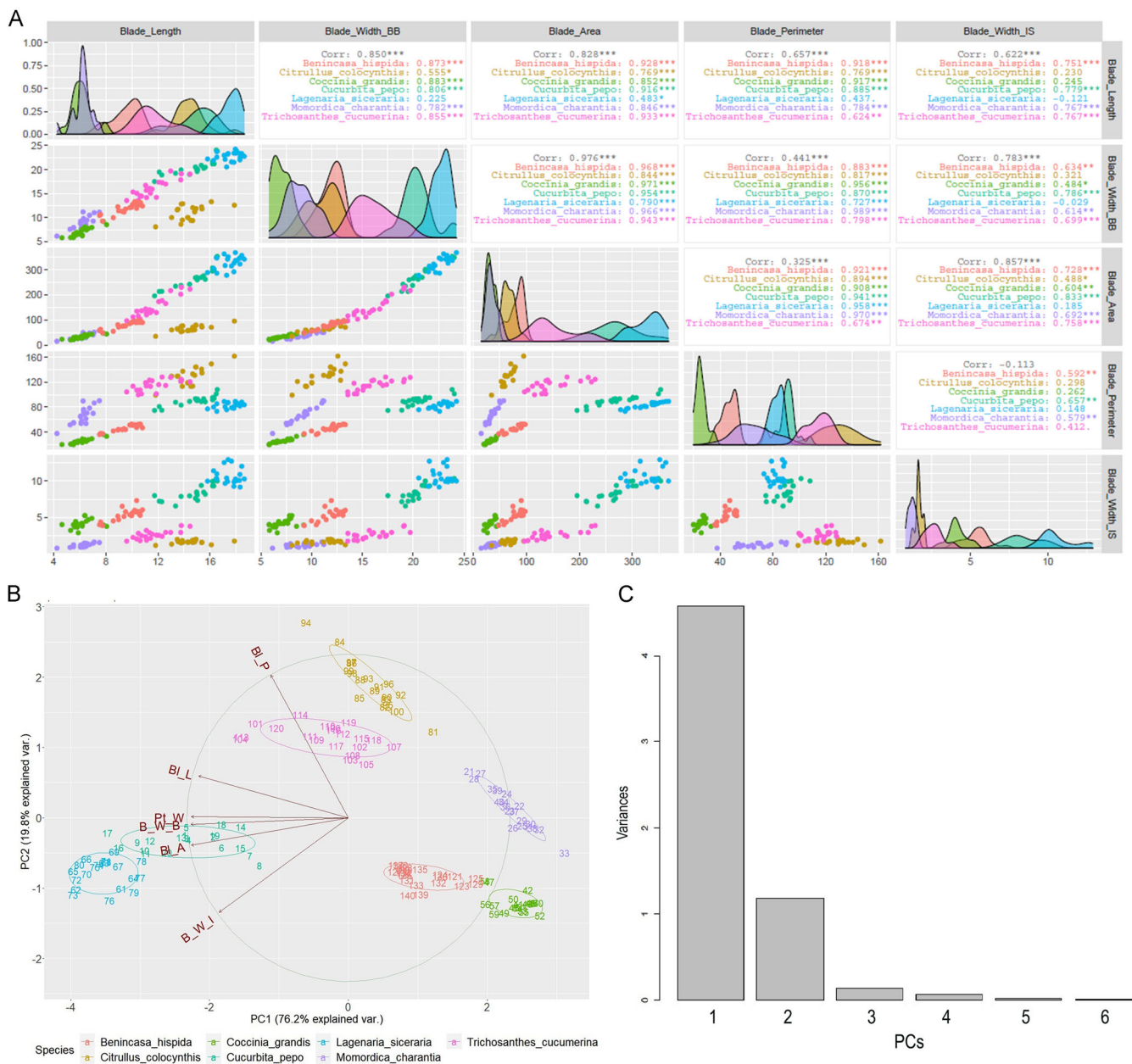
FIGURE 2 Comparison of all leaf blade data sets plotted with leaf blade variables only. (A) Scatterplot and coefficient matrix of the leaf blade data. The data set excluding teeth variables shows statistically significant positive correlations between all leaf blade variables within species. (B) Principal component analysis of the leaf data set showing the space of best fit for each species. B_W_B = blade width BB, B_W_I = blade width IS, Bl_L = blade length, Bl_A = blade area, Bl_P = blade perimeter, and Pt_W = petiole width. (C) Scree plot of the principal component analysis

The IS blade width is an important informative landmark of the leaf. It represents the distance between the position 1 sinuses (the first sinus from the leaf apex) on either side of the leaf symmetry. This trait shows significant correlation with the BB blade width (Figure 4A). Similarly, the area (or total blade cover) of the leaf is another discriminative landmark. Only *M. charantia* and *C. grandis* have overlapping leaf area sizes (Figure 4B), just as they overlapped in the leaf length and leaf width traits. Likewise, leaf perimeter allows further discrimination between species, especially those that share overlap in other landmarks

(Figures 4C, 4D). The pairs of *C. pepo* and *L. siceraria*, *T. cucumerina* and *C. colocynthis*, and *C. grandis* and *M. charantia* each overlapped in leaf area, but could be distinguished by their perimeter values.

Tooth traits analysis

A similar set of analyses to those carried out on the leaf landmarks data set was also carried out on the teeth landmarks data set. The teeth data set was further processed
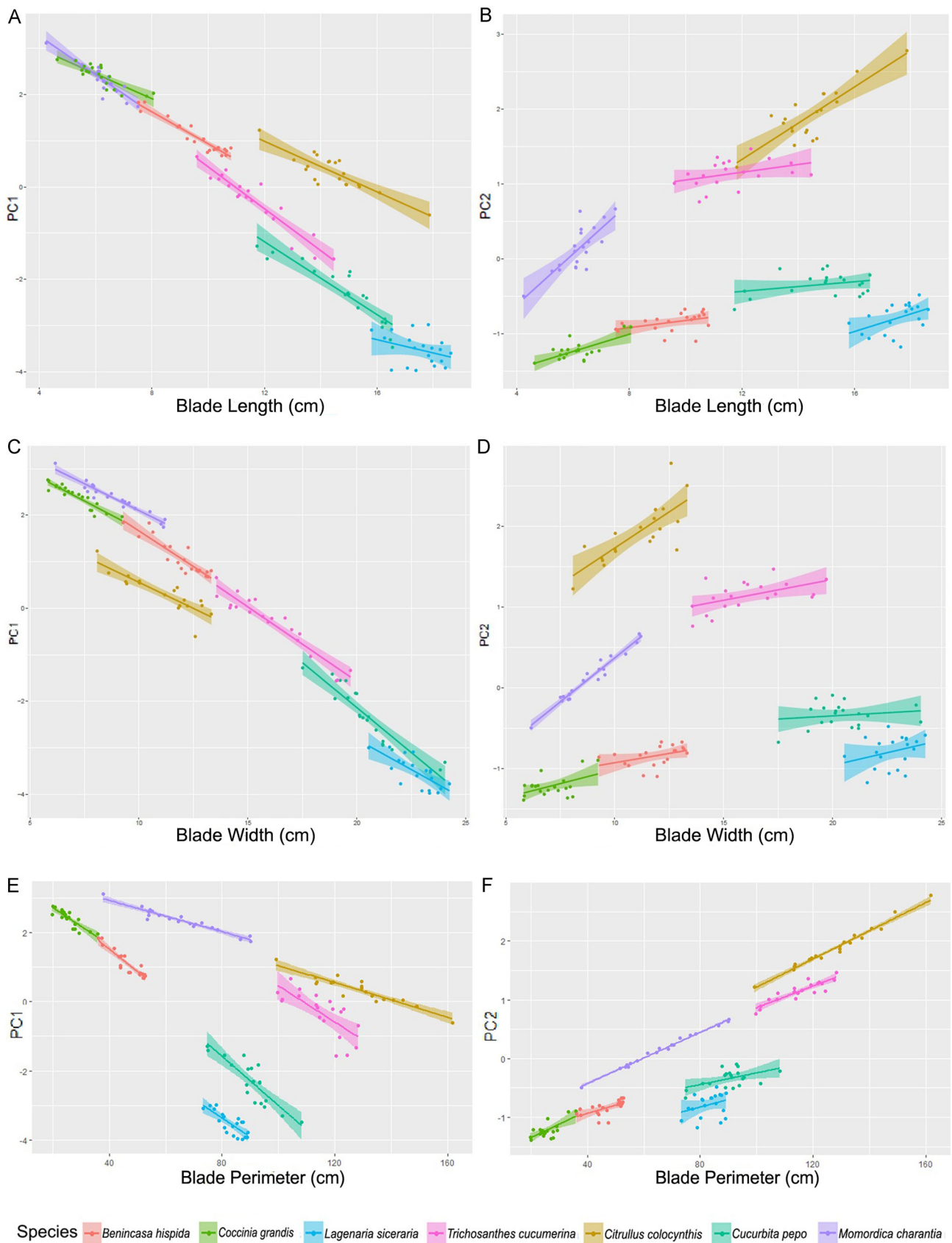
FIGURE 3 Principal component analysis of leaf variables to determine the variations between species. Principal components 1 (PC1) and 2 (PC2) were plotted against the leaf length (A, B), leaf width (C, D), and leaf perimeter (E, F). These traits were negatively morrelated with PC1 and positively correlated with PC2
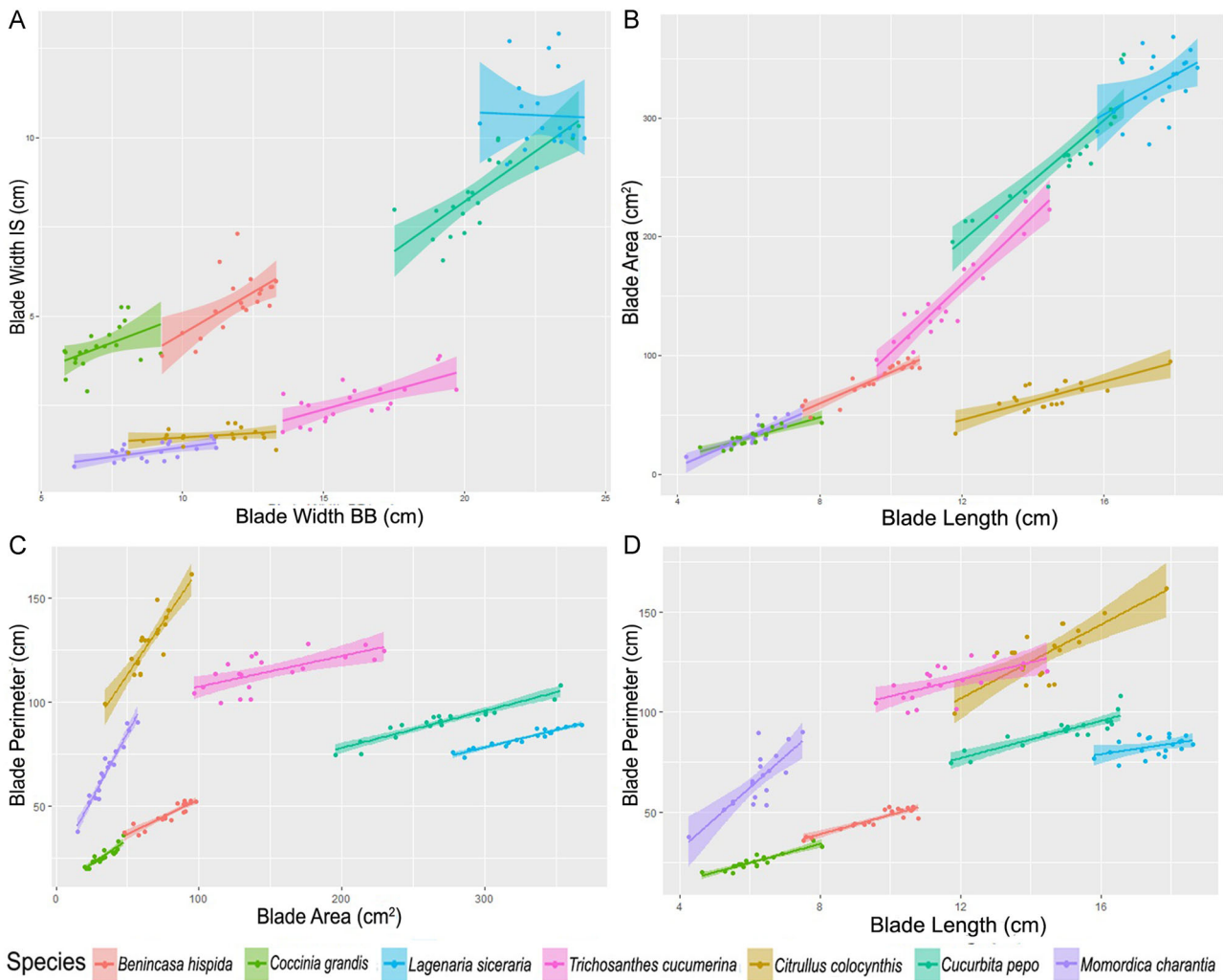
FIGURE 4 Comparison of leaf blade parameters to assess their correlation between species. (A) Blade width BB vs. blade width IS. (B) Blade length vs. blade area. (C) Blade area vs. blade perimeter. (D) Blade length vs. blade perimeter
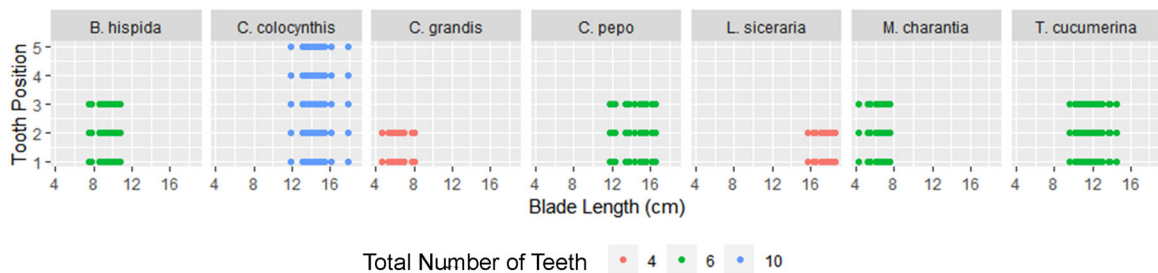

FIGURE 5 Total number of teeth for each species and the position of the teeth in relation to the blade length varies across the seven species studied

from 140 samples into 840 samples, i.e., the total number of teeth per leaf for each species multiplied by the number of specimens. This was done to treat each tooth position as an entity of its own (the leaf tip does not count as a tooth, it is the apex; Oso, 2020). The number of teeth per species varied between four and 10 (Figure 5), with all species showing unique tooth structures, and the total number of teeth per species and the tooth position in relation to the blade length

varied across the different species. The tooth variables considered during our PCA were tooth position, tooth width, tooth height (latitude), tooth length (median), tooth area, perimeter, tooth position from leaf base, and tooth position from leaf tip.

Figure 6 presents scatterplots of principal components 1–4 with all leaf and tooth variables, 13 in all, with PC1 responsible for the most variation. The inclusion of the
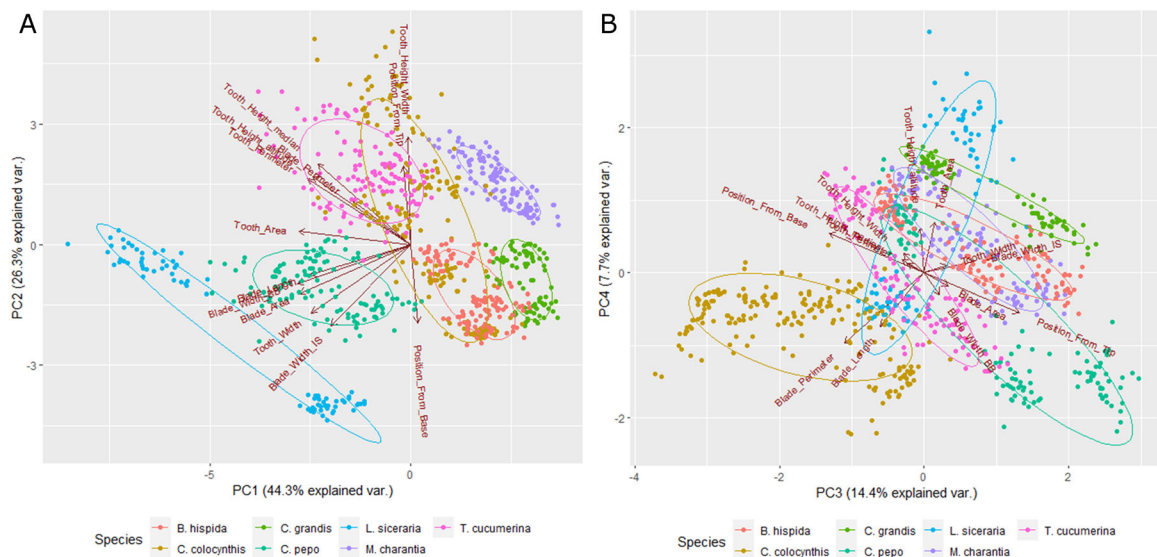
FIGURE 6 Principal component analysis of all data sets, including the tooth and leaf blade variables. (A) Principal component 1 (PC1) contributed the most to the variation within the group studied (44.3% of variation), while PC2 contributed 26.3%. (B) PC3 contributed the least of the three most important principal components, at 14.4% of the variation

tooth variables caused overlaps in the grouping, which could be attributed to the teeth positions on either side of the blade's longitudinal axis. However, as the species were known a priori, they were still separated correctly into seven groups, with each group representing a single species, and with the tooth position then separating each species into different subgroups based on the number of teeth. The correlation coefficient of the teeth landmarks was then plotted (Figure 7A) and expanded in a pairwise scatterplot of all landmarks (Appendix S1). The distance of tooth position from the leaf tip and the distance of tooth position from the leaf base are the most negatively correlated because they are opposite measurements. The blade height to width ratio is also negatively correlated against all leaf variables except the leaf perimeter. When all variables from an all-inclusive blade data set (including teeth landmarks) were plotted as a correlation matrix against PC1, PC2, and PC3, important landmarks were found to contribute to the variations in each of the principal components (Figure 7B).

The perimeter of each tooth relative to the blade perimeter is a highly variable feature (Figure 8A). While the tooth perimeter at position 1 was low in some species (*L. siceraria*, *C. colocynthis*, and *C. grandis*), it was higher in others (*C. pepo*, *M. charantia*, and *T. cucumerina*); tooth perimeter was also seen to vary at positions 2–5 across all species in our study. There was also a positive correlation between the tooth perimeter and the blade perimeter: the longer the blade perimeter, the longer the tooth perimeter. A similar situation was observed in the tooth area relative to the blade area (Figure 8B). This is a result of the differences in tooth size at each position on the leaf. The size (width and height) of each tooth on either side of the blade's longitudinal axis were similar (Figure 8C); i.e., the size of the tooth at position 1 on the left side of the leaf is similar to the tooth at position 1 on the right side, the size of the tooth

at position 2 on the left side of the leaf is similar to the tooth at position 2 on the right side, and so on.

DISCUSSION

There has been an evolution in the visualization and analysis of leaf shapes in recent times, from ImageJ (Abramoff et al., 2004) to LEAFPROCESSOR (Backhaus et al., 2010), MorphoJ (Klingenberg, 2011), MorphoLeaf (Biot et al., 2016), and MASS (Chuanromanee et al., 2019), with different scientists demonstrating their effective application to similar studies. These tools are simple to use and allow for replicability in application. While other programs rely on a combination of multiple software packages for landmarking, outlining, and analysis, coupled with the import and export of file formats, MorphoLeaf is reliable for landmark selection, in-software image analysis, superimposition, and visualization. It performs all of these tasks as a plugin in FreeD (Andrey and Maurin, 2005) and comes with an S-viewer for the 2D visualization of the shapes.

The seven species used in this study had leaf shapes ranging from simple to complex (Figure 1A), and the shape variation was visibly separated into seven distinct species groups, with all landmarks important for this grouping. When MorphoLeaf is applied to more species within the family, the species cluster into reliable groups that reveal the influence of leaf landmarks and teeth variations on evolutionary data within the group, even for divergent plants that have been modified as a result of adaptation-inducing variables such as the environment. ImageJ (Abramoff et al., 2004), a similar tool for landmark extraction, has been used along with other applications for the subsequent processing and analysis of leaf shapes (Corney et al., 2012), generating data that were used to predict the species
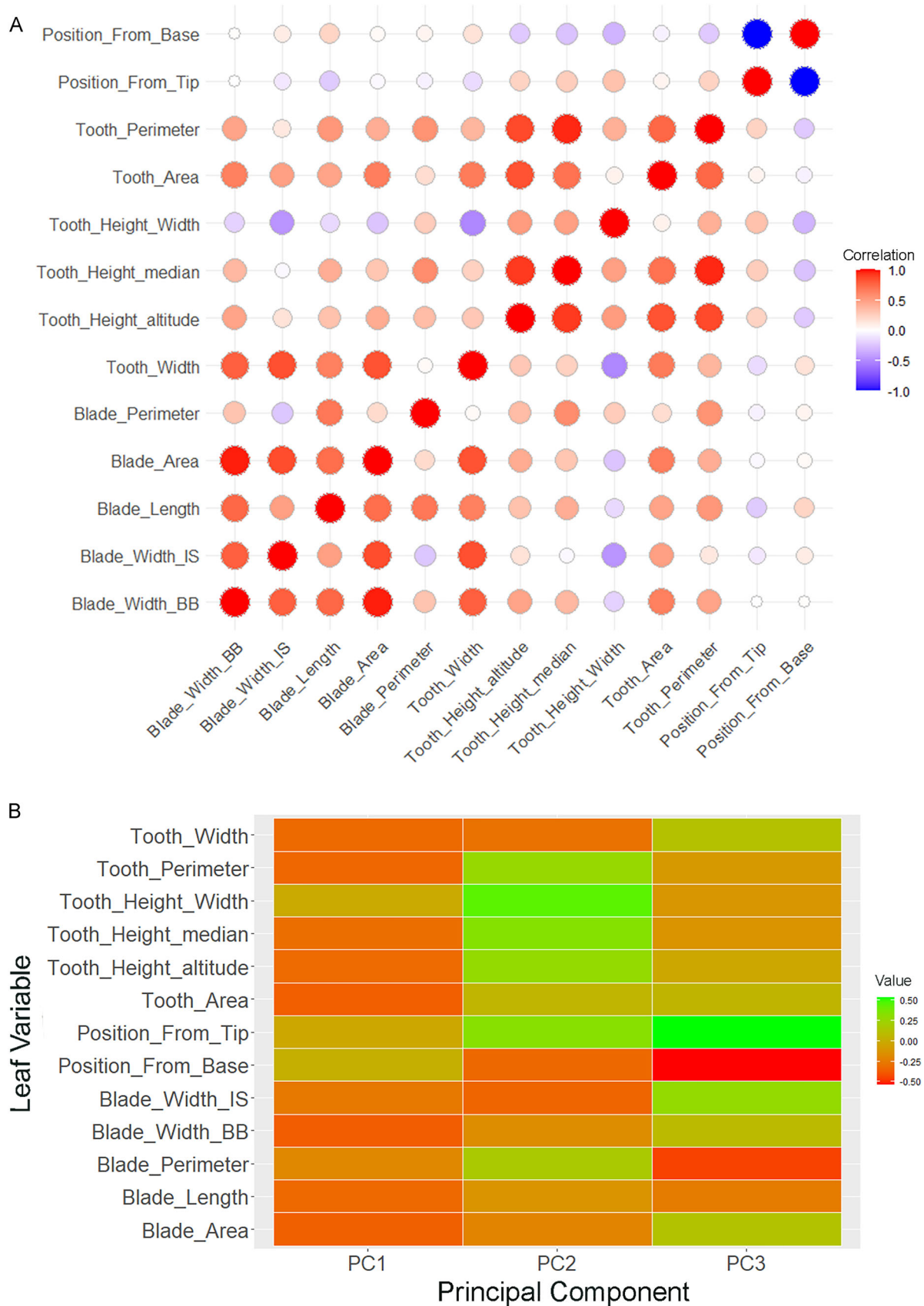
FIGURE 7 Correlation plots of leaf and teeth variables. (A) Correlation coefficient matrix of all teeth landmarks, from negatively correlated variables (−1) to positively correlated variables (+1). (B) Correlation matrix of all informative variables vs. the three informative principal components (PCs)
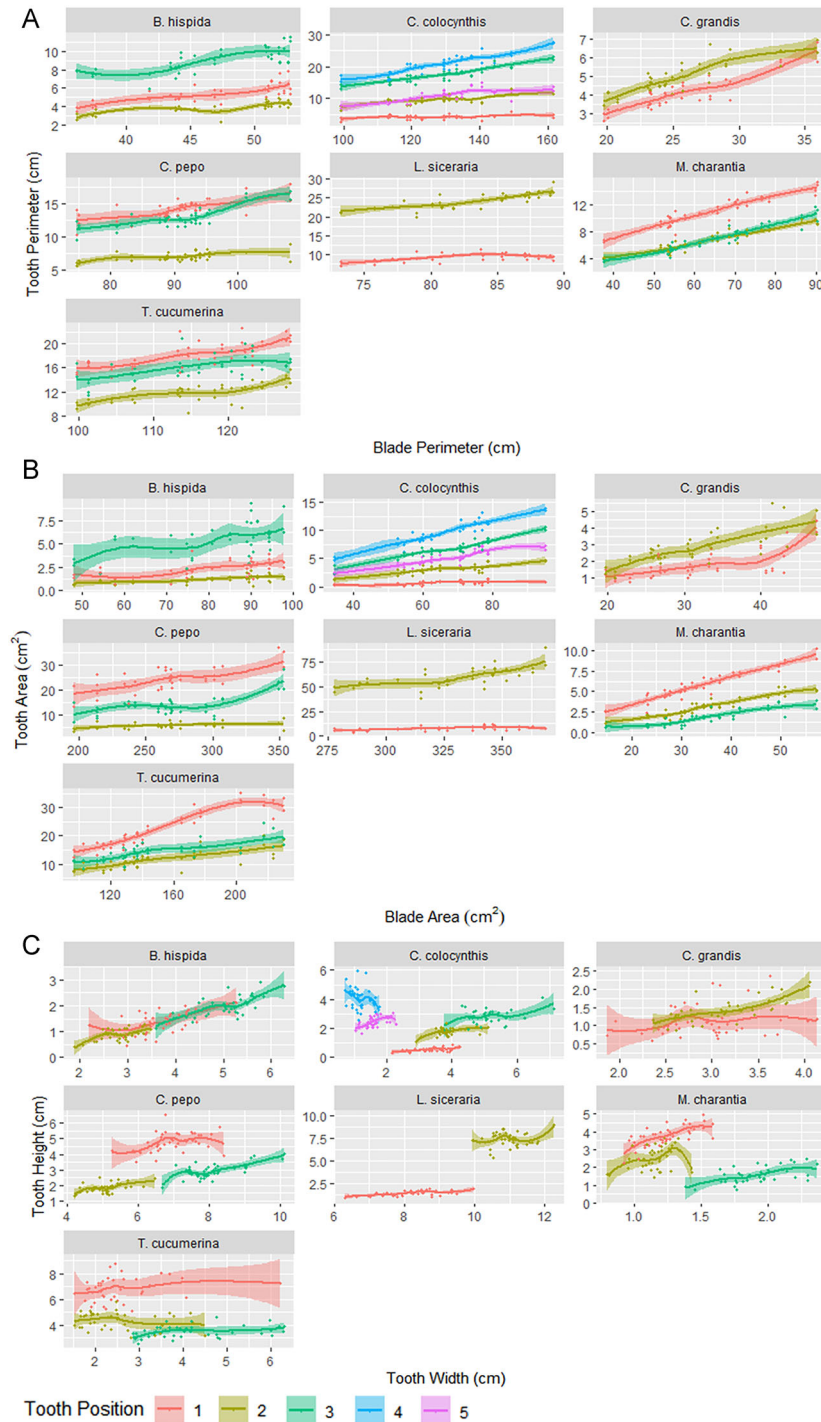
FIGURE 8 Comparison of teeth vs. leaf blade parameters to assess their correlation between species. (A) Perimeter of each tooth relative to the blade perimeter is a highly variable landmark. (B) Tooth area contributes significantly to the blade area. (C) Tooth height and width vary by tooth position within and between species

position and the role of leaf landmarks in the evolution and development of *Passiflora* L. (Chitwood and Otoni, 2017) and *Vitis* L. (Chitwood et al., 2016).

This study applies an image-based interspecific GMM analysis (digital morphometry), and the potential effectiveness of these methods when applied to a larger group can be observed in the differentiation reported between the species. The different patterns in the leaf perimeter, leaf area, tooth area, and tooth perimeter indicate that different node positions will be observed (on the phylogenetic tree) when these methods are applied to the analysis of a larger group.

It is possible that a GMM-based clustering would show identical groupings to a gene-based phylogeny, as some level of success was achieved in revealing the evolution of shapes within *Oxalis* L. when geomorphometric data were combined with molecular data (Morello et al., 2018). Importantly, however, GMM-based clustering contributes to a more robust result regarding the evolutionary and functional significance of the variations in leaf shape across plant groups. These variations may be caused by climatic variables or genetic modifications, with different shapes and sizes representing the effects of temperature, rainfall, exposure to

sunlight, and mutations, providing scientists with reasons to hypothesize the evolutionary importance of leaf shape (Kidner and Umbreen, 2010; Nicotra et al., 2011; Edwards et al., 2017; Gallaher et al., 2019).

Previous morphometric studies in the family Cucurbitaceae have included few species, largely focusing on qualitative characters or measurements of characters. These characters are important but limited to the identification of species and the morphological characterization of varieties, not the robust delimitation of higher taxonomic ranks. Our analysis suggests that a PCA of the leaf outlines and teeth clearly segregates the seven Cucurbitaceae species studied here into seven groups in agreement with the existing species delimitation. Each of the variables contributed to this separation, with blade area, blade perimeter, tooth area, tooth perimeter, the measure of the distance from tooth position to the tip, and the measure of the distance from tooth position to the base being important and informative factors when quantitatively determining the placement of each species into the correct group. This should not, however, be used in isolation, but given the same weight as traditionally used characters (e.g., leaf type, apex, base, and venation) in understanding the evolution within a plant group. The relationship between vasculature and leaf blade is also of great importance as it contributes up to 15 additional informative landmarks (Chitwood and Otoni, 2017).

Developmental studies on different genotypes (using specific cultivars and varieties) within a species have shown that heteroblastic situations (usually caused by genetic or environmental factors) result in divergence from the basic leaf shapes (Chitwood et al., 2012; Ostria-Gallardo et al., 2016); therefore, intrinsic differences within each species cannot be ignored. The samples used in the present study were collected from different populations; thus, the different ecological variables in the environments from which they were collected could be (at least partially) responsible for these differences. This does not diminish the importance of the base leaf outline of each species, rather it underscores the need for sampling across multiple populations to avoid bias during taxonomic work. Furthermore, sampling multiple populations ensures that all factors and characters are taken into consideration to provide a full picture of the variation within a species when tracking evolution within a plant group or performing other bioinformatic work. The ease of use and efficiency of digital morphometrics should encourage botanists to collect a comprehensive pool of data, thus allowing them to draw more accurate inferences, as they seek to answer their research questions (Luca and Annamaria, 2019).

There have already been some morphometric studies (Josephine et al., 2015; Ekeke and Agogbua, 2018), phylogenetic studies (Zhang et al., 2006; Renner and Pandey, 2013; Misra et al., 2017; Chomicki et al., 2019), and recently phylogenomic studies (Guo et al., 2020) on the Cucurbitaceae, but none referred to the important informative characters listed earlier. Although digital morphometry is only beginning to reveal these additional characters, it is important that they are used in subsequent studies. Digital morphometrics reveals that morphological data can improve the results of bioinformatic research on interspecific and intraspecific plant populations, particularly when genetic and environmental data are incorporated (Klein et al., 2017; Migicovsky et al., 2018).

Although MorphoLeaf was designed using intraspecific models, this study shows its success in interspecific leaf shape analysis and the quantitative tracking of trait diversity and structural homology. This is further supported by evidence from GMM analyses of leaf variation in four species of *Quercus* L., revealing a strong correlation between their leaf shape and taxonomy (Viscosi et al., 2009). MorphoLeaf's success in identifying intraspecific and interspecific leaf shape variations, just like the other applications mentioned earlier, allows evolutionary biologists to use complementary traits to identify and classify plant species. These tools are encouraging further exploration of the ability of image-based geomorphometry to provide phenotypic data that can supplement existing morphometric and molecular data for phylogenies, and GMM analyses have already been key in developing various applications, guides, devices, and ultimately a central virtual herbarium for plant identification (Agarwal et al., 2006; Belhumeur et al., 2008; Cope et al., 2012; Jamil et al., 2015; Schlautman et al., 2020).

As the shift from traditional morphometrics to digital morphometrics becomes more widespread, there will be more applications and technological solutions that will allow mass botanical digitization and overcome difficulties in plant identification. Mass digitization provides natural history collectors, including but not limited to herbarium and field experts, with opportunities to mobilize data to explore different research hypotheses, solve scientific problems, reduce data gaps and biases, and for bioinformatic purposes (Beaman and Cellinese, 2012; Soltis, 2017; Soltis et al., 2018; Lorieul et al., 2019; Kattge et al., 2020). By integrating systematics and digital morphometrics, taxonomists and herbarium curators can leverage the availability of additional data generated from modern technologies in the digital identification of plants. MorphoLeaf will also make possible subsequent works focusing on its application to species discrimination within a single genus, and applications involving herbarium specimens and other leaf types, particularly complex leaves.

AUTHOR CONTRIBUTIONS

A.A.J. supervised the experiments and reviewed the manuscript; O.A.O. designed the experiments, performed the overall experiments, wrote the first draft of the manuscript, and revised the final version. All authors approved the final version of the manuscript.

OPEN RESEARCH BADGES

This article has been awarded Open Materials and Open Data badges. All materials and data are publicly accessible via the Open Science Framework at https://doi.org/10.5281/ZENODO.5147936. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki

DATA AVAILABILITY STATEMENT

The data set presented in this manuscript, including the raw data set of the leaves and teeth, as well as the R script used for analysis of both data sets, is available at Zenodo (https://doi.org/10.5281/zenodo.5147936; Oso, 2020).

REFERENCES

Abramoff, M., P. Magalhaes, and S. Ram. 2004. Image processing with ImageJ. *Biophotonics International* 11: 36–43.

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. https://doi.org/10.1111/2041-210X.12035

Agarwal, G., P. Belhumeur, S. Feiner, D. Jacobs, W. J. Kress, R. Ramamoorthi, N. A. Bourg, et al. 2006. First steps toward an electronic field guide for plants. *Taxon* 55: 597–610. https://doi.org/10.2307/25065637

Andrey, P., and Y. Maurin. 2005. Free-D: An integrated environment for three-dimensional reconstruction from serial sections. *Journal of Neuroscience Methods* 145: 233–244. https://doi.org/10.1016/j.jneumeth.2005.01.006

Backhaus, A., A. Kuwabara, M. Bauch, N. Monk, G. Sanguinetti, and A. Fleming. 2010. Leafprocessor: A new leaf phenotyping tool using contour bending energy and shape cluster analysis. *New Phytologist* 187: 251–261. https://doi.org/10.1111/j.1469-8137.2010.03266.x

Beaman, R. S., and N. Cellinese. 2012. Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys* 209: 7–17. https://doi.org/10.3897/zookeys.209.3313

Belhumeur, P., D. Chen, S. Feiner, D. Jacobs, and W. Kress. 2008. Searching the world's herbaria: A system for visual identification of plant species. *In* D. Forsyth, P. Torr, and A. Zisserman [eds.], Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol. 5305. Springer, Berlin, Germany.

Biot, E., M. Cortizo, J. Burguet, A. Kiss, M. Oughou, A. Maugarny-Calès, B. Gonçalves, et al. 2016. Multiscale quantification of morphodynamics: MorphoLeaf software for 2D shape analysis. *Development* 143: 3417–3428. https://doi.org/10.1242/dev.134619

Borges, L. M., V. C. Reis, and R. Izbicki. 2020. Schrödinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Methods in Ecology and Evolution* 11: 1296–1308. https://doi.org/10.1111/2041-210X.13450

Chitwood, D. H., L. R. Headland, R. Kumar, J. Peng, J. N. Maloof, and N. R. Sinha. 2012. The developmental trajectory of leaflet morphology in wild tomato species. *Plant Physiology* 158: 1230–1240. https://doi.org/10.1104/pp.111.192518

Chitwood, D. H., L. L. Klein, R. O'Hanlon, S. Chacko, M. Greg, C. Kitchen, A. J. Miller, and J. P. Londo. 2016a. Latent developmental and evolutionary shapes embedded within the grapevine leaf. *New Phytologist* 210: 343–355. https://doi.org/10.1111/nph.13754

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: giw008. https://doi.org/10.1093/gigascience/giw008

Chitwood, D. H., S. M. Rundell, D. Y. Li, Q. L. Woodford, T. T. Yu, J. R. Lopez, D. Greenblatt, et al. 2016b. Climate and developmental plasticity: Interannual variability in grapevine leaf morphology. *Plant Physiology* 170:1480–1491.

Chitwood, D. H., and N. R. Sinha. 2016. Evolutionary and environmental forces sculpting leaf development. *Current Biology* 26: R297–R306. https://doi.org/10.1016/j.cub.2016.02.033

Chomicki, G., H. Schaefer, and S. S. Renner. 2019. Origin and domestication of Cucurbitaceae crops: Insights from phylogenies, genomics and archaeology. *New Phytologist* 226:1240–1255. https://doi.org/10.1111/nph.16015

Chuanromanee, T. S., J. I. Cohen, and G. L. Ryan. 2019. Morphological Analysis of Size and Shape (MASS): An integrative software program for morphometric analyses of leaves. *Applications in Plant Sciences* 7: e11288. https://doi.org/10.1002/aps3.11288

Cope, J. S., D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin. 2012. Plant species identification using digital morphometrics: A review. *Expert Systems with Applications* 39: 7562–7573. https://doi.org/10.1016/j.eswa.2012.01.073

Corney, D. P. A., H. L. Tang, J. Y. Clark, Y. Hu, and J. Jin. 2012. Automating digital leaf measurement: The tooth, the whole tooth, and nothing but the tooth. *PLoS ONE* 7: e42112. https://doi.org/10.1371/journal.pone.0042112

Dkhar, J., and A. Pareek. 2014. What determines a leaf's shape? *EvoDevo* 5: 47. https://doi.org/10.1186/2041-9139-5-47

Edwards, E. J., D. S. Chatelet, E. L. Spriggs, E. S. Johnson, C. Schlutius, and M. J. Donoghue. 2017. Correlation, causation, and the evolution of leaf teeth: A reply to Givnish and Kriebel. *American Journal of Botany* 104: 509–515. https://doi.org/10.3732/ajb.1700075

Edwards, E. J., E. L. Spriggs, D. S. Chatelet, and M. J. Donoghue. 2016. Unpacking a century-old mystery: Winter buds and the latitudinal gradient in leaf form. *American Journal of Botany* 103: 975–978. https://doi.org/10.3732/ajb.1600129

Ekeke, C., and J. U. Agogbua. 2018. Morphological and anatomical studies on *Trichosanthes cucumerina* L. (Cucurbitaceae). *International Journal of Plant and Soil Science* 25: 1–8. https://doi.org/10.9734/ijpss/2018/44982

Fu, G., X. Dai, J. Symanzik, and S. Bushman. 2017. Quantitative gene–gene and gene–environment mapping for leaf shape variation using tree-based models. *New Phytologist* 213: 455–469. https://doi.org/10.1111/nph.14131

Gallaher, T. J., D. C. Adams, L. Attigala, S. V. Burke, J. M., Craine, M. R. Duvall, P. C. Klahs, et al. 2019. Leaf shape and size track habitat transitions across forest–grassland boundaries in the grass family (Poaceae). *Evolution* 73: 927–946. https://doi.org/10.1111/evo.13722

Guo, J., W. Xu, Y. Hu, J. Huang, Y. Zhao, L. Zhang, C. H. Huang, and H. Ma. 2020. Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. *Molecular Plant* 13: P1117–1133. https://doi.org/10.1016/j.molp.2020.05.011

Hernández-Esquivel, K. B., E. M. Piedra-Malagón, G. Cornejo-Tenorio, L. Mendoza-Cuenca, A. González-Rodríguez, E. Ruiz-Sanchez, and G. Ibarra-Manríquez. 2020. Unraveling the extreme morphological variation in the neotropical *Ficus aurea* complex (subg. *Spherosuke*,

sect. *Americanae*, Moraceae). *Journal of Systematics and Evolution* 58: 263–281. https://doi.org/10.1111/jse.12564

Ichihashi, Y., J. A. Aguilar-Martínez, M. Farhi, D. H. Chitwood, R. Kumar, L. V. Millon, J. Peng, et al. 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proceedings of the National Academy of Sciences, USA* 111: E2616–2621. https://doi.org/10.1073/pnas.1402835111

Itgen, M. W., S. K. Sessions, L. D. Wilson, and J. H. Townsend. 2019. Integrative systematic revision of *Bolitoglossa celaque* (Caudata: Plethodontidae), with a new species from the Lenca Highlands of Honduras. *Herpetological Monographs* 33: 48–70. https://doi.org/10.1655/HERPMONOGRAPHS-D-19-00001.1

Jamil, N., N. A. C. Hussin, S. Nordin, and K. Awang. 2015. Automatic plant identification: Is shape the key feature? *Procedia Computer Science* 76: 436–442. https://doi.org/10.1016/j.procs.2015.12.287

Josephine, A., E. Chimezie, and E. O. Bosa. 2015. Morpho-anatomical characters of *Zehneria capillacea* (Schumach) C. Jeffrey and *Zehneria scabra* (L.F.) Sond Cucurbitaceae. *African Journal of Plant Science* 9: 457–465. https://doi.org/10.5897/ajps2015.1306

Kattge, J., G. Bönisch, S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, S. Tautenhahn, et al. 2020. TRY plant trait database – Enhanced coverage and open access. *Global Change Biology* 26: 119–188. https://doi.org/10.1111/gcb.14904

Kidner, C. A., and S. Umbreen. 2010. Why is leaf shape so variable? *International Journal of Plant Developmental Biology* 4: 64–75.

Klein, L., and H. Svoboda. 2017. Comprehensive methods for leaf geometric morphometric analyses. *Bio-Protocol* 7: e2269. https://doi.org/10.21769/bioprotoc.2269

Klein, L. L., M. Caito, C. Chapnick, C. Kitchen, R. O'Hanlon, D. H. Chitwood, and A. J. Miller. 2017. Digital morphometrics of two North American grapevines (*Vitis*: Vitaceae) quantifies leaf variation between species, within species, and among individuals. *Frontiers in Plant Science* 8: e373. https://doi.org/10.3389/fpls.2017.00373

Klingenberg, C. P. 2011. MorphoJ: An integrated software package for geometric morphometrics. *Molecular Ecology Resources* 11: 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Li, M., H. An, R. Angelovici, C. Bagaza, A. Batushansky, L. Clark, V. Coneva, et al. 2018. Topological data analysis as a morphometric method: Using persistent homology to demarcate a leaf morphospace. *Frontiers in Plant Science* 9: e553. https://doi.org/10.3389/fpls.2018.00553

Lorieul, T., K. D. Pearson, E. R. Ellwood, H. Goëau, J. F. Molino, P. W. Sweeney, J. M. Yost, et al. 2019. Toward a large-scale and deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras. *Applications in Plant Sciences* 7: e1233. https://doi.org/10.1002/aps3.1233

Luca, G., and G. Annamaria. 2019. Effectiveness of modern leaf analysis tools for the morpho-ecological study of plants: the case of *Primula albenensis*. *Nordic Journal of Botany* 37: njb.02386. https://doi.org/10.1111/njb.02386

Manacorda, C. A., and S. Asurmendi. 2018. Arabidopsis phenotyping through geometric morphometrics. *GigaScience* 7: giy073. https://doi.org/10.1093/gigascience/giy073

Migicovsky, Z., M. Li, D. H. Chitwood, and S. Myles. 2018. Morphometrics reveals complex and heritable apple leaf shapes. *Frontiers in Plant Sciences* 8: e2185. https://doi.org/10.3389/fpls.2017.02185

Misra, S., A. K. Srivastava, S. Verma, S. Pandey, S. S. Bargali, T. S. Rana, and K. N. Nair. 2017. Phenetic and genetic diversity in Indian *Luffa* (Cucurbitaceae) inferred from morphometric, ISSR and DAMD markers. *Genetic Resources and Crop Evolution* 64: 995–1010. https://doi.org/10.1007/s10722-016-0420-1

Morello, S., A. B. Sassone, and A. López. 2018. Leaflet shape in the endemic South American *Oxalis* sect. *Alpinae*: An integrative approach using molecular phylogenetics and geometric morphometrics. *Perspectives in Plant Ecology, Evolution and Systematics* 35: 22–30. https://doi.org/10.1016/j.ppees.2018.09.003

Nicotra, A. B., A. Leigh, C. K. Boyce, C. S. Jones, K. J. Niklas, D. L. Royer, and H. Tsukaya. 2011. The evolution and functional significance of leaf shape in the angiosperms. *Functional Plant Biology* 38: 535–552. https://doi.org/10.1071/FP11057

Oso, O. 2020. Seven Curcubitaceae leaves MorphoLeaf morphometric data and statistical code. Available at Zenodo repository: https://doi.org/10.5281/ZENODO.5147936 [posted 2 November 2020; accessed 2 November 2020].

Ostria-Gallardo, E., A. Ranjan, D. H. Chitwood, R. Kumar, B. T. Townsley, Y. Ichihashi, L. J. Corcuera, and N. R. Sinha. 2016. Transcriptomic analysis suggests a key role for *SQUAMOSA PROMOTER BINDING PROTEIN LIKE*, *NAC* and *YUCCA* genes in the heteroblastic development of the temperate rainforest tree *Gevuina avellana* (Proteaceae). *New Phytologist* 210: 694–708. https://doi.org/10.1111/nph.13776

Page, L. M., B. J. MacFadden, J. A. Fortes, P. S. Soltis, and G. Riccardi. 2015. Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience* 65: 841–842. https://doi.org/10.1093/biosci/biv104

Pérez-Miranda, F., O. Mejía, A. A. González-Díaz, N. Martínez-Méndez, E. Soto-Galera, G. Zúñiga, and O. Říčan. 2020. The role of head shape and trophic variation in the diversification of the genus *Herichthys* in sympatry and allopatry. *Journal of Fish Biology* 96: 1370–1378. https://doi.org/10.1111/jfb.14304

Punyasena, S. W., and S. Y. Smith. 2014. Bioinformatic and biometric methods in plant morphology. *Applications in Plant Sciences* 2: e1400071. https://doi.org/10.3732/apps.1400071

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: http://www.R-project.org/ [accessed 21 November 2020].

Renner, S., and A. Pandey. 2013. The Cucurbitaceae of India: Accepted names, synonyms, geographic distribution, and information on images and DNA sequences. *PhytoKeys* 20: 53–118. https://doi.org/10.3897/phytokeys.20.3948

Rohlf, F. J. 2015. The tps series of software. *Hystrix* 26: 9–12. https://doi.org/10.4404/hystrix-26.1-11264

Schlautman, B., L. Diaz-Garcia, and S. Barriball. 2020. Morphometric approaches to promote the use of exotic germplasm for improved food security and resilience to climate change: A kura clover example. *Plant Science* 290: e110319. https://doi.org/10.1016/j.plantsci.2019.110319

Soltis, P. S. 2017. Digitization of herbaria enables novel research. *American Journal of Botany* 104: 1281–1284. https://doi.org/10.3732/ajb.1700281

Soltis, P. S., G. Nelson, and S. A. James. 2018. Green digitization: Online botanical collections data answering real-world questions. *Applications in Plant Sciences* 6: e1028. https://doi.org/10.1002/aps3.1028

Terhune, C. E., A. D. Sylvester, J. E. Scott, and M. J. Ravosa. 2020. Internal architecture of the mandibular condyle of rabbits is related to dietary resistance during growth. *Journal of Experimental Biology* 223: jeb220988. https://doi.org/10.1242/jeb.220988

Viscosi, V., O. Lepais, S. Gerber, and P. Fortini. 2009. Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: A comparison of traditional and geometric morphometric methods. *Plant Biosystems* 143: 564–574. https://doi.org/10.1080/11263500902723129

Wäldchen, J., M. Rzanny, M. Seeland, and P. Mäder. 2018. Automated plant species identification—Trends and future directions. *PLoS Computer Biology* 14: e1005993. https://doi.org/10.1371/JOURNAL.PCBI.1005993

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, New York, USA. https://doi.org/10.1007/978-0-387-98141-3

Willis, C. G., E. R. Ellwood, R. B. Primack, C. C. Davis, K. D. Pearson, A. S. Gallinat, J. M. Yost, et al. 2017. Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology and Evolution* 32: 531–546. https://doi.org/10.1016/j.tree.2017.03.015

Zhang, L. B., M. P. Simmons, A. Kocyan, and S. S. Renner. 2006. Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system

evolution. *Molecular Phylogenetics and Evolution* 39: 305–322. https://doi.org/10.1016/j.ympev.2005.10.002

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Appendix S1. Scatterplot of the Cucurbitaceae teeth data set.