

Comprehensive Methods for Leaf Geometric Morphometric Analyses

Laura L. Klein^{1,*} and Harlan T. Svoboda^{2,*}

¹Department of Biology, Saint Louis University, St. Louis, Missouri, USA; ²Department of Environmental and Plant Biology, Ohio University, Athens, Ohio, USA

*For correspondence: lklein10@slu.edu; hs923311@ohio.edu

[Abstract] Leaf morphometrics are used frequently by several disciplines, including taxonomists, systematists, developmental biologists, morphologists, agronomists, and plant breeders to name just a few. Leaf shape is highly variable and can be used for identifying species or genotypes, developmental patterning within and among individuals, assessing plant health, and measuring environmental impacts on plant phenotype. Traditional leaf morphometrics requires hand tools and access to specimens, but modern efforts to digitize botanical collections make digital morphometrics a readily accessible and scientifically rigorous option. Here we provide detailed instructions for performing some of the most informative digital geometric morphometric analyses available: generalized Procrustes analysis, elliptical Fourier analysis, and shape features. This comprehensive procedure for leaf shape analysis is comprised of six main sections: A) scanning of material, B) acquiring landmarks, C) analysis of landmark data, D) isolating leaf outlines, E) analysis of leaf outlines, and F) shape features. This protocol provides a detailed reference for applying landmark and outline analysis to leaf shape as well as describing leaf shape features, thus empowering researchers to perform high throughput phenotyping for diverse applications.

Keywords: Digital leaf morphometrics, Landmarks, Generalized Procrustes Analysis, Leaf outlines, Elliptic Fourier Descriptors, Shape features, Aspect ratio, Circularity

[Background] There are a variety of approaches to digital leaf shape morphometrics, including outline or Fourier analysis, contour signatures, landmark analysis, shape features, fractal dimensions, and texture analysis (Cope *et al.*, 2012). Among these analyses, landmark and Fourier analysis together perform exceptionally well at distinguishing between groups among leaf shapes (McLellan and Endler, 1998; Hearn, 2009). Landmark analysis is ideal for capturing aspects of shape that are consistent among all leaves within a given dataset. The selection of landmarks should include points that are biologically homologous and adequately represent the morphology of the leaf (see more pointers in Bookstein, 1991 and Zelditch *et al.*, 2004). If leaves in the dataset do not have evolutionarily conserved shape features, ‘pseudo-landmarks’ can instead be placed (Chitwood and Sinha, 2016); that is, landmarks can be placed at equidistant points along the leaf outline relative to homologous points that act as anchors. Landmarks can then be analyzed using Generalized Procrustes Analysis (GPA), which normalizes shape data (annotated by landmarks) at equal scale, allowing for an accurate comparison of shapes regardless of their size. Outline analysis offers a more broadly applicable phenotyping method in that Elliptical Fourier Descriptors (EFDs) are used to build shape descriptors of the leaf

outline (Kuhl and Giardina, 1982; Iwata and Ukai, 2002). While sensitive to noise, EFDs are ideal for large leaf datasets that have subtle differences between shapes. Shape features are an additional, simple method of outline analysis that can include the perimeter to area ratio, aspect ratio, and circularity measurements (Cope *et al.*, 2012). In this protocol, we focus on aspect ratio and circularity, as they detect signatures of lobing and serration. Aspect ratio is the ratio of the major axis to the minor axis of a fitted ellipse, in which case values close to '1' are more circular in shape regardless of lobing. Circularity is the ratio of the leaf area to perimeter outline. This measurement is useful for discriminating leaves with lobing and serration, with low circularity values indicating significant lobing and serration. This protocol is designed such that researchers can choose between all three methods (GPA, EFD, and shape features) based on which analyses best fit their data.

Materials and Reagents

1. Herbarium specimens and/or fresh leaf material

Equipment

1. Computer that can run Microsoft® Windows® XP (or later) and/or Mac® OS X® 10.4 (or later)
2. Flatbed photo scanner (Epson Expression, model: 10000XL)

Software

1. Adobe® Photoshop® CS4 (or later)
2. Epson® Scan Utility v3.4.9.6 (<https://support.epson.com/>)
3. Java™ (<https://java.com/>)
4. ImageJ (<https://imagej.nih.gov/ij/>)
5. Microsoft® Excel® 2011 (or later)
6. SHAPE v1.3 (<http://lmb.ab.a.u-tokyo.ac.jp/~iwata/shape/>)
SHAPE is built for Windows but if using a Mac, Wine and Winebottler (<http://winebottler.kronenberg.org/>) are required
7. R (<https://r-project.org/>)
Packages: 'shapes,' 'ggplot2,' 'devtools,' 'ellipse,' and 'roxygen2'
8. RStudio (<https://rstudio.com/products/rstudio/>)
RStudio is an optional user interface for R

Procedure

Note: Examples of R scripts and ImageJ macros referenced throughout the protocol can be freely downloaded from GitHub (link: <https://github.com/htsvoboda/LeafGeometricMorphometrics.git>; Note 1).

A. Scanning fresh leaves or herbarium specimens

1. For fresh leaves: Place leaves, with petioles removed, flat on the scanner bed. Multiple leaves can be placed on the scanning bed at once, so long as they do not overlap. If the background of the scanner is not already white, place a white piece of paper on top of the leaves; subsequent analyses work best with a solid white background (Note 2).
2. For herbarium specimens: carefully place the entire sheet face-down on the scanner bed. Any loose material should be placed in a fragment packet on the sheet.
3. Scans should be saved as .jpg or .tif files and named with a unique identifier (e.g., accession number).

Note: It is best to scan at 400 dpi (or a higher resolution) as this improves the quality of the image.

B. Acquiring and exporting landmarks

1. Create a spreadsheet in Excel with the following column names: 'order,' 'label,' 'x,' and 'y.' This file will be referred to as the 'master spreadsheet.'
2. Open ImageJ
For the first use: Select 'Analyze > Set Measurements...' and check only the 'Display label' checkbox. 'Redirect' and decimal place parameters can be left at their defaults.
3. From the ImageJ menu bar, select the 'Point selection' tool (Figure 1; Note 3).



Figure 1. The ImageJ menu bar. Here, the 'Point selection' tool button is selected.

4. Open the first image to be landmarked. In the menu bar, select 'File > Open...'
5. Using the 'Point selection' tool in multi-point mode, begin placing landmarks on your predetermined landmark points (see Figure 2; Notes 4 and 5).

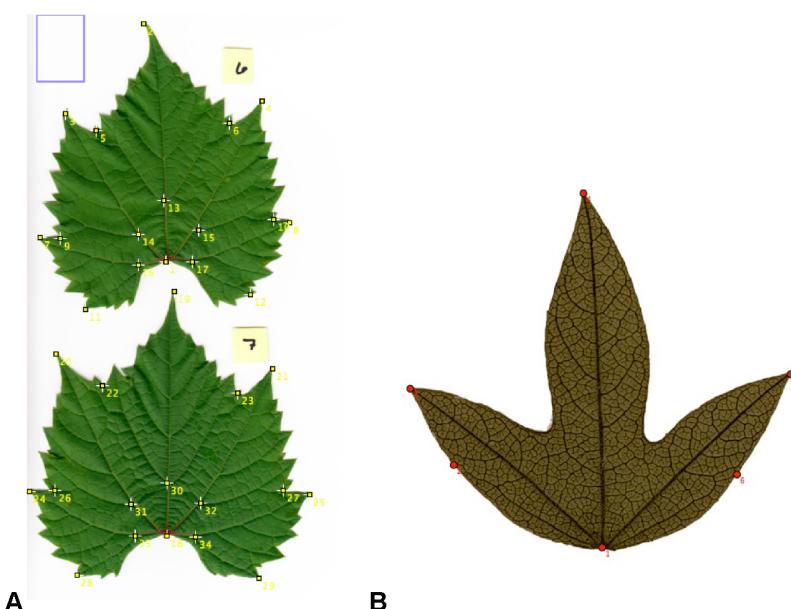


Figure 2. Placement of landmarks on some representative leaves. A. Seventeen landmarks placed on two fresh leaves of *Vitis riparia*; B. Six landmarks placed on a leaf of *Passiflora campanulata* from a herbarium specimen.

- Once all landmarks have been applied to a leaf (or to multiple leaves per image), view the landmark data. In the menu bar, select 'Analyze > Measure.' A new window, called 'Results' (Figure 3), will appear with the x, y coordinates for each landmark.

Results		
Label	X	Y
1	804	1464
2	666	114
3	396	702
4	1164	672
5	228	600
6	1350	534
7	174	1302
8	1398	1212
9	60	1338
10	1506	1224
11	330	1734
12	1284	1644
13	780	1104
14	630	1308
15	954	1266
16	624	1464
17	948	1440
18	792	3006
19	840	1632
20	444	2148
21	1206	2214
22	174	1992
23	1386	2058
24	132	2748
25	1476	2754
26	1.C1B.b.jpg	36
27	1.C1B.b.jpg	2748
28	1.C1B.b.jpg	1608
29	288	2766
30	1.C1B.b.jpg	3234
31	1326	3270
32	810	2700
33	1.C1B.b.jpg	594
34	1.C1B.b.jpg	2820
35	1.C1B.b.jpg	984
36	1.C1B.b.jpg	2820
37	1.C1B.b.jpg	600
38	1.C1B.b.jpg	3012
39	1.C1B.b.jpg	966

Figure 3. Example of the 'Results' window produced by ImageJ. This is viewed by selecting 'Analyze > Measure' in the menu bar.

7. Copy (CTRL + C) and paste (CTRL + V) these data from the ‘Results’ window directly into the ‘master spreadsheet’ (see Figure 4).

	A	B	C	D	
1	order	Label	X	Y	
2	1	78.1.a.jpg	366	767	
3	2	78.1.a.jpg	936	1051	
4	3	78.1.a.jpg	674	787	
5	4	78.1.a.jpg	532	1039	
6	5	78.1.a.jpg	770	743	
7	6	78.1.a.jpg	566	1117	
8	7	78.1.a.jpg	542	633	
9	8	78.1.a.jpg	326	973	
10	9	78.1.a.jpg	570	585	
11	10	78.1.a.jpg	302	1025	
12	11	78.1.a.jpg	360	625	
13	12	78.1.a.jpg	230	823	
14	13	78.1.a.jpg	474	841	
15	14	78.1.a.jpg	452	755	
16	15	78.1.a.jpg	390	861	
17	16	78.1.a.jpg	404	721	
18	17	78.1.a.jpg	340	811	
19	18	78.1.a.jpg	494	1629	
20	19	78.1.a.jpg	1530	1637	
21	20	78.1.a.jpg	968	1915	
22	21	78.1.a.jpg	960	1367	
23	22	78.1.a.jpg	1108	1255	
24	23	78.1.a.jpg	1074	2009	
25	24	78.1.a.jpg	602	1275	
26	25	78.1.a.jpg	610	2007	
27	26	78.1.a.jpg	604	1195	
28	27	78.1.a.jpg	600	2069	
29	28	78.1.a.jpg	318	1469	
30	29	78.1.a.jpg	334	1815	
31	30	78.1.a.jpg	692	1631	
32	31	78.1.a.jpg	606	1517	
33	32	78.1.a.jpg	618	1773	
34	33	78.1.a.jpg	484	1551	
35	34	78.1.a.jpg	488	1737	
36	35	78.1.a.jpg	1116	2726	

Figure 4. Example format of the ‘master spreadsheet’

8. Close the ‘Results’ and image windows and repeat this process with each image file until all leaves have been landmarked. ‘File > Open Next’ can be used if all of the images are in the same folder.
9. Landmark coordinates for every leaf will be pasted as-is into the ‘master spreadsheet’. The column ‘order’ may need to be adjusted at the end to reflect a continuous numerical set.
10. Before analyzing the landmark data, the ‘master spreadsheet’ will need to be converted from Excel format (.xlsx) to a Tab Delimited Text format (.txt).
 - a. Click ‘File > Save As...’ and select the .txt option.
11. Data within the .txt file will need to be reformatted such that the rows are single leaves and the adjacent columns represent the landmark data (x, y coordinates) in sequential order. This can be done in R using the code from GitHub (file: ‘Protocol_stepB-11.R’; R Core Team, 2017).

12. Import the reformatted file (referred to as ‘reformatted.txt’) into Excel to verify the data was properly written.
 - a. Click ‘File > Import...’ and choose ‘space’ and/or ‘tab’ as the delimiter (Note 6).
13. Supplemental columns and information (*i.e.*, species, genotype, sex, *etc.*) can be added at this point to help with downstream analyses (Figure 5).

	A	B	C	D	E	F	G	H	I	J	K	
1	order	species	Pl.number	genotype	sex	location	label	x1	y1	x2	y2	x3
2	1	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	312	606	324	96	
3	2	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	444	1344	492	768	
4	3	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	486	2142	486	1572	
5	4	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	468	3294	432	2664	
6	5	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	1590	1236	1566	102	
7	6	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.a.jpg	1626	2886	1488	1434	
8	7	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.b.jpg	1062	1344	1260	174	
9	8	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.b.jpg	1104	2910	1026	1674	
10	9	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.c.jpg	1044	1410	1158	246	
11	10	Vitis_riparia	588653.3	Okoboji	Female	C10A	2.C10A.c.jpg	1374	2742	1248	1908	
12	11	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.a.jpg	354	798	450	78	
13	12	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.a.jpg	444	1914	576	930	
14	13	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.a.jpg	918	3150	966	1908	
15	14	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.a.jpg	1596	1698	1446	252	
16	15	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.b.jpg	834	1632	690	120	
17	16	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.b.jpg	1536	3276	1536	1968	
18	17	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.c.jpg	660	1314	606	78	
19	18	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.c.jpg	642	2766	576	1662	
20	19	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.c.jpg	1716	2652	1596	1128	
21	20	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.d.jpg	762	1368	810	144	
22	21	Vitis_riparia	588354.7	B_75	Male	C10B	2.C10B.d.jpg	1452	2718	1458	1794	
23	22	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.a.jpg	261	693	183	210	
24	23	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.a.jpg	360	1680	492	933	
25	24	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.a.jpg	516	2943	438	1953	
26	25	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.a.jpg	1623	1224	1763	200	
27	26	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.a.jpg	1799	2612	1783	1792	
28	27	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.b.jpg	676	1220	492	172	
29	28	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.b.jpg	584	2979	536	1947	
30	29	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.b.jpg	1832	1519	1852	427	
31	30	Vitis_riparia	588439.4	GVIT_775	Female	C1A	2.C1A.b.jpg	1880	2823	1772	1887	
32	31	Vitis_riparia	588653.4	Okoboji	Female	C1B	2.C1B.a.jpg	372	705	456	150	
33	32	Vitis_riparia	588653.4	Okoboji	Female	C1B	2.C1B.a.jpg	393	1680	447	957	
34	33	Vitis_riparia	588653.4	Okoboji	Female	C1B	2.C1B.a.jpg	633	3006	852	2094	
35	34	Vitis_riparia	588653.4	Okoboji	Female	C1B	2.C1B.a.jpg	1861	1416	1618	612	

Figure 5. The reformatted ‘master spreadsheet’ (‘reformatted.txt’) as seen in Excel. The spreadsheet now displays each row as a leaf with its landmarks distributed across the columns. Additional information can be added to the reformatted spreadsheet that may be useful in downstream analyses (*e.g.*, columns B-F).

14. Check landmarks for accuracy. This can be done by plotting the landmark coordinates in R with the package ‘ggplot2’ (Figure 6; Wickham, 2009). The corresponding R script can be downloaded from GitHub (file: ‘Protocol_stepB-14.R’).

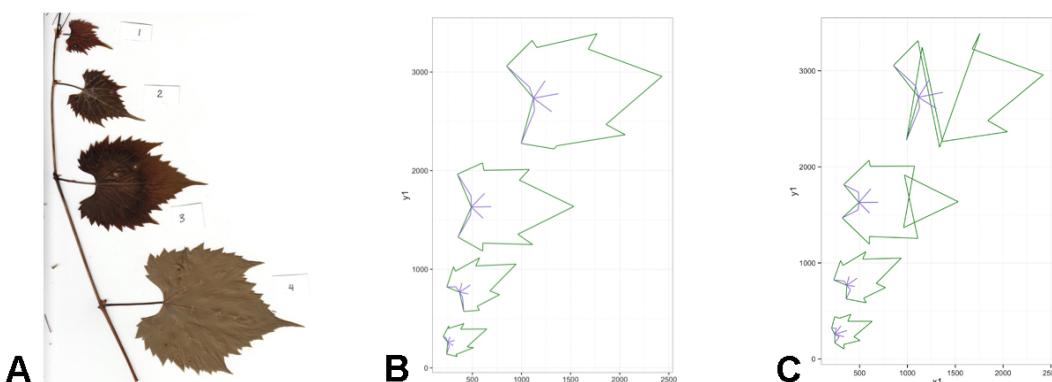


Figure 6. Landmarks are checked for accuracy using the R package ggplot. A. Example of specimen scan; B. The companion plotted landmark coordinates to the scan (A); C. Incorrectly placed landmarks will be immediately apparent visually. Note that (B) and (C) are inverted to that of (A), as pixel coordinates map inversely in a linear regression.

15. If there appears to be an obvious error in the position of the landmarks for any image (as in Figure 6C), redo landmark placement (step B5) for the affected leaf/file and paste the updated coordinates into the ‘master spreadsheet.’ Repeat steps B11-B12 to reformat and check landmark accuracy for the new spreadsheet.

C. Analysis of landmarks: Generalized Procrustes Analysis (GPA)

1. To perform GPA using the R package ‘shapes,’ (Dryden, 2017) the input file must consist of only landmark data.
 - a. Import (‘File > Import...’) the ‘reformatted.txt’ file into Excel.
 - b. Remove all column headers and any other column information.
 - c. In the menu bar, select ‘File > Save As...’ and rename the file to distinguish that it contains only x, y coordinates (e.g., ‘reformatted_coords.txt’).
2. The data can now be processed in R using the ‘shapes’ package.
 - a. The analysis produces Procrustes principal component scores and percent variance explained, Eigenleaves, Eigenvalues, among other informative outputs. The R script includes code that will write files containing PC scores and percentages for further analysis. Example R script can be found on GitHub (file: ‘Protocol_stepC-2-a.R’).
 - b. Because leaf order is preserved in the output files, any additional information (e.g., individual leaf identity, species, genotype, etc.) can be re-added to these output files by adding additional columns to the file, then pasting the additional information for each leaf from the ‘reformatted.txt’ file into the PC score file.
 - c. The leaves can be visualized in morphospace, using the packages ‘ggplot2,’ ‘devtools,’ (Wickham and Chang, 2015) ‘ellipse,’ (Murdoch *et al.*, 2007) and ‘roxygen2’ (Wickham *et al.*, 2015). Example code can be found on GitHub (file: ‘Protocol_stepC-2-c.R’). An example ordination can be seen in Figure 7.

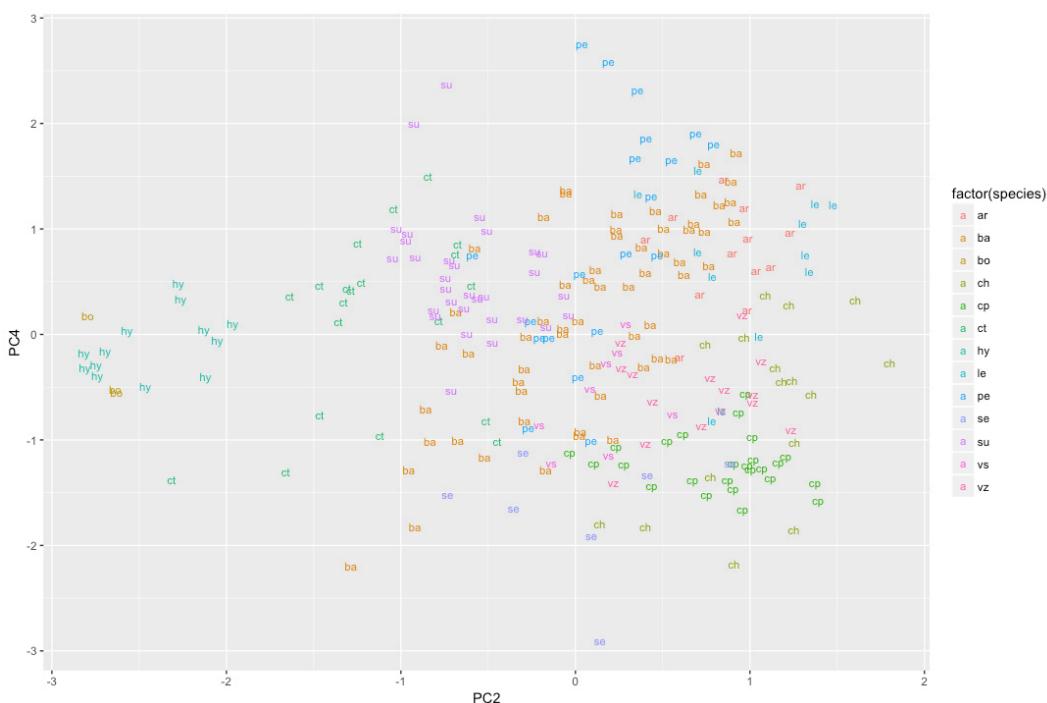


Figure 7. PCA ordination resulting from a Generalized Procrustes Analysis using an example landmark dataset. In this example, the leaves were labeled and color-coded according to the species identity.

D. Isolating leaves from scans

1. For scans of fresh leaves:
 - a. Open ImageJ.
 - b. Create a macro to isolate leaf images as binary images.
 - i. In the menu bar, select 'Plugins > New > Macro.'
 - ii. A new window will appear titled 'Macro.txt.' Input the text from the script found on GitHub (file: 'Protocol_stepD-1-b-ii.txt') and leave this window open. We will refer to this macro as the 'multiple open' macro (Note 7).
 - c. Next, create a second macro that will select individual leaves from a scan.
 - i. With the ImageJ toolbar as the active window, navigate to the menu bar and select 'Plugins > New > Macro.'
 - ii. A second window will appear titled 'Macro.txt.' Input the text from the script found on GitHub (file: 'Protocol_stepD-1-c-ii.txt'). Leave this macro window open. We will refer to this macro as the 'select' macro.
 - d. Open the first image to be analyzed.
 - i. With the ImageJ toolbar as the active window, navigate to the menu bar and select 'File > Open.'
 - e. Execute the 'multiple open' macro by highlighting all code line and clicking either '⌘ + R' (for a Mac) or 'CTRL + R' (for a PC) (also see Note 8).

- f. The leaves within the image should now be thresholded in black and white.
- g. Isolate single leaves to extract and save as separate files by selecting the 'Wand (tracing)' tool from the ImageJ menu toolbar (Figure 8).



Figure 8. The ImageJ menu toolbar. Here, the 'Wand (tracing)' tool is selected.

- i. Select an individual leaf by clicking on it and confirm its outline is properly highlighted.
 - ii. Highlight all code lines and execute the 'select' macro.
 - iii. A prompt will appear to save a binary image of a single leaf. Name the file appropriately and save in a different folder with other binary .jpeg files generated during this process for later use.
 - h. Repeat steps D1a-D1h until all leaves in the image have been isolated.
 - i. When each leaf in the open file has been successfully converted to binary images, add the text "run("Open Next");" as the first line of script in the 'multiple open' macro.
 - j. Highlight all code lines, including the new line, then execute. The addition of the 'Open Next' command will now open the next image file in the folder.
 - k. A prompt will appear to save the changes made to the current open, binary image. Do not save changes, as this will alter the original image scan.
 - l. The next scan will open. Repeat the steps in D1 until all files have been converted to binary images of one leaf per image.
2. For scans of herbarium specimens:
 - a. Open Photoshop CS4.
 - b. If the 'Tools' menu bar (Figure 9) is not already on the main screen, manually open it by selecting 'Window > Tools.'

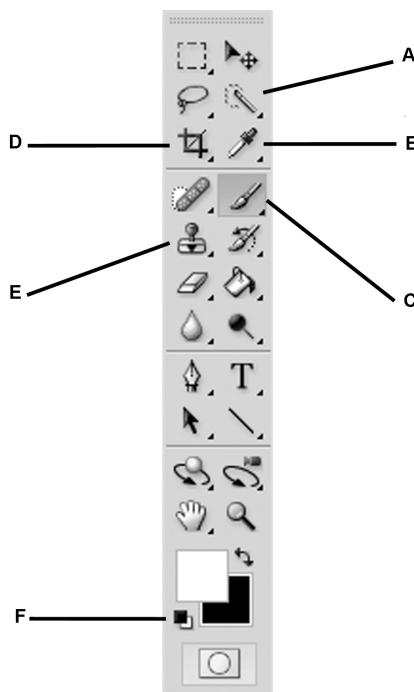


Figure 9. The ‘Tools’ menu in Photoshop. A. Quick Selection Tool; B. Eyedropper Tool; C. Brush Tool; D. Crop Tool; E. Clone Stamp Tool; F. Default Foreground/Background Color Tool.

- c. From the menu, select ‘File > Open...’ and select the first scan to be processed.
- d. Identify an appropriate leaf (*i.e.*, one that is in good condition, flat, and does not have many [*or any*] structures intersecting it).
- e. If the leaf is not in an upright position, rotate it using ‘Image > Image Rotation’ until the apex of the leaf is pointing upward and the base or petiole is pointing downward.
- f. Use the ‘Crop Tool’ (Figure 9D) to isolate this leaf from most of its surroundings, but leaving a buffer around each side of the leaf.
- g. If any structures (*i.e.*, stems, tendrils, flowers, other leaves, *etc.*) intersect the leaf of interest, these must be removed for further analyses.
 - i. Click on the ‘Default Foreground/Background Color Tool’ button (Figure 9F), making sure that ‘white’ is indicated in the foreground box (as seen in Figure 9).
 - ii. Select the ‘Brush Tool’ (Figure 9C) and ensure that under the Options Menu (‘Window > Options’) the ‘Mode’ is set to ‘Normal’ and the ‘Opacity’ is set to ‘100%’.
 - iii. Using the ‘Brush Tool’, select an appropriate brush diameter and then paint over any structures or tissues that intersect with the leaf of interest (Figure 10).

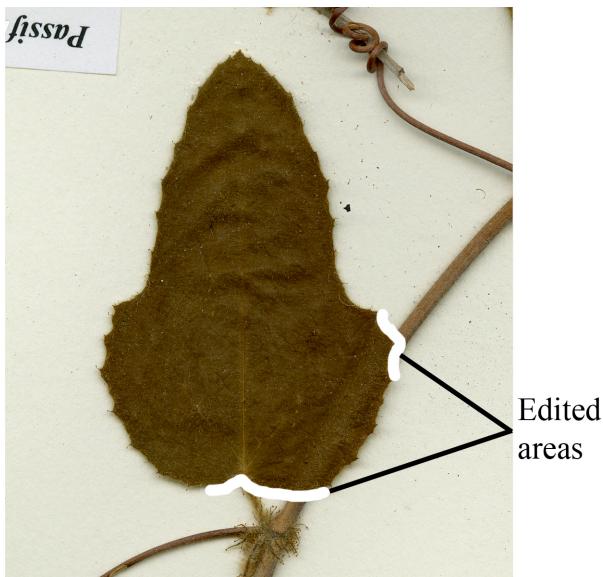


Figure 10. An example of areas edited out using the ‘Brush Tool.’ By using this tool in Photoshop, intersecting structures have been painted over to remove them.

- h. Click on the ‘Quick Selection Tool’ (Figure 9A). Left-click on the leaf and drag along the blade until the entire lamina, and only the lamina, is outlined.
- i. Right-click in the selected area and then click ‘Select Inverse’ from the new menu.
- j. Switch back to the ‘Brush Tool’ and increase the brush diameter. Paint over the background (with white) until only the leaf remains.
- k. Right-click the image and select ‘Select Inverse’ again.
- l. Any holes, tears, or anomalies on the leaf surface need to be filled in so as to match the color of the lamina. This can be done using either the ‘Clone Stamp Tool’ (Figure 9E) or the ‘Eyedropper Tool’ (Figure 9B).
 - i. To use the ‘Clone Stamp Tool,’ hold the Alt/Option key and click an intact area on the leaf. Now use the tool to fill in any damaged spots.
 - ii. To use the ‘Eyedropper Tool,’ left-click on an intact area of the leaf near the damaged area to extract the color. Switch to the ‘Brush Tool’ and paint over the damaged area(s) to match the color of the rest of the leaf.
- m. In some cases, especially when using .tif files, it may be necessary to ‘flatten’ the layers of the image by clicking ‘Layers > Flatten Image’ before saving. This will merge all of the edits into a single, savable image.
- n. Save each leaf (‘File > Save as...’) with a slightly different filename, but still keeping track of the original scan that it came from (e.g., ‘MO1624745_leaf1,’ ‘MO1624745_leaf2,’ etc.).
- o. Repeat steps D2a-D2n for multiple leaves per scan, and for multiple scans.
- p. To convert isolated leaf outlines to black and white images for SHAPE analysis (step D3), follow steps D1a-D1f.

3. The program SHAPE (Iwata and Ukai, 2002) uses binary leaf outline image files in BMP format. Leaf images can be converted in ImageJ by creating a batch macro.
 - a. Open ImageJ.
 - b. Create a macro to easily convert many images at once.
 - i. From the ImageJ menu click ‘Process > Batch > Macro...’
 - ii. Select the appropriate ‘Input’ folder containing the images to be converted.
 - iii. Select an ‘Output’ folder to contain the new binary images.
 - iv. Choose ‘BMP’ from the ‘Output Format’ drop-down menu.
 - v. Input the text from the script found on GitHub (file: ‘Protocol_stepD-3-b-v.txt’) into the large blank space provided.
 - c. View the output folder to check that the .bmp files were properly converted (Note 9).

E. Analysis of outlines: Elliptical Fourier Descriptors (EFDs)

1. Use the SHAPE software to convert image outlines to chain code. We encourage users to become familiar with the SHAPE User Manual in order to better explore parameter choice.
 - a. Open the ChainCoder program within SHAPE.
 - i. Select ‘Files > Select Image File(s).’
 - ii. In the following window (Figure 11), select the folder of BMP files. The BMP files will then appear in the field ‘File(s).’
 - iii. Select all images and select the double arrow button to transfer the desired files into the ‘Selected File(s)’ field. Press ‘OK.’

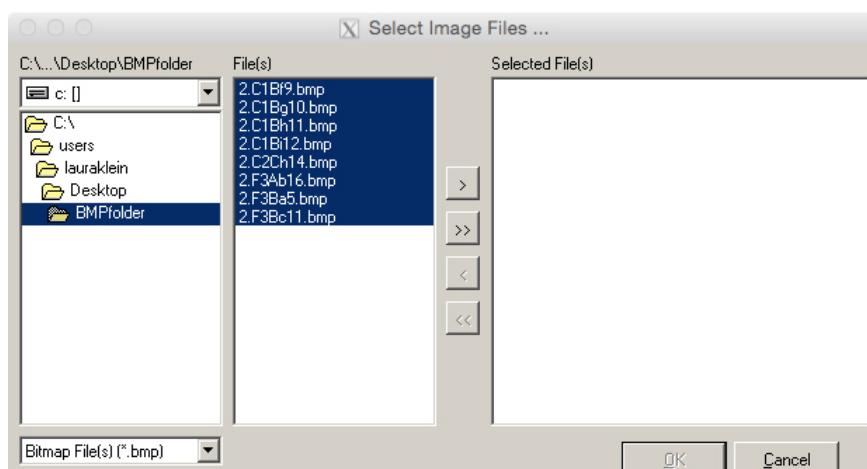


Figure 11. Selection of .bmp files for chain coding in the software SHAPE. Use the double arrow button to move all files to the ‘Selected File(s)’ field.

- b. Before beginning the analyses, select the ‘Config’ tab (Figure 12).
 - i. Set ‘Object Color’ to ‘Dark (Black),’ and ‘Scale Included’ to ‘No.’ Leave the other fields at their defaults.

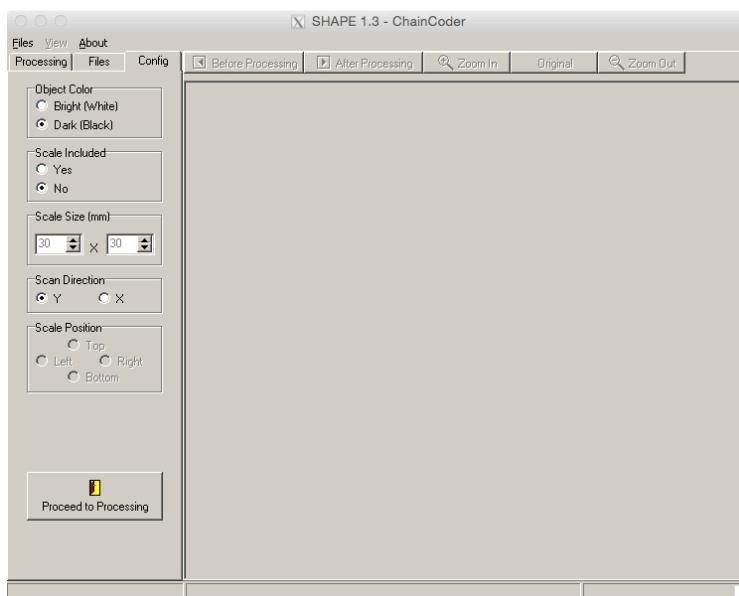


Figure 12. The ‘Config’ tab of the ChainCoder program. This allows for configuration of parameters before beginning the chain coding process.

- c. Select the ‘Processing’ tab to begin processing photos.
 - i. Select ‘Load Image’ (Note 10).
 - ii. Deselect the ‘Select Area’ box.
 - iii. Select ‘Gray Scale.’
 - iv. Select ‘Make Histogram.’
 - v. Select ‘Binarize Image.’
 - vi. Check the ‘Ero Dil Filter’ and ‘Dil Ero Filter’ boxes and select to what degree to ‘erode’ and ‘dilate’ the outline; this option adds and subtracts the amount of pixels from the image to give a smoother border.
 - vii. Select the ‘Labeling Object’ button. A new window named ‘Chain Code Data’ will appear, allowing the user to indicate which objects, over a certain number of pixels, should be isolated for analysis.
 - viii. Select ‘Chain Coding.’ This will add the chain code to the user selection.
 - ix. Select ‘Save to File.’ For the first image, you will have to name the file. Subsequently, chain codes for further images will be saved as processed (Note 11).
- d. Repeat steps E1a-E1c for all further images. If many images need to be processed, hold down the ‘enter’ key (or put a weight on it until the image processing is finished).
2. Convert chain code to normalized EFDs.
 - a. Open the CHC2NEF program within SHAPE. This program converts the chain code created above to normalized EFDs.
 - b. A new window will appear (Figure 13). Select the chain code file produced in step E1c.ix (‘CHC File Name’), then create a name for the new NEF file that will be generated in the following steps (‘NEF File Name’).

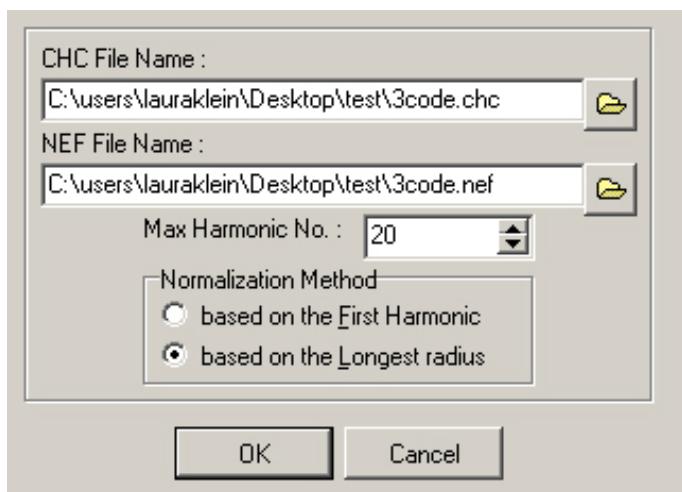


Figure 13. The CHC2NEF window. Select the chain code input file (.chc) and name the resulting normalized elliptical Fourier descriptors file (.nef).

- c. Set the 'Max Harmonic No.' Higher harmonic numbers lead to better shape approximations, but usually 20 is sufficient to recapitulate leaf shape accurately.
- d. Select the Normalization Method to be 'based on the longest radius.' This is the way the image will initially be oriented, and this option allows better subsequent manipulation to align properly.
- e. Click 'OK.'
- f. A new window will appear. Click 'Start !!'
- g. Orient the image so that all images are similarly aligned. This is an arbitrary designation, but needs to be consistent among all images.
 - i. There are a number of buttons to assist in this process (see Figure 14). The image can be turned by any number of degrees in either direction, and arrow buttons can be adjusted by rotation degree. Alternatively, images can rotate by 180°.

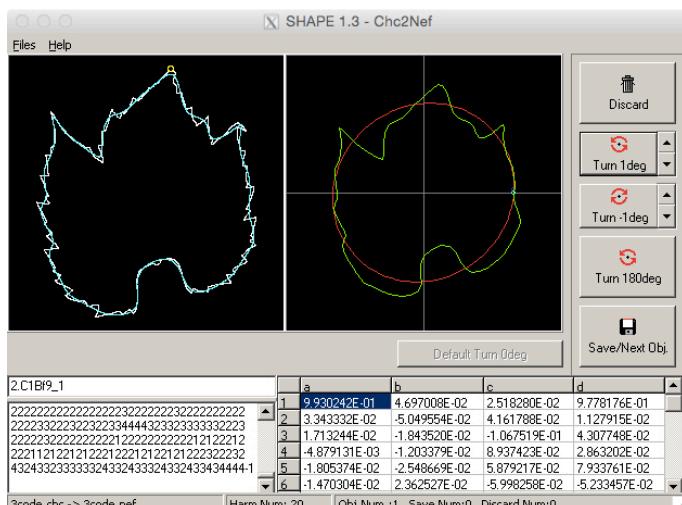


Figure 14. Orienting leaf image chain code to normalized elliptical Fourier descriptors.

Utilize the three turning buttons at the right side of the screen to orient the leaf image appropriately. NEF code will reflect these changes (bottom of window).

- ii. Once aligned, click ‘Save/Next Obj.’ and repeat until all images have been normalized.
3. Visualize Elliptical Fourier Descriptors (EFDs) using Principal Component Analysis (PCA). These analyses can be done within the SHAPE program itself (see below; Note 12).
 - a. Open the ‘PrinComp’ program within SHAPE.
 - b. A new window will appear. In the menu bar, select ‘Files > Open Nef File’ and select the desired .nef file.
 - c. An additional window will appear (Figure 15) with parameters to determine for the PCA.

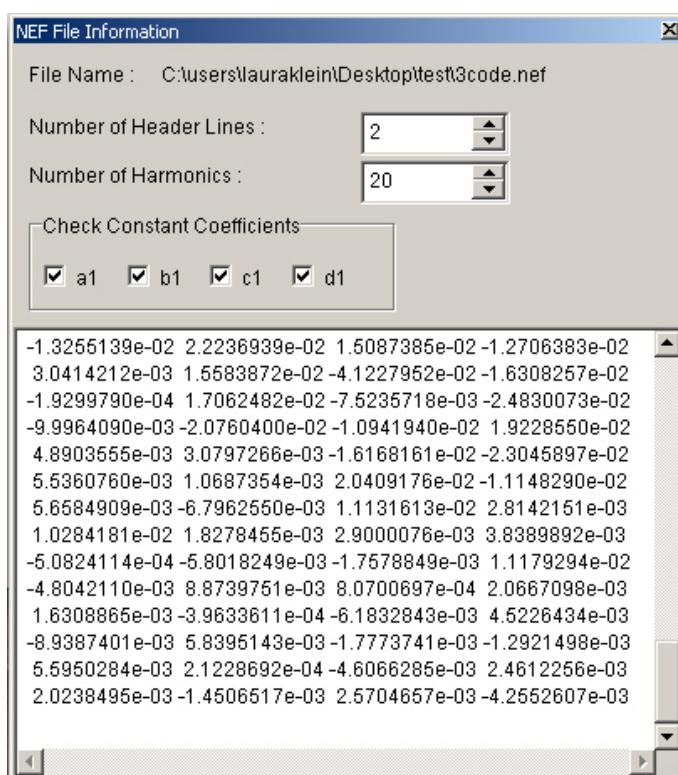


Figure 15. The ‘NEF File Information’ window. Desired PCA parameters can be set for the normalized Elliptical Fourier Descriptors.

- i. ‘Number of Header Lines’ is automatically set according to the NEF file.
- ii. Select an appropriate number of harmonics on which to perform the PCA (the default is 20).
- iii. Select the desired coefficients to keep constant.
 - 1) To analyze both symmetric and asymmetric variance: select ‘a-d.’
 - 2) To analyze symmetric variance: select ‘a’ and ‘d.’
 - 3) To analyze asymmetric variance: select ‘b’ and ‘c.’
- d. Perform the PCA by clicking the ‘Principal Component Analysis’ button (see Figure 16).

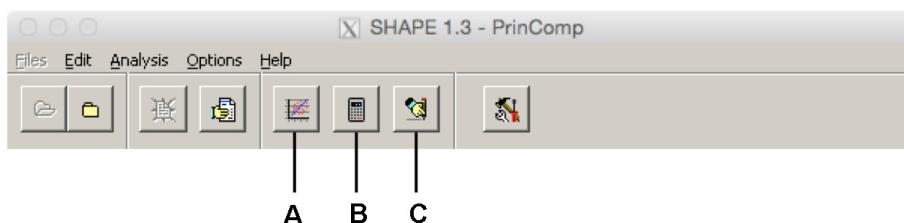


Figure 16. The PrinComp program toolbar. A. ‘Principal Component Analysis’ button performs the PCA; B. ‘Calculate Principal Component Scores’ button generates PC scores for the dataset; C. ‘Reconstruct Principle Component Contours’ button generates Eigenleaves for visualization.

- e. Verify that parameters are correct in a new window, and click ‘OK.’
- f. Once the PCA results have been generated, a new window will appear to save the results file (.pcr file) in the desired folder.
- g. A new window will appear with information from the PCA (Figure 17). There are a number of tabs with useful information about the analysis.

Information of Principal Component Analysis					
Analysis Info.	Mean	Std	Variance-Covariance Matrix	Eigenvalues	Eigenvectors
1	9.632598E-00	7.980617E-00	-1.605471E-00	9.564618E-00	
2	4.300023E-00	-8.397836E-00	4.352083E-00	1.527422E-00	
3	1.281298E-00	-1.773130E-00	-5.441934E-00	3.188441E-00	
4	-1.229856E-00	-1.300024E-00	6.675275E-00	3.082880E-00	
5	-1.246420E-00	-3.048998E-00	1.496227E-00	5.530305E-00	
6	-8.363277E-00	1.975330E-00	-4.544305E-00	-3.741592E-00	
7	-9.127031E-00	3.977788E-00	7.134427E-00	-1.056209E-00	
8	9.742094E-00	-4.099123E-00	-2.594093E-00	-1.362424E-00	
9	2.274701E-00	5.259143E-00	5.646689E-00	-1.434905E-00	
10	-5.497473E-00	-1.711926E-00	9.327247E-00	1.816947E-00	
11	1.187666E-00	1.084397E-00	5.549783E-00	1.910393E-00	
12	4.658812E-00	-2.971391E-00	-6.877223E-00	-3.542407E-00	
13	2.382005E-00	-3.894921E-00	3.960957E-00	-5.313793E-00	
14	-3.832806E-00	-4.623620E-00	-4.683221E-00	5.249300E-00	
15	-3.413748E-00	-4.676442E-00	-1.452148E-00	8.949904E-00	
16	3.274583E-00	4.343511E-00	2.129310E-00	-1.750571E-00	
17	3.232446E-00	-4.666203E-00	-8.534966E-00	6.655027E-00	
18	-2.754313E-00	3.957500E-00	3.852536E-00	-4.285007E-00	
19	3.183106E-00	-2.848343E-00	-2.200872E-00	3.349695E-00	
20	-2.584712E-00	4.011000E-00	-4.355703E-00	-4.166149E-00	

Figure 17. The ‘Information of Principal Component Analysis’ window. This window will appear after the program has completed the PCA run, providing useful information about the results of the analysis in various tabs.

- i. At the bottom of the window, select the ‘Make Report’ button.
- ii. A new window titled ‘Report Option Dialog’ (Figure 18) will appear to select analysis information that will be written into a report (.txt file). Confirm the ‘Eigenvalues &

Eigenvalue Proportions' box is checked, as this contains the percent variance explained by each principal component (PC).

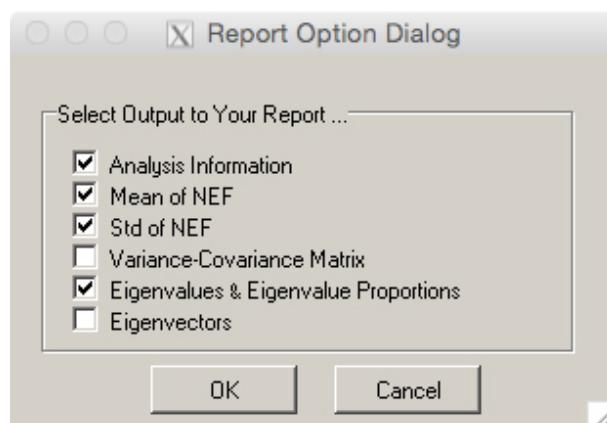


Figure 18. Select report information for printing. Boxes checked in this window will include relevant information in a report that will be written as a .txt file.

- iii. Click 'OK.' A new window will appear with the text file containing a PCA report.
- h. To retrieve the PC values for further analysis, click the 'Calculate Principal Component Scores' button in the PrinComp toolbar (Figure 16B) to create a PC score file.
 - i. A new window will appear to name and select the output file. Click 'OK.'
- i. To visualize the 'Eigenleaves' and what each PC represents, click the 'Reconstruct Principal Component Contours' button in the PrinComp toolbar (Figure 16C).
 - i. A dialogue box will appear to select the number of components to reconstruct, options being 'Reconstruct Effective Components Only' or 'Select Manually...' (Figure 19). This will be the number of components that will be visualized.



Figure 19. The 'Reconstruct Contours' window. In order to visualize the Eigenleaves, select the desired number of components to be reconstructed.

- j. Save the resulting ‘PC contours’ file. New windows will appear (opens automatically in SHAPE’s PrinPrint program; Figure 20), one with a graphic showing Eigenleaves, another to select draw options.
4. The PrinPrint program can be used to view the ‘PC contours’ file at a later time.

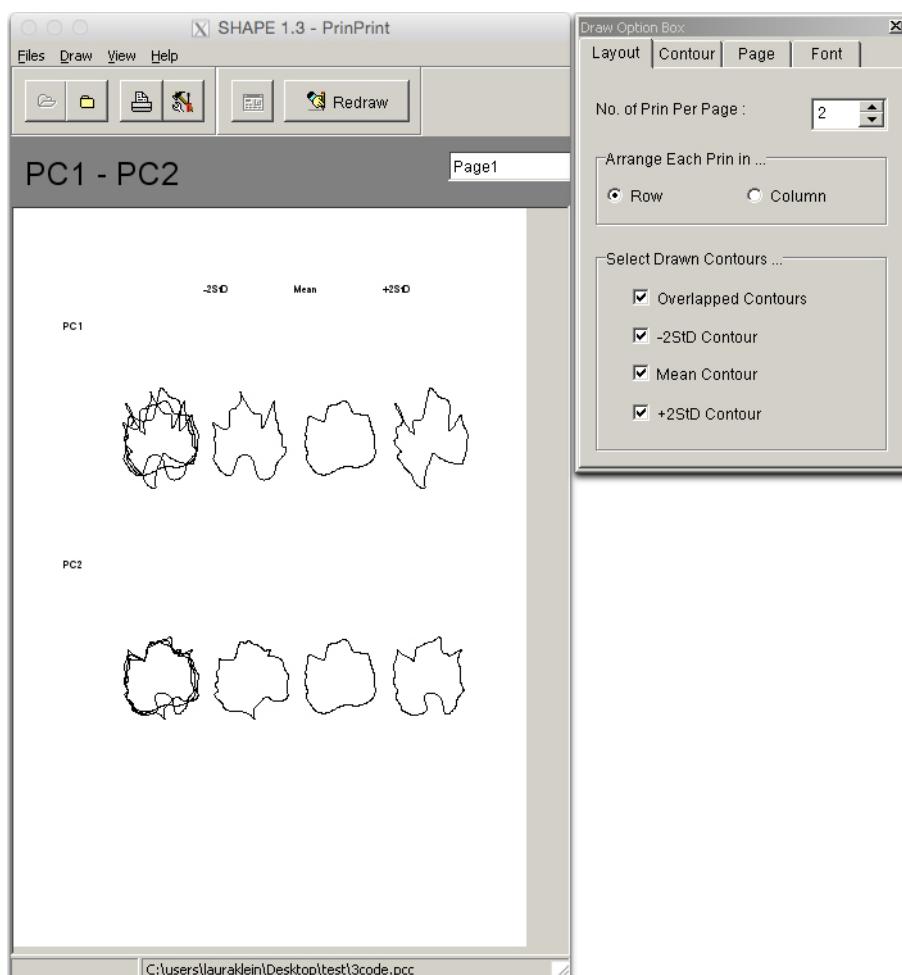


Figure 20. Eigenleaves visualized in the ‘PrinPrint’ program

F. Shape features: Aspect ratio and circularity

1. Open ImageJ.
2. Navigate to the menu bar and select ‘Analyze > Set Measurements...’
3. Check the ‘Area,’ ‘Shape Descriptors,’ and ‘Display label’ boxes.
4. Navigate to the menu bar and select ‘Process > Batch > Macro...’
 - a. A new window will appear titled ‘Batch Process.’ Select the appropriate input (i.e., binary leaf image files) and output folders (Note 13).
 - b. In the macro field, input the ImageJ code found on GitHub (file: ‘Protocol_stepF-4-b.txt’).
 - c. Select ‘Process.’
 - d. A new window will appear titled ‘Results.’ Save this report.

5. The resulting data can be visualized using linear regression (Figure 21).

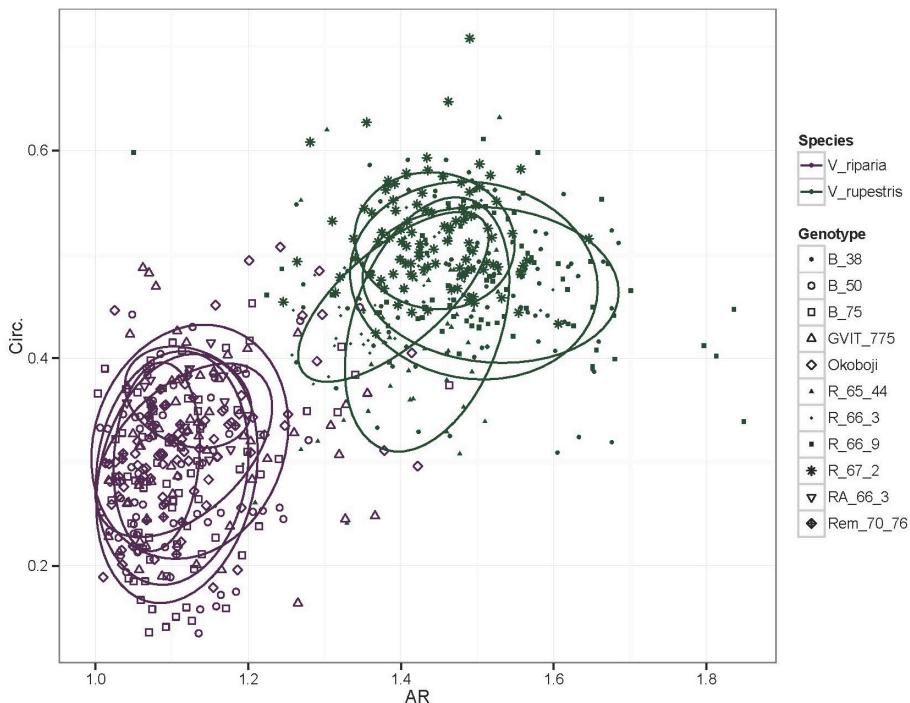


Figure 21. Linear regression of Aspect Ratio (AR) and Circularity (Circ.) data. In this example, multiple leaves from several genotypes of *Vitis riparia* (purple) and *V. rupestris* (green) have been measured for AR and Circ., with the resulting data visualized in this linear regression. Low AR and Circ. values capture the deeper lobing and significant serration of *V. riparia* leaves compared to that of *V. rupestris* leaves.

Data analysis

Data analysis techniques are fundamental to the purpose of this protocol and are integrated within the procedure (*i.e.*, C, E3, F5), as they are often challenging for new users to develop *de novo*. However, our methods represent a sampling of available methods for digital morphometric analysis. We encourage users to explore the literature and available programs to develop a method that is best suited for their material and the particular scientific inquiry.

Notes

1. In the scripts available from GitHub, any text preceded with '#' should be considered user notes—these will not be read by the computer if the whole script is copied and pasted into the R or macro consoles.
2. Because Generalized Procrustes Analysis (GPA) and outline analysis do not require scaling, it is not necessary to include a ruler in scans for morphometric analyses; however, we recommend this as good practice.

3. If the point tool is in the single ‘Point’ setting, simply right-click on the button to change it to the ‘Multi-point’ setting.
4. Landmarks must be placed in the same order on all leaves.
5. If an error was made during placement, points can be deleted by holding the CTRL key and clicking on the point. It is also possible to move points by clicking on the imprecise point, then moving it to the desired location.
6. It may be necessary to adjust column headers or other information that might be erroneously askew from the conversion process.
7. If there is only one fresh leaf per scan, refer to the directions in step D1e-D1f. Repeat until all images have been converted to black and white.
8. For the first image file of the dataset, do not use the first line of code (e.g., run(“Open Next”)).
9. BMPs are large files, and it may be convenient to use an external hard drive or cloud storage client to store and use them from this point forward.
10. Viewing the first image in the viewing field may require using the ‘Zoom Out’ button in the top right corner to visualize properly.
11. This produces one file for all of the images’ chain code. If chain-coding cannot be completed in one sitting, the file can be updated to include the remaining files at a later date.
12. Alternatively, using the R package ‘Momocs’ (Bonhomme *et al.*, 2014) converts NEFs to objects for a variety of graphical visualizations, including PCA.
13. The desired output will not be a folder of files, rather the measurement report that can be saved as a single file. Therefore, the output folder can be the same as the input folder. To minimize file size, select a small file format (we suggest .jpeg) in the ‘Output Format’ drop down menu.

Acknowledgments

This protocol was developed in part for the publications of Chitwood *et al.* (2016) and Klein *et al.* (2017). The authors are grateful to Dr. Dan Chitwood for his comments, guidance, and expertise in using and developing these methods. We would also like to thank Dr. Allison Miller lab undergraduate members for their comments and suggestions on the protocol, as well as three reviewers who helped improve this manuscript. A Saint Louis University Graduate Research Assistantship to LLK and an Ohio University Original Work Grant to HTS supported this work.

References

1. Bonhomme, V., Picq, S., Gaucherel, C. and Claude, J. (2014). [Momocs: Outline analysis using R](#). *J Stat Softw* 56(1): 1-24.
2. Bookstein, F. L. (1991). [Morphometric Tools for Landmark Data: Geometry and Biology](#). Cambridge University Press, pp: 435.

3. Chitwood, D. H., Klein, L. L., O'Hanlon, R., Chacko, S., Greg, M., Kitchen, C., Miller, A. J. and Londo, J. P. (2016). [Latent developmental and evolutionary shapes embedded within the grapevine leaf](#). *New Phytol* 210(1): 343-355.
4. Chitwood, D. H. and Sinha, N. R. (2016). [Evolutionary and environmental forces sculpting leaf development](#). *Curr Biol* 26(7): R297-306.
5. Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P. and Wilkin, P. (2012). [Plant species identification using digital morphometrics: A review](#). *Expert Syst Appl* 39(8): 7562–7573.
6. Dryden, I. L. (2017). [Shapes: Statistical Shape Analysis](#). R package version 1.1-8.
7. Hearn, D. J. (2009). [Shape analysis for the automated identification of plants from images of leaves](#). *Taxon* 58(3): 934-954.
8. Iwata, H. and Ukai, Y. (2002). [SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors](#). *J Heredity* 93: 384-385.
9. Klein, L. L., Caito, M., Chapnick, C., Kitchen, C., O'Hanlon Regan, Chitwood, D. H. and Miller, A. J. (2017). [Digital morphometrics of two North American grapevines \(*Vitis*: Vitaceae\) quantifies leaf variation between species, within species, and among individuals](#). *Front Plant Sci* 8: 373.
10. Kuhl, F. P. and Giardina, C. R. (1982). [Elliptic features of a closed contour](#). *Comput Vision Graphs* 18: 236-258.
11. McLellan, T. and Endler, J. A. (1998). [The relative success of some methods for measuring and describing the shape of complex objects](#). *Sys Biol* 47(2): 264-281.
12. Murdoch, D., Chow, E. D. and Celayeta, J. F. (2007). [ellipse: Functions for drawing ellipses and ellipse-like confidence regions](#). R package version 0.3-5.
13. R Core Team. (2017). [R: a language and environment for statistical computing](#). R foundation for statistical computing. Vienna.
14. Wickham, H. (2009). [ggplot2: Elegant Graphics for Data Analysis](#). Springer.
15. Wickham, H. and Chang, W. (2015). [devtools: Tools to Make Developing R Packages Easier](#). R package version 1(0): 185.
16. Wickham, H., Danenberg, P. and Eugster, M. (2015). [roxygen2: In-source documentation for R, 2015](#). R package version 5(1).
17. Zelditch, M. L., Swiderski, D. L., Sheets, H. D. and Fink, W. L. (2004). [Geometric Morphometrics for Biologists: A Primer](#). Elsevier Academic Press, pp: 437.