

Protocol for leaf image acquisition and analysis, version 2

Edwin Correa, Walton Green*, Carlos Jaramillo, Maria Carmen Ruiz Jaen, Cristina Salvador, David Siabatto, Joe Wright

*contact: Walton Green <wagreen@bricol.net>

1. INTRODUCTION

This is a protocol for acquiring (and processing) images of leaves for automated morphometric analysis. We have implemented an earlier version of this protocol more or less successfully on a set of about 1300 images from the CTFS 50-hectare plot on Barro Colorado Island, Panama. This protocol reflects our experiences in that collection. There are several options discussed below from which to choose, but once choices have been made it is important to be consistent about applying the options. Automatic processing only saves time if data is collected so that human processing is restricted to certain steps early in the collection process, so it is generally not a good idea to collect data thinking ‘this is wrong, but I’ll correct it later’. With automatic processing, there is frequently no later. We will perform some automatic checks and eliminate any files which do not pass these checks, but we hope that once the raw images are collected we will be able to avoid a person ever opening them again one-by-one.

2. IMAGE ACQUISITION

Leaf images can be acquired using a digital camera or a flat-bed scanner; both methods have problems, and we have more fully explored the use of a scanner, but provide two protocols below in case scanning proves impossible in certain cases

2.1. Scanning. Scanning seems to be the safest procedure for obtaining an image. Problems are its slow speed and sometimes unavoidable shadows around edges. Shadows can be eliminated by the use of a back-lit (transmitted light) scanner if one is available, but this degrades the quality (exposure) of the actual image. [Will check this more fully.]

1. If petiole is present, cut petiole and lamina apart. [modification from v.1]
2. Place each lamina and petiole on the scanner such that the abaxial side (ventral side or bottom) of the lamina is towards the glass.
3. Add a 1 cm by 4 cm blue scale bar with its long axis parallel to the midvein of the lamina. We found it convenient to tape the scale bar provided in Figure 1 to the scanner glass so that it was included in each scan and arrange the lamina parallel to it. Experimentation with different scale bars may help reduce shadows around the scale bar. There should therefore be exactly three objects in each image unless the petiole is absent or too short to be removed, in which case there will be exactly two objects.

[In the current version of the analysis it is vital that the scale bar be 1 cm by 4 cm, but its orientation is not used.] In order to identify leaves with an aspect ratio (length to width ratio) less than 1 it is necessary to align the scale bar with the midvein of the lamina.



FIGURE 1. Scale Bar

3. Scan to an 8-bit color image at 300 dpi with moderate jpeg compression. [In v.1 we used .tif files and 8-bit grey-scale images, but recommend moderate jpeg compression for the future (to minimize file size). Color is not currently used, but could be important in the future.]

4. If the leaf is too large for the scanner bed, it should be cut into pieces small enough to be scanned and then reconstituted electronically so that the final image file contains only the complete lamina image, complete petiole image (if petiole is not negligible), and one scale bar (oriented parallel to the midvein of the lamina).

5. Transmitted light and reflected light scans can be mixed to get the best combination of no shadows and well-exposed leaf venation. [Further experiments with transmitted light scanning to be performed soon.]

6. Compound leaflets should be treated in the same way that simple leaves are treated. If the compound leaflets are very small, a number of them can be scanned together and then cut apart digitally into separate image files, each with scale bar, leaflet lamina, and petiolule. In general, 5 leaflets should be scanned for each compound leaf. [Joe is 5 enough? You may want to add more in here concerning weighing the other leaflets and petiole of a compound leaf.]

2.2. Photography [alternate procedure]. Photography is faster than scanning and can provide better results but requires more attention to details of the setup; the process is not as mechanical as scanning, so the worst photograph is much worse than the worst scan, and the best photograph a little better than the best scan. Problems include reflections from glass used to press the leaf flat, shadows with incident lighting if the leaf is not pressed flat enough, a ‘halo’ effect around opaque edges if back lighting is used and the leaf is not pressed flat enough, and over- or under-exposure of the photograph..

The leaf can be lit either from the back (back lighting) or from the front (incident lighting); the background behind the leaf can be black, white, or some other solid, textureless color. Overall, the best results we obtained were from back-lit, light-field photography, i.e. placing the leaf on a light box covered by a sheet of glass, though certain features were better shown by dark-field (dark-ground) lighting and some leaf images were improved by adding low-angle, diffuse incident fill light to the back-lighting. The procedure we found most reliable is as follows:

1. Separate leaf lamina and petiole as above.
2. Place the petiole and lamina adaxial side down (abaxial side towards the camera lens) on a light box, with a 1 cm by 4 cm rectangular blue scale bar, parallel to the leaf midvein.
3. Cover with a sheet of plate glass and dim room lights to prevent reflections of light from the covering glass plate.

4. Focus and photograph leaf with moderate jpeg compression from a point directly over the lamina + petiole + scale bar frame. Depending on the camera, there may be different images sizes and jpeg compression ratios available; it is not possible to give a general rule for selecting these, but the resulting photograph should clearly show at least tertiary (intercostal) venation and should generally be less than 1 MB in size (unless many photographs have to be digitally stuck together). If possible, exposure should be taken in spot mode on the leaf lamina itself; otherwise different exposures should be tried till an acceptable image of the lamina is produced. If the lamina is nearly opaque, this may require long exposure times so a tripod, copy stand, or other camera mount may be needed. Aperture, shutter speed, and lens focal length (zoom) can be varied, but approximate file size and compression ratio should be kept constant.

3. FILE MANAGEMENT

1. Each image should get a site name in capital letters plus single integer number with 5 digits and its name followed by the file extension, '.jpg'. Therefore the third leaf scanned from Barro Colorado Island, Panama will be 'BCI00003.jpg'.

Images of parts of leaves that will be digitally reconstructed, should be given the same name with '.01', '.02', etc. inserted after the number. So if the fourth leaf from BCI had to be cut into three pieces, they would be labelled 'BCI00004.01.jpg', 'BCI00004.02.jpg', and 'BCI00004.03.jpg'. The final images should be named exactly the same way as images that were not digitally reconstructed, so the composite of the fourth leaf will be 'BCI00004.jpg'. All images should be stored in a flat file hierarchy (all in the same folder). This folder can be compressed using bzip2, gzip, stuffit, or winzip and sent by ftp to [ftp site to be entered later]. All files can also be backed up to DVD+R disk or saved on a hard-disk or flash drive.

The integer file numbers need not be consecutive but must (obviously) be unique.

2. If images of partial leaves cannot be reconstructed or if an image of multiple compound leaflets cannot be cut apart, these preliminary files should be put in a separate directory called TEMPORARY. The file name for an image of multiple compound leaflets should look the same as for a simple leaf, i.e. 'BCI00005.jpg', even if it will later have to be cut up into 'BCI00005.01.jpg', 'BCI00005.02.jpg', etc.

3. Along with the images, it is necessary to keep a spread-sheet (e.g. in Excel) with the following columns: 1. image file name (as described above), 2. date collected (in yyyy-mm-dd format), 3. date imaged (in yyyy-mm-dd format), 4. Compound or Simple (the upper case letter 'C' or 'S'), 5. tree stem code, 6. other code (like a, b, c, etc., for several leaves from a single tree stem), 7. comments

In the case of compound leaves, '6. other code' should be used to label multiple leaflets.

4. Note that the '5. tree stem code' plus the '6. other code' must also be linked to another spread-sheet that gives the other contextual information like taxonomic identification and stem location in the plot.

4. IMAGE PROCESSING

There are two methods of image processing that can be applied to these images, both using freely available, open-source image analysis software. The first uses ImageJ, a relatively user-friendly, interactive, window-based program. The second uses a command-line analysis package called EBImage written in the R graphics/statistics language. Final analysis will use EBImage, but ImageJ can be useful for checking and illustrating results before the final automatic data collection script is run.

4.1. Algorithm. The approximate processing algorithm runs as follows. Current versions of the scripts (versions 2) in both ImageJ (java) and R as well as this document are available from <http://www.bricol.net/imageanalysis>.

1. Open the image.
2. Calculate true size information from the resolution and pixel dimensions.
3. Down-sample it to 300 pixels wide.
4. Convert it to an 8-bit grey scale.
5. Normalize the histogram.
6. Threshold it to create a binary mask.
7. Scrub small (less than 9 pixel) dirt from the mask.
8. Index the remaining objects in the mask.
9. Calculate measurements for each object (filling in all holes).
10. Write the measurements of the three largest objects to a file.
11. Save the mask.
12. Close the image.

4.2. ImageJ notes. The image analysis program ImageJ can be downloaded from <http://rsb.info.nih.gov/ij/>. Installation instructions are available at the website.

4.3. EBImage notes. The graphics/statistics programming language and interpreter R can be downloaded from <http://www.r-project.org/>. Installation instructions are available at the website. After installation, the package EBImage must also be installed, e.g. by running the command `> install.packages('EBImage')`.

5. PROBLEMS

The following images show some of the problems that we have encountered; some are avoidable; some not.

5.1. Compound and Deeply-Lobed Leaves. Frequently compound and deeply-lobed leaves have leaflets or lobes that overlap each other leaves when pressed flat. [Compound leaves dealt with above: what about overlapping deeply lobed?]

Especially in the case of multiply compound leaves, conifers, and other taxa with reduced leaves, the leaf(let)s may be too small...[Do we have a fail-safe option if there's something we haven't thought about? Photograph?]

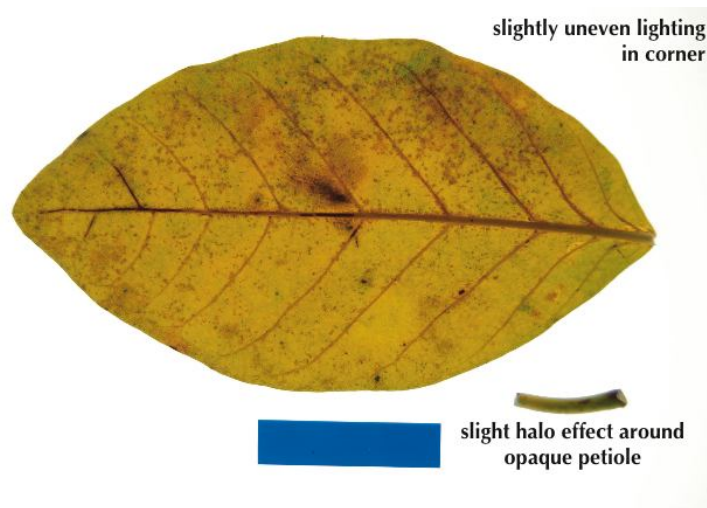


FIGURE 2. A good image

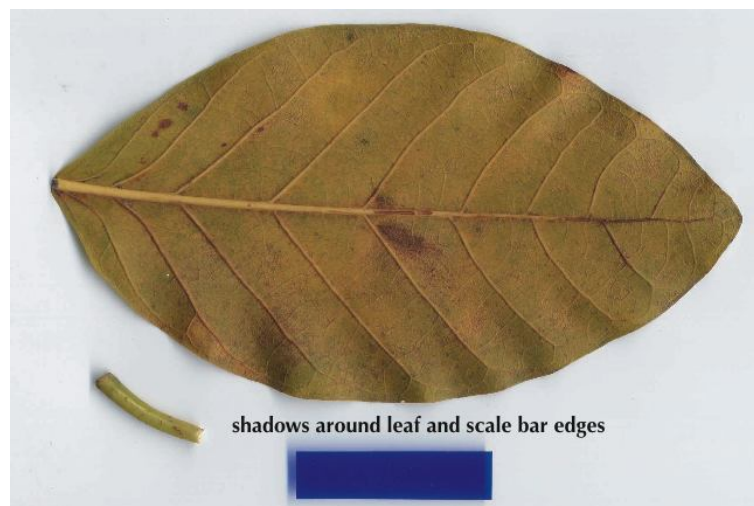


FIGURE 3. An acceptable image

5.2. **Large Leaves.** When leaves are slightly too large for a single scan or photograph, they can be processed in pieces, as described above, but there are some leaves (like palms) which may be too large for such a procedure. In such cases, see the section on compound leaves above.

5.3. **3-Dimensional Leaves.** Many leaves are 3-dimensional in life position and pleat or crinkle when pressed flat under glass. In general, this can be ignored, but when it is

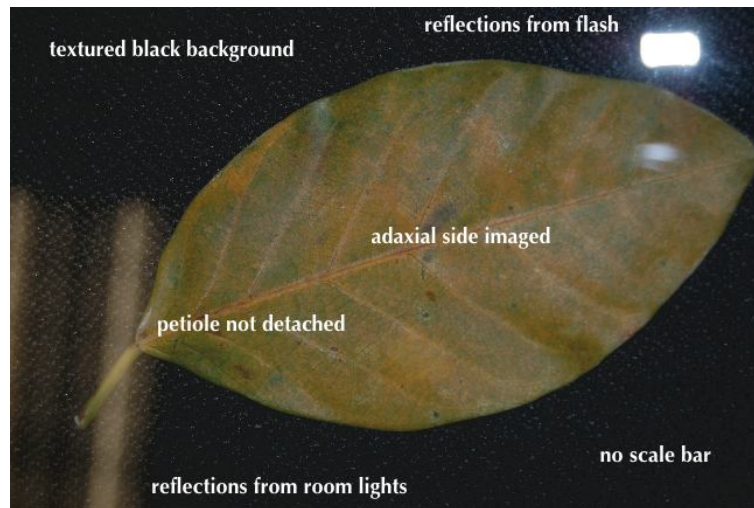


FIGURE 4. An unusable image.

particularly noticeable should be mentioned in the comments column 7 described above under 'file management'.

The same is true of winged petioles, or percurrent lamina where the petiole is not distinct. In such cases a judgement must be made about whether to remove the petiole at an arbitrary point or scan in one piece as if the petiole was missing.