

ON THE NATURE OF THINGS: ESSAYS

New Ideas and Directions in Botany

High-throughput phenotyping¹

Malia A. Gehan² and Elizabeth A. Kellogg²

Anyone who has written a species description knows the slow process of measuring the length and width of plant parts with a ruler and ocular micrometer, counting hairs or branches, or assessing the color of fruits. Anyone who has studied plant communities has counted seedlings, measured leaf area, or laid out plots and counted their contents. Until recently, however, optimizing the speed of the process has not been a high priority. If it takes an hour to measure one herbarium specimen, how might that be reduced to minutes? If it takes 10 undergraduates a week to record plant communities along a transect, how might one undergraduate accomplish the same work in an afternoon?

Lower-cost, automated and semiautomated methods for data acquisition and analysis are now being developed, enabled by inexpensive cameras and computers with open-source software. Most recent applications have been in crops and model organisms, but the tools can be extended to systematics and ecology, fields that often require huge amounts of specimen data. In this essay we describe a few available tools to encourage readers to consider ways to increase the throughput of their own research. While the term high-throughput phenotyping could apply to any morphological, physiological, or biochemical phenotype, here we focus on morphology or other phenotypes (e.g., drought response) that can be captured using images. “High throughput” was defined by Fahlgren et al. (2015a) as “hundreds of plants per day”, but for many projects even tens of plants per day would be a massive leap forward.

DATA ACQUISITION

The appropriate method of image acquisition is determined first by the biological question and scale (macroscopic or microscopic) and second by the budget. Different camera types can provide different information (Fahlgren et al., 2015a). For example, shape and size of a herbarium specimen can be captured with a digital camera

recording visible light. Conversely, drought response may be better investigated with near-infrared imaging. State-of-the-art imaging technologies are being assessed in the field by the ARPA-E TERRAREF project (<http://terraref.org/>), a major goal of which is to determine what additional biological information can be gained from pricey (e.g., hyperspectral and thermal) cameras compared to basic RGB cameras. Regardless of camera type, speed, or scale of data, the measurement needs to fit the scientific question. In addition, enough experimental and image metadata must be captured for downstream data analysis. Standards for metadata to include with phenotyping experiments have been established by two large European networks, transPLANT (Transnational Infrastructure for Plant Genomic Science) and EPPN (European Plant Phenotyping Network) (Ćwiek-Kupczyńska et al., 2016), and are summarized on the miappe website (Minimum Information About a Plant Phenotyping Experiment; <http://www.miappe.org/>; Ćwiek-Kupczyńska et al., 2016). The North American Plant Phenotyping Network (NAPPN) held its inaugural meeting in 2016 and prioritized reviewing and improving data standards. The NAPPN is still in a formative stage and would greatly benefit from more members from ecology and systematics, fields that were largely unrepresented among the geneticists, engineers, and computational biologists at the inaugural meeting.

Budget also dictates methods. Large-scale, custom high-throughput phenotyping platforms are offered for the field and laboratory by companies such as LemnaTec, Phenokey, PhenoSpex, Photon System Instruments, Wiwam, and We Provide Solutions. In the herbarium, conveyor belts can speed acquisition of specimen images (Tegelberg et al., 2014). Such commercial systems can require facility managers, annual maintenance, and a steady user pool, and thus may be too expensive for many researchers and institutions. Do-it-yourself solutions, for which tools and resources are increasingly available, are much cheaper. For example, an ordinary digital camera on a copy stand over a light box can be attached to a remote computer and photos taken directly onto the hard drive (Chitwood et al., 2014). Using this approach, 9500 leaf images were captured by two people in 3 days (D. Chitwood, personal communication). The Maker Movement, a culture interested in open-source homemade technology, has made good use of low-cost microcontrollers (i.e., Arduinos, Teensy, BeagleBone) and computers (i.e., Raspberry Pi) that allow amateur technologists to build and

¹ Manuscript received 31 January 2017; revision accepted 17 March 2017.
Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, Missouri 63132 USA

² Authors for correspondence: mgehan@danforthcenter.org; ekellogg@danforthcenter.org
doi:10.3732/ajb.1700044

prototype tools for many applications. These Maker Technologies can control familiar equipment such as flat-bed scanners and digital cameras (see <http://maker.danforthcenter.org/>). Anyone comfortable doing Web searches is well equipped to set up a Raspberry Pi computer and camera, and online forums (and other laboratories!) are more than willing to lend a hand. We have guided very young researchers (ages nine and up) to set up their own time-lapse imaging stations, so we estimate that it would take a novice Maker an afternoon to set up a Raspberry Pi camera for time-lapse or other sorts of imaging. The PhenoTiki project is one recent example, using the Raspberry Pi to capture and analyze leaf shape, growth, color and number for plants in a growth chamber (Minervini et al., 2017). In many cases, image acquisition is no longer the bottleneck in a plant phenotyping project. Instead, image analysis is more often the area where researchers may need to develop computational skills or find an appropriately skilled collaborator. Currently available programs vary in accessibility for new users.

DATA ANALYSIS

Image analysis is an active and challenging field of computer science that is rapidly providing tools applicable to biological problems. A common first step in manipulating a photograph is extracting the relevant portion of the image (e.g., a plant, leaf, or spikelet) from the background, which can be done in programs such as ImageJ (Schneider et al., 2012), PlantCV (Fahlgren et al., 2015b), or MatLab

(MathWorks). The first two of these are open source, whereas MatLab is a commercial product. Basic use of ImageJ is relatively simple and can be extended with macros. MatLab is a programming language of its own, and PlantCV requires some programming skill in Python. The latter two are particularly flexible in the kinds of images that can be handled and the information that can be extracted.

In ImageJ or landmark-based approaches such as TPSDig (<http://life.bio.sunysb.edu/morph/soft-dataacq.html>), traits can be measured without necessarily extracting the target portions of the image first, and both are faster than recording measurements by hand. We have used both tools on photographs of grass spikelets (C. A. McAllister [Principia College] and E. A. Kellogg, unpublished data) (Fig. 1). While arranging the spikelets for imaging was decidedly low throughput, standard measurements were captured reliably and rapidly by two undergraduates. If human intervention is necessary to measure or capture data, as in this spikelet example, crowdsourcing can be an option (Ellwood et al., 2015). For example, the Microplants project from the Field Museum (<http://microplants.fieldmuseum.org/>) is a citizen science project that focuses on labeling image data. More broadly, Amazon's Mechanical Turk platform (<https://www.mturk.com/mturk/welcome>) can be used to increase throughput. Further, manually labeled images are a necessary first step for training machine learning algorithms to automate the process. Software built specifically for large-scale high-throughput processing of images includes the Integrated Analysis Platform (Klukas et al., 2014), ImageHarvest (Knecht et al., 2016), and PlantCV (Fahlgren et al., 2015b). A herbarium specimen analyzed with PlantCV is shown in Fig. 2, from which leaflet length, color, and area, as well as other measurements can be retrieved (see http://plantcv.readthedocs.io/en/latest/output_measurements/). PlantCV and ImageHarvest are available through Github, where developers can get a unique document identifier (doi number) and thus cite the code as a research product; in addition, Github offers a built-in means for the community to add code.

Reproducibility, precision, and accuracy are also important considerations. Automated image analysis is repeatable and reproducible, reducing human bias and increasing precision. Images can be reanalyzed if improved analytical methods become available. However, accuracy needs to be evaluated for each application. In Fig. 2, for example, the semiautomated measurements miss small sections of the leaflets, and the length of the terminal leaflets might not exactly follow the mid-vein; it is up to the user to decide

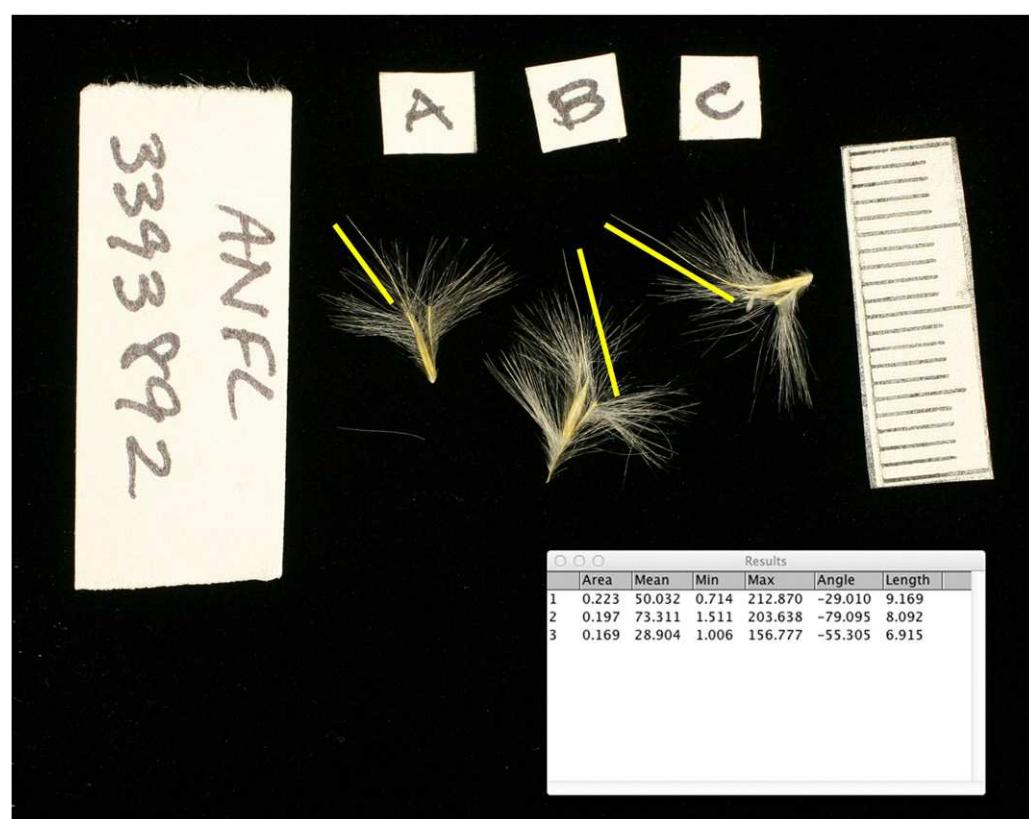


FIGURE 1 Spikelets of *Andropogon floridanus* (Godfrey 79233, MO), photographed in the herbarium of the Missouri Botanical Garden using a standard digital camera on a stand. Specimen is labeled on the left, and scale is on the right. Yellow lines indicate the length of the awns, and the inserted table lists the measurements calculated by ImageJ (Schneider et al., 2012). Photo by Sarah Clewell and Christine McAllister.



FIGURE 2 Herbarium specimen of *Solanum lycopersicum* analyzed with PlantCV (Fahlgren et al., 2015b). The outermost margins of the leaflets are outlined, showing that they have been identified from the surrounding objects and plant structures. The terminal leaflets are distinguished from the others by blue outlines. Perpendicular red lines indicate the length and width of the bounding rectangle of the leaflet, and intersect at the center of mass. The convex hull of the terminal leaflet is also outlined in red. Lengths of the terminal leaflets are 3.88, 4.44, and 4.49 (average 4.3) cm, and areas 7.93, 7.67, and 9.20 (average 8.3) cm². Specimen image retrieved from Tropicos.org (Missouri Botanical Garden, 07 March 2017, <http://www.tropicos.org/Image/100123249>).

if these inaccuracies are a problem and to adjust accordingly. In this context, comparison of traditional methods of data capture with high-throughput methods is difficult because the entire workflow differs. For example, the measurement data for the spikelets in Fig. 1 can be captured manually with a dissecting microscope and ocular micrometer in about 15 min. However, it is hard to do this kind of measuring for more than a few hours a day (ca. 8–12 specimens), and the microscopist must be familiar with plant structure so the task requires training and supervision.

In principle, image files can be mined for phenotypes other than those for which they were collected, but this may not be true in practice. For example, the image in Fig. 2 contains a good quality specimen in which the leaves are not overlapping, representing an ideal layout for image analysis. If subsequent specimens were laid out in the same approximate position, the same analysis pipeline could be used. However, large alterations in layout would require alterations in image analysis and/or modifications of the pipeline. Identifying analysis software before beginning a project will alleviate later challenges if particular image layouts, file structure, or formats are needed during data acquisition. A valuable online resource for commercial and open-source plant image analysis software can be found at the Plant Image Analysis website (<http://www.plant-image-analysis.org/>; Lobet et al., 2013), which lists tools that measure whole plants or organs as well as ones that focus on analyzing anatomical and histological images.

DATA STORAGE

An unsolved problem is how to store and index accumulating image data for further public use. Image files are large and not easily compressed. Furthermore, it can be hard to find images without a central repository. Besides Morphbank (Morphbank: Biological Imaging, 2017; www.morphbank.net/), Morphobank (O'Leary and Kaufman, 2012; <https://www.morphobank.org/>), iDigBio (<https://www.idigbio.org/>), NEON (www.neonscience.org/), and BisQue on Cyverse (www.cyverse.org/bisque), large image data sets are indexed at an online database at www.plant-image-analysis.org/dataset (Lobet et al., 2013). A major goal of the NAPPN, like the EPPN or IPPN (International Phenotyping Network), will be to unify the community around sets of standards that will allow indexing and searching of data.

PROSPECTS FOR THE FUTURE

Increasingly powerful tools are available for high-throughput data acquisition and analysis, especially for those willing to make their own equipment and apply open-source software. Using the latest phenomics software can require some comfort at the command-line, which can be a barrier for some researchers, but this type of impediment is true of genomics tools as well. Also like genomics, some ability to write computer code can be helpful. A growing community with diverse backgrounds will help to ensure that the tools and tutorials meet the needs of a greater pool of researchers. Many tools and approaches developed for crops (see for example, Hawkesford and Lorence, 2017) can be applied more or less directly

to plots of wild plants. The example in Fig. 2 shows that a tool such as PlantCV can be applied to specimens with minimal manual input. Speed of data capture and analysis can now be considered in early project design, and approaches originally designed for agricultural and translational research can be extended to the many other fields that make up plant biology.

ACKNOWLEDGEMENTS

The authors thank the editor for the opportunity to contribute this paper and Editor-in-Chief Pamela Diggle and three reviewers for comments that greatly strengthened the manuscript. This work was funded in part by grants from the National Science Foundation to E.A.K. (DEB-1457748 and IOS-1413824) and to M.A.G (IOS-1202682, EPSCoR IIA-1355406, IIA-1430427, and IIA-1430428).

LITERATURE CITED

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

Ćwiek-Kupczyńska, H., T. Altmann, D. Arend, E. Arnaud, D. Chen, G. Cornut, F. Fiorani, et al. 2016. Measures for interoperability of phenotypic data: Minimum information requirements and formatting. *Plant Methods* 12: 44.

Ellwood, E. R., B. A. Dunckel, P. Flemons, R. Guralnick, G. Nelson, G. Newman, S. Newman, et al. 2015. Accelerating the digitization of biodiversity research specimens through online public participation. *Bioscience* 65: 383–396.

Fahlgren, N., M. A. Gehan, and I. Baxter. 2015a. Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology* 24: 93–99.

Fahlgren, N., M. Feldman, M. A. Gehan, M. S. Wilson, C. Shyu, D. W. Bryant, S. T. Hill, et al. 2015b. A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in *Setaria*. *Molecular Plant* 8: 1520–1535.

Hawkesford, M., and A. Lorence (eds). 2017. Plant phenotyping: increasing throughput and precision at multiple scales. *Functional Plant Biology* (Special Issue) 44(1).

Klukas, C., D. Chen, and J.-M. Pape. 2014. Integrated analysis platform: An open-source information system for high-throughput plant phenotyping. *Plant Physiology* 165: 506–518.

Knecht, A. C., M. T. Campbell, A. Caprez, D. R. Swanson, and H. Walia. 2016. Image Harvest: An open-source platform for high-throughput plant image processing and analysis. *Journal of Experimental Botany* 67: 3587–3599.

Lobet, G., X. Draye, and C. Périlleux. 2013. An online database for plant image analysis software tools. *Plant Methods* 9: 38.

Minervini, M., M. V. Giuffrida, P. Perata, and S. A. Tsaftaris. 2017. Phenotiki: An open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. *Plant Journal*.

Morphbank: Biological Imaging. 2017. Website <http://www.morphbank.net/>. Florida State University, Department of Scientific Computing, Tallahassee, Florida, USA [accessed 24 November 2016].

O'Leary, M. A., and S. G. Kaufman. 2012. MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. Website <http://www.morphobank.org>.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9: 671–675.

Tegelberg, R., T. Mononen, and H. Saarenmaa. 2014. High-performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. *Taxon* 63: 1307–1313.