

# An opinion on imaging challenges in phenotyping field crops

Derek Kelly<sup>1</sup> · Avimanyou Vatsa<sup>2</sup> · Wade Mayham<sup>2</sup> · Linh Ngô<sup>3</sup> ·  
Addie Thompson<sup>4</sup> · Toni Kazic<sup>2</sup>

Received: 20 April 2015 / Revised: 10 October 2015 / Accepted: 19 October 2015 / Published online: 29 December 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** Almost all the world's food is grown in open fields, where plant phenotypes can be very different from those observed in greenhouses. Geneticists and agronomists studying food crops routinely detect, measure, and classify a wide variety of phenotypes in fields that contain many visually distinct types of a single crop. Augmenting humans in these tasks by automatically interpreting images raises some important and nontrivial challenges for research in computer vision. Nonetheless, the rewards for overcoming these obstacles could be exceptionally high for today's 7 billion people, let alone the 9.6 billion projected by 2050 (United Nations Department of Economic and Social Affairs, Population Division, World Population Prospects: The 2012 Revision). To stimulate dialog between researchers in computer vision and those in genetics and agronomy, we offer our views on three computational challenges that are central to many phenotyping tasks. These are disambiguating one plant from another; assigning an individual plant's organs to it; and identifying field phenotypes from those shown in archival images.

We illustrate these challenges with annotated photographs of maize highlighting the regions of interest. We also describe some of the experimental, logistical, and photographic constraints on image collection and processing. While collecting the data sets needed for algorithmic experiments requires sustained collaboration and funding, the images we show and have posted should allow one to consider the problems, think of possible approaches, and decide on the next steps.

**Keywords** Maize phenotypes · Field phenotyping · Segmentation · Registration · Plant disambiguation · Organ assignment · Phenotype identification · Species recognition

## 1 Introduction

Increasing food security now and for the future relies heavily on identifying and understanding beneficial phenotypes in crop plants. A phenotype is a visible feature of an organism. Some phenotypic variations are desirable improvements in agricultural crops, such as increased disease resistance, better yield in poorer soils or under drought, and improved nutritional content. All these phenotypes vary tremendously among members of a species, and all are targeted for improvement by many investigators. Current agricultural methods will be insufficient to keep pace with the projected growth in population and the need for improved nutrition [1–3]. A step change in the yield and quality of food crops, and in the sustainability of their production, is urgently needed. Nearly all of the world's food is grown in farm fields. World-wide vegetable production in greenhouses in 2015 totalled 414,127 ha (1.02 million acres), while in 2014 in the United States alone, acreage harvested for just corn and soybeans totalled 67.26 million hectares (166.2 million acres) [4,5]. Research has shown that greenhouse experiments are poor predictors

---

✉ Toni Kazic  
kazict@missouri.edu

<sup>1</sup> Interdisciplinary Plant Group, Informatics Institute, Missouri Maize Center, University of Missouri, Columbia, MO, USA

<sup>2</sup> Interdisciplinary Plant Group, Department of Computer Science, Informatics Institute, Missouri Maize Center, University of Missouri, 201 Engineering Bldg, West, Columbia, MO 65201, USA

<sup>3</sup> Genetics Area Program, Interdisciplinary Plant Group, Missouri Maize Center, University of Missouri, Columbia, MO, USA

<sup>4</sup> Department of Agronomy, Purdue University, West Lafayette, IN, USA

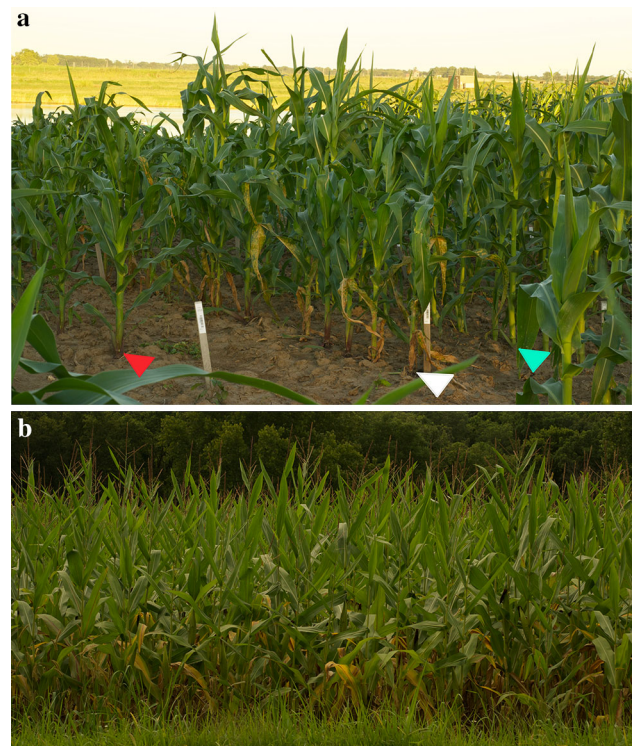
of field performance for grain yield and drought tolerance [6,7]. So agronomically important phenotypes must be studied in field experiments: for food, it is the field that counts.

The foundation of crop improvement is to detect and characterize potentially beneficial phenotypes. This is done in experimental fields around the world, each of which is planted with hundreds or thousands of genetically different varieties of a crop. In this situation, the desired plant may be only one out of tens or hundreds of thousands in a field. Today, experienced observers, working alone or in small teams, scrutinize thousands of plants in a single season, albeit with inter- and intra-observer variations [8,9]. Many agronomically important phenotypes are signaled by changes in the plant's morphology (the size, shape, color, and spatial position of the plant and its organs) over time [10]. Monitoring these phenotypes requires either human examination or imaging of plants in situ. Automating the capture and analysis of plant images would spare human effort for more complex tasks and improve the quantitation of the phenotypes. But as we will see, field plants do not pose nicely for the camera: they crowd together, irregularly occlude each other, grow unevenly, shade parts of each other and the soil, lay on the ground, and hide their phenotypically informative organs. Substitutes for human expertise, even for relatively simple tasks, will require algorithmic approaches that can cope with such issues.

Because phenotyping involves the analysis of multiple individual plants, the machine vision challenges posed by crop fields used in genetic and agronomic research are quite different from those presented by fields in production agriculture or yield trials. In the latter, a field will be planted with a single variety that will exhibit much less phenotypic variation than in the genetically diverse populations of research fields. The difference between the two is illustrated in Fig. 1.

In a uniformly planted field, many parameters can be measured “disembodied and in bulk”: the average values of the greens for chlorophyll content; or the average position of the blue/green sky/plant boundary for plant height; or the average deviation of a leaf from the vertical for leaf angle [11–16]. Traditional assessments of crop health by aerial and satellite vehicles rely on existing techniques, which are now being applied to ground-based images in production situations. Active research, detailed in an excellent recent review and elsewhere in this issue, focusses on phenotyping in greenhouses, where the problems of occlusion and image standardization are much less acute [17]. In contrast, field phenotyping, still in its infancy, is a new frontier for machine vision [6,18,19].

Here, we offer our views on three challenges for research in computer vision that are common to many phenotyping tasks in the field. Our perspective is that of agronomists, geneticists, and computational biologists who photograph maize plants and phenotypes in the field for characteriza-



**Fig. 1** Two different field situations. **a** A research field. Each row is planted with a genetically and phenotypically different variety of maize, as evidenced by differences in height and color. A row of shorter maize is marked with a red triangle; a row of maize with yellow–orange lesions is marked with a white triangle; and a row of taller maize is marked with a cyan triangle. **b** A production agriculture field planted with a single variety of maize. In this field, average leaf angle could probably be estimated from measuring the angle between a vertical axis and planes defined by each leaf (especially those in the upper half of the field), similar to the work of reference [11]

tion, and extract phenotypic information from the images [20,21]. We selected these challenges based on three sources of information. The first source is the methods described in the literature of plant genetics and breeding (for example, reference [22]). Second, we have had many conversations with our plant science colleagues who work with many different species, but especially maize, on their phenotyping tasks and which ones they would most like to have automated. We have also been fortunate to watch our colleagues at work in the field. Finally, our own work in maize has involved many of the phenotyping tasks mentioned in this paper, and thinking about how one might automate these tasks led us to consider the machine vision challenges that would need to be surmounted. Nonetheless, the selection and abstraction of the challenges is ours alone, and our interlocutors are blameless.

The challenges are:

*Disambiguation* Individuating plants from the mass of green is essential for many phenotyping tasks: one must know which plant has what phenotype. We define disambiguation

biguation as the algorithmic segmentation of one plant from its neighbors.

**Assignment** As the plants grow, the appearance of a field changes from orderly rows of physically well-separated little plants to a jungle of leaves, stems, and reproductive organs. Which body parts are from the same plant? We define assignment as the algorithmic assembly of visually separated organs into the correct plant.

**Identification** Scientists use thousands of archival images in the literature and databases to identify the phenotypes they see in the field and to determine when a phenotype may be novel. We define identification as the detection and classification of phenotypes by comparing a plant's features to those of related plants and to archival images.

We illustrate the challenges with images of maize fields photographed from the ground. The images show what can be readily captured today, using either human photographers or with minimal automation. The images shown here are downsampled from the high-resolution ones posted online. The images in the online material were shot with either Nikon D80 with an AF MicroNikkor 60 mm lens (denoted DSLR in the figure legends), an iPad2, or unknown cameras. These images, and those posted online, are not the traditional data sets that one might use in algorithm development and testing. Collecting large data sets that combine experimentation with imaging techniques and technologies and include the collection and annotation of ground-truth information requires adequately supported collaborations between biological and computational scientists. Realistically, the scale and complexity of the data set workers in machine vision need are simply beyond the capacity of biological scientists to “squeeze in” during the very busy field season without compromising funded experiments. Instead, our goal is to illustrate the challenges in field phenotyping and provide enough images to let those working in computer vision see if the problems are interesting enough to pursue in collaboration with biological scientists. We believe that the best work in field phenotyping will require sustained, long-term, and mutually beneficial collaborations, and we wish to encourage those. The good news is that many biologists are seeking help with image processing and phenotyping tasks.

We first describe maize to provide some context for the challenges. A typical plant is shown in Fig. 2. *Z. mays* is a major crop world-wide; many phenotypes are visible to the eye; and the plants are large enough that intra-plant spatial differences are easily detected [22]. Over a hundred years of intensive study of this important cereal crop, mostly in farm fields, has identified many genetically different varieties of maize. Their phenotypes vary widely in size; shape;

**Fig. 2** Disambiguation of a typical maize plant at flowering. In (a), the plant in its field context, with rows behind it and weeds around it. In (b), the plant has been partially isolated using a cloth background, an unusual and laborious photographic technique. In (c), the plant has been manually isolated in the image by selecting it and setting the remaining context pixels to off-white. The male reproductive organ, the tassel, is indicated by the *red triangle*, and the female reproductive organ, the ear, by the *blue triangle*. Photographed with a DSLR



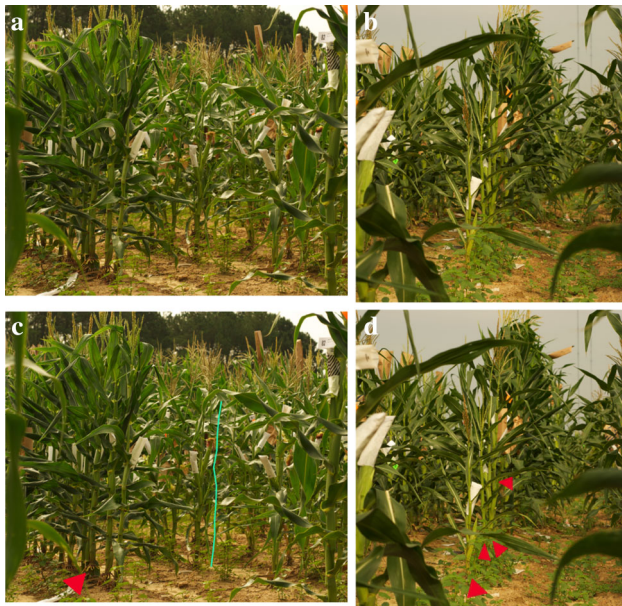
color; number, placement, and types of organs; the rates at which the plant grows, develops, and dies; the yield of kernels and other useful parts; and their responses to different environments [23] (In the United States, maize is colloquially called “corn”, a term used elsewhere in the world for any cereal grain). Starting algorithm development with maize is particularly advantageous because the plants are larger, less densely planted, and more distinct from many weeds than rice or wheat.

## 2 Three challenges for computer vision

### 2.1 Disambiguation of plants

Disambiguation segments plants, or key plant organs, from the mass of green in the field. Two common field tasks are to count the number of crop plants in each row and to detect unusual stem shapes. Consider the images in Fig. 3.

The two different vantage points—diagonally across multiple rows and along a row—balance different types of plant occlusion against confusion by background plants. Shooting multiple rows reduces occlusion within a row, but does not eliminate it: the foreground row contains three very closely spaced plants (red triangle). Segmenting stems by selecting for contiguous, vertical dark green areas would need to adapt to zigzagging of the stem (cyan line), and the occasional very bent stem (see images 1/series/DSC\_0449–451.NEF, posted online, for an example). Similarly, the rows in the image are parallel to each other, so that the angle of the line defined by the intersection of the rows' stems and the soil is relatively constant, once the background weeds are eliminated. Determining in which row a plant lies might require some estimate of depth of field in different parts of the image, allowing for



**Fig. 3** Rows of maize shot from two different vantage points. **a** Looking diagonally across several rows of maize from a fortuitously empty spot in the field. **b** Looking along a row. **c** The same image as in **(a)**, now with three very closely spaced plants in the foreground row marked with a *red triangle*, and an example of a zigzagged stem is marked with a *cyan line*. This particular field was planted by machine; irregular spacing of plants will routinely occur with either manual or machine planting. **d** The same image as in **(b)**, but now with *red triangles* marking the first four stems. In all panels, the emerging ears are covered with white or striped shoot bags; the brown paper bags at the top cover tassels; and smaller weeds are visible. Photographed with a DSLR

the varying sizes of the plants. In contrast, looking along a row increases occlusion, including from plant organs near the camera, but simplifies determining the row.

A more labor-intensive alternative is to shoot stills or video along a row from several different vantage points, and then reconstruct the row after registering the plants. Figure 4 shows one such series of images for the same row. In the images in Fig. 4, registration is simplified by the presence of shoot and tassel bags (small white and large brown, respectively). This would not be true much of the time. As the camera proceeds down the row, portions of the rows behind the row of interest appear.

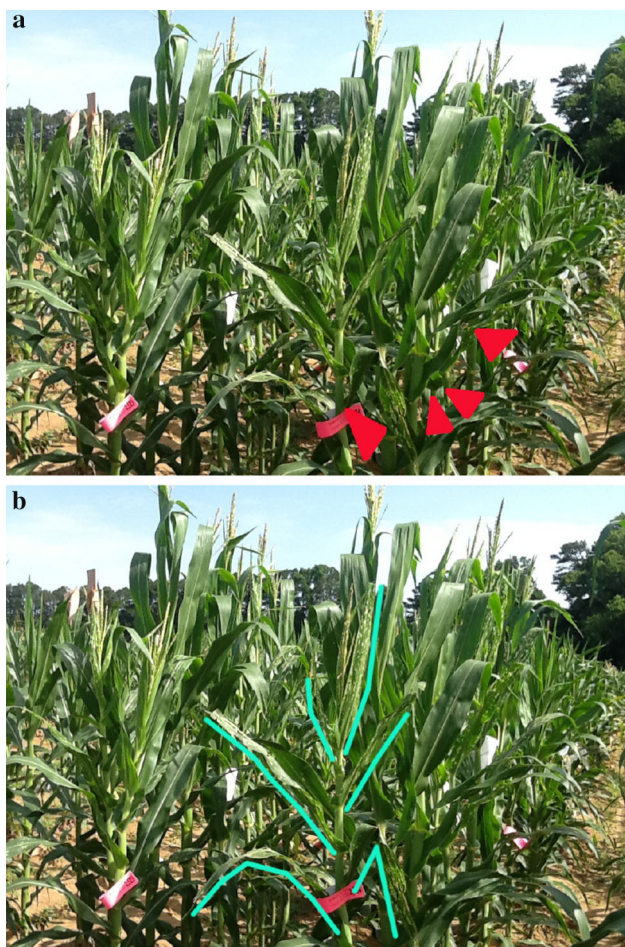
For phenotypes that can be determined in a uniform stand of plants, disambiguation can be bypassed by looking at populations of organs. An example is shown in the right in Fig. 1, panel **(b)**. 3D reconstructions of the outer edges of soybean stands have detected changes in leaf angle without assigning leaves to plants [11]. Current approaches that detect flowering tassels rely on evaluation of the hyperspectral reflectance of the canopy [24,25]. Nonetheless, there will be many situations where disambiguation is important; constructing crude 3D models of plants might help assign unoccluded organs. Isolated maize and rose plants have been reconstructed using photogrammetry and the Microsoft Kinect, respectively [26,27]. The photogrammetry was obtained from consumer DSLR cameras, and the Kinect can be run on battery power (Guilherme DeSouza, personal communication).



**Fig. 4** A row shot from different vantage points along the row. *Panel a*, a view of the entire row for orientation. *Panels b–d* show a series of close-ups shot at different vantage points within the row, starting at the right end of the row. Photographed with a DSLR

## 2.2 Assignment of organs to plants

It is usually not enough to know a field contains a phenotypically different plant: biologists need to know which particular plant has the phenotype of interest. If one considers a plant as a collection of organs, then the challenge of assignment is to form the correct set of organs from an individual plant, for any number of plants. Figure 5 shows a row containing plants with an unusual leaf phenotype. In the image, the plants with spotted leaves are marked by red triangles. How many plants have spotted leaves? Which plants are those? The answers depend on associating the leaves to other plant parts, most likely stems. Two tricks humans use might be helpful in algorithm development. The first is to look at the junctions the leaves make with a stem, starting from a leaf and following its path to the stem (or other organ). This is illustrated in panel (b) of Fig. 5 for first plant in the row. The second is to wiggle a stem and watch for coupled motions of its organs. In the material posted online, we include an iPad2 video of sev-



**Fig. 5** Panel a, a row containing four plants displaying a spotted leaf phenotype, marked by red triangles. Panel b illustrates one way humans assign organs to plants: the cyan arcs trace some paths between stem and leaves for a plant of interest. Shot with an iPad2

eral rows that illustrate several distinct motion components (IMG\_4655.MOV). Exploiting video would require a good understanding of the relationships between different motions and the plant parts that display them [28].

## 2.3 Phenotype identification

Biologists depend heavily on archival images from the literature to learn to identify different species and phenotypes. Much scientific value lies in detecting novel phenotypes. Comparing archival images (or one's memory of those) to the plants in front of one is the key visual step in recognizing phenotypes and determining their novelty.

Distinguishing weeds from crop plants is very important in production agriculture. Compared to the problem of identifying phenotypes from archival images, this simpler goal has already received considerable attention [12–16]. Figure 6 illustrates two situations in which such algorithms might be applied. The panel (a) shows several clear differences between weeds and crop, including height, position, color, and plant structure. Panel (b) seems more problematic for current algorithms: the weeds are more sparsely and irreg-



**Fig. 6** Two weedy situations. In panel a, a dense growth of short weeds fills the space between two rows. In panel b, a sparse growth of weeds is mixed with two semidwarf maize plants (marked by red triangles) and normal plants. Photographed with a DSLR



**Fig. 7** Two different leaf phenotypes. *Top row in panel a*, a clear spot-like phenotype; *panel b*, a disease lesion mimic phenotype. *Bottom row* archival images of the same phenotypes from MaizeGDB [29]. *Panel c*, a *csp1-NA1173* showing the same clear spots; *panel d*, an *lls1* mutant showing very similar lesions. The *top row* was shot with an iPad2; the camera(s) for the *bottom row* is unknown

ularly spaced and the maize includes plants of normal and much shorter heights.

The more challenging version of this problem is to exploit the information in archival images to identify phenotypes. Figure 7 shows two very different leaf spot phenotypes. Similar phenotypes can be found in the online resources of MaizeGDB, and we have included a link to a zip file of mutant images collected by Gerald Neuffer in the Appendix's Table 1.

Each biologist's image was taken to illustrate a particular phenotype, without considering computational processing. Neither the images of Fig. 7 nor those in MaizeGDB or the literature are standardized in composition, photographic technique, or annotation. Some images isolate individual plants or organs; others include several plants in the same frame for comparison; still others show rows. Finding common and distinguishing features among large sets of images, with each phenotype represented by a relatively small number of images, will be quite challenging. Nonetheless, many funda-

mental elements are repeated among the images, increasing the sample size despite compositional diversity. Learning to recognize organs such as leaves, stems, tassels, and ears would open the door to identifying many phenotypes, and with refinement might be extended to smaller scale phenotypes.

### 3 Constraints on image collection and processing

The images shown here and in the online material are sobering from the standpoint of computer vision. The challenges described above all need algorithms that are robust to the images one can actually take today and that yield biologically useful data. Because image collection is not the primary task of the biologist, it must be simple, easy, and fast. So the most common camera used in the field is a point-and-shoot (often a smart phone). Very few image sets are consistent in composition, internal standards, photographic parameters, or lighting, and rarely do images include a calibration standard. Sustained collaborations between biologists and computer vision scientists might change this situation.

The coming era of robotic collection of images from the air and the ground will surely increase the number of images, but these images may pose similar challenges to computer vision. Already, there is considerable experimentation with large, tractor-based platforms that carry a set of sensors; traditional remote sensing; and aerial vehicles [6, 15, 30, 31]. Robots offer a wider range of imaging frequencies and techniques, opening new algorithmic possibilities [15, 32].

#### 3.1 Not every photographic issue can be ameliorated

Many things that would simplify a computational problem change or eliminate the phenotypes of biological interest. For example, increasing the space among plants to simplify disambiguation decreases plant height: crowded plants must grow taller to capture enough sunlight [33]. Good places for diagonal shots across rows, such as in Fig. 3, are rare in most fields: plants must be held out of the line of sight (see images 333–360 in the online material for some standard corn photography trickery) or simply mown down.

Other photographic difficulties are manageable at the cost of development effort, personnel, time, machinery, or all of these. For example, irregular shading could complicate segmentation and assignment. Figure 8 shows several examples of images that are easy to collect, but could be algorithmically challenging.

#### 3.2 The phenotype of interest strongly influences data collection procedures

In phenotype identification, the nature of the target phenotype will strongly influence image composition and the scale and



**Fig. 8** Examples of irregular shading. *Panel a* the light green patches on the leaf vanish as the leaf is illuminated by the sun on the right. Color mottling due to uneven illumination is visible on the leaf below. The identifying tag has been whited out to preserve the investigator's privacy. *Panel b* the leaf is evenly illuminated, but the soil is not; so the white–brown leaf boundary one might use to mask the leaf tends to vanish in the sunny patch. *Panel c* shadows cast on leaves, stems, soil, and bag by the plants could complicate segmentation. Photographed with an iPad2

rate of image collection. Subtle changes at small spatial scales suggest close-ups, simplifying masking out extraneous parts of the image. Whole-plant phenotypes, such as height, could be imaged in wide fields of view capturing multiple rows, but now the issues of disambiguation and assignment recur. Tracking changes in a phenotype over time means imaging the same plants' organs, and achieving photographic consistency is much more laborious.

A complementary approach to field images is to remove plant organs or products and image these in the laboratory, either by photography or scanning. Such destructive sampling speeds image collection and facilitates more consistent image composition and lighting, permitting simpler algorithmic approaches. This approach has been used to size and count kernels, measure ear dimensions, or identify leaf lesions [21, 34–36] and (Nathan Miller, personal communication). Specialized laboratory equipment, such as microscopes and systems to image roots grown in transparent media, is used to generate image series for morphometric measurements and 3D reconstruction [20, 37].

For any phenotype, imaging demands good engineering of the data collection regime, whether the images are stills, time-lapse, or video; and whether manually or robotically collected. Sample sizes must be adequate to ensure reasonable levels of statistical confidence in the results, so image collection procedures need to be fast enough to be feasible. Since the rate of phenotype development can vary widely, pilot experiments may be needed to determine a reasonable sampling protocol.

### 3.3 How much biological knowledge is really needed?

Lurking behind these challenges is the question of how much biological knowledge is needed to tune collection schemes and identify regions of interest and phenotypes. In some cases, the knowledge needed is fairly minimal. For example, approximate models of the plant's anatomy, perhaps one for each organ or plant feature of interest, could be used to produce best fits in assignments. Robotically “feeling” imaged plants along their stems would help in tuning such models and fitting them to images. Another example is imaging along a row, where the ambiguities that must be resolved to produce good registration change. Knowing how consistent the biological structures are could help with selecting regions to align. In other cases, more knowledge of both the target phenotype and the appearance of normal plants is needed: detecting broken stems, insect bites, or cankers requires some sort of model of expected plant morphology. Shredded leaves, such as those produced by the *Shr\*<sup>-</sup>N2477*, *Shr\*<sup>-</sup>N2483*, and *shr1-JH87* mutant alleles, exemplify how

a phenotype difficult to directly image might yield to a clever proxy based on biological knowledge. In plants with these mutations, the leaf decomposes into long thin strips, joined at their ends, that occupy a large volume [38–40]. Measuring the reduced green area or smaller amplitude, higher period motions in the volumes the leaves are expected to occupy might be good proxies for detecting the shredded phenotype.

Changes in multiple dimensions may signal a phenotype of interest or be the result of normal plant development. Which combinations of dimensions are most informative varies with the phenotype, and may not be fully generalizable. Since even genetically identical plants do not look or behave exactly the same, recognizing the significant variations requires the biologist look at many plants and remember their appearance, in the context of the known biological relationships. Such knowledge of inconsequential variation could be useful in thresholding changes.

### 3.4 Use by biologists

Diffusion of algorithms that solve these challenges into the biological community will hinge on how easily they can be incorporated into field workflows. High-throughput phenotyping depends just as much on organizing the entire workflow and maintaining the provenance of physical objects and data as it does on computer vision [6, 30, 41]. Currently, one difficulty is that workflow management for field experiments is in its infancy. Field phenotyping magnifies the organizational challenges compared to greenhouse-based systems, which usually include pre-packaged workflow and data management systems [42]. There have been several generations of both workflow and interoperability systems, but to the best of our knowledge their application so far has been limited to molecular data collected in the laboratory [43–48].

In the face of such moving targets, a brief description of the non-image information biologists collect may provide some perspective. Data collection, transfer, storage, pre-processing, phenotype extraction, and generation of quantitative data are essential steps in the phenotyping workflow. Data include locational information on the fields, rows, and plants (both GPS and relative positions); reference points for measurements; weather and other environmental data; field sensors; genotypic and physiological data from the laboratory; and detailed protocols for collecting each type of data. The ability to easily cross-reference data, images, and descriptions from other projects and servers around the world will be increasingly important. The present state of the art is mostly clicking, with model organism databases supplying some cross-referencing as their resources permit (Mary Schaeffer, personal communication). All of these require planning on the front end to determine the structure of the data collected and the desired connections to be made; to preserve provenance information throughout the workflow; to maxi-

mize the scalability of the databases and computation servers; to define the quantified phenotypes; and to ensure all participants are trained. Shared cyberinfrastructure, such as the iPlant project, will prove crucial in support and in training investigators [49].

## 4 Online image sets

In cooperation with several maize geneticists, MaizeGDB, and iPlant, we have made several sets of images available. Table 1 of the Appendix lists the URLs, photographic subject, one or more computational challenges one might explore with these images, and the images' contributors. Navigating to the root URL will show either a directory of image files (nearly all) or the zipped file (the Neuffer phenotype images at MaizeGDB).

While the images do not include benchmarks, they should provide a preliminary venue for experimentation when considered with this paper. We have included a variety of ground-based images, and MaizeGDB's images are annotated with phenotypic descriptions that identify the target phenotype in the image. Browsing the image sets lets one rapidly explore potential problems and approaches. In many cases, we have included multiple images of the same subject in case the slight motions of the subject offer some algorithmic possibilities.

## 5 Prospects

We do not minimize the difficulties of these challenges. Solving these, directly or by having better ideas, will require the collaboration of a wide variety of specialists and interdisciplinary workers. The rewards for even modest improvements in our ability to characterize phenotypes in the field at higher speeds and better discrimination are both very great and very timely. Crop improvement is necessary in increasing food security, though many socioeconomic factors must also change to meet the expanding needs of the world's people [1]. High-throughput phenotyping in the field is pivotal to crop improvement. Come help.

**Acknowledgments** We are grateful to Frank Baker, Peter Balint-Kurti, Guilherme DeSouza, Nathan Miller, Martha Narro, Gerry Neuffer, John Portwood, Mac, Mary Schaeffer, James Schnable, Ann E. Stapleton, and Vinny for enlightening discussions. Drs. Balint-Kurti and Stapleton graciously allowed us to photograph their fields in the summer, 2015 field season. The wonderful iPlant team helped make our images publicly available. We gratefully acknowledge grants from the U. S. National Science Foundation (MCB-1122130), the University of Missouri Research Board to T. K.; and from the Vietnam Education Foundation Fellowship and the International Peace Fellowship to L. N.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)



ons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## 6 Appendix

See Table 1.

**Table 1** Synopsis of online images

URL and image file	Description	Possible computational goal
1/field_evaluatn/*		
DSC_0003.tif	Close-up of maize row	Disambiguation, assignment
DSC_0005.tif	A field of maturing corn: multiple rows with and without mutants; overhead wires and other background noise	Detection mutant phenotypes, developmental stage
dsc_0014.tif	Close-up of maturing ear, base of leaf, and stem	Assignment, organ identification
dsc_0021.tif	Close-up of younger ear, base of two leaves, and stem	Assignment, organ identification
dsc_0028.tif	Close-up of top of plant with tassel about to emerge from the whorl and upper leaves in "prayer position"	Assignment, organ identification
dsc_0031.tif	Close-up of exposed tassel before anther emergence (dehiscence)	Assignment, organ identification
dsc_0034.tif	Close-up of another tassel against the sky, with a different configuration of surrounding leaves	Assignment, organ identification
dsc_0035.tif	Close-up of tassel center spike with numerous anthers	Organ identification
dsc_0130.iff	Well-isolated close-up of tassel against the sky	Organ identification
IMG_1373.JPG	Row of very young plants	Disambiguation, assignment
IMG_1494.JPG	Row of mature corn with mutant plant	Disambiguation, assignment, identification
1/morphometry/*		
DSC_0023.tif	Edge view of field with multiple rows of maize of two different heights and color	Disambiguation, assignment, identification

**Table 1** continued

URL and image file	Description	Possible computational goal
DSC_0272.tiff <sup>†</sup>	Isolated plant against a blue background	Assignment, organ identification
DSC_0287.tif	Close-up of row of plants showing bottom of stems and lower leaves	Disambiguation, assignment
DSC_3877.tif	Close-up of tassel with anthers	Assignment, organ identification
DSC_2322.JPG	Shade avoidance: edge view of field with multiple rows of maize of two different heights and color; irrigation pipe and alley	Disambiguation, assignment
1/phenotypes/*		
DSC_0105.tif	Row of dwarf and wild-type plants; behind it, a row of yellow mutant plants	Disambiguation, assignment, phenotype identification
DSC_0109.tif	Close-up of dwarf plant with plant tag; adjacent wild-type plant	Assignment, organ and phenotype identification
DSC_0319.tiff <sup>†</sup>	Row with purple mutant plant in foreground	Disambiguation, assignment, organ and phenotype identification
IMG_1390.JPG	Close-up of two young plants, one albino, the other wild-type	Assignment, phenotype identification
IMG_1403.JPG	Two rows of young plants, wild-type and height mutants	Disambiguation, assignment, phenotype identification
IMG_1409.JPG	Plant with striped leaves	Disambiguation, assignment, phenotype identification
IMG_1449.JPG	Bird's nest in tassel; a confusable for identification	Disambiguation, assignment
IMG_2016.JPG	Row with wild-type and dead mutant plants	Disambiguation, assignment, identification
IMG_2020.JPG	Another row with wild-type and dead mutant plants	Disambiguation, assignment, identification
1/disease/*		
DSC_0137.tif	Blighted tassel and upper leaf; blue background	Identification
IMG_0662.JPG	Lesioned leaves of young plant; hand and irrigation pipe	Assignment, phenotype identification
DSC_5694.tif	Cicada and plant tag on leaf; confusables	Phenotype identification

Table 1 continued

URL and image file	Description	Possible computational goal
IMG_1941.JPG	Field with wild-type and dying plants	Disambiguation, assignment, phenotype identification
1/time_course/*.NEF	In situ images of leaf developing lesions; fiducials on leaf; blue background, XRite color checker, identifying tags, carpet tape	
1.NEF	First; darker white balance	Registration, phenotype identification
2.NEF	Second; better white balance	
3.NEF	Third; lesions growing	
4.NEF	Fourth; lesions growing, new lesions forming	
5.NEF	Last; multiple lesions	
1/series/DSC_0*.NEF		
155, 156	Shootbagable ear	Part of ear growth series; organ detection
162–170	Young ear	Part of ear growth series; organ detection
171–175	Stem with concealed shoot	Part of ear growth series; organ detection
176, 177	Over-exposed tassel against row and sky	An example of bad photographic technique
178–183	Better exposure; shows low branching angle and branch number, also yellow anthers	Tassel variation; organ detection
184–192	Anther close-up	Anthers and glumes; organ detection
193–196	Different tassel geometry and color (very M14-like); image crowded with other rows	Tassel variation; organ detection
197–202	Tassel w/spread geometry and yellow anthers; blurring of bkrd due to focal depth	Tassel variation, photo techniques; organ detection
203–208	Close-up of tassel, bkrd very blurred	Tassel variation, photo technique; organ detection
209–264	Old ear, crossed ear, crowded visual field; slight relative motion in some leaves	Development of ear series; disambiguation
265–285	Same as above, but different depths of field	Depth of field series; possible 3D reconstruction

Table 1 continued

URL and image file	Description	Possible computational goal
286–289	Several relatively sparse rows with low weeds, shot along row, fixed vantage point (288 is Fig. 3b)	Row appearance; disambiguation
290–297	Multiple rows, transverse shot, with sparse planting and low weeds	Weeds vs corn; identification
298–300	Two rows flanking heavier weeds	Rows, weeds; identification, disambiguation
301–310	Parts of five rows, uneven planting density, shoot-bagged ears, some weeds	Plant density and orientation; disambiguation, assignment
311–318	Down a file of rows, sparse weeds	Plant density; disambiguation, assignment
319–332	Sparse weeds, shoot-bagged ears, weird leaf arrangement in center	Plant density, ear heights; assignment
333–352	Photographic trickery needed for isolated plant shot; maturing ear	Time-consuming photographic techniques
353–355	More trickery, also shoot bags, cross bags	Time-consuming photographic techniques
356–360	More trickery	Time-consuming photographic techniques
361–371	6 Dense rows, shoot bags and tassel bags set up (369 is Fig. 3a)	Disambiguation, assignment, reproductive status; organ detection
372–382	Well-isolated row, shot along the row, few weeds, shows density, ear height, confusing shoot bag on ground (372 is Fig. 4a)	Disambiguation, assignment
383–385	Great shoot bags showing ear height (383 is Fig. 4b)	Photographic technique; disambiguation, assignment
386–396	Shot along row (386 is Fig. 4c; 393 is Fig. 4d)	Disambiguation, assignment
397–402	Two shoot-bagged ears	Development of ear series; organ detection
403–413	Along a shoot-bagged row	Plant density; disambiguation, assignment
414–426	Along a row, shoot bags and cross bag	Disambiguation, assignment

Table 1 continued

URL and image file	Description	Possible computational goal
427–429	Close-up of cross bag	Control for assignment
430–448	A little vantage point series along the same row, into next row, close-up	Control for assignment
449–451	Not every plant whose stem is in a row is really in that row!	Assignment
452–458	Weeds, cross bags, semidwarf, multiple rows (456 is Fig. 6b)	
459–467	Dense weeds between rows (461 is Fig. 6a)	Identification
468–470	Ideal maize plant before weeding	Introduction to maize; disambiguation
471–487	Ideal maize plant with crummy backdrop (483 is Fig. 2b)	Introduction to maize; disambiguation
488–493	Ideal maize plant in context (490 is Fig. 2a)	Introduction to maize; disambiguation
494–496	End-on view of field, with purple border plants in front	Changing vantage points; disambiguation
497–499	Border rows with vertical elements in background	Changing vantage points; disambiguation
500–504	Diagonal shot across border rows	Density; disambiguation
505–531	Two different phenotypes, side by side and close-ups of each, against field and sky	Identification
1/originals/*	Original images used in the manuscript	

Table 1 continued

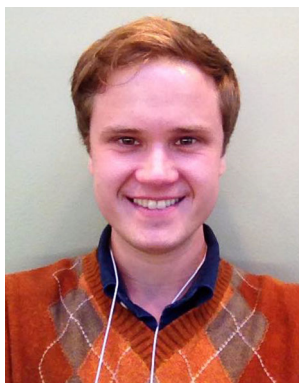
URL and image file	Description	Possible computational goal
DSC_0001.NEF	Figure 1b	
DSC_0011.NEF	Figure 1a	
DSC_0288.NEF	Figure 3b	
DSC_0369.NEF	Figure 3a	
DSC_0372.NEF	Figure 4a	
DSC_0383.NEF	Figure 4b	
DSC_0386.NEF	Figure 4c	
DSC_0393.NEF	Figure 4d	
DSC_0456.NEF	Figure 6b	
DSC_0461.NEF	Figure 6a	
DSC_0483.NEF	Figure 2b	
DSC_0490.NEF	Figure 2a	
IMG_4496.jpg	Figure 8a	
IMG_4546.JPG	Figure 8b	
IMG_4560.JPG	Figure 7b	
IMG_4569.JPG	Figure 8c	
IMG_4614.JPG	Figure 7a	
IMG_4616.JPG	Figure 5a	
csp_mgdb.jpg	Figure 7c	
lls1_mgdb.jpg	Figure 7d	
1/video/IMG_4655.MOV	Border rows moving in a slight breeze, multiple motion components	Disambiguation, assignment
2 <sup>††</sup>	Archival images of thousands of maize phenotypes	Identification

All or part of the URLs are abbreviated as 1 ([http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/tonikazic/field\\_phenotyping\\_repo](http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/tonikazic/field_phenotyping_repo)) and 2 ([http://ftp.maizegdb.org/MaizeGDB/FTP/Neuffer\\_Mutant\\_Images/Neuffer\\_Mutant\\_Images.zip](http://ftp.maizegdb.org/MaizeGDB/FTP/Neuffer_Mutant_Images/Neuffer_Mutant_Images.zip)). Images contributed by other investigators are marked with superscripts: <sup>†</sup> for Kristen Leach and David Braun, University of Missouri; <sup>††</sup> for M. Gerald Neuffer, University of Missouri, and MaizeGDB. All images with the DSC prefix were shot with a Nikon D80 DSLR camera equipped with a MicroNikkor AF 60mm lens. Many others were shot with an iPad2; and the archival images at MaizeGDB were shot with a variety of cameras. Images with the prefix IMG were shot with an iPad2. Each image posted by the authors preserves the EXIF information with the remaining photographic details

## References

1. OECD-FAO.: OECD-FAO Agricultural Outlook 2010–2019. Tech. Rep., Organization for Economic Cooperation and Development and U. N. Food and Agriculture Organization, 2010. <http://www.oecd-ilibrary.org/agriculture-and-food/oecd-fao-agricultural-outlook-2010agoutlook-2010-en>
2. Daschle, T., Clay, J., Luck, J., Oniang'o, R.K., Penn, J.B., and Sanchez, P.A.: Agricultural Innovation in the 21st Century. Special Report. Tech. rep., DuPont Advisory Committee on Agricultural Innovation and Productivity, 2013. <http://www.slideshare.net/DuPont/agricultural-innovation-productivity-for-the-21st-century-17044720?fromaction=save>
3. Global Harvest Initiative.: 2014 Global Agricultural Productivity Report. Global Harvest Initiative (2013)
4. Cuesta Roble Consulting.: International Greenhouse Vegetable Production-Statistics. Cuesta Roble Consulting (2015). <http://www.cuestaroble.com/statistics.htm>
5. United States Department of Agriculture, National Agricultural Statistics Service.: Crop Production: 2014 summary. United States Dept. of Agriculture, National Agricultural Statistics Service, Washington, DC (2015)
6. Araus, J.L., Cairns, J.E.: Field high-throughput phenotyping: the new crop breeding frontier. *Trend. Plant Sci.* **19**, 52–61 (2014)
7. Campos, H., Cooper, M., Habben, J.E., Edmeades, G.O., Schussler, J.R.: Improving drought tolerance in maize: a view from industry. *Field Crops. Res.* **90**, 19–34 (2004)
8. Pearl, R.: The personal equation in breeding experiments involving certain characters of maize. *Biol. Bull.* **21**, 339–366 (1911)
9. Singh, A.S., Masuku, M.B.: An insight in statistical techniques and design in agricultural and applied research; lousy english. *World J. Agric. Sci.* **8**, 568–584 (2012)
10. Stuber, C.W., Edwards, M.D., Wendel, J.F.: Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop. Sci.* **27**, 639–648 (1987)
11. Biskup, B., Schar, H., Schurr, U., Rascher, U.: A stereo imaging system for measuring structural parameters of plant canopies. *Plant Cell Environ.* **30**, 1299–1308 (2007)
12. Burgos-Artizzu, X.P., Ribeiro, A., Guijarro, M., Pajares, G.: Real-time image processing for crop/weed discrimination in maize fields. *Comput. Electron. Agric.* **75**, 337–346 (2011)
13. Kiani, S., Jafari, A.: Crop detection and positioning in the field using discriminant analysis and neural networks based on shape features. *J. Agric. Sci. Technol.* **14**, 755–765 (2012)
14. Montalvo, M., Pajares, G., Guerrero, J.M., Romeo, J., Guijarro, M., Ribeiro, A., Ruz, J.J., Cruz, J.M.: Automatic detection of crop rows in maize fields with high weeds pressure. *Exp. Sys. Appl.* **39**, 11889–11897 (2012)
15. Peña, J.M., Torres-Sánchez, J., de Castro, A.I., Kelly, M., López-Granados, F.: Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. *PLoS One* **8**, e77151 (2013)
16. Romeo, J., Pajares, G., Montalvo, M., Guerrero, J.M., Guijarro, M., Ribeiro, A.: Crop row detection in maize fields inspired on the human visual perception. *Sci. World J.* **2012**, 484390 (2012)
17. Minervini, M., Schar, H., Tsafaris, S.A.: Image analysis: the new bottleneck in plant phenotyping. *IEEE Sig. Proc. Mag.* **32**(4), 126–131 (2015). doi:10.1109/MSP.2015.2434233
18. Furbank, R.T., Tester, M.: Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **16**, 635–644 (2011)
19. Großkinsky, D.K., Svendsgaard, J., Christensen, S., Roitsch, T.: Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* **66**, 5429–5440 (2015)
20. Thompson, A.M., Crants, J., Schnable, P.S., Yu, J., Timmermans, M.C.P., Springer, N.M., Scanlon, M.J., Muehlbauer, G.J.: Genetic control of maize shoot apical meristem architecture. *Genes Genom. Genet.* **4**, 1327–1337 (2014)
21. Kelly, D., Vatsa, A., Mayham, W., Kazic, T.: Extracting complex phenotypes from images. *Mach. Vis. Appl.* pp 1–12 (2015). doi:10.1007/s00138-015-0718-6
22. Bennetzen, J.L., Hake, S.C. (eds.). *Handbook of Maize: Its Biology*. vol. 1. Springer, New York (2009)
23. Neuffer, M.G., Edward H. Coe, Jr., Wessler, S.R.: *Mutants of Maize*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (CHS) (1997)
24. Vinã, A., Gitelson, A.A., Rundquist, D.C., Keydan, G., Leavitt, B., Schepers, J.: Monitoring maize (*Zea mays* L.) phenology with remote sensing. *Agron. J.* **96**, 1139–1147 (2004)
25. Kaleita, A.L., Steward, B.L., Ewing, R.P., Ashlock, D.A., Westgate, M.E., Hatfield, J.L.: Novel analysis of hyperspectral reflectance data for detecting onset of pollen shed in maize. *Trans. Am. Soc. Agric. Biol. Eng.* **49**, 1947–1954 (2006)
26. Frasson, R.P., Krajewski, W.F.: Three-dimensional digital model of a maize plant. *Agric. For. Meteorol.* **150**, 478–488 (2010)
27. Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, É., Chapeau-Blondeau, F.: On the use of depth camera for 3D phenotyping of entire plants. *Comput. Electron. Agric.* **82**, 122–127 (2012)
28. Dixon, M., Abrams, A., Jacobs, N., Pless, R.: On analyzing video with very small motions. In: *IEEE Computer Society (ed.)*, The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011, New York, pp. 425–432, IEEE Computer Society Press (2011)
29. MaizeGDB.: MaizeGDB. U. S. Department of Agriculture, Agricultural Research Service (2003–present). <http://www.maizegdb.org/>
30. White, J.W., et al.: Field-based phenomics for plant genetics research. *Field Crops Res.* **133**, 101–112 (2012)
31. Liebisch, F., Kirchgessner, N., Schneider, D., Walter, A., Hund, A.: Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Meth.* **11**, 9 (2015)
32. Li, L., Zhang, Q., Huang, D.: A review of imaging techniques for plant phenotyping. *Sensors* **14**, 20078–20111 (2014)
33. Sawers, R.J.H., Sheehan, M.J., Brutnell, T.P.: Cereal phytochromes: targets of selection, targets for manipulation? *Trends Plant Sci.* **10**, 138–143 (2005)
34. Martin, D.P., Rybicki, E.P.: Microcomputer-based quantification of maize streak virus symptoms in *Zea mays*. *Phytopathology* **88**, 422–427 (1988)
35. Foard, A., et al.: Collection of quantitative images of leaves in the field and greenhouse. *Maize Genet. Coop. News.* **82**, 13–15 (2008)
36. Barbedo, J.G.A.: An automatic method to detect and measure leaf disease symptoms using digital image processing. *Plant Dis.* **98**, 1709–1716 (2014)
37. Spalding, E.P., Miller, N.D.: Image analysis is driving a renaissance in growth measurement. *Curr. Opin. Plant Biol.* **16**, 100–104 (2013)
38. MaizeGDB.: MaizeGDB Phenotype Record Page: 107757. MaizeGDB (2015–present). [http://www.maizegdb.org/data\\_center/phenotype?id=107757](http://www.maizegdb.org/data_center/phenotype?id=107757)
39. Neuffer, M.G.: First Report of Shredded Leaf Phenotype. MaizeGDB (1995). [http://www.maizegdb.org/data\\_center/reference?id=106353](http://www.maizegdb.org/data_center/reference?id=106353)
40. Trimmell, M., Albertsen, M.C., Noble Jr, S.W.: New leaf mutation shr\*<sup>-</sup>JH87, shredded leaf. *Maize Genet. Coop. News.* **74**, 36 (2000)
41. Kazic, T.: Ten simple rules for experiments' provenance. *PLoS Comput. Biol.* **11**, e1004384 (2015). doi:10.1371/journal.pcbi.1004384
42. LemnaTec GmbH.: LemnaTec. LemnaTec GmbH (2012–present). <http://www.lemnatec.com>

43. Altinas, I., Berkley, C., Jaeger, E., Jones, M.B., Ludäscher, B., Mock, S.: Kepler: an extensible system for design and execution of scientific workflows. In: Proceedings of the 16th international conference on scientific and statistical database management (SSDBM'04), Santa Barbara, CA, IEEE (2005)
44. BioMOBY.org.: Moby. BioMOBY.org (2005). <http://www.biomoby.org>
45. myGrid Developers.: myGrid. mygrid.org.uk (2014–present). <http://www.mygrid.org.uk>
46. Goble, C.A., Stevens, R., Ng, G., Bechhofer, S., Paton, N.W., Baker, P.G., Peim, M., Brass, A.: Transparent access to multiple bioinformatics information sources. *IBM Syst. J.* **40**, 532–552 (2001)
47. Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M.R., Li, P., Oinn, T.: Taverna: a tool for building and running workflows of services. *Nucl. Acids Res.* **34**, W729–W732 (2006)
48. The Pegasus Project.: Pegasus Workflow Management System. Information Sciences Institute, University of Southern California (2015–present). <http://Pegasus.isi.edu>
49. iPlant Collaborative.: iPlant Collaborative. Empowering a New Plant Biology. Cold Spring Harbor Laboratory and University of Arizona (2008–present). <http://iplantcollaborative.org/>



**Derek Kelly** received his B.S. in Anthropology and Biological Sciences from the University of Missouri, Columbia in 2013. He is currently a Ph.D. student in the Genomics and Computational Biology Graduate Group at the University of Pennsylvania under Dr. Sarah Tishkoff. His interests include human evolutionary genomics, regulation of gene expression, and statistical genetics.



**Avimanyou Vatsa** is pursuing his Ph.D. and received his M.S. (Computer Science) with a minor in Statistics from the University Of Missouri, Columbia in 2015. He obtained his M.Tech. (computer engineering) with Hons. from Shobhit University, Meerut and B. Tech. (I.T.) from V.B.S. Purvanchal University, Jaunpur, (U.P.), India, in 2009 and 2001, respectively. He was an assistant professor for ten years in several engineering colleges and universities in India. He was a member

of academic and administrative bodies and coordinated many technical fests and national conferences at the college and university level. He also worked as a software engineer in the software industry. He is a member and reviewer on the editorial boards of several international and national journals in networks and security. His area of research includes low- and high-dimensional data (big data), image processing, mobile ad hoc networks, computer networks, and network security.



**Wade Mayham** received his B.S. in Computer Science from the University of Missouri, Columbia in 2015. Currently, he is working as a freelance software developer and will pursue a Masters of business administration in 2016. His interests include artificial intelligence, computer networks, and virus research and defense.



**Linh Ngô** is a graduate student in the Genetics Area Program at the University of Missouri, Columbia. She received her B.S. in Crop Science from the joint program between Hanoi University of Agriculture and the University of California, Davis in 2011. Her research interests include forward genetics of complex phenotypes and analysis of ethnobotanical data.



**Addie Thompson** is a post-doctoral research associate in the Department of Agronomy at Purdue University. She received her Ph.D. in 2014 in Applied Plant Sciences: Plant Breeding and Molecular Genetics from the University of Minnesota, and B.S. in Genetics from Iowa State University in 2008. Addie's area of expertise is quantitative genetics and trait mapping in maize. Her current studies also include high-throughput phenotyping in maize and sorghum.



**Toni Kazic** is a computational biologist and an associate professor in the Department of Computer Science, University of Missouri. She is also a member of that University's Interdisciplinary Plant Group, Informatics Institute, Missouri Maize Center, and Genetics Area Program. Her research is focused on understanding complex phenotypes and using these to infer biological networks and optimization of selection strategies. A geneticist by training, she is finishing her ninth year working in her corn field and thinking about the computational and mathematical challenges maize poses.

ishing her ninth year working in her corn field and thinking about the computational and mathematical challenges maize poses.