

SOFTWARE

Open Access



# GRABSEEDS: extraction of plant organ traits through image analysis

Haibao Tang<sup>1\*</sup>, Wenqian Kong<sup>2</sup>, Pheonah Nabukalu<sup>3</sup>, Johnathan S. Lomas<sup>4</sup>, Michel Moser<sup>5</sup>, Jisen Zhang<sup>6</sup>, Mengwei Jiang<sup>7</sup>, Xingtian Zhang<sup>7</sup>, Andrew H. Paterson<sup>2\*</sup> and Won Cheol Yim<sup>4\*</sup>

## Abstract

**Background** Phenotyping of plant traits presents a significant bottleneck in Quantitative Trait Loci (QTL) mapping and genome-wide association studies (GWAS). Computerized phenotyping using digital images promises rapid, robust, and reproducible measurements of dimension, shape, and color traits of plant organs, including grain, leaf, and floral traits.

**Results** We introduce GRABSEEDS, which is specifically tailored to extract a comprehensive set of features from plant images based on state-of-the-art computer vision and deep learning methods. This command-line enabled tool, which is adept at managing varying light conditions, background disturbances, and overlapping objects, uses digital images to measure plant organ characteristics accurately and efficiently. GRABSEED has advanced features including label recognition and color correction in a batch setting.

**Conclusion** GRABSEEDS streamlines the plant phenotyping process and is effective in a variety of seed, floral and leaf trait studies for association with agronomic traits and stress conditions. Source code and documentations for GRABSEEDS are available at: <https://github.com/tanghaibao/jcvi/wiki/GRABSEEDS>.

**Keywords** Image analysis, Phenotype, Seed traits, High throughput, QTL mapping

\*Correspondence:

Haibao Tang  
tanghaibao@gmail.com  
Andrew H. Paterson  
paterson@uga.edu  
Won Cheol Yim  
wyim@unr.edu

<sup>7</sup> National Key Laboratory for Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, Guangdong, China

<sup>1</sup> Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Haixia Institute of Science and Technology and College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

<sup>2</sup> Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30605, USA

<sup>3</sup> The Land Institute, 2440 E Water Well Road, Salina, KS 67401, USA

<sup>4</sup> Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557, USA

<sup>5</sup> Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland

<sup>6</sup> State Key Lab for Conservation and Utilization of Subtropical Agro-Biological Resources, Guangxi Key Lab for Sugarcane Biology, Guangxi University, Nanning 530004, Guangxi, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Quantitative Trait Locus (QTL) mapping and genome-wide association studies (GWAS) are crucial in unraveling the genetic underpinnings of phenotypic traits in various plant parts, such as seeds, flowers, leaves, fruits, and nuts [1–5]. Research consistently shows that genome-wide association studies (GWAS) are effective for identifying and precisely mapping QTLs associated with complex agricultural traits. Precise morphological measurements are essential in genetic research, yet they present challenges due to their labor-intensive and time-consuming nature [6, 7]. Additionally, there can be considerable variability in measurement and handling methods which contributes to data inconsistency. Non-uniform operating protocols among data collectors further impacts reliability and reproducibility [8]. Thus, improving the efficiency of these measurement techniques is crucial to effectively utilize the extensive genetic resources available in both experimental and agricultural plant systems [9]. These improvements are essential for enhancing our understanding of plant genetics and successfully applying this knowledge to progress in agriculture [10].

The application of computer-aided image analysis is revolutionizing large-scale phenotyping experiments by incorporating image processing and machine learning techniques. This advanced approach enables the extraction of critical features such as size, shape, and color from digital images, facilitating the mapping of quantitative traits vital for agricultural research [11]. In particular, machine learning algorithms such as convolutional neural networks, are broadening applications in high-throughput plant phenotyping and accelerating the elucidation of gene functions associated with traits in model plants [12, 13]. Notably, integration of image processing in various agricultural settings has illustrated how robotics and computer vision are useful for automatic phenotyping [14].

Among the major plant organs, seeds are a primary focus due to their fundamental role in the agricultural production of grains and legumes. Seed attributes such as weight, texture, and shape are closely linked to crop yield and germination effectiveness and are therefore subject of intensive breeding programs [15]. Automated seed identification through image processing algorithms has important applications in the quality control of seed production and for harvest classification [16].

A variety of image processing applications are available, each designed for recognizing different biological elements. For instance, ImageJ is an image processing program developed at the National Institutes of Health which was designed to manipulate and process microscopy images [17]. BISQUE provides a web-based

framework to create, share, and analyze images, with options for tailored analyses [18]. CellProfiler excels in quantifying structures like yeast colonies and mouse tumors, and also in evaluating tissue topology [19]. SHERPA is adept at processing vast numbers of diatom micrographs through image segmentation techniques [20]. Nonetheless, these tools are primarily developed for microscopy imagery rather than standard digital photography, with limited capabilities in color adjustment or label recognition. The use cases in agriculture necessitate user-generated algorithms or macros for batch operations, which can be challenging for non-specialists [19, 20].

In recent years, substantial progress has been made in plant phenotyping through the use of digital images, highlighting the increasing importance of digital tools in plant science [21]. For example, Plant phenotyping using Computer Vision (PlantCV) provides a comprehensive and flexible toolkit for complex plant image analysis [22]. Machine learning has also been integrated into advanced phenotyping systems, enabling precise plant segmentation and analysis [15, 23–25]. Techniques that combine thermal and visible images for detecting plant stress, offering comprehensive evaluations of plant health, were introduced [26–28]. The application of close-range hyperspectral imaging has been pivotal in digital phenotyping, particularly in addressing recent challenges such as illumination correction, which is crucial for accurate phenotypic analysis [29]. Additionally, the Digital Imaging of Root Traits (DIRT)/3D is a groundbreaking platform that uses image-based 3D technology for phenotyping root traits [30]. Moreover, the development of cost-effective, Raspberry Pi-powered imaging systems has enabled high-throughput phenotyping, suitable for a wide range of plant applications [31]. An online database has been created for plant image analysis software, aimed specifically at meeting the needs of plant science [32, 33]. These advancements underscore the continuous evolution of computer vision technologies in plant phenotyping, reflecting the growing sophistication of imaging techniques in this field.

We present GRABSEEDS, an advanced software tool specifically engineered to accelerate the identification and phenotyping of plant seeds, leaves and flowers, with demonstrated efficacy across a wide range of grains and legumes [16]. GRABSEEDS is equipped with a robust, command-line interface designed for high-throughput batch processing, ensuring rapid and accurate performance under diverse and challenging conditions, such as variable lighting, complex backgrounds, and closely clustered or overlapping seeds. This tool integrates cutting-edge image processing techniques, including adaptive color correction and intelligent label recognition, to fully

automate the seed phenotyping workflow. GRABSEEDS not only streamlines the identification, sorting, labeling, and measuring processes but also provides deep insights into the genetic architecture of related traits. By accurately characterizing critical plant organ features, GRABSEEDS enhances varietal identification and supports the discovery of key genetic traits, offering invaluable contributions to plant breeding and agricultural research.

## Implementation

### Image processing

The image processing pipeline implemented in GRABSEEDS has four core components including edge detection, object (seed) identification, image cropping and text label recognition. In this study, the “objects” specifically refer to the seeds that we seek to identify, but they could represent other general items of interest. GRABSEEDS supports several digital images, including PNG and JPEG formats. The JPEG format is the most common file output format from digital cameras. The core image processing routines within GRABSEEDS are based on the *scikit-image* Python library [34].

*Edge detection:* GRABSEEDS incorporates a suite of edge detection algorithms, such as Canny, Sobel, Roberts and Otsu’s methods [34]. The Canny edge detector, set as the default option, operates on the principle of utilizing the derivative of a Gaussian distribution to calculate image gradients. By adjusting the Gaussian’s variance, or sigma (configurable, default=1), the method effectively diminishes noise interference in the image, enhancing the Canny detector’s resilience against cluttered backgrounds.

*Object identification:* following edge detection, a closing operation is applied to mend any gaps or ‘cracks’ in the outlines of potential objects, which might result from background noise or insufficient lighting. This process ensures that all areas enclosed and 4-connected (linked to their adjacent pixels on the top, bottom, left, and right) are recognized as distinct objects. By default, any object touching the image’s border is automatically excluded from consideration. For more difficult cases, a deep learning model, Meta AI’s Segment Anything (SAM) is also available to generate the mask for each object [35].

*Image cropping:* to omit non-target elements (such as background features) from analysis, images could be cropped by directly slicing them. This technique is particularly beneficial in batch processing scenarios, allowing for the exclusion of specific areas, like those containing text labels, from the final analysis.

*Text label recognition:* GRABSEEDS uses the Google tesseract-OCR to identify and extract the text in the label [36]. To enhance the speed and accuracy of text recognition, users have the option to input a cropped area

specifically containing the label. In the batch processing, where the label’s location remains largely consistent, this approach significantly speeds up the label extraction process.

### Optimizing accuracy

There are several common issues with image quality that affect the accuracy of object recognition, including noisy background, blurred edges, heterogeneous and often overlapping objects. GRABSEEDS resolves these problems through tuning of key parameters through command line options. These adjustments typically involve balancing the software’s sensitivity and specificity for object identification and segmentation, thereby enhancing the accuracy of recognition.

*Noisy background:* for images with noisy backgrounds, such as those featuring cloth or other materials with rough textures, users can adjust the sigma values for Gaussian de-noising within the Canny edge detector. Increasing sigma values allows the algorithm to better handle background noise, although it might also increase the likelihood of overlooking smaller seeds due to the smoothing effect. Additionally, there is a feature that allows changing the background to a color that contrasts with the seed color and its complementary color, facilitating better seed detection.

*Addressing blur edges:* in cases of low lighting that lead to blurred edges, resulting in open edges that do not properly enclose regions identifiable as seeds, GRABSEEDS offers the option to set a closing morphology with a specific kernel size (the default being 2 pixels). This setting helps close ‘cracks’ with a radius of up to 2 pixels. In blurry pictures, the kernel size could be increased to refine the edge of the objects but has the additional risk of falsely connecting objects that are further apart from one another.

*Managing heterogeneous sizes:* certain objects in the background might be recognized as seeds. The “size” of certain object is defined by the number of contiguous pixels covered by the object. This could be effectively filtered out by setting minimum or maximum size thresholds. This feature directly excludes features that are incorrectly recognized as target objects based on their size.

*Separating overlapping objects:* watershed segmentation is used to separate the touching seeds [34]. This feature uses the furthest points from the detected edges as markers, and the ‘flooding’ of basins from these markers separates overlapping objects along a delineated ‘watershed’ line. This approach allows GRABSEEDS to accurately identify and delineate individual seeds even when they are in close contact, ensuring precise phenotypic measurements [37]. For the more difficult cases, the deep

learning model SAM is also available to separate background with foreground [35], but at a cost of higher computational cost.

**Visual debugging**

GRABSEEDS offers a visual debugging tool by generating a PDF document that overlays object details on top of the original image for analysis (Fig. 1). This document is structured into four sections, including the original image, edge detection results, object detection outcomes, and a list of detected objects (Fig. 1). These panels are useful in exploring the best parameters for fine-tuning the object detection. Within the Object detection panel, the contours of the identified objects are drawn, as well as the two major axes showing the length and width of the seed, respectively. In the Object list, several identified objects are listed for visual validation. This advanced level of visual debugging aids significantly in adjusting parameters for image batches. Given the consistency of camera settings across a batch, optimizing parameters based on a small sample can reliably improve accuracy across the entire set.

**Calibration**

In scenarios where photo sessions span multiple days, it might be important to calibrate before each batch of images to maintain a unified standard of feature extraction. The calibration serves several purposes. First, calibration allows the calculation of “pixel-to-inch ratio”, then the seed length and width can be converted to physical lengths such as inch or centimeter (cm). Second, calibration normalizes the effect of lighting and corrects the RGB code. Through calibration, results remain consistent despite variations in camera settings,

lighting conditions, or distance between the lens and the subject table. This step is especially vital for accurately assessing size and color traits.

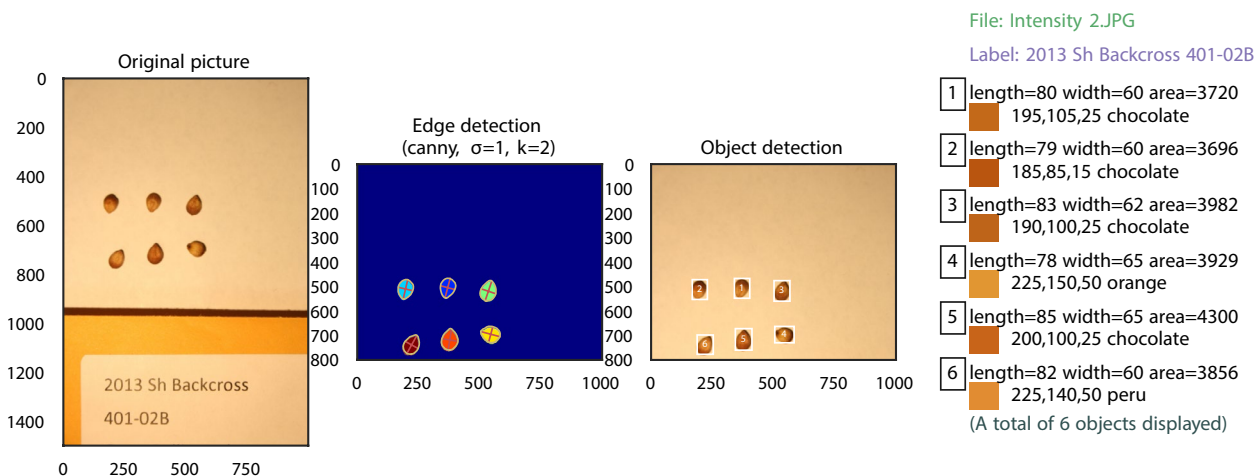
For precise calibration, the use of a “ColorChecker” to perform the calibration, which is a palette of colors with 24 prearranged color samples [38]. Users can make a ColorChecker by simply printing it out [39], and then snap a picture of the printout. Users then measure the individual boxes on the paper and record the size in squared cm units. The ColorChecker is identified by 24 boxes, the boxes are then aligned to a 6×4 grid using K-means clustering. Color correction employs a linear color transformation or “channel re-mixing” [38]:

$$\begin{pmatrix} R_{target} \\ G_{target} \\ B_{target} \end{pmatrix} = \begin{pmatrix} RR & GR & BR \\ RG & GG & BG \\ RB & GB & BB \end{pmatrix} \begin{pmatrix} R_{observed} \\ G_{observed} \\ B_{observed} \end{pmatrix}$$

This is a model that has nine parameters. Given two color points, the distance is defined by:

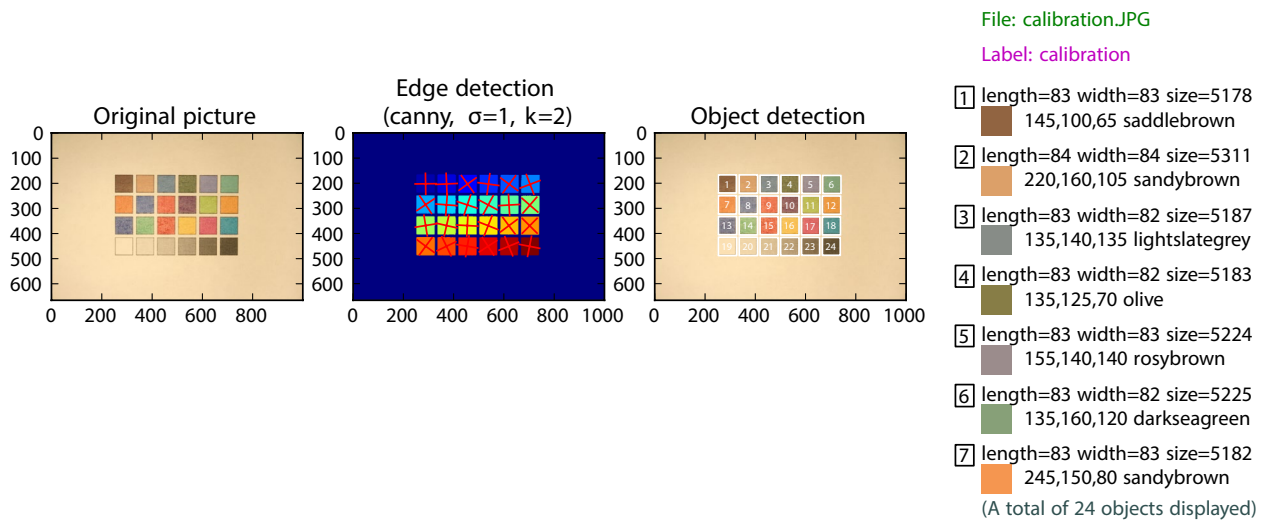
$$\Delta(C, C') := \sqrt{\sum_{i=1}^{24} [(R_i - R_i')^2 + (G_i - G_i')^2 + (B_i - B_i')^2]}$$

We then seek to find the best estimates for the nine parameters that minimizes the above distance between the ‘observed’ and the ‘target’. GRABSEEDS applies a Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm to solve the minimization problem [40]. The linear transform can then be applied to correct the ‘observed’ colors back to their original hue. Like seed image processing, color correction can be visually debugged (Fig. 2). Ensuring proper calibration is critical so that GRABSEEDS accurately recognizes most of the 24



**Fig. 1** Visual output from GRABSEEDS on sorghum seeds. The four panels (from left to right) are original picture, picture after edge detection, picture after seed detection, and a list of identified seeds and text label. The visual output provides a rapid debugging method to ensure accuracy





**Fig. 2** Visual output from GRABSEEDS on ColorChecker for calibration. The four panels (from left to right) are original picture, picture after edge detection, picture after box detection, and a list of identified boxes with their observed color. The observed color of each box is then compared with the expected color to calculate the color correction

squares, maintaining the integrity of feature extraction and analysis.

**Metrics extraction**

During batch processing, GRABSEEDS produces a composite PDF for visual inspection and generates detailed metrics for each detected object in a spreadsheet or as CSV (comma-separated values) files, facilitating subsequent statistical analysis (Fig. 3). For each object, axis length, area, circularity, and color are assigned to each recognized seed. The spreadsheet also tracks the information extracted from the EXchangeable Image file Format (EXIF) header such as the time that the photo was taken and a useful subset of camera settings. A total of 16 main attributes are extracted for each identified seed.

Dimensional data track the distance and shape measurements, acknowledging that many seeds exhibit an oval form. Thus, both the major and minor axes of the seeds are recorded. We also track shape in the ‘circularity’ metric, which is a numeric value between 0 and 1 (with

1 being perfectly circular and 0 resembling a line or an elongated shape), is calculated as below [17]:

$$Circularity = 4\pi \times \frac{Area}{Perimeter^2}$$

A richer set of shape embeddings can be derived from Elliptic Fourier Descriptors (EFDs) [41, 42]. The EFDs are a list of numbers corresponding to the calculations of Fourier power (default order of 10). The shape EFDs have the advantage of normalizing complex contours for comparisons, and are especially suitable for simplifying periodic contours, e.g. flower or leaf edges.

Seed color is determined using RGB values across red, green, and blue channels. GRABSEEDS outputs familiar color names from the "web colors" palette, such as "lightslategrey", "sandybrown", "olive", "darkseagreen", matching each seed to the closest web color (Fig. 2). The similarity between the web color and extracted color is based on the Delta E (CMC) function between any two

ImageName	DateTime	Accession	SeedNum	Location	Area	Circularity	Length(px)	Width(px)	ColorName	RGB	Length(cm)	Width(cm)	CorrectedColorName	CorrectedRGB
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	1	369 504	3662	0.87	79	59	peru	205,120,40	0.39	0.29	sandybrown	255,170,90
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	2	197 510	3628	0.88	78	60	chocolate	190,90,15	0.39	0.3	chocolate	239,132,53
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	3	536 519	3937	0.87	82	61	chocolate	200,110,30	0.41	0.3	sandybrown	252,157,76
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	4	545 694	3876	0.88	77	64	orange	230,150,50	0.38	0.32	orange	255,210,109
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	5	378 718	4232	0.87	84	64	chocolate	200,100,25	0.42	0.32	chocolate	251,145,68
Intensity_1.JPG	2014.07.01 22:57:52	2013 Sh Backcross 401O2B	6	222 738	3797	0.86	82	60	peru	230,140,50	0.41	0.3	sandybrown	255,198,108
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	1	370 505	3714	0.87	80	60	chocolate	195,105,25	0.4	0.3	sandybrown	246,151,68
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	2	197 510	3698	0.88	79	60	chocolate	185,85,15	0.39	0.3	chocolate	232,125,52
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	3	535 519	3984	0.87	83	62	chocolate	185,90,20	0.41	0.31	chocolate	232,131,59
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	4	545 695	3930	0.89	78	65	orange	225,150,50	0.39	0.32	orange	255,210,108
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	5	378 718	4299	0.87	85	65	chocolate	200,100,25	0.42	0.32	chocolate	251,145,68
Intensity_2.JPG	2014.07.01 22:58:01	2013 Sh Backcross 401-02B	6	222 739	3852	0.86	82	60	peru	225,140,50	0.41	0.3	sandybrown	255,197,107

**Fig. 3** Sample spreadsheet from GRABSEEDS on sorghum seeds. A comprehensive set of dimension, shape and color metrics are extracted from the seeds, useful for QTL analyses of seed traits

colors, which approximates human perception of color differences. Assigning discrete color names enables easier grouping and comparison of seeds by color, offering an advantage over the direct application of RGB numerical values.

## Results

### Camelina seeds

GRABSEEDS has demonstrated its robust capabilities in the image provided, specifically highlighting the phenotyping of Camelina seeds. Notable features that have been exemplified include Text Label Recognition, where the tool has successfully identified and read the label "349-3" associated with the seeds (Fig. 4). Image Cropping has been utilized to focus on the seeds, excluding extraneous background elements. Object Identification has been adeptly executed, with individual seeds being recognized and quantified despite their proximity to one another. Edge Detection, a critical component of the process, has been fine-tuned with a sigma value of 3 and a kernel size of 2, which successfully delineates the edges of the Camelina seeds against the Noisy Background. The feature also effectively addresses Blur Edges that might result from lower lighting conditions or camera focus issues, ensuring each seed is enclosed accurately for further analysis.

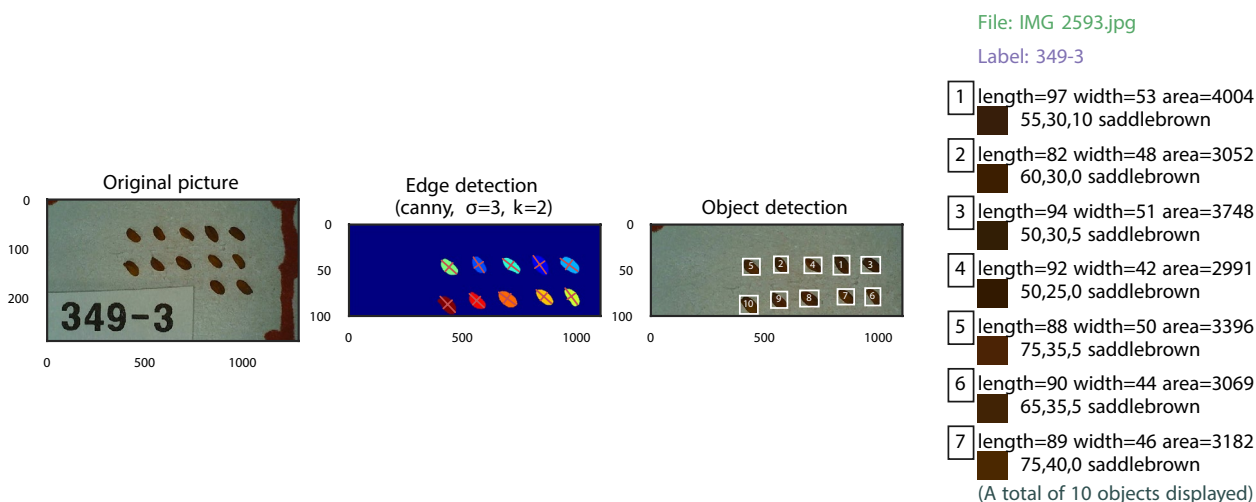
This integrated approach by GRABSEEDS, managing various challenges such as noisy or blurred backgrounds, reinforces its utility in accurately phenotyping seeds for genetic and agricultural research. The program extracts detailed metrics—length, width, and area, alongside the mostly uniform color designation of 'saddlebrown', for each Camelina seed. The lengths of the seeds span from 82 to 97 pixels, widths from 42 to 53 pixels, and areas

from 2991 to 4004 pixels. With a total of 10 objects analyzed, the tool's proficiency in processing and quantifying a cluster of seeds in a single frame is emphasized, reflecting GRABSEEDS's robust functionality in accurately extracting and highlighting the phenotypic characteristics of Camelina seeds.

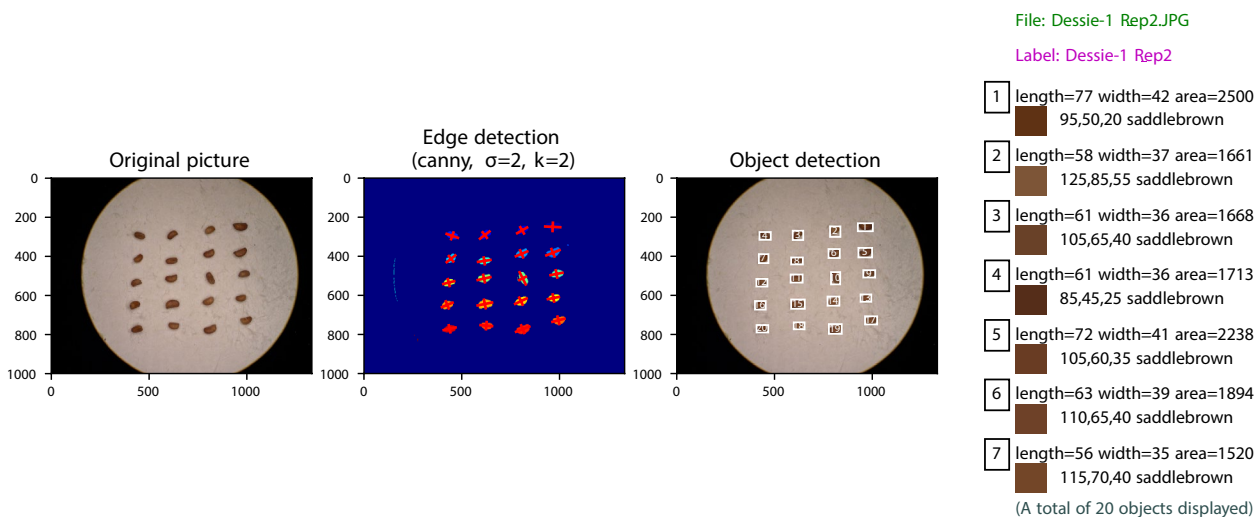
### Teff seeds

GRABSEEDS also processed images displaying several teff seeds arrayed randomly (Fig. 5). The software's edge detection, configured with a sigma value and kernel size both set to 1, effectively outlined the seeds on a blue background. The object detection has accurately enclosed each seed with a bounding box, confirming their individual identification. Among the seeds analyzed, seven detailed metrics have been provided, revealing variations in size with lengths ranging from 56 to 73 pixels and widths from 35 to 44 pixels. The areas covered by these seeds span from 1541 to 2473 pixels, and all are marked with the color 'saddlebrown', with RGB values ranging between (110,85,65) and (130,100,75). A total of 29 objects were distinguished in the analysis, showcasing GRABSEEDS's capability to perform text label recognition, handle noisy backgrounds, address blurry edges, and extract critical seed metrics in a consistent and automated manner.

The batch processing reveals a comprehensive phenotype analysis of teff seeds, capturing dimensions and color variations in significant detail. These results offer an in-depth analysis of teff seed phenotypes, detailing the intricacies of size and color. Each seed is meticulously cataloged with a unique SeedNum and exact coordinates, reflecting areas between 1442 and 2769 pixels that showcase the array of seed sizes.



**Fig. 4** Visual output from GRABSEEDS on camelina seeds



**Fig. 5** Visual output from GRABSEEDS on teff seeds

The circularity metric provides insight into the seed shapes, which vary from elongated to nearly round.

A calibrated transformation, using a PixelCMratio set at 60.14, translates the pixel measurements of seed length and width into centimeters. This critical scaling, which yields lengths between 0.91 and 1.31 cm and widths from 0.55 to 0.75 cm, ensures the accuracy of physical dimensions regardless of the camera setup used during the capture. Additionally, calibration rigorously corrects color distortions from varied lighting or camera specifics, employing RGB transformation to ensure fidelity to the seeds' true colors. Initial colors, mostly 'saddlebrown,' after correction, may become 'peru' or 'darkgoldenrod,' as evidenced by the adjusted RGB values. This adjustment not only changes the nominal color but also reflects correction based on the imaging conditions, illustrating the capability of GRABSEEDS to accurately represent phenotypic traits.

**Petunia flowers**

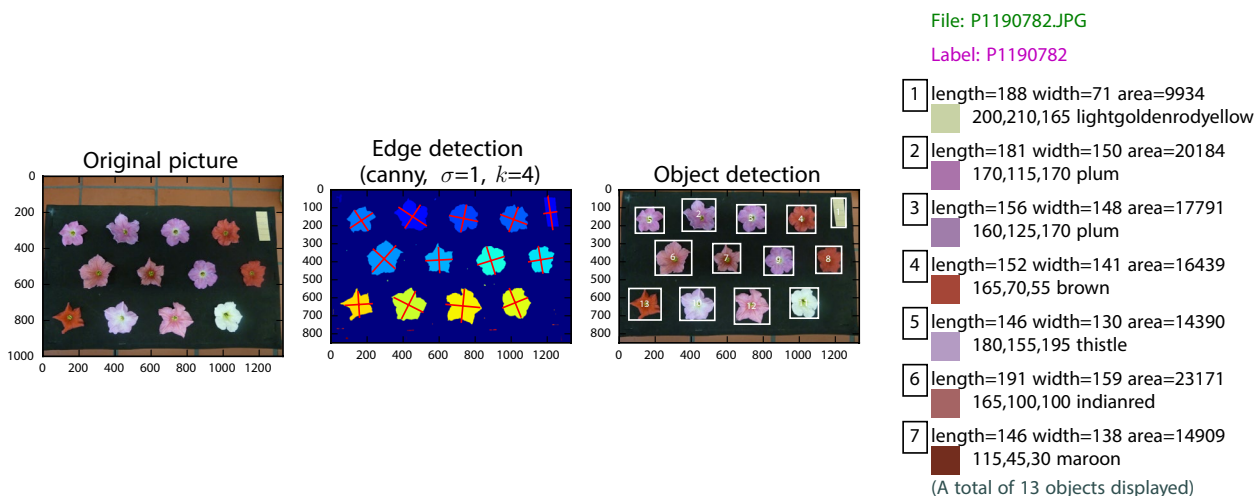
To extract color traits for QTL analysis, images of flowers were taken from mapping populations along with anthocyanin measurements. Despite the name of GRABSEEDS, flower images can also be readily processed with little modifications of protocols. Usually, images were taken from 12 flowers at a time with a guide stripe as a size reference (on the right). Using GRABSEEDS, several flowers can be isolated on the images along with the stripe (Fig. 6). The petunia flowers contain a diverse range of floral shapes (best represented as harmonic EFDs) and petal colorings.

**Tobacco leaves**

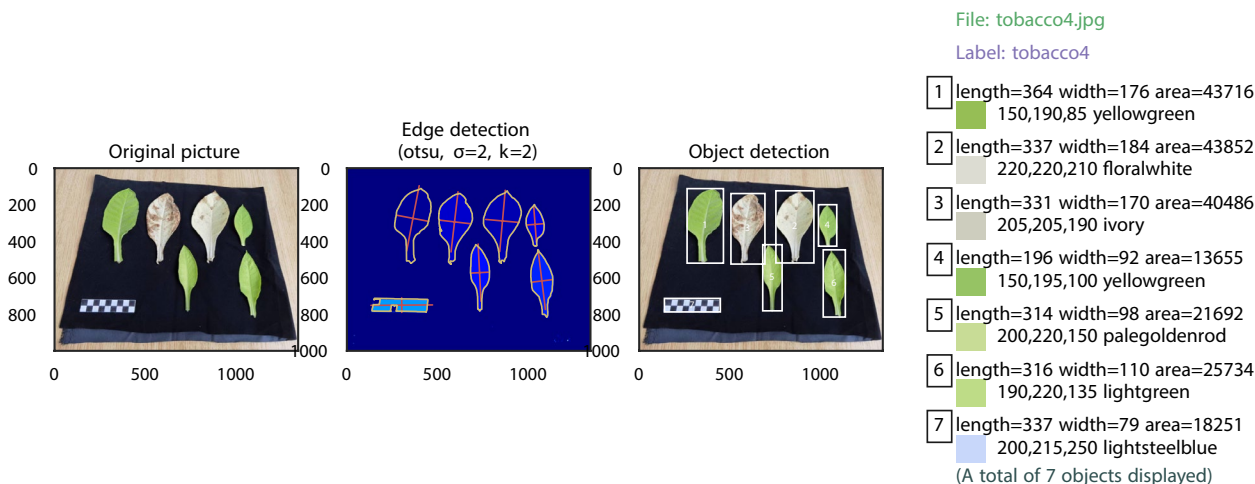
Diseased leaves often have a different shape, size, or color than healthy leaves. To identify the diseased leaves, we measured the color, area, length, and width of *Nicotiana tabacum* leaves. We collected photographs of the lesser and more severe cases of the disease as well as normal leaves. For example, leaf 1 is an example of normal leaf with yellow green color, while leaf 2 and 3 are tobacco leaves carrying disease, which can show much more pale colors, such as floral white and ivory (Fig. 7). By establishing a linear relationship between leaf color and disease severity as measured by GRABSEEDS (Fig. 7), it is possible to identify leaves at an early stage of the disease, thereby preventing potential yield losses.

**Sorghum kernels**

In a recent study, we examined the genetic and phenotypic variation and kernel size, shape and color for two sorghum BC<sub>1</sub>F<sub>2</sub> populations derived from *Sorghum bicolor* BTx623 and *Sorghum halepense* Gypsum 9E [42]. A total of 246 BC<sub>1</sub>F<sub>2</sub> families and the parents were phenotyped in terms of seed size (area, length, width, and aspect ratio), shape (circularity) and color (RGB) (Fig. 1). Associations were found between several kernel traits, including area, length, width, color, and shape EFDs with certain linkage group intervals in both populations. All the kernel traits can be collectively analyzed and viewed in a Principal Component Analysis (PCA) plot to illustrate the phenotype space of sorghum kernels (Fig. 8).



**Fig. 6** Visual output from GRABSEEDS on petunia flowers



**Fig. 7** Visual output from GRABSEEDS on tobacco leaves. Leaves 2 and 3 carry more pale numbers compared to the remaining healthy tobacco leaves, that are primarily identified as yellow green

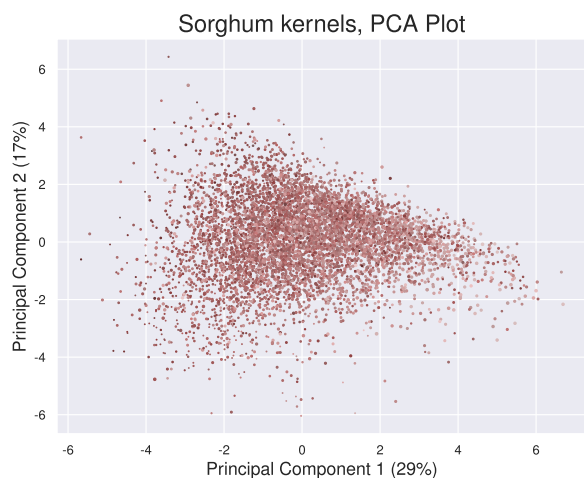
### Discussion

Achieving high recognition accuracy is a primary goal in the development of GRABSEEDS. Accuracy, which is measured by both sensitivity (true positive rate) and specificity (true negative rate), is essential for reliable seed phenotyping. Although GRABSEEDS offers a seamless and automatic pipeline for seed phenotyping, the sensitivity of seed recognition can be challenged by low-quality images, which may be affected by poor lighting, low contrast, or noisy backgrounds.

To overcome these issues, GRABSEEDS integrates advanced filtering techniques, including the Canny Gaussian filter, adjustable Gaussian sigma values, and optimized closing procedures, which together enhance

image quality and boost recognition accuracy. Specificity is further refined by employing size filters that effectively distinguish seeds from non-seed objects in the background. Additionally, the watershed algorithm used by GRABSEEDS excels in segmenting overlapping seeds, a frequent challenge in densely packed samples, by accurately separating seeds that are in close contact. GRABSEEDS offers significant improvements over tools like ImageJ, BISQUE, CellProfiler, and SHERPA [17–20], which, while powerful in their respective domains, are not optimized for seed phenotyping. Unlike ImageJ and BISQUE, which require extensive customization for seed analysis, GRABSEEDS is designed specifically for agricultural applications, providing built-in capabilities to





**Fig. 8** Principal Component Analysis (PCA) plot for sorghum kernel measurements. Each scatter dot represents a sorghum kernel collected from a total of 249 varieties. The size of the dot shows the relative area size of the kernel, while the color of the dot matches the extracted color from each kernel

handle challenges such as variable lighting and overlapping seeds.

In comparison to previous approaches [43–45], GRABSEEDS offers a significant improvement in handling the segmentation of adhered or overlapping seeds. Traditional methods have often struggled with maintaining accuracy in such complex scenarios, whereas GRABSEEDS, through its advanced watershed segmentation algorithm, successfully mitigates these challenges. More difficult cases can be solved with the deep learning model SAM [35], but with a higher computational cost.

This represents a noteworthy advancement in seed phenotyping, demonstrating GRABSEEDS' robustness and innovation in accurately extracting phenotypic traits even in difficult conditions. Compared to tools like SmartGrain and SeedGerm [43, 44], which focus on specific tasks like shape measurement or automated imaging, GRABSEEDS integrates these functions into a single, user-friendly platform optimized for a wide range of seed phenotyping tasks. Although we do not include direct comparisons with other tools for segmenting adhered or overlapping seeds, as these tools often lack such functionality, the superior segmentation capabilities of GRABSEEDS clearly demonstrate its effectiveness in seed phenotype extraction. Machine learning-based software can extract multiple phenotypic features from plant seeds, enabling detailed analysis and efficient separation based on qualities such as clarity and vigor [13, 15]. In contrast, GRABSEEDS stands out by offering a streamlined, user-friendly

approach to seed phenotyping. It simplifies the process by focusing on the most critical aspects such as seed size and color, allowing for rapid, accurate measurements without the need for complex model training. GRABSEEDS is designed to be accessible and efficient, providing quick results while maintaining the accuracy needed for reliable seed analysis. Its simplicity and speed make it an ideal choice for users who need an effective yet straightforward tool for seed phenotype extraction, without the overhead of machine learning complexities. We still support the out-of-the-box use of state-of-the-art machine learning models, such as SAM [35], but processing efficiency remains our primary focus.

GRABSEEDS' accuracy in seed recognition can be compromised by poor-quality images, such as those with low lighting, low contrast, or noisy backgrounds. While the tool incorporates advanced filters to mitigate these issues, it may still struggle with suboptimal image conditions, potentially affecting the reliability of results. To maximize the precision and efficiency of GRABSEEDS in large-scale phenotyping experiments, we offer a few recommendations below.

Leveraging the tool's sophisticated algorithms, users can optimize outcomes by ensuring that seeds occupy a significant portion of the image frame, which enhances the efficacy of the size filter and minimizes misidentification of seeds as background noise. By employing high-resolution imaging and precise focus, GRABSEEDS can more accurately distinguish seeds from smaller artifacts, further refining its segmentation performance. Moreover, to fully capitalize on GRABSEEDS' adaptive image processing capabilities, it is advisable to control environmental variables such as lighting and background texture. Reducing shadows and using a low-texture background minimizes the noise input to the Gaussian de-noising algorithm, thereby improving the clarity of seed edges and overall detection accuracy. While the watershed algorithm within GRABSEEDS is highly effective at separating closely packed seeds, it is recommended that seeds be appropriately spaced within the image to further enhance the algorithm's ability to accurately delineate individual seed boundaries, ensuring precise size and shape measurements even in dense samples. Although GRABSEEDS is highly effective for seed phenotyping (with leaves and flowers also supported as demonstrated), it is specialized for this purpose and may not be as versatile as more general image analysis tools that can be adapted to a broader range of applications. This specialization could be a limitation for users needing a more flexible tool for different types of biological image analysis such as Image J.

## Conclusions

GRABSEEDS is an advanced software platform that harnesses state-of-the-art image processing techniques to extract a comprehensive set of key metrics from seeds, leaves and flowers. Engineered for exceptional robustness, GRABSEEDS performs reliably under a wide range of challenging conditions, including variable lighting, noisy backgrounds, and diverse seed sizes and colors. The algorithms within GRABSEEDS have been rigorously validated across a broad spectrum of seed types, including *Camelina*, teff, sorghum, flower traits in *petunia*, and leaf traits in tobacco, showcasing its versatility and dependability. By integrating cutting-edge technologies, GRABSEEDS redefines high-throughput phenotyping, delivering unmatched accuracy and efficiency. This powerful tool empowers researchers to conduct large-scale phenotyping studies with ease, making it an essential asset in the pursuit of deeper insights into plant biology. GRABSEEDS not only streamlines the phenotyping process but also facilitates the discovery of genetic traits, thereby driving significant innovation in plant science and agricultural research.

### Availability and requirements

Project name: GRABSEEDS. Project home page: <https://github.com/tanghai-bao/jcvi/wiki/GRABSEEDS>. Operating system (s): Linux. Programming language: Python. License: Freeware, royalty-free, non-exclusive. Any restrictions to use by non-academics: none.

### Author contributions

HT implemented the GRABSEEDS software. HT, WK, PN, JSL, MM, MJ and WCY tested the software. HT and AHP designed the software. HT, JZ, XZ, AHP and WCY wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by funding from the National Key Research and Development Program (2021YFF1000104) to HT and (2021YFF1000101) to JZ.

### Availability of data and materials

The example images and datasets generated are available at: <https://github.com/tanghaibao/jcvi/wiki/GRABSEEDS>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 21 May 2024 Accepted: 6 September 2024

Published online: 12 September 2024

## References

- Bac-Molenaar JA, Fradin EF, Becker FF, Rienstra JA, van der Schoot J, Vreugdenhil D, Keurentjes JJ. Genome-wide association mapping of

- fertility reduction upon heat stress reveals developmental stage-specific QTLs in *Arabidopsis thaliana*. *Plant Cell*. 2015;27(7):1857–74.
- He G, Traore SM, Binagwa PH, Bonsi C, Prakash CS. Date palm quantitative trait loci. In: Al-Khayri JM, Jain SM, Johnson DV, editors. The date palm genome omics and molecular breeding. Cham: Springer; 2021. p. 155–68.
- Kooke R, Kruijer W, Bours R, Becker F, Kuhn A, van de Geest H, Buntjer J, Doeswijk T, Guerra J, Bouwmeester H. Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in *Arabidopsis*. *Plant Physiol*. 2016;170(4):2187–203.
- O'Connor K, Hayes B, Hardner C, Nock C, Baten A, Alam M, Henry R, Topp B. Genome-wide association studies for yield component traits in a macadamia breeding population. *BMC Genomics*. 2020;21(1):1–12.
- Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet*. 2014;10(9):e1004573.
- Bretani G, Shaaf S, Tondelli A, Cattivelli L, Delbono S, Waugh R, Thomas W, Russell J, Bull H, Igartua E. Multi-environment genome-wide association mapping of culm morphology traits in barley. *Front Plant Sci*. 2022;13:926277.
- Li F, Liu Z, Chen H, Wu J, Cai X, Wang H, Wang X, Liang J. QTL mapping of leaf-related traits using a high-density bin map in *Brassica rapa*. *Horticulturae*. 2023;9(4):433.
- Ree MJ, Carretta TR. The role of measurement error in familiar statistics. *Organ Res Methods*. 2006;9(1):99–112.
- Underwood J, Wendel A, Schofield B, McMurray L, Kimber R. Efficient in-field plant phenomics for row-crops with an autonomous ground vehicle. *Journal of Field Robotics*. 2017;34(6):1061–83.
- Borevitz JO, Chory J. Genomics tools for QTL analysis and gene discovery. *Curr Opin Plant Biol*. 2004;7(2):132–6.
- Moore CR, Gronwall DS, Miller ND, Spalding EP. Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3*. 2013;3(1):109–18.
- Mochida K, Koda S, Inoue K, Hirayama T, Tanaka S, Nishii R, Melgani F. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaScience*. 2019;8(1):gij153.
- Tu K, Wu W, Cheng Y, Zhang H, Xu Y, Dong X, Wang M, Sun Q. ASeed: an automated image analysis software for high-throughput phenotyping and quality non-destructive testing of individual plant seeds. *Comput Electron Agric*. 2023;207:107740.
- Fontejn H, Afonso M, Lensink D, Mooij M, Faber N, Vroegop A, Polder G, Wehrens R. Automatic phenotyping of tomatoes in production greenhouses using robotics and computer vision: from theory to practice. *Agronomy*. 2021;11(8):1599.
- Duc NT, Ramlal A, Rajendran A, Raju D, Lal SK, Kumar S, Sahoo RN, Chinnusamy V. Image-based phenotyping of seed architectural traits and prediction of seed weight using machine learning models in soybean. *Front Plant Sci*. 2023;14:1206357.
- Varma VS, Kanaka DK, Keshavulu K. Seed image analysis: its applications in seed science research. *Int Res J Agric Sci*. 2013;1(2):30–6.
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671–5.
- Kvilekval K, Fedorov D, Obara B, Singh A, Manjunath BS. Bisque: a platform for bioimage analysis and management. *Bioinformatics*. 2010;26(4):544–52.
- Lamprecht MR, Sabatini DM, Carpenter AE. Cell Profiler: free, versatile software for automated biological image analysis. *Biotechniques*. 2007;42(1):71–5.
- Kloster M, Kauer G, Beszteri B. SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects. *BMC Bioinform*. 2014;15:218.
- Das Choudhury S, Samal A, Awada T. Leveraging image analysis for high-throughput plant phenotyping. *Front Plant Sci*. 2019;10:508.
- Gehan MA, Fahlgren N, Abbasi A, Berry JC, Callen ST, Chavez L, Doust AN, Feldman MJ, Gilbert KB, Hodge JG. PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ*. 2017;5:e4088.
- Tross MC, Gaillard M, Zwiener M, Miao C, Grove RJ, Li B, Benes B, Schnable JC. 3D reconstruction identifies loci linked to variation in angle of individual sorghum leaves. *PeerJ*. 2021;9:e12628.

24. Miao C, Guo A, Thompson AM, Yang J, Ge Y, Schnable JC. Automation of leaf counting in maize and sorghum using deep learning. *Plant Phenome J*. 2021;4(1):e20022.
25. Lee U, Chang S, Putra GA, Kim H, Kim DH. An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis. *PLoS ONE*. 2018;13(4):e0196615.
26. Leinonen I, Jones HG. Combining thermal and visible imagery for estimating canopy temperature and identifying plant stress. *J Exp Bot*. 2004;55(401):1423–31.
27. Peters RD, Noble SD. Characterization of leaf surface phenotypes based on light interaction. *Plant Methods*. 2023;19(1):26.
28. Gong H, Yang M, Wang C, Tian C. Leaf phenotypic variation and its response to environmental factors in natural populations of *Eucommia ulmoides*. *BMC Plant Biol*. 2023;23(1):562.
29. Mishra P, Lohumi S, Ahmad Khan H, Nordon A. Close-range hyperspectral imaging of whole plants for digital phenotyping: recent applications and illumination correction approaches. *Comput Electron Agric*. 2020;178:105780.
30. Liu S, Barrow CS, Hanlon M, Lynch JP, Bucksch A. DIRT/3D: 3D root phenotyping for field-grown maize (*Zea mays*). *Plant Physiol*. 2021;187(2):739–57.
31. Tovar JC, Hoyer JS, Lin A, Tielking A, Callen ST, Elizabeth Castillo S, Miller M, Tessman M, Fahlgren N, Carrington JC. Raspberry Pi-powered imaging for plant phenotyping. *Appl Plant Sci*. 2018;6(3):e1031.
32. Lobet G, Draye X, Périlleux C. An online database for plant image analysis software tools. *Plant Methods*. 2013;9:1–8.
33. Lobet G. Image analysis in plant sciences: publish then perish. *Trends Plant Sci*. 2017;22(7):559–66.
34. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453.
35. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y et al. Segment anything. In: 2023. [arXiv:2304.02643](https://arxiv.org/abs/2304.02643).
36. Kay A (2007) Tesseract: an open-source optical character recognition engine. *Linux J*. Retrieved 28 September 2011.
37. Koyuncu CF, Arslan S, Durmaz I, Cetin-Atalay R, Gunduz-Demir C. Smart markers for watershed-based cell segmentation. *PLoS ONE*. 2012;7(11):e48664.
38. Behringer A. Camera array calibration with color rendition charts. Berlin: Humboldt University of Berlin; 2013.
39. McCamy CS, Marcus H, Davidson JG. A color-rendition chart. *J Appl Photogr Eng*. 1976;2(3):95–99.
40. Fletcher R. Newton-Like Methods. In: Fletcher R, editor. *Practical methods of optimization*. Hoboken: Wiley; 2000. p. 44–79.
41. Kuhl FP, Giardina CR. Elliptic fourier features of a closed contour. *Comput Graphics Image Process*. 1982;18(3):236–58.
42. Nabukalu P, Kong W, Cox TS, Pierce GJ, Compton R, Tang H, Paterson AH. Genetic variation underlying kernel size, shape, and color in two inter-specific *S. bicolor* × *S. halepense* subpopulations. *Genet Resour Crop Evolut*. 2022;69(3):1261–81.
43. Tanabata T, Shibaya T, Hori K, Ebana K, Yano M. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol*. 2012;160(4):1871–80.
44. Colmer J, O'Neill CM, Wells R, Bostrom A, Reynolds D, Websdale D, Shiralagi G, Lu W, Lou Q, Le Cornu T, et al. SeedGerm: a cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination. *New Phytol*. 2020;228(2):778–93.
45. Halcro K, McNabb K, Lockinger A, Socquet-Juglard D, Bett KE, Noble SD. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. *Plant Methods*. 2020;16:49.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.