DR JI ZHOU (Orcid ID : 0000-0002-5752-5524)

Article type : Methods Paper

Methods

SeedGerm: a cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination

Joshua Colmer[1*], Carmel M. O'Neill[2*], Rachel Wells[2*], Aaron Bostrom[1], Daniel Reynolds[1], Danny Websdale[1], Gagan Shiralagi[2], Wei Lu[3], Qiaojun Lou[4], Thomas Le Cornu[1], Joshua Ball[1], Jim Renema[5], Gema Flores Andaluz[5], Rene Benjamins[5], Steven Penfield[2], Ji Zhou[6,7]

[1]Engineering biology, Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK; [2]Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK; [3]College of Engineering, Nanjing Agricultural University, Nanjing 210095, Jiangsu China; [4]Shanghai Agrobiological Gene Center, Shanghai Academy of Agricultural Sciences, Shanghai, 201106, China; [5]Syngenta Seeds B.V., Enkhuizen, 1601 BK, Netherlands; [6]State Key Laboratory of Crop Genetics & Germplasm Enhancement, Plant Phenomics Research Center, Jiangsu Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and Ministry, Nanjing Agricultural University, Nanjing 210095, China; [7]Cambridge Crop Research, National Institute of Agricultural Botany, Cambridge CB3 0LE, UK.

[*]These authors contributed equally to this work.

Correspondence should be addressed to:

1) Ji Zhou

Tel: +0086 (0) 25 84395921 (China); Email: ji.zhou@njau.edu.cn

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/NPH.16736

2) Steven Penfield
Tel: +44 (0) 1603 450862 (UK); Email: steven.penfield@jic.ac.uk

3) Rene Benjamins
Tel: +31(0)622479690 (Netherlands); Email: rene.benjamins@syngenta.com

Received: *22 April 2020*
Accepted: *25 May 2020*

Authors' ORCID:
Joshua Colmer, orcid: 0000-0002-0511-685X
Rachel Wells, orcid: 0000-0002-1280-7472
Aaron Bostrom, orcid: 0000-0002-7300-6038
Daniel Reynolds, orcid: 0000-0001-5846-0016
Joshua Ball, orcid: 0000-0003-4840-3768
Steven Penfield, orcid: 0000-0001-7749-8298
Ji Zhou, orcid: 0000-0002-5752-5524

Summary

- Efficient seed germination and establishment are important traits for field and glasshouse crops. Large-scale germination experiments are laborious and prone to observer errors, leading to the necessity for automated methods. We experimented with five crop species, including tomato, pepper, Brassica, barley, and maize, and concluded an approach for large-scale germination scoring.
- Here, we present the SeedGerm system, which combines cost-effective hardware and open-source software for (1) seed germination experiments, (2) automated seed imaging, and (3) machine-learning based phenotypic analysis. The software can process multiple image series simultaneously and produce reliable analysis of germination- and establishment-related traits, in both comma-separated values (CSV) and processed images (PNG) formats.

- In this article, we describe the hardware and software design in detail. We also demonstrate that SeedGerm could match specialists' scoring of radicle emergence. Germination curves were produced based on seed-level germination timing and rates rather than a fitted curve. In particular, by scoring germination across a diverse panel of *Brassica napus* varieties, SeedGerm implicates a gene important in abscisic acid (ABA) signalling in seeds.
- We compared SeedGerm with existing methods and concluded that it could have wide utilities in large-scale seed phenotyping and testing, for both research and routine seed technology applications.

Keywords

Seed germination, seed imaging, germination scoring, phenotypic analysis, machine learning, big data biology, crop seeds

Introduction

Seeds are essential for human beings, not only as important food sources, but also for efficient crop production. High-vigour seeds with better seed germination and seedling emergence rates can ensure reliable emergence under varied agricultural conditions and hence are key to yield potential and uniformity (TeKrony & Egli, 1991). A common scoring method for seed germination is to assess radicle protrusion, which quantifies the speed and frequency of germination (Finch-Savage & Bassel, 2016). Traditionally, the task was accomplished by seed technologists through visual inspections on colour and morphological changes during physiological processes of seed germination (Lin, 1999); however, this approach is labour-intensive and subjective (Joosen *et al.*, 2010; Demilly *et al.*, 2015).

Routine germination scoring still commonly relies on human observation, which has practically constrained the frequency, scale, and accuracy of such experiments (Reyazul *et al.*, 2015; Jahnke *et al.*, 2016; Zhang *et al.*, 2018). This bottleneck has led to many attempts to automate both seed imaging and associated phenotypic analysis, resulting in several research-based solutions such as GERMINATOR and the package, *pheno*Seeder, and the MultiSense tool (Ducournau *et al.*, 2005; Joosen *et al.*, 2010; Demilly *et al.*, 2015; Jahnke *et al.*, 2016; Ligterink & Hilhorst, 2016; Keil *et al.*, 2017). More recently, advanced computer-vision (CV) and machine-learning (ML) techniques are being applied to germination assays, including the Rice Seed Germination Evaluation System (RSGES) for assessing the germination status of Thai rice species using an artificial neural network (ANN) classifier (Lurstwut & Pornpanomchai, 2017); machine-vision based analysis on visible and X-ray images for evaluating soybean seed quality based on physical purity, viability and vigour (Mahajan *et al.*, 2018); deep learning (DL) algorithms such as U-Net and ResNet for segmenting and classifying rice seed germination status (Nguyen *et al.*, 2018); linear discriminant analysis and multispectral imaging combined for classifying cowpea seeds into categories of ageing, germination, and normality (Elmasry *et al.*, 2019), and a high-throughput micro-CT-RGB (HCR) phenotyping system for dissecting the rice genetic architecture from seedling (Wu *et al.*, 2019).

The above solutions include customised hardware devices (e.g. bespoke germination trays, image sensors and seed handling system) and tailored analytic software built on MATLAB Toolbox, ImageJ/Fiji, Microsoft Excel macros, image analysis libraries (e.g. VideometerLab3 and

OpenCV), and ML/DL libraries (e.g. PyTorch). Although not fully automated, they have been successfully applied to impute germination traits from the acquired seed images, including the quantification of morphological traits (e.g. size and shape), cumulative germination rates (e.g. time to 50% germination, T_{50}, and the proportion of seeds germinated at the conclusion of an experiment, G_{max}), and quality traits such as viability and vigour (Ducournau *et al.*, 2005; Jahnke *et al.*, 2016; Mahajan *et al.*, 2018). Nevertheless, the throughput, automation level, and the range of traits of the above solutions are still limited, such that seed imaging and associated germination-related traits analyses still require human interference.

The emergence of plant phenomics in recent years has brought new perspectives to seed science research (Dell'Aquila, 2009; Watson *et al.*, 2018). By combining cost-effective digital imaging and environment sensors, organ-level plant growth and development can be recorded with detailed imagery, at a very high frequency (Tardieu *et al.*, 2017; Pieruschka & Schurr, 2019; Reynolds *et al.*, 2019b). In particular, many CV and ML combined analytic methods have been developed to enable the automation of organ-level phenotypic analysis, including leaves, roots, and reproductive organs (Pound *et al.*, 2017; Sadeghi-Tehran *et al.*, 2017; Xiong *et al.*, 2017; Zhou *et al.*, 2017a; Yasrab *et al.*, 2019). By combining colour, texture, morphologies, and growth patterns, seed germination can be quantified in a dynamic and objective manner, based on which large-scale and reproducible evidence can be produced to enable new biological discoveries for seed physiology (Teixeira *et al.*, 2007; Demilly *et al.*, 2015; Reyazul *et al.*, 2015; Lurstwut & Pornpanomchai, 2017; Elmasry *et al.*, 2019). Furthermore, the automation of seed germination scoring presents a good opportunity to initiate the standardisation of seed science research. Not only can seed quality and vigour be digitally assessed, but in addition, biological experiments under varied conditions can be cross-referenced quantitatively to increase the confidence of our research outcomes.

Here, we introduce SeedGerm, a platform designed for automating seed imaging and high-throughput germination analysis for a variety of crop seeds. SeedGerm incorporates cost-effective hardware components for seed imaging and experimental conditions (e.g. ambient temperature and humidity) acquisition, as well as ML-based analytic software for measuring both germination- and establishment-related traits during the germination process. Utilising SeedGerm, we are able to quantify the performance of seed lots based on individual seeds rather than a fitted germination

curve. The analytic software embedded in SeedGerm is able to process multiple image series at the same time and export analysis results in both comma-separated values (CSV) files and processed images (e.g. germination masks, in PNG format), at both seed and panel levels (normally one genotype per germination panel). We also demonstrate that SeedGerm matches seed specialists' observations for the scoring of radicle emergence timing for crop species such as tomato, pepper, Brassica, barley and maize seeds, which can also be used as a research tool to identify the genetic basis of germination differences between varieties.

Materials and Methods

Seed batch production and storage

Seed lots were produced in commercial production and stored at 12^0C and 35% relative humidity (RH) until use. For seed production from the 88 *B. napus* Diversity Fixed Foundation Set (DFFS) lines used in this study, plants were vernalised (8-h photoperiod, 5 ^{o}C) for 6 weeks at the four-leaf stage and grown in a polytunnel. Seeds were used within three months after harvesting. Seed batches from independent mother plants constituted biological replicates. High-quality seed batches of tomato and Brassica were utilised to generate lower quality batches. To this end, a sub batch was taken from these, which was heat-treated for three days at 70 ^{o}C.

Seed germination conditions

A typical experimental setup uses standard A3-sized filter paper, dark blue seed testing paper used in the germination chambers supplied by Munktell Ahlstrom (Grade 194, Bärenstein Germany), substrate to accommodate six sets of 64 individual seeds (384 seeds in total, in six germination panels) for tomato and Brassica seeds. For barley, we carried out experiments with three extended germination panels, with 40 seeds per panel and 120 seeds in total. Due to the size of maize seeds, the entire germination box was used to host 35 seeds per experiment. For pepper seeds, 81 seeds were used in a given panel, resulting in a total of 486. To facilitate sound germination classification, a minimum of A4-sized filter paper is recommended to allow sufficient space between seeds, but further divisions could also be made to separate different genotypes.

Typical automated seed imaging was set with an hour interval and normally conducted between 5 and 10 days depending on the crop species. For example, *B. napus* seeds were germinated on

saturated filter paper in SeedGerm boxes in constant white light at 10°C (in a cold-room or a growth chamber). A standard seed testing took 7-14 days, with two key traits (i.e. germination frequency and seed vigour) frequently checked by experienced seed technologists. To screen the 88 *B. napus* DFFS lines, seeds were gridded in panels of 50 seeds, with six panels per germination box and five replicates per line. A fully randomised experimental design was followed. In a routine experiment, each SeedGerm box contains two layers of white filter (Grade 3644, Hahnemuehle Germany), with a single sheet of blue seed germination paper on top. A fixed volume of water (i.e. sterile de-ionised water, 350mls) was added to the filter paper stack prior to the start of the experiment. To ensure even absorption across the filter paper, the wetted paper was allowed to stand for 2 hours after the addition of water (i.e. a further 30mls), before gridding the seeds and starting the experiment.

Hardware design

To carry out high-quality seed imaging to record physiological processes of germination in a continuous manner, we have designed two types of hardware apparatus: (1) a relatively low-cost translucent plastic germination box mounted with a fixed camera for routine germination experiments, and (2) a more expensive bespoke mini-gantry imaging system built on the top of a transparent polyethylene box for long-term experiments. Both designs are shown in **Figure 1**, where the image sensors used in the fixed design are high-definition (HD) *Pi* camera modules (i.e. 5 megapixel, MP, with a maximum 2592x1944 pixels per image, **Fig. 1a**) and the mini-gantry design is equipped with an 8 MP HD USB camera, with undistorted wide-angle lens and a maximum 4160×3120 pixels per image (**Fig. 1b**). As the focus of the moving USB camera is adjustable, the latter design has been used for a variety of experiments to explore the physiological processes between germination and seedling (e.g. a 15-day experiment for wheat seeds, **Fig. 1b**). Also, some digital sensors have been installed in the SeedGerm device, recording ambient humidity and temperature on an hourly basis. As advised by previously published work (Schumann *et al.*, 1995; Afzal *et al.*, 2017), transparent polypropylene used to build SeedGerm devices has been tested repeatedly and did not have effects on germination and seedling growth. A brief outline of the hardware design and cost of the SeedGerm device can be seen in **Supporting Information Note S1**.

Both SeedGerm hardware designs are controlled by low-cost single-board computers (i.e. *Raspberry Pi 2* or *Pi 3* computers). In a given experiment, users can set up seed imaging via a graphic user interface (GUI) based software application (i.e. the imaging module) running on *Pi* computers embedded in the SeedGerm hardware, through which imaging parameters such as resolution and interval can be programmed. The GUI control software is cross-platform and was developed using Python's native GUI package, Tkinter (Shipman, 2013), and has been described previously (Zhou *et al.*, 2017b). It also allows users to define metadata for each experiment, including species, genotypes, experiment duration, and the naming convention for the acquired images. A number of experiments can be monitored simultaneously (**Figs. 1a&1c**). The data collation and management are controlled by Linux-based *crontab* scheduling, at near real-time. Users can visually inspect experiments (e.g. tomato in **Fig. 1c** and wheat in **Fig. 1b**) from their own computers or smart devices using a virtual private network (VPN) or remote desktop software.

Open-source software system

Besides the seed imaging module, the SeedGerm software system also contains a light-weight data management module and a ML-based analysis module (**Fig. 2a**). An image acquired by the imaging module is firstly saved on the SeedGerm hardware's local storage. Then, the image is checked according to its size and clarity; if the size and clarity are greater than a predefined threshold value, it will be transferred to a gateway computer via wired (Ethernet) or wireless network connections. This synchronisation task is carried out on an hourly basis, between many SeedGerm devices and the gateway computer, where images from different devices are collated in folders named after their associated experiments defined through the seed imaging module. Between the gateway machine and onsite storage (e.g. a dedicated workstation or high-performance computing infrastructure, HPC), data synchronisation tasks are normally accomplished overnight, when onsite network traffic is less busy. The data management module is administered by either *crontab* scheduling on Linux (Debian 9.0 onwards) or Bash scripting on Windows (Windows 7 onwards), which have been described in our previous work (Reynolds *et al.*, 2019a).

For different end-users, automated phenotypic analysis can be conducted either centrally on HPC or in a distributed manner on a workstation. Each experiment generates a time-lapse image series, which is uploaded to the onsite storage progressively during the experiment. Then, users

can use the ML-based phenotypic analysis module (i.e. analysis software) to either analyse these images through a command-line interface on HPC clusters, or tailored GUI-based software on a normal workstation computer. Both approaches output similar analysis results, including the quantification of germination- and establishment-related traits in CSV files, as well as a sequence of processed images (e.g. dynamic seed masks and panel segmentation images) in PNG format.

GUI-based analysis software

For the ML-based phenotypic analysis module, the workflows for both GUI and command-line approaches are fundamentally identical. We therefore use the more accessible GUI software to introduce the analysis procedure, which has been designed to execute on either Windows (i.e. the .exe executable, Windows 10 tested) or Mac OS (i.e. the .app file, version 10 onwards). The analysis software packages can be downloaded from our GitHub repository. The initial GUI contains an empty window with a menu bar and users can add experiments via the "Add experiment" window (**Fig. 2b**), through which users can enter a given experiment's name, select an image series for processing, and choose a crop species such as Brassica, maize, pepper, tomato, or cereals. New plant species can be trained and added to the software through the Modules directory, an approach that is independent of the core analysis algorithm. Users need to briefly define the germination experiment associated with the selected image series, including the number of panels in a given SeedGerm device, Rows and Columns of seeds in each panel. In particular, users can define the Start and End image IDs to initiate and terminate the phenotypic analysis, because the background in early images can be over-saturated due to excess water soaked by the filter paper, whereas late images can contain too many overgrown seedling and roots (e.g. images between the fourth and 167th image will be analysed in **Fig. 2b**). Default values of the Start and End images are the first and last image of the selected series.

In order to deal with varied image quality and features caused by lighting, crop species, and different establishment phases, a number of ML-based algorithms have been implemented in the software. Users can select the ML technique from the "BG remover" dropdown to remove the background pixels, which includes U-Net (Ronneberger *et al.*, 2015), Gaussian mixture model (GMM) (Stauffer & Grimson, 2003), and stochastic gradient descent (SGD) (Bottou & Bousquet, 2008), which are explained in the following sections. After an experiment is added, users are required to set YUV colour-space ranges (Y stands for the brightness, U and V for colour

components (Szeliski, 2010)) to delineate the background (i.e. filter paper) in the first image of the selected series (**Fig. 2c**). By adjusting the sliding bars in the "Set YUV ranges" window, backgrounds are mostly retained, representing different types of filter paper used in diverse experiments. After defining YUV values, users can click the "Process images" item to start the phenotypic analysis (**Fig. 2c**). Similar to our previous work (Zhou *et al.*, 2017a), the analysis software has also employed parallel computing to process multiple experiments simultaneously, with up to 12 image series have been analysed at a time on an average computer (Intel Core i5, 8GB RAM) and over 120 series on HPC (**Fig. 2d**). This implementation has enabled a multi-threading analysis running on HPC clusters for greater throughput.

Finally, when the analysis is completed, germination traits (e.g. T_{25}, T_{50}, T_{75}, G_{max}, and germination timing curves for each panel), morphological traits (e.g. area, width and length, extent, convex area, and circularity for each seed), and a range of processed images (showing the germination procedure and labelling individual seeds) are produced (**Fig. 2e**). Users can click "View results" on the shortcut menu to display the analysis outputs, as well as download a range of processed images (**Supporting Information Video S1**) and the analysis results in CSV files, containing phenotypic analysis at the image (overall results), the panel (i.e. a given genotype), and the seed levels (see **Supporting Information Note S2**).

Core analysis algorithm

The core analysis algorithm for SeedGerm includes three key parts: (1) ML-based background remover, (2) feature extraction and germination detection, and (3) traits measurement (**Fig. 3**). To establish a more general algorithm to analyse different types of seeds robustly, we have used a mixture of deep learning (DL, i.e. U-Net) and supervised ML (i.e. GMM and SGD) to divide background (filter paper) and foreground (seeds) pixels. For example, after users set the YUV values to retain background pixels, the selected BG remover is trained based on features of the background (e.g. RGB, contrast, intensity values) in the image (**Fig. 3a**). Then, the YUV values are applied to representative images across the image series (i.e. images at the beginning, middle, and end of the series) to segment background pixels, which allows the ML model to learn background features at different establishment stages during a given experiment, without overfitting the classifier for a specific crop species or a particular experimental setting. Finally, the

trained classifier (i.e. the background remover) is applied to each image in the series, producing background masks excluding any seed for each germination panel (**Fig. 3b**).

After producing background masks, they are inverted so that only seed-related objects can be retained. SGD has been chosen as the main learning algorithm for germination scoring and was used for our routine germination experiments because it performs well when the seed-background contrast is high when seeds can be clearly delineated from the surrounding background pixels. Unlike SGD, the GMM model is used for image series with low seed-background contrast when seeds are slightly out-of-focus. It is slower, but more robust when the background is complex (e.g. roots from different seeds are crossing). For images with acceptable quality but under changeable lighting conditions, we tend to use U-Net, a recent convolutional neural network (CNN) for semantic segmentation, excelling fully convolutional network (Jonathan *et al.*, 2015) by adding skip connections and extra upsampling layers to provide both local and global information. The implementation of U-Net is exploratory, with the aim of using deep learning techniques to improve analysis for unseen datasets (e.g. treated seeds and new plant species).

For feature extraction and germination detection (**Figs. 3c,d**), we applied descriptive statistical moments, i.e. Hu Moments (Hu, 1962) to describe a given seed's area and its centroid position, which are invariant to the scale and rotation changes of seeds due to imbibition in early germination stages. Features such as minor axis length (seed width), major axis length (seed length), length and width ratio, perimeter, delta of Hu moments (i.e. difference of a seed's Hu moments between two consecutive images), delta of seed area, and delta of seed length and width, have been computed to monitor a seed's morphological changes. The aforementioned features from the pre-germination stage are extracted from the first 20% of images in the series and are then combined to form a training matrix (**Fig. 4c**) to train a classification model with the assumption that the label for all the seeds is non-germinated. The detection model used is called novelty detection (Schölkopf *et al.*, 2001), a one-class support vector machine (SVM) established on the training matrix generated from the first 20% of images and is then applied to determine the germination status of each seed in the image series. Based on the training data, a decision function is generated by the model to enclose pre-germination feature vectors, i.e. white circles enclosed by the red-coloured contour in the embedding p-dimensional space (**Fig. 3d**); then, as the germination experiment progresses, feature vectors are recomputed. When a seed begins to germinate, its

feature vector should gradually leave the boundary of the initial observation region (i.e. abnormal with a given confidence in the germination assessment, black circles outside the red-coloured contour). The seed's probability of germination will increase as well. The novelty detection model scores germination for all detected seeds, resulting in cumulative germination rates for each seed lot in a given germination panel (**Fig. 3e**). Since the novelty detection model is reinitialised and retrained for each experiment using the first 20% pre-germination images of the selected image series as training data, the detection model is dynamic and hence the risk of overfitting is low. The implementation of the above algorithms can be seen in **Supporting Information Note S3**.

Morphological traits analyses

The last key component of the analysis software is the measurement of morphological features, prior to true leaf production from apical meristems. This approach utilises the *measure* module in Scikit-Image (van der Walt *et al.*, 2014), which is enumerated briefly below:

1. "Seed Area" is the total number of pixels in the region of a segmented seed, which quantifies the size of a seed together with its associated radicals, if it is germinated. This trait can be used to define the size change of seeds during germination (e.g. imbibition).
2. "Seed Perimeter" measures the length of the contour line that encloses a given seed and, if germinated, its associated radicals. This trait can be used to verify the change of the seed size and radicle emergence during germination.
3. "Seed Major/Minor Ratio" measures the width and length (W/L) ratio of the ellipse that encloses a seed and, if germinated, its associated root regions. This trait can be used to define the shape change of a given seed during germination.
4. "Seed Convex Hull Area" measures the area of the smallest polygon that can enclose a seed and, if germinated, its associated root regions. This trait can be used to define the change of the seed and root coverages in a germination panel.
5. "Seed Extent" is the ratio of the total number of pixels contained by a seed to the total pixels contained in its bounding box. This trait is useful to assess the seed establishment rate because the area of the bounding box should increase faster than the seed area.
6. "Seed Circularity" calculates the roundness of a seed, and, if germinated, its associated root regions. If the seed is a perfect circle, its circularity reading is 1; a line segment would have a circularity of 0. The circularity is defined as $4 * \pi * Area/Perimeter^2$, where *Area* is Seed

Area (in pixels) and *Perimeter* is Seed Perimeter (in pixels). This trait is also used to differentiate different crop seeds and their germination rates.

Results

Germination and morphological traits

A set of germination experiments have been conducted to test and improve the SeedGerm platform. The analysis results of an experiment with 384 tomato seeds (six genotypes), which have been placed on six panels in a customised germination box, with one genotype per panel (64 seeds) can be seen in **Figure 4**. The imaging interval is 60 minutes and 186 images have been acquired in total, within eight days (**Fig. 4a**). Analysis outputs include two types of traits: (1) germination traits quantified using 1^{st}~186^{th} images (**Fig. 4b**), including cumulative germination curves, T_{50} germination rates to assess the uniformity of germination, and G_{max} to quantify the proportion of seeds germinated at the end of the experiment; and, (2) morphological traits quantified using 1^{st}~160^{th} images (**Fig. 4c**), including seed area, width and length (W/L) ratio, and circularity. By combining both traits, we can identify morphological changes of six genotypes at the pre-gemination stage (before the 106^{th} image). As soon as the germination process started, the cumulative germination curves and associated morphological features became divergent between genotypes. It is observable that there is a strong correlation between the germination curves and the seed area curves, fitting in the developmental procedure when radicals coming out of seeds can dramatically increase the W/L ratio, and the more roots the lower W/L ratio and circularity. The above quantification exhibits the usefulness of combining both germination and morphological traits to verify and improve the detection accuracy.

Additionally, we also used the analysis outputs to evaluate germination uniformity or variability, an important trait requiring complex formulas to compute previously (Ranal & De Santana, 2006). For example, box-and-whisker plots are provided amongst the result files to demonstrate statistical dispersion of T_{50} germination rates (**Fig. 4b**), showing the difference between 25^{th} and 75^{th} percentiles of each genotype, as well as the median time to 50% germination of all genotypes. For example, genotype 6 seeds (G6) possess lower germination variability and better germination uniformity, which is verified by narrower percentile ranges and similar median values across tested seed batches. We have removed a number of late images (after T_{75}) when presenting the

morphological traits (**Fig. 4c**), which is due to substantial measurement variations caused by overlapped roots at late stages. The analysis results of the experiment can be seen in **Supporting Information Datasets S1 & S2**.

Germination analysis for different crop seeds

To demonstrate the robustness and generalisation of the SeedGerm system, we have applied SeedGerm to score germination for a range of crop seeds. The germination analyses for four selected crop species are tomato, pepper, maize and barley (**Fig. 5**). Seed images at three different experiment stages can be seen in the first three columns of images in **Figure 5**. After conducting time-series seed imaging, we used SeedGerm software to measure germination and morphological traits. Each germination panel (enclosed by dotted rectangles coloured in red in **Fig. 5**) contains one genotype. Seeds in the panel were monitored continuously, with dissimilar durations due to varied research objectives, for example, 165 hours (7 days) for tomato (Groot & Karssen, 1992), 180 hours (8 days) for pepper (Smith & Cobb, 1991), 138 hours (6~7 days) for maize (Flórez *et al.*, 2007), and 138 hours (6~7 days) for barley (Al-Karaki, 2001). These experiments were also checked by specialists daily, so that manual and SeedGerm scores can be compared and verified.

The tomato seed germination experiments were conducted in six panels (i.e. six genotypes), with 64 seeds per panel and 384 seeds monitored in total (**Fig. 5a**). Six cumulative germination curves have been produced based on hourly measurements for a seven-day period. We could clearly identify small differences amongst these genotypes between T_{50} and T_{75}, when germination rates diverted. Similarly, germination variances could also be quantified for pepper and barley experiments (**Figs. 5b&c**). The three barley genotypes monitored exhibited a wide variety of cumulative germination, similar to what has been reported previously (Matthews & Khajeh-Hosseini, 2007). Due to the size of maize seeds, we conducted one experiment per germination box (35 seeds per box, **Fig. 5d**). Still, SeedGerm software can perform sound measurement even when the number of germination experiments is changed. The above panel- and seed-level germination measures were exported and saved in several CSV files (see **Supporting Information Datasets S3-6**).

New morphological traits included in the SeedGerm analysis are seed convex area, seed extent and seed circularity, which have been used to quantify dynamics of germination of different crop

seeds as they were difficult to assess using traditional approaches (TeKrony & Egli, 1991; Dell'Aquila, 2009). For example, using the seed convex area trait, we found that maize had the quickest establishment rate after T_{50}, while other crop seeds were very similar (**Fig. 5e**). Due to substantial variations caused by too many overlapped radicals at the late germination stage, end image IDs for the above analysis are different. Similarly, panel- and seed-level morphological measures are saved in CSVs (see **Supporting Information Datasets S7-10**).

Validation of the SeedGerm platform

To validate analysis produced by SeedGerm, we have used a range of validation methods to comprehensively compare human and SeedGerm scores. A multitude of metrics were produced (**Table 1**), including Pearson's correlation metric (*r*) to measure the strength of the linear relationship between SeedGerm and manual scoring for cumulative germination rates. For all tested crop species, SeedGerm's cumulative predictions yield a Pearson's correlation greater than 0.98 (column two in **Table 1**), indicating a strong linear correlation and goodness of fit. Pearson's correlation (*r*) was used to evaluate the linear relationship between SeedGerm's true positive germination timings and their respective timings scored by seed scientists (column three in **Table 1**). In addition to the correlation metrics, we have also calculated the mean absolute error (MAE, column four in **Table 1**) to interpret the average error in hours of the germination time compared with manual scores for each germinated seed. The MAE measures forecast error in SeedGerm's prediction against human scoring (the true value), showing a satisfactory error range. Lastly, the F1 score (1 indicates a perfect set of classifications and 0 means all false negatives or false positives, (Sasaki, 2007)), a classification metric similar to accuracy but more appropriate for imbalanced datasets, was used to incorporate the number of true positives, false positives, and false negatives into a single score for evaluating the germination classifications made by SeedGerm. Based on F1 scores (column five in **Table 1**), we can conclude that SeedGerm performed well across all tested crop species. The above methods evaluate both SeedGerm's final germination scoring as well as the germination timing of each seed, covering germination rate, timing and the uniformity respectively.

To visualise the correlation between SeedGerm scoring and seed specialists' counting, we have used 19 time series (over 4,000 images in total) to perform the correlation, with three series of maize (129 seeds in total), six series of tomato (384 seeds), six series of Brassica (384 seeds), one

series of pepper (81 seeds), and three series of barley (120 seeds). Manual scoring was performed using the image series, where cumulative germinated seed counts for each image and the image ID for when each seed germinates were recorded. There is a strong correlation between SeedGerm's scoring and that of the manual observers, which can be seen in **Figure 6**. A predicted equals actual line (coloured red) is included (**Fig. 6a**) to show how SeedGerm's cumulative scores deviate from the manual scores. Additionally, line plots contrasting cumulative seed-by-seed scoring between SeedGerm and specialists' counting are shown in **Figure 6b**. SeedGerm's scoring is largely identical in comparison with manual counting, except for it tending to overestimate the number of germinated seeds in crowded experiments such as the later establishment stages for Brassica and tomato experiments.

SeedGerm as a research tool

To test the ability of SeedGerm to be used as a research tool in routine biological experiments, we used the *B. napus* Diversity Fixed Foundation Set (Harper et al., 2012) to detect genetic differences in seed germination. After setting replicate seed batches of each variety, biological replicates of 50 seeds were sowed in SeedGerm boxes in a randomised design. SeedGerm scored the germination parameters of 88 varieties with a range of germination behaviours, with some showing strong dormancy, while most seed lots germinated to high levels, but with varying kinetics. SeedGerm scored the T_{10}, T_{50}, T_{90} and G_{max} after 8 days (**Figs. 7a&b**). To test the accuracy of the SeedGerm outputs, 60 seed lots were also scored by a manual observer based on images. The agreement was strong (**Supporting Information Dataset S11**), except for T_{90} in varieties requiring the longest time to germination, where SeedGerm has a weak tendency to score seeds as germinated before the manual observation.

The SeedGerm outputs were then used for associative transcriptomic (AT) analysis, as described previously (Harper *et al.*, 2012). The AT found no significant associations between T_{10}, T_{50} and T_{90} and polymorphisms in *B. napus*. However, we found a strong association between G_{max} and genotype on chromosome A5, with both SNPs and gene expression markers (Harper *et al.*, 2012) associated with the trait in this region (**Figs. 7c-e**). This is distinct from those loci identified in previous studies (Hatzig *et al.*, 2015, 2018), but significant, even after correcting for multiple testing. This region spans approximately 340kb and contains at least 69 known transcribed genes, one of which is a *B. napus* orthologue of the known germination regulator, protein phosphatase 2C

known as *HIGH ABA INDUCED 3* (*HAI3*) (Yoshida *et al.*, 2006; Bhaskara *et al.*, 2012), which has a role in seed sensitivity to abscisic acid. Although more work is needed to precisely identify the underlying gene of interest, it is evident that the SeedGerm platform is capable of automating phenotypic analysis of seed germination with sufficient accuracy to perform standard genetic analysis of seed dormancy and germination rate.

Discussion

Automated seed phenotyping

Plant phenomics is a fast-developing research area focusing on obtaining meaningful phenotypic information to enable scientists to address diverse biological questions, from cellular organisms to populations in the field (Tardieu *et al.*, 2017; Zhou *et al.*, 2018; Furbank *et al.*, 2019; Yang *et al.*, 2020). To study seed germination and seedling vigour, many academic and industrial attempts have been made, including research-based tools such as Germinator, *pheno*Seeder, MultiSense and RSGES, as well as commercial solutions such as the PhenoSeeder platform (developed by Forschungszentrum Jülich, Germany), SeedAIXPERT and Germination Scanalyzer (www.lemnatec.com/products/seed-screening), and Seeds Automatic Germination Analyzer (SAGA, France, no longer trading). These methods are capable of carrying out seed imaging, advanced 3D seed morphological analysis (i.e. *pheno*Seeder), and germination related traits analyses; however, their applications are limited due to their costs, availability, automation level, analysis throughput, and the technical scalability.

In this study, we present the SeedGerm system, a platform that combines automated seed imaging and vision-based phenotypic analysis with cost-effective hardware to enable high-throughput analysis of seed germination experiments for a variety of crop species. Based on more than three years' experiments and system improvements, we believe that our system is easy-to-access and capable of carrying out scalable seed germination scoring for the following reasons: its low-cost and easy-to-build hardware design, its flexibility to incorporate different experiments, its open-source and modular software design, its scalability of traits analyses, and the availability of user-friendly GUI software, source code and design documents.

The SeedGerm hardware design

In comparison with high-end seed phenotyping devices such as *pheno*Seeder (Jahnke *et al.*, 2016) and its commercial version, our hardware design follows a low-cost and easy-to-build strategy. The material used to build the device can be easily accessed (see **Supporting Information Note S1**). The germination box was made of either translucent plastic for the fixed design or transparent polyethylene for the more expensive gantry design, which can ensure reliable germination and seedling growth. To provide biologically relevant data from imbibition to seedling, an overhead image sensor (e.g. a *Pi* camera module or an HD USB camera) was installed to acquire high-quality seed image series during the entire germination procedure. To increase the throughput of seed imaging, we used *Raspberry Pi* computers (e.g. *Pi* 2 Model B or *Pi 3* Model B+) to control imaging and collect metadata via SeedGerm's seed imaging module. Because a fixed-camera design such as GERMINATOR (Joosen *et al.*, 2010), the MultiSense tool (Keil *et al.*, 2017), and RSGES (Lurstwut & Pornpanomchai, 2017) is relatively limited in automation, we therefore developed a mini-gantry design to improve the automation level to monitor more germination experiments. A single-axis camera movement framework has been built to move the camera module to specific positions to monitor different experiments, with programmed intervals controlled by the *Pi* computer. By extending the length of the gantry system and the size of the germination box, more experiments can be included. The mini-gantry design consists of a pulley system, belts, 3D-printed supports, an extendable steel bar for rails, and a stepper motor with driver circuit.

To assess and compare germination performance for different seed batches with varied treatments is often laborious and prone to errors. Previous work (Ligterink & Hilhorst, 2016; Mahajan *et al.*, 2018) relies on CV and ML techniques to calibrate obtained images to ensure the soundness and the experiment. However, because experiment conditions (e.g. temperature and humidity) are also key to seed germination, we therefore have installed affordable environmental sensors (e.g. combined ambient temperature and humidity) and a fluorescent lighting device in the SeedGerm hardware to facilitate continuous experiment monitoring.

We followed open hardware suggestions (Gibney, 2016; Czedik-Eysenberg *et al.*, 2018) to improve the flexibility of SeedGerm. The hardware design is freely available to the community and allows changes for other research requirements. For example, by replacing the image sensor

with multi- or hyper- spectral cameras, seed germination can be studied beyond visible bands. Also, adding a side image sensor (Humplík *et al.*, 2015) or 3D imaging (Roussel *et al.*, 2016) in the SeedGerm system can support the analysis of seedling growth with vertical information. Such hardware improvements can be carried out without any restriction, which is likely to provide flexible options for seed research rather than mainly relying on costly commercial solutions such as Germination Scanalyzer or SAGA. Although SeedGerm is low-cost, its design is capable of carrying out high-quality (e.g. each pixel equals to 0.15-0.2 mm) and automated seed imaging to provide sufficient visual evidence and sensor data for biological experiments. Also, the low-cost feature is prone to increase the scalability of SeedGerm, as more devices can be built relatively cheaply to accommodate more experiments, which is hard to achieve previously.

The SeedGerm software design

There is a growing need for standardising plant phenotyping in recent years, resulting in the ISA-Tab format (Sansone *et al.*, 2012), minimal Information About Plant Phenotyping Experiments (MIAPPE) (Ćwiek-Kupczyńska *et al.*, 2016), and ontology approaches to enable comparative phenomics research (Oellrich *et al.*, 2015). Much previous work (Demilly *et al.*, 2015; Nguyen *et al.*, 2018; Wu *et al.*, 2019) in seed phenotyping has employed bespoke data collection processes and data formats, limiting external researchers and laboratories to utilise and support these methods. Hence, when designing SeedGerm's software system, we chose to standardise the collection of image and sensor datasets following the ontological suggestions. Additionally, to calibrate images acquired by different SeedGerm devices, users were required to enter metadata to define their experiments, including experiment ID, genotype, biological replicates, treatment, and experiment duration; then, imaging intervals, image resolution, white balance, exposure mode, and shutter speed were controlled automatically by the imaging module to largely standardise the data collection.

To increase the scalability of the phenotypic analysis, we chose to implement our algorithms in Python instead of MATLAB as previously reported (Jahnke *et al.*, 2016; Elmasry *et al.*, 2019). The reasons are that Python is easy-to-understand, cross-platform, and self-contained (Millman & Aivazis, 2011), which is supported by a wide range of open-source libraries such as Scikit-Image (van der Walt *et al.*, 2014), OpenCV (Howse, 2013), Scikit-Learn (Pedregosa *et al.*, 2011), and Keras/TensorFlow (Rampasek & Goldenberg, 2016). Publicly available development kits have

enabled us to extend and upgrade our software relatively easily. For example, new crop species and traits can be added to the core analysis algorithm through new modules, where guideline seed morphological features can be predefined. Also, we followed the modular software design, so that modules developed for one species can be shared by other functions in analysis and parallel computing.

Recently, deep learning has become a powerful technique used by some seed germination analysis software (Mahajan *et al.*, 2018; Nguyen *et al.*, 2018; Halcro *et al.*, 2020), for which it was applied to extract features, segment seeds, and classify germination status. Although DL is relatively easy to implement through Python presently, the reasons we chose a combined CV and ML approach are: 1) DL requires a very large amount of training datasets to perform better than supervised ML and CV-based methods; for features that need to be engineered frequently such as varied seed germination experiments, DL might not be suitable. 2) normally we need to build a dedicated DL model of each species; hence, it is time-consuming and ineffective to employ DL techniques for analysing a large number of crop species. 3) DL is likely to be overfitting for particular experiment settings and becomes problematic when conditions are changed. To allow our solution to be adopted by a broader research community that has varied experimental settings, we chose supervised GMM, SGD and novelty detection learning techniques based on generalised feature selection. More importantly, by designing the ML models to reinitialise and retrain with background features at different establishment stages for each experimental setting, the learning models embedded in SeedGerm are dynamic and can be updated for each experiment, enabling us to avoid overfitting the learning models for a specific crop species or a particular experiment.

By employing CV algorithms, SeedGerm can also measure cumulative germination rates and seed morphologies such as size, width and length, extent and circularity to assess seed quality and seedling vigour, from germination to seedling. For example, we have measured imbibition using the change of seed size, radical protrusion based on seed major/minor ratio, and germination speed through seed extent. If new biological questions are proposed, new traits and features could be designed jointly by biologists and computer scientists, instead of relying on DL techniques blindly. Because the SeedGerm software can be easily extended and accessed, we believe it is scalable and easy-to-access.

Applications of the SeedGerm system

Our work has demonstrated that SeedGerm is capable of scoring germination and measuring morphological changes automatically, for five major crop species and between different genotypes. The results show that SeedGerm could be employed to score germination frequency and seedling vigour, based on which the preformation of seed batches can be assessed. These traits were regularly checked by experienced seed engineers and scientists in order to provide certificates of seed germination and establishment performance in seed testing and seed insurance (Khurana & Singh, 2001; Dell'Aquila, 2009). Hence, it is evident that SeedGerm has the potential to provide a replacement for manual assessment of germination frequency and radical emergence activities. Furthermore, as many traits measured by SeedGerm are highly correlated with seed performance and the effectiveness of post-harvest seed enhancement processes, SeedGerm is likely to contribute towards seeds certification, guidance on sowing density, or even seed insurance in the future.

Besides routine seed testing on germination frequency, the applications of SeedGerm could also be expanded to the seed vigour (i.e. how fast and uniform radical emergence) through monitoring morphological traits, which are important for estimating canopy closure, weed suppression, and crop yields through seed research (Attree *et al.*, 1992; Nelson *et al.*, 2012; Paparella *et al.*, 2015). Beyond existing traits analyses, the continuous phenotypic analysis can extend our insights into the entire physiological procedure of germination to understand phenotypic effects of individual seed and seed batches under dissimilar treatments. Furthermore, we also set up a range of experiments to score germination rates and timing across a diverse panel of *B. napus* varieties to demonstrate the biological relevance of SeedGerm as a research tool to measure the effect of genetics. We showed that SeedGerm outputs can be used successfully for GWAS, identifying an association on *B. napus* chromosome A5 that explains the difference between high and low germinating varieties in the panel (**Figs. 7b-e**). Although the GWAS study identified associations over a 100kb region, this region does contain one gene BnaA05g27660D, a homologue of Arabidopsis *AHG3*, known to regulate ABA signalling during germination in Arabidopsis (Yoshida *et al.*, 2006), which would be a strong candidate for further study. The low-germinating allele is only present in older spring varieties including Bronowski and Duplo, suggesting that it has been consistently selected against by modern oilseed rape breeders. Hence, we believe that

SeedGerm has a great potential to have significant utilities in seed germination scoring and seed testing, for both research and routine seed technology applications.

Issues of SeedGerm

It is also important to point out some edge cases where the system has struggled. Due to camera position and lighting problems, some image series were of poor quality. Although we have added software calibration to allow users to improve the classification accuracy on the low-quality datasets (e.g. colour features), the analysis could still suffer. For such datasets, only through manually selecting image IDs (see **Supporting Information Note S4**) could we realistically reduce variance and improve the analysis accuracy. SeedGerm's scoring tends to overestimate the number of germinated seeds in crowded experiments such as the later stages for Brassica and pepper experiments. To provide reproducible measures of the uniformity, timing and germination rates, we have scored large numbers of seed samples and found that distancing the seeds from each other with at least 1 cm apart has improved the analysis accuracy noticeably for crowded experiments. As different crop seeds have very diverse morphologies, some morphological measures cannot be easily transferred from one species to the other, which indicates the application of Online-Leaning or Transfer Learning mechanisms (Wen *et al.*, 2017) could be potentially beneficial in future development. Although the learning models embedded in SeedGerm are dynamic for each experiment, the cost of such a design is that additional computational resources are required, demanding users to build a decent computer (i7 CPU with 16GB memory) to perform analysis. Notably, to maintain the reliability of the parallel computing, we do not recommend more than eight tasks to be paralleled on an average computer, because processing multiple image series simultaneously requires a high demand of computing resources and some Python functions have been locked because they are not thread-safe during multi-thread processing.

Conclusion

In conclusion, limitations of current seed imaging and scoring approaches have prevented automated and scalable analysis of seed germination. In this paper, we present the SeedGerm system that integrates cost-effective hardware and user-friendly software for performing seed imaging and ML-based analysis for measuring establishment- and germination-related traits. The

system has been applied to many germination experiments for five crop species, through which we could assess the performance of seed batches quantitatively. Morphological traits such as seed size, width and length, extent and circularity were also measured to provide insights into the physiological procedure of seed germination. We demonstrate that SeedGerm matches seed specialists' observations for the scoring of radicle emergence timing and its biological relevance in identifying a gene important in ABA signalling in seeds with associative transcriptomics. We trust that the SeedGerm system could have wide utilities in seed testing and germination scoring, for both research and industrial applications.

Availability and requirements

Project name: SeedGerm

Project release page and source code: https://github.com/Crop-Phenomics-Group/SeedGerm/releases

Testing image series: barley (109 MB), Brassica (638 MB), corn (257 MB), pepper (572 MB), and tomato (563 MB).

GUI software: SeedGerm.exe (128 MB)

Operating system(s): platform-independent

Requirements: Python 3.7, Scikit-Image, OpenCV, Scikit-Learn, Keras

Abbreviations

Abscisic acid (ABA), artificial neural network (ANN), associative transcriptomic (AT), background (BG), central processing unit (CPU), comma separated values (CSV), computed

tomography (CT), computer vision (CV), convolutional neural network (CNN), deep learning (DL), diversity fixed foundation set (DFFS), Gaussian mixture model (GMM), graphic user interface (GUI), high definition (HD), high-performance computing (HPC), identifier (ID), machine learning (ML), megapixel (MP), portable network graphics (PNG), red green blue (RGB), Seeds Automatic Germination Analyzer (SAGA), single nucleotide polymorphism (SNP), stochastic gradient descent (SGD), support vector machine (SVM), three dimensional (3D), universal serial bus (USB), width and length (W/L).

Author Contribution

J.Z., S.P. and J.C. wrote the manuscript, J.Z., R.B., S.P., C.M.O'N., D.R., T.L.C., W.L., and J.B. designed the germination hardware used in the paper, J.Z., J.C., T.L.C., A.B., and D.W. designed the SeedGerm analysis algorithms. J.C., T.L.C., A.B. and D.W. implemented the software system and established the learning models under J.Z.'s supervision, S.P, and R.W. designed the biological experiments, C.M.O'N., G.S., Q.L. performed biological experiments and manual scoring under S.P.'s supervision. S.P., R.W., and R.B. provided biological expertise. R.B., J.R., and G.F.A. designed and performed biological experiments at Syngenta. J.C. and J.B. packaged the GUI executables. J.Z., J.C., C.M.O'N., S.P., G.S., T.L.C., A.B. W.L. and J.B. tested the software. J.Z., J.C., A.B., D.W., C.M.O'N., R.W., and S.P. performed the data analysis and improved the platform. All authors read and approved the final manuscript. J.C., C.M.O'N. and R.W. contributed equally to this work.

Funding

J.Z. was partially funded by the United Kingdom Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council's (BBSRC) Designing Future Wheat Strategic Programme (BB/P016855/1) to Graham Moore and BBS/E/T/000PR9785 to J.Z.; J.C. was supported by BBSRC's National Productivity Investment Fund CASE Award (BB/S507441/1 to JZ), hosted at Norwich Research Park Biosciences Doctoral Training Partnership (BB/M011216/1), in collaboration with R.B. at Syngenta; D.R. was partially supported by the Core Strategic Programme Grant (BB/CSP17270/1) at the Earlham Institute; A.B., T.L.C, and D.W were partially supported by NRP's Translational Fund (GP072/JZ1/D) and Syngenta's industrial collaboration fund (GP104/JZ1/D) awarded to J.Z. W.L. and J.Z. were also supported by Jiangsu Collaborative Innovation Center for Modern Crop Production. This work has been

supported by the UK Biological and Biotechnology Research Council (BBSRC) via grant BB/P013511/1 to the John Innes Centre.

Acknowledgements

The authors would like to thank all members of the Zhou laboratory at the Nanjing Agricultural University, the National Institute of Agricultural Botany (Cambridge Crop Research), and Earlham Institute, as well as the Penfield Group at John Innes Centre for fruitful discussions and cross-disciplinary collaborations. We thank Mr Mark Scoles in the Desktop Support team at the department of Norwich BioScience Institutes Partnership (NBIP) Computing and Mr John Humble in the Equipment Services team at the NBIP Facilities department for excellent technical support in networking and hardware manufacture. We also thank researchers at the John Innes Centre and the University of East Anglia for constructive suggestions. We gratefully acknowledge the support of NVIDIA Corporation with the award of the Quadro GPU used for this research. The authors declare no competing financial interests.

References

Afzal I, Bakhtavar MA, Ishfaq M, Sagheer M, Baributsa D. **2017**. Maintaining dryness during storage contributes to higher maize seed quality. *Journal of Stored Products Research* **72**: 49–53.

Al-Karaki GN. **2001**. Germination, sodium, and potassium concentrations of barley seeds as influenced by salinity. *Journal of Plant Nutrition* **24**: 511–522.

Attree SM, Pomeroy MK, Fowke LC. **1992**. Manipulation of conditions for the culture of somatic embryos of white spruce for improved triacylglycerol biosynthesis and desiccation tolerance. *Planta* **187**: 395–404.

Bhaskara GB, Nguyen TT, Verslues PE. **2012**. Unique Drought Resistance Functions of the *Highly ABA-Induced* Clade A Protein Phosphatase 2Cs. *Plant Physiology* **160**: 379–395.

Bottou L, Bousquet O. **2008**. The Tradeoffs of Large Scale Machine Learning. *Advances in Neural Information Processing Systems* **20**: 161–168.

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

Czedik-Eysenberg A, Seitner S, Güldener U, Koemeda S, Jez J, Colombini M, Djamei A. **2018**. The 'PhenoBox', a flexible, automated, open-source plant phenotyping solution. *New Phytologist* **219**: 808–823.

Dell'Aquila A. **2009**. New Perspectives for Seed Germination Testing Through Digital Imaging Technology. *The Open Agriculture Journal* **3**: 37–42.

Demilly D, Ducournau S, Wagner MM-H, Dürr C. **2015**. Digital imaging of seed germination. In: Gupta SD, Ibaraki Y, eds. Plant Image Analysis Fundamentals and Applications. Boca Raton: CRC Press, 147–165.

Ducournau S, Feutry A, Plainchault P, Revollon P, Vigouroux B, Wagner MH. **2005**. Using computer vision to monitor germination time course of sunflower (*Helianthus annuus* L.) seeds. *Seed Science and Technology* **33**: 329–340.

Elmasry G, Mandour N, Wagner MH, Demilly D, Verdier J, Belin E, Rousseau D. **2019**. Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods* **15**: 1–16.

Finch-Savage WE, Bassel GW. **2016**. Seed vigour and crop establishment: Extending performance beyond adaptation. *Journal of Experimental Botany* **67**: 567–591.

Flórez M, Carbonell MV, Martínez E. **2007**. Exposure of maize seeds to stationary magnetic fields: Effects on germination and early growth. *Environmental and Experimental Botany* **59**: 68–75.

Furbank RT, Jimenez-Berni JA, George-Jaeggli B, Potgieter AB, Deery DM. **2019**. Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytologist* **223**: 1714–1727.

Gibney E. **2016**. 'Open-hardware' pioneers push for low-cost lab kit. *Nature* **531**: 147–148.

Groot SPC, Karssen CM. **1992**. Dormancy and germination of abscisic acid-deficient tomato seeds: Studies with the sitiens mutant. *Plant Physiology* **99**: 952–958.

Halcro K, McNabb K, Lockinger A, Socquet-Juglard D, Bett KE, Noble SD. **2020**. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. *Plant Methods* **16**: 1–13.

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I. **2012**. Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology* **30**: 798–802.

Hatzig S, Breuer F, Nesi N, Ducournau S, Wagner M-H, Leckband G, Abbadi A, Snowdon RJ. **2018**. Hidden Effects of Seed Quality Breeding on Germination in Oilseed Rape (*Brassica napus* L.). *Frontiers in Plant Science* **9**: 1–14.

Hatzig S V., Frisch M, Breuer F, Nesi N, Ducournau S, Wagner M-H, Leckband G, Abbadi A, Snowdon RJ. **2015**. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Frontiers in Plant Science* **6**: 1–13.

Howse J. **2013**. *OpenCV Computer Vision with Python*. Birmingham, UK: Packt Publishing Ltd.

Hu MK. **1962**. Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory* **8**: 179–187.

Humplík JF, Lazár D, Husičková A, Spíchal L. **2015**. Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - A review. *Plant Methods* **11**: 1–10.

Jahnke S, Roussel J, Hombach T, Kochs J, Fischbach A, Huber G, Scharr H. **2016**. pheno Seeder - A Robot System for Automated Handling and Phenotyping of Individual Seeds . *Plant Physiology* **172**: 1358–1370.

Jonathan L, Evan S, Trevor D. **2015**. Fully Convolutional Networks for Semantic Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 3431–3440.

Joosen RVLL, Kodde J, Willems LAJJ, Ligterink W, Van Der Plas LHWW, Hilhorst HWMM. **2010**. Germinator: A software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *Plant Journal* **62**: 148–159.

Keil P, Liebsch G, Borisjuk L, Rolletschek H. **2017**. MultiSense: A Multimodal Sensor Tool Enabling the High-Throughput Analysis of Respiration. In: Walker JM, ed. Plant Respiration and Internal Oxygen. New York: Humana Press, 47–56.

Khurana E, Singh JS. **2001**. Ecology of seed and seedling growth for conservation and restoration of tropical dry forest : A review. *Environmental Conservation* **28**: 39–52.

Ligterink W, Hilhorst HW. **2016**. High-throughput scoring of seed germination. In: Kleine-Vehn J. SM, ed. Plant Hormones, Methods in Molecular Biology. New York: Humana Press, 57–72.

Lin Y. **1999**. The Pex16p Homolog SSE1 and Storage Organelle Formation in *Arabidopsis* Seeds. *Science* **284**: 328–330.

Lurstwut B, Pornpanomchai C. **2017**. Image analysis based on color, shape and texture for rice seed (*Oryza sativa* L.) germination evaluation. *Agriculture and Natural Resources* **51**: 383–389.

Mahajan S, Mittal SK, Das A. **2018**. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *Journal of Food Science and Technology* **55**: 3949–3959.

Matthews S, Khajeh-Hosseini M. **2007**. Length of the lag period and metabolic repair explain vigour differences in seed lots of maize (*Zea mays*). *Seed Science and Technology* **35**: 200–212.

Millman KJ, Aivazis M. **2011**. Python for scientists and engineers. *Computing in Science and Engineering* **13**: 9–12.

Nelson DC, Flematti GR, Ghisalberti EL, Dixon KW, Smith SM. **2012**. Regulation of Seed Germination and Seedling Growth by Chemical Signals from Burning Vegetation. *Annual Review of Plant Biology* **63**: 107–130.

Nguyen TT, Hoang VN, Le TL, Tran TH, Vu H. **2018**. A vision based method for automatic evaluation of germination rate of rice seeds. *2018 1st International Conference on Multimedia Analysis and Pattern Recognition, MAPR 2018 - Proceedings* **2018-Janua**: 1–6.

Oellrich A, Walls RL, Cannon E, Cannon SB, Cooper L, Gardiner J, Gkoutos G V, Harper L, He M, Hoehndorf R, *et al.* 2015. An ontology approach to comparative phenomics in plants. *Plant Methods* **11**: 10.

Paparella S, Araújo SS, Rossi G, Wijayasinghe M, Carbonera D, Balestrazzi A. **2015**. Seed priming: state of the art and new perspectives. *Plant Cell Reports* **34**: 1281–1293.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al.* 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**: 2825–2830.

Pieruschka R, Schurr U. **2019**. Plant Phenotyping: Past, Present, and Future. *Plant Phenomics* **2019**: 1–6.

Pound MP, Atkinson JA, Townsend AJ, Wilson MH, Griffiths M, Jackson AS, Bulat A, Tzimiropoulos G, Wells DM, Murchie EH, *et al.* 2017. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience* **6**: 1–10.

Rampasek L, Goldenberg A. **2016**. TensorFlow: Biology's Gateway to Deep Learning? *Cell Systems* **2**: 12–14.

Ranal MA, De Santana DG. **2006**. How and why to measure the germination process? *Revista Brasileira de Botanica* **29**: 1–11.

Reyazul M, Rouf N, Choudhary B, Singh I, Khandy A, Bawa V, Sofi P, Wani A, Kumari S, Shalu Jain A, *et al.* 2015. Harnessing Genomics Through Phenomics. In: Kumar J, Pratap A, Kumar S, eds. Phenomics in Crop Plants: Trends, Options and Limitations. New Delhi: Springer, 273–283.

Reynolds D, Ball J, Bauer A, Davey R, Griffiths S, Zhou J. **2019a**. CropSight: A scalable and open-source information management system for distributed plant phenotyping and IoT-based crop management. *GigaScience* **8**: 1–11.

Reynolds D, Baret F, Welcker C, Bostrom A, Ball J, Cellini F, Lorence A, Chawade A, Khafif M, Noshita K, *et al.* 2019b. What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Science* **282**: 14–22.

Ronneberger O, Fischer P, Brox T. **2015**. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 234–241.

Roussel J, Geiger F, Fischbach A, Jahnke S, Scharr H. **2016**. 3D surface reconstruction of plant seeds by volume carving: Performance and accuracies. *Frontiers in Plant Science* **7**: 1–13.

Sadeghi-Tehran P, Virlet N, Sabermanesh K, Hawkesford MJ. **2017**. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods* **13**: 1–16.

Sansone SASA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, *et al.* 2012. Toward interoperable bioscience data. *Nature genetics* **44**: 121.

Sasaki Y. **2007**. The truth of the F-measure. *Teach Tutor mater* **1**: 1–5.

Schölkopf B, Platt J, Shawe-Taylor J, Smola A, Williamson R. **2001**. Estimating the support of a high-dimensional distribution. *Neural computation* **13**: 1443–1471.

Schumann AW, Little KM, Eccles NS. **1995**. Suppression of seed germination and early seedling growth by plantation harvest residues. *South African Journal of Plant and Soil* **12**: 170–172.

Shipman JW. **2013**. *Tkinter 8.5 reference: a GUI for Python*. New Mexico.

Smith PT, Cobb BG. **1991**. Accelerated Germination of Pepper Seed by Priming with Salt Solutions and Water. *HortScience* **26**: 417–419.

Stauffer C, Grimson WEL. **2003**. Adaptive background mixture models for real-time tracking. : 246–252.

Szeliski R. **2010**. *Computer Vision : Algorithms and Applications*. Springer Science & Business Media.

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. **2017**. Plant Phenomics, From Sensors to Knowledge. *Current Biology* **27**: R770–R783.

Teixeira PCN, Coelho Neto JA, Rocha H, De Oliveira JM. **2007**. An instrumental set up for seed germination studies with temperature control and automatic image recording. *Brazilian Journal of Plant Physiology* **19**: 99–108.

TeKrony D, Egli D. **1991**. Relationship of Seed Vigor to Crop Yield: A Review. *Crop Science* **31**: 816–822.

van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. **2014**. Scikit-image: image processing in Python. *PeerJ* **2**: 1–18.

Watson A, Ghosh S, Williams M, Cuddy WS, Simmonds J, Rey M-D, Hatta MAM, Hinchliffe A, Steed A, Reynolds D, *et al.* 2018. Speed breeding: a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**: 23–29.

Wen L, Gao L, Li. X. **2017**. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **49**: 136–144.

Wu D, Guo Z, Ye J, Feng H, Liu J, Chen G, Zheng J, Yan D, Yang X, Xiong X, *et al.* 2019. Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *Journal of Experimental Botany* **70**: 545–561.

Xiong X, Duan L, Liu L, Tu H, Yang P, Wu D, Chen G, Xiong L, Yang W, Liu Q. **2017**. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **13**: 1–15.

Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J. **2020**. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant* **13**: 187–214.

Yasrab R, Atkinson JA, Wells DM, French AP, Pridmore TP, Pound MP. **2019**. RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. *GigaScience* **8**: 1–16.

Yoshida T, Nishimura N, Kitahata N, Kuromori T, Ito T, Asami T, Shinozaki K, Hirayama T, T. H. **2006**. *ABA-hypersensitive germination3* encodes a protein phosphatase 2C (AtPP2CA) that strongly regulates abscisic acid signaling during germination among Arabidopsis protein phosphatase 2Cs. *Plant Physiology* **140**: 115–126.

Zhang C, Si Y, Lamkey J, Boydston RA, Garland-Campbell KA, Sankaran S. **2018**. High-throughput phenotyping of seed/seedling evaluation using digital image analysis. *Agronomy* **8**: 1–14.

Zhou J, Applegate C, Alonso AD, Reynolds D, Orford S, Mackiewicz M, Griffiths S, Penfield S, Pullen N. **2017a**. Leaf-GP: An open and automated software application for measuring growth phenotypes for arabidopsis and wheat. *Plant Methods* **13**: 1–30.

Zhou J, Reynolds D, Websdale D, Cornu T Le, Gonzalez-Navarro O, Lister C, Orford S, Laycock S, Finlayson G, Stitt T, *et al.* 2017b. CropQuant: An automated and scalable field phenotyping platform for crop monitoring and trait measurements to facilitate breeding and digital agriculture. *bioRxiv*: 161547.

Zhou J, Tardieu F, Pridmore T, Doonan J, Reynolds D, Hall N, Griffiths S, CHENG T, ZHU Y, WANG X, *et al.* 2018. Plant phenomics: history, present status and challenges. *Journal of Nanjing Agricultural University* **41**: 580–588.

Tables

Table 1 Table of validation metrics used to compare between manual counting and SeedGerm scoring.

	r **(Cumulative Rate)**	***r*** **(Image ID)**	**MAE (hours)**	**F1**
Barley	0.981072	0.803827	13.27500	0.961039
Brassica	0.992307	0.885759	9.140625	0.935779
Maize	0.993662	0.873848	3.542857	0.985507
Pepper	0.999013	0.952276	6.025000	0.993631
Tomato	0.992766	0.888033	4.903226	0.991736

r denotes Pearson's correlation, MAE denotes mean absolute error, F1 denotes F1 score.

Figure legends

Figure 1: Two types of SeedGerm hardware with wired and wireless connectivity used for acquiring seed germination image series for different crop species.

(a) A set of imaging devices (dimensions: H31, W39, D48 cm) installed in a cold-room for acquiring seed germination images during experiments. Image sensors used in the fixed design are controlled by *Raspberry Pi* computers (in the small window) and data is transferred via Ethernet connection. **(b)** The mini-gantry design is equipped with an undistorted wide-angle lens USB camera affixed to the top of the germination box (dimensions: H25, W30, D45 cm) for recording images of long-term germination experiments and data is transferred via WiFi connection. **(c)** Stylised view of SeedGerm devices shows the connection with a gateway machine to allow for device configuration and imaging data transfer. Single panel of Brassica seeds monitored in a short experiment illustrates the germination procedure at various establishment stages.

Figure 2: The software components and data flows of the SeedGerm software together with GUI-based analysis software designed for processing multiple germination image series using supervised machine learning, computer vision and parallel computing.

(a) Three software modules designed for automated seed imaging, data management and machine-learning based phenotypic analysis. **(b)** The input parameters that can be set by users before automated phenotypic analysis. **(c)** The background pixels in germination panels (i.e. filter paper) identified using user-configured YUV colour-space ranges. **(d)** After processing a number of experiments, germination and morphological traits are quantified and ready to be downloaded. **(e)** Cumulative germination curves, uniform and G_{max} plots produced to score seed quality and seed vigour together with dynamic seed masks recording entire germination procedures for different genotypes in a germination box.

Figure 3: Crop analysis algorithm embedded in the SeedGerm analysis software for automated seed germination scoring and phenotypic analysis.

(a) Location of individual experiments in a given germination box, which is identified from the first image by panel segmentation using the YUV colour-space ranges. **(b)** Background removal models trained using labelled background (filter paper) and foreground (seeds) pixels, which are subsequently applied to retain seed-related objects. **(c)** Descriptive statistical moments (i.e. Hu

moments) are used to measure seed morphological features for each seed in a given germination panel. Features across the image series are combined to form a set of training vectors. **(d)** A novelty detection model (one-class SVM) is trained based on the training vectors and is then applied to the time series to classify the germination status for each seed, germinated (closed circles) and non-germinated (open circles). **(e)** Germination scoring and morphological traits for all seeds are collated to produce cumulative germination curves and germination timing plots.

Figure 4: Germination-related and morphological traits quantified by the SeedGerm software.

(a) The time-lapse image series of six tomato genotypes (384 seeds) acquired in an eight-day experiment. **(b)** The quantification of germination related traits, including cumulative germination curves, T_{50} germination rates, and G_{max} final germination rate. For the whisker box plot in the middle, boxes show the quartile values (25-75%) of the T_{50} germination data, with vertical lines (coloured orange) at the median. The whiskers extended from the boxes show the range of the data (i.e. the lower quartile minus 1.5 multiplied by the interquartile range, the upper quartile plus 1.5 multiplied by the interquartile range). Data beyond the whiskers are considered as outliers (open circles). **(c)** The quantification of morphological traits, which include seed area, width and length ratio, extent, and circularity. Coloured shading areas denote confidence intervals, between the 15^{th} and 85^{th} percentiles of the data. The three red dashed lines indicate the corresponding image IDs when germination has started, 50% of seeds have germinated, and 75% of seeds have germinated.

Figure 5: Germination scoring for four crop seeds with varied experimental settings together with comprehensive morphological traits measurements.

(a) 384 tomato seeds (six genotypes) used for seed germination experiments, producing six cumulative germination curves on hourly measures during a seven-day period. **(b)** 486 pepper seeds (six genotypes) used for germination experiments, producing six cumulative germination curves on hourly measures during an eight-day period. **(c)** 120 barley seeds (three genotypes) used for germination experiments, producing three cumulative germination curves on hourly measures during a six-day period. **(d)** 35 maize seeds (one genotype) used for a germination experiment, producing a cumulative germination curve on hourly measures during a six-day period. **(e)** Morphological measurements produced by plotting hourly changes against the duration of experiments, so that all experiments can be compared on similar bases, including a number of

traits such as seed area (in pixels), W/L ratio (0~1), seed circularity (0~1), and convex area (in pixels). Coloured shading areas denote confidence intervals, between the 15th and 85th percentiles of the data.

Figure 6: Pearson's correlation (*r*) performance metric to evaluate SeedGerm's scoring using manual cumulative germination rates and seed-by-seed germination scoring for five crop species.

(a) For all tested species, SeedGerm's predictions display a strong linear correlation and goodness of fit ($r > 0.98$) based on cumulative germination rates. Each point represents the number of seeds classified as germinated in a panel in an image, meaning multiple populations are plotted together. The red line displays SeedGerm's cumulative count equalling the cumulative manual count. The number of panels associated with each scatterplot is denoted as *p*. **(b)** Seed-by-seed scoring between SeedGerm and specialists' counting plotted to demonstrate the reliability of SeedGerm scoring as well as its tendency to predict additional germinated seeds at the end of crowded experiments.

Figure 7: SeedGerm applied to detect genetic differences of 88 varieties of *Brassica napus* with a range of germination behaviours. SeedGerm scored the T_{10}, T_{50}, T_{90} and G_{max} after 8 days.

(a) Germination time (in hours) for T_{10}, T_{50}, and T_{90} cumulative germination rates for 88 *B. napus* varieties. **(b)** G_{max} cumulative germination rates for 88 *B. napus* varieties. **(c)** A SNP association between G_{max} and genotype on chromosome A5, blue dashed line indicates significant associations with an FDR of 0.1. **(d)** Box plot to show germination scores for variety sets with each allele score for the most significant SNP on chromosome A5. Boxes show the quartile values (25-75%) of the maximum germination data, with horizontal lines at the median and crosses at the mean. The whiskers extended from the boxes show the range of the data. Data beyond the whiskers are plotted as outliers (open circles). **(e)** Analysis with correlated gene expression markers showed that the expression of at least two genes in this region are correlated with germination, demonstrating sufficient accuracy of SeedGerm to perform standard genetic analysis of seed dormancy and germination rate. Blue dashed line indicates significant associations with an FDR of 0.005.

Supporting Information

Notes S1 SeedGerm hardware design and cost

Notes S2 The analysis results of SeedGerm

Notes S3 Code fragments for background-removal and germination novelty detection algorithms

Notes S4 The selection of image IDs before SeedGerm analysis

Datasets S1 Cumulative germination data

Datasets S2 Morphological traits

Datasets S3 Barley cumulative data

Datasets S4 Tomato cumulative data

Datasets S5 Pepper cumulative data

Datasets S6 Maize cumulative data

Datasets S7 Barley morphological data

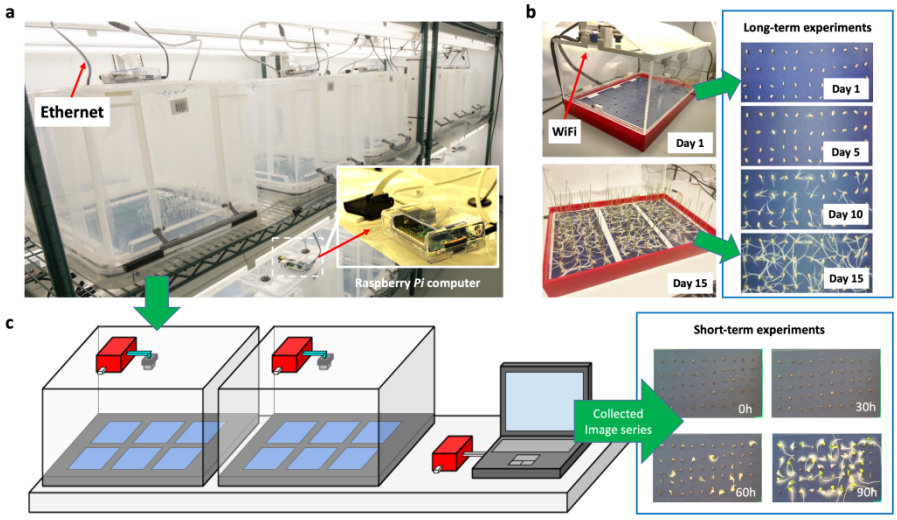
Datasets S8 Tomato morphological data

Datasets S9 Pepper morphological data

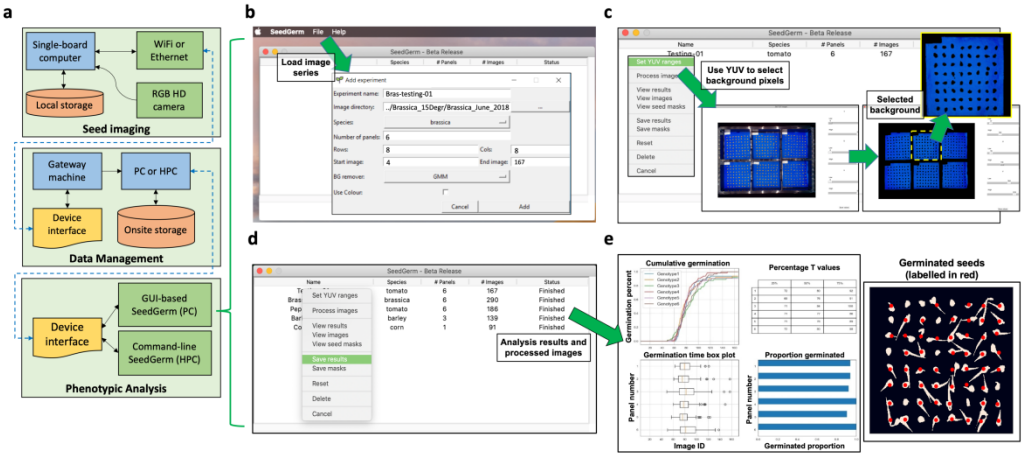
Datasets S10 Maize morphological data

Datasets S11 *B. napus* SeedGerm and manual comparison

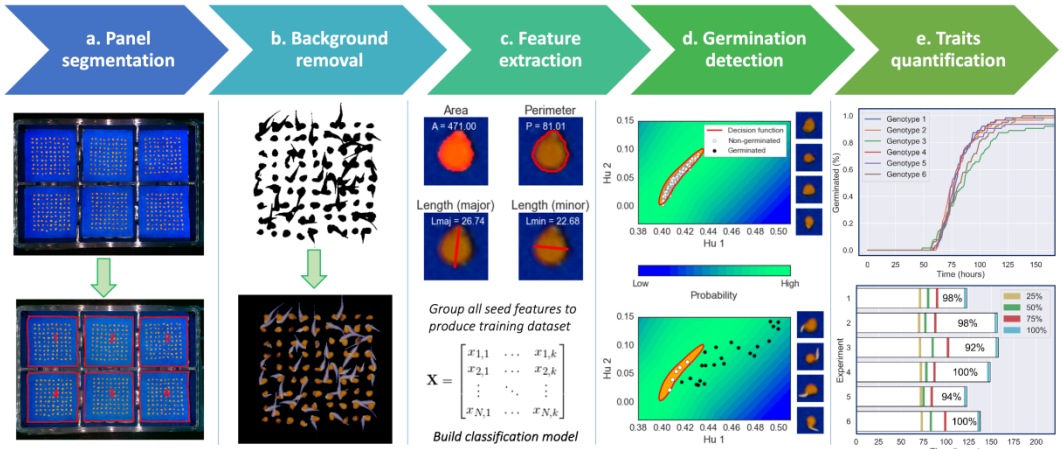
Video S1 Time series of seed germination and automated scoring



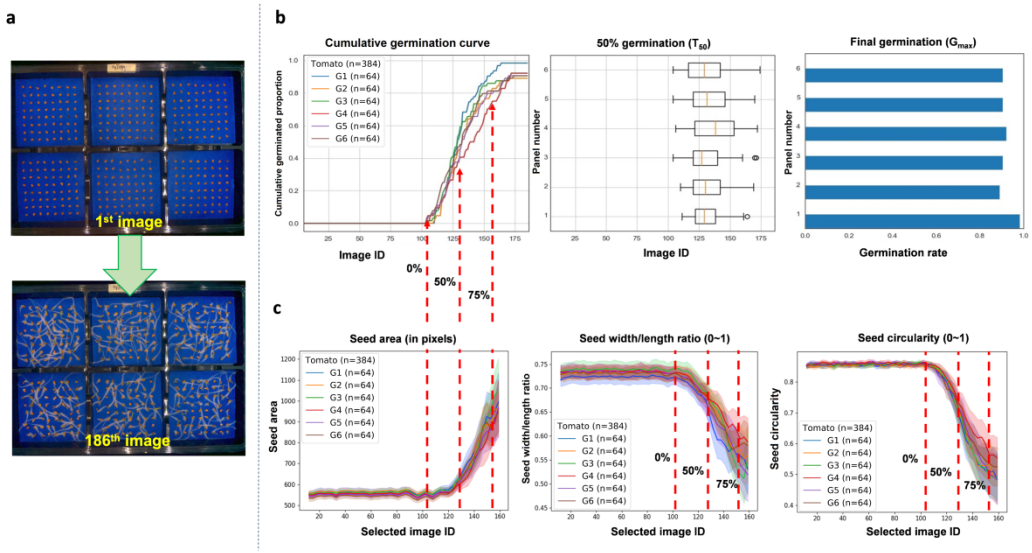
nph_16736_f1.png



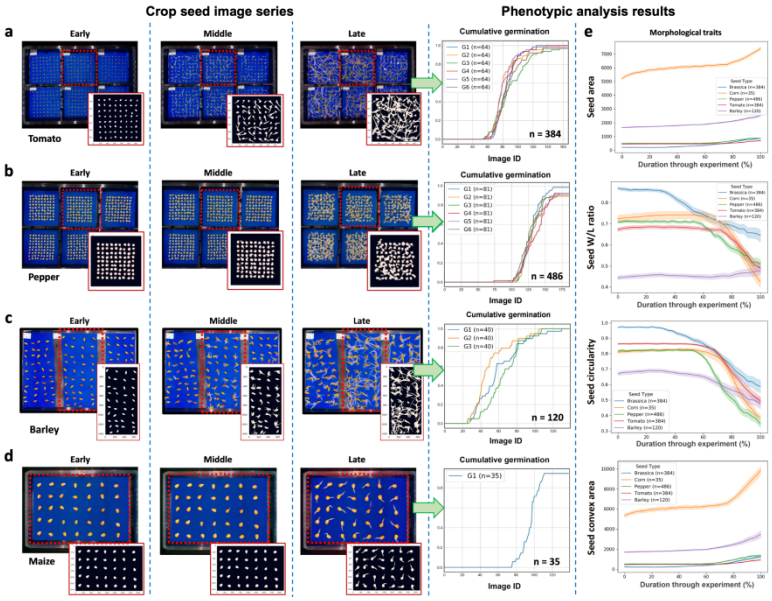
nph_16736_f2.png



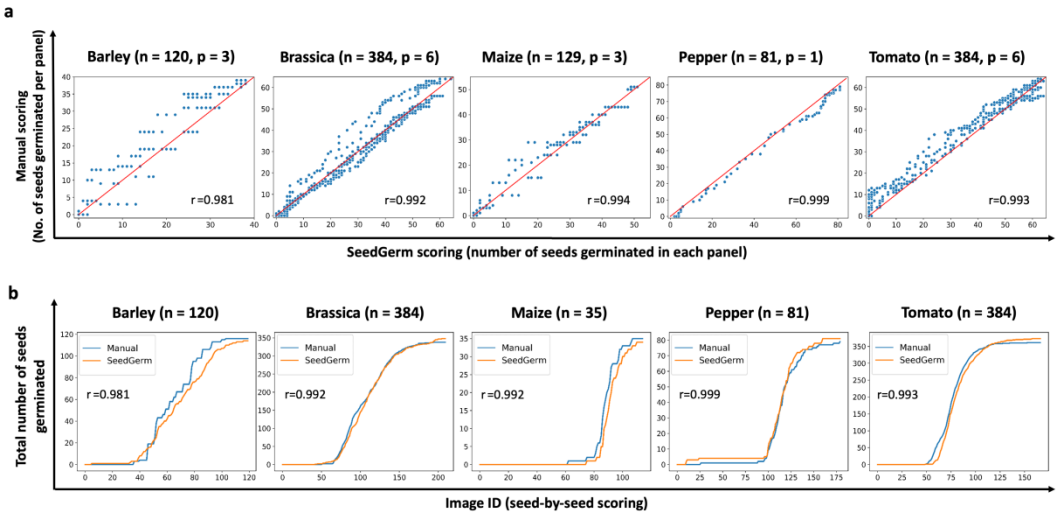
nph_16736_f3.png



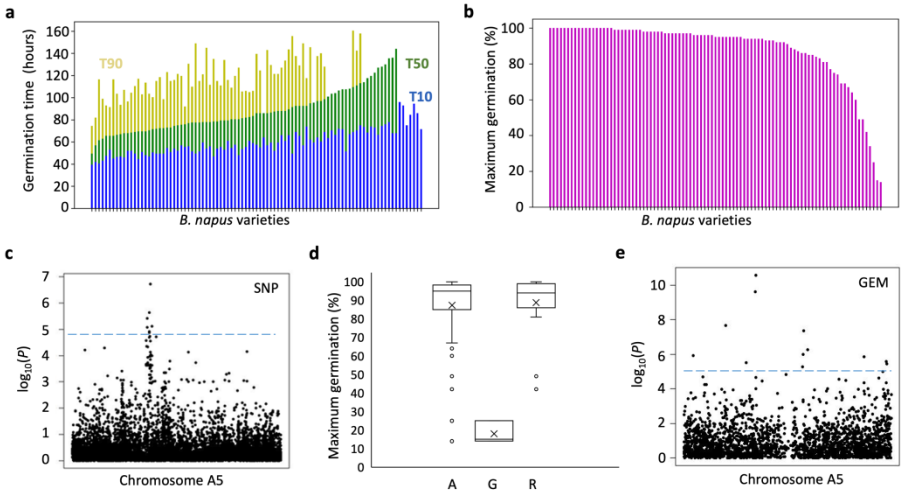
nph_16736_f4.png



nph_16736_f5.png



nph_16736_f6.png



nph_16736_f7.png