

# GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination

Ronny V. L. Joosen<sup>1,\*</sup>, Jan Kodde<sup>2</sup>, Leo A. J. Willems<sup>1</sup>, Wilco Ligterink<sup>1</sup>, Linus H. W. van der Plas<sup>1</sup> and Henk W.M. Hilhorst<sup>1</sup>

<sup>1</sup>Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands, and

<sup>2</sup>Bioscience, Plant Research International, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

Received 18 August 2009; revised 25 November 2009; accepted 9 December 2009.

\*For correspondence (fax +31317484740; e-mail Ronny.Joosen@wur.nl).

## SUMMARY

Over the past few decades seed physiology research has contributed to many important scientific discoveries and has provided valuable tools for the production of high quality seeds. An important instrument for this type of research is the accurate quantification of germination; however gathering cumulative germination data is a very laborious task that is often prohibitive to the execution of large experiments. In this paper we present the GERMINATOR package: a simple, highly cost-efficient and flexible procedure for high-throughput automatic scoring and evaluation of germination that can be implemented without the use of complex robotics. The GERMINATOR package contains three modules: (i) design of experimental setup with various options to replicate and randomize samples; (ii) automatic scoring of germination based on the color contrast between the protruding radicle and seed coat on a single image; and (iii) curve fitting of cumulative germination data and the extraction, recap and visualization of the various germination parameters. The curve-fitting module enables analysis of general cumulative germination data and can be used for all plant species. We show that the automatic scoring system works for *Arabidopsis thaliana* and *Brassica* spp. seeds, but is likely to be applicable to other species, as well. In this paper we show the accuracy, reproducibility and flexibility of the GERMINATOR package. We have successfully applied it to evaluate natural variation for salt tolerance in a large population of recombinant inbred lines and were able to identify several quantitative trait loci for salt tolerance. GERMINATOR is a low-cost package that allows the monitoring of several thousands of germination tests, several times a day by a single person.

**Keywords:** germination, automated scoring, curve fitting, Arabidopsis, image analysis.

## INTRODUCTION

Seeds are the most sophisticated means of propagation created by plant evolution. They are indispensable for human society as food sources and as starting materials for new crops. Seed physiology and technology have provided valuable tools for the production of high quality seeds, various seed treatments and optimal storage conditions. In fundamental research, seeds are studied exhaustively and systems biology approaches are undertaken to fully explore dormancy and germination (Penfield and King, 2009). An important instrument to indicate the performance of a seed lot is the accurate quantification of germination by gathering cumulative germination data.

Completion of germination is defined as the protrusion of the radicle through the endosperm and seed coat (Bewley, 1997). The uptake of water of the dry seed during imbibition is triphasic and consists of a rapid initial uptake (phase I)

followed by a plateau phase (phase II) and a further increase in water uptake (phase III). During phase III the embryo axis elongates and breaks through the testa. In Arabidopsis the testa is dead tissue whereas the endosperm layer is living tissue. The action of several cell wall-modifying proteins is required to enable a break through the endosperm. For accurate scoring of seed germination a careful discrimination should be made between the testa and endosperm rupture because this lag phase may vary among germination conditions and treatments (Liu *et al.*, 2005; Finch-Savage and Leubner-Metzger, 2006; Muller *et al.*, 2006).

Although very often used, the total percent germination after a nominated period of time is not very explanatory. It lacks information about start, rate and uniformity of germination, which are essentially parameters of a normally distributed seed population, for many traits such as

dormancy, stress tolerance and seed aging. Information about germination at various time intervals is required to calculate a cumulative germination curve, but the number of samples that can be handled with manual counting is usually the limiting factor. Moreover, *Arabidopsis* seeds are small, requiring the use of a binocular or magnifying glass. Therefore, a fast and reliable automated procedure would enable high-throughput screens and unlock the full potential of seed science research.

*Arabidopsis thaliana* is a popular model plant for seed science and provides insight in common physiological processes which can be translated to economically important crops (Lin *et al.*, 1999). The availability of mutants, ecotypes, inbred populations and sequence information enables the molecular-genetic analysis of many seed germination-related traits. For example, mutant analysis has identified seeds with reduced dormancy and altered flavonoid biosynthesis, as well as altered germination tolerances to stresses like salt and osmotic potential, desiccation, heat and cold (Shirley *et al.*, 1995; Leon-Kloosterziel *et al.*, 1996; Espinosa-Ruiz *et al.*, 1999; Hong and Vierling, 2000; Wehmeyer and Vierling, 2000; Kim *et al.*, 2005). Screens for natural variation in the various available inbred populations revealed, among other traits, loci involved in dormancy, storability, glucosinolate production, salt tolerance, storage oil production and mineral content (Schaar *et al.*, 1997; Bentsink *et al.*, 2000; Kliebenstein *et al.*, 2001; Quesada *et al.*, 2002; Hobbs *et al.*, 2004; Vreugdenhil *et al.*, 2004). Mutant analysis and quantitative trait loci (QTL) localization is complemented with the exhausting inventory provided by transcriptomics, proteomics and metabolomics approaches (Gallardo *et al.*, 2001; Cadman *et al.*, 2006; Routaboul *et al.*, 2006; Goda *et al.*, 2008). Taken together, it is evident that *Arabidopsis* has become an important and valuable tool for seed scientists but that high-throughput detailed phenotyping for effects on seed germination could extend its prospects.

When studying natural variation for germination performance in an inbred population or performing mutant screens, the number of germination assays required is tremendous. In this type of large experiments it is difficult to manually score germination at multiple time points per day for a number of days or weeks. Although major progress for semi-automated scoring of seed germination of *Lactuca sativa*, using a flat-bed scanner, was made by Teixeira *et al.* (2007), the setup suggested by these authors only allows a limited number of samples. Also the system with a camera above a Jacobsen table for *Helianthus annuus* seeds as described by Ducournau *et al.* (2005) does not accommodate high-throughput screens without expensive robotics, as it requires proper alignment between two consecutive images. The setup that we developed enables large screens without the need for expensive robotics. We are making use of germination trays which are kept in climatized cabinets.

Digital photographs are made from these trays at flexible time intervals and automatically analyzed by our Germinator scripts. The power of this procedure is that it does not score germination based on the difference between two consecutive pictures but instead uses the information from two different color threshold analyses on a single picture, which circumvents alignment problems.

Interpretation of germination performance can be accomplished by extracting the relevant parameters from the germination-time curve. We have used a method described by El-Kassaby *et al.* (2008) to mathematically fit the germination curve using the four-parameter Hill function (4PHF). This function allows extraction of biologically relevant parameters such as maximum percentage of germination ( $G_{max}$ ), time to reach 50% germination  $t_{50}$ ,  $t_x$  = time to reach a user-defined percentage of germination and uniformity of germination (for example  $U_{7525}$ : time interval between 25% and 75% of viable seeds to germinate). Integration of the area under the curve (AUC) provides a value that enumerates these parameters and often shows a high discriminative power between samples. To enable the quick analysis of many cumulative germination curves in large experiments we developed a curve-fitting module which results in a clearly formatted output that summarizes the biological relevant parameters, describing germination behavior. The curve-fitting module enables analysis of any type of cumulative germination data and is not restricted to any plant species.

Various experiments were performed to test and validate the procedure. We used a 1-h interval measurement to quantify the germination of *Arabidopsis* accessions *Landsberg erecta* (Ler) and *Columbia* (Col). We compared manual with automatic counting and assessed the accuracy of the curve fitting at different time intervals. Furthermore, we tested the procedure for germination of *Arabidopsis* Col. at different concentrations of NaCl. The application of salt stress results in different levels of maximal germination percentage, germination rate and uniformity of germination which provides an ideal test for the flexibility and accuracy of our automatic germination scoring procedure. Next, an *Arabidopsis* recombinant inbred population consisting of 165 lines was used to show the power of high-throughput germination phenotyping. Plant salt tolerance is a complex trait that is polygenic and hence difficult to dissect and manipulate. We used the GERMINATOR package to score and analyze germination in control versus salt conditions (which equals ~2000 germination assays), and tested the genetic variation for salt tolerance. Finally, we analyzed germination of seeds from a *Brassica* spp. recombinant inbred line with a huge variation in seed color to show that the GERMINATOR package might be applicable to many more species. Recently it was shown in maize that measuring the rate of germination is a good indicator for relative vigour and field performance, which underlines the importance of

high-throughput methods for scoring germination in commercial crop testing as well (Khajeh-Hosseini *et al.*, 2009).

## RESULTS

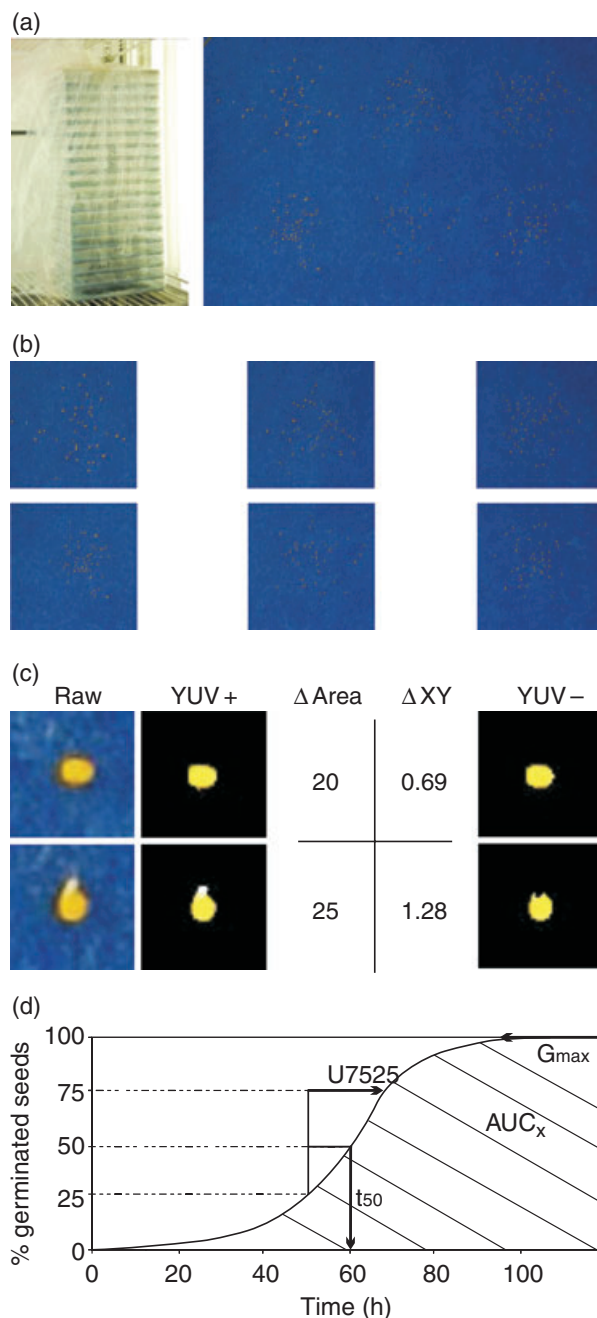
We have divided the process of analyzing germination in three basic steps: experimental setup, image analysis and data analysis. These three steps are represented by different modules of the GERMINATOR package and can be carried out independently. However, especially in large scale experiments a solid administration is crucial. Therefore we complemented the package with a Microsoft Office Excel visual basic script that creates an overview of the experiments performed with active links to the generated output (Germinator\_menu1.0.xls).

### Experimental setup

To enable as much automation as possible we have standardized the whole experimental setup. In the first module of the GERMINATOR package (Germinator\_table 1.0.xls) the user can define the number of samples, treatments, and repetitions and whether a randomized setup is desired. These choices result in an 'experiment setup' (ES) table which can be used to set up the experiment. The exact starting times of the individual tests can be added to the ES tables. Multiple time ranges within one experiment are allowed and can be handled by both the automatic scoring and curve fitting scripts. We use transparent germination trays that can be stacked in an incubator with light from the sides (Figure 1a, see Experimental procedures for details). The content of these trays, consisting of a blue filter paper with six samples of seeds, are manually photographed at different time intervals. The blue filter paper is used to obtain optimal contrast between seed, radicle and filter paper. All images are automatically named with tray number, date and time. These data are used to automatically match the pictures to the correct tray, different treatments and samples and extract information about the time intervals as mentioned in the ES tables.

### Image analysis

In large scale experiments the number of images can become prohibitive; therefore an automated procedure for image analysis is required. First, the images are batch preprocessed in Adobe Photoshop CS3 with the action 'crop.atn', which divides each image into six individual pictures and saves them under a unique name (Figure 1b). Subsequent image analysis is performed with ImageJ and is based on segmentation by color-thresholding (Figure 1c). Using visual scripting for IMAGEJ (Baecker and Travo, 2006) two batch scripts were developed which perform contrast enhancement, color-threshold, invert image, particle analysis and reporting of the results. Every image is analyzed twice; firstly with a color threshold that only selects the yellow/brown seed ( $Y_{100-255}U_{0-80}V_{130-255}$ ) and secondly with



**Figure 1.** The workflow of the automated scoring of Arabidopsis germination. (a) A pile of plastic germination trays in a climatized cabinet and the raw image from each tray. (b) An Adobe Photoshop action crops the raw picture in six individual pictures. (c) Scoring of germination is based on the double color thresholding of a single image; indicated are the raw image, the  $\Delta$ area and the  $\Delta$ XY (both in pixels) between the color threshold that selects seedcoat only (YUV-) and the color threshold that selects both seedcoat and radicle (YUV+). (d) Cumulative germination data is used as input for the curve fitting module. Multiple germination parameters are automatically extracted.  $G_{max}$  indicates the maximum germination capacity of a seed lot. The  $t_{50}$  is the time required for 50% of viable seeds to germinate ( $t_{50}$ ). Uniformity ( $U_{7525}$ ) of germination is the time interval between 75% and 25% of viable seeds to germinate. The area under the curve ( $AUC_x$ ) is the integration of the fitted curve between  $t = 0$  and a user-defined endpoint (x).

a color-threshold that selects everything but the background ( $Y_{100-255}U_{0-130}V_{80-255}$ ). For this reason, fungal contamination during the experiment should be prevented as much as possible because this may cause false-positive scoring of germination. The YUV model defines a color space in terms of one luma (Y) and two chrominance (UV) components. By using this color space we obtained the best separation between seed coat and protruding radicle. In both analyses the XY position (average of X and Y coordinates of all the pixels in the selection) and size (area + perimeter in pixels) for each individual seed are extracted to output tables, which will be saved in tab-delimited format. The output tables are analyzed with the help of a Microsoft Excel visual basic script (Germinator\_table 1.0.xls) that compares the XY position and the size of each individual seed. Seeds are scored as not germinated when both the difference between XY position and size of the two color thresholds are within a user defined limit. To prevent artifacts caused by clustered seeds, a size restriction is added. The total number of seeds is extracted from the first image; this number is used in the later time points to calculate the number of germinated seeds based on the detection of non-germinated seeds. To set accurate thresholds for both XY-position and size differences we developed a 'parameter screen' function as part of the Germinator table script that empirically compares manual versus automatic counts and determines the most optimal settings. The germination data and time intervals are transported to the initial ES tables. These final cumulative germination tables can automatically be loaded into the third Germinator module (Germinator\_curve-fitting1.0.xls), which performs curve fitting and parameter extraction (Figure 1d).

#### Data analysis

Using the visual basic module from the Microsoft Excel package we developed a script which performs automated curve fitting on cumulative germination data using the Solver add-in (Germinator\_curve-fitting1.0.xls). The Solver is used in combination with the least sum of squares method to find the right parameters to fit the curves to the four-parameter Hill function (El-Kassaby *et al.*, 2008):

$$y = y_0 + \frac{ax^b}{c^b + x^b}$$

where  $y$  is the cumulative germination percentage at time  $x$  (h),  $y_0$  is the intercept on the  $y$  axis ( $\geq 0$ ),  $a$  is the maximum cumulative germination percentage ( $\leq 100$ ),  $b$  is controlling the shape and steepness of the curve and  $c$  is the time required for 50% of viable seeds to germinate ( $t_{50}$ ). Initial values for the parameters  $a$  and  $c$  are extracted from the cumulative germination count and  $b$  is set to 20. With these initial values the solver performs an iterative process (max 10 000) until the sum of squares between the measured cumulative germination and the calculated curve does not

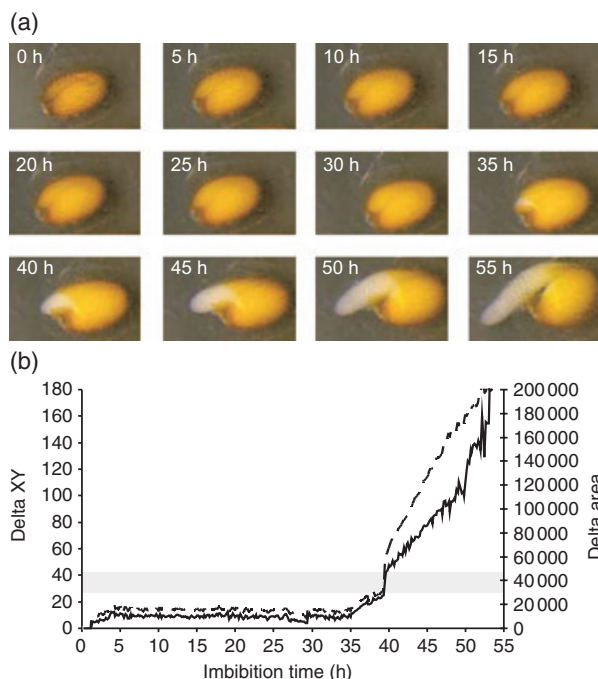
decrease any further. Because in rare cases the first iteration does not result in optimal parameters a second iteration is performed using the results of this first iteration as starting values. The iteration resulting in the lowest sum of squares is taken as the final result. The user can define a threshold for the minimum number of germinated seeds, since curve fitting on very small amounts of germinated seeds will not be very informative. In these situations the script returns a 'false' and the data are not used in the statistical analysis. Uniformity ( $U_{b-a}$ ) of germination is the time interval between  $a\%$  and  $b\%$  of viable seeds to germinate. Users can define values used for  $a$  and  $b$ . The area under the curve (AUC) is the integration of the fitted curve between  $t = 0$  and a user-defined endpoint, which results in a parameter that combines information on maximum germination,  $t_{50}$  and uniformity. As described by (El-Kassaby *et al.*, 2008) the AUC can also be used to calculate a dormancy index (DI), by subtracting the AUC after dormancy release (e.g. by cold stratification) with the AUC of dormant seeds. By the same analogy the AUC can be used to measure the effect of any stress treatment and calculate a stress index (SI). The Germinator curve fitting script will summarize the results by calculating averages and standard errors for repeated samples, performing Student's  $t$ -test, and provides a clearly formatted output including graphs for the different germination parameters.

#### Accuracy and flexibility

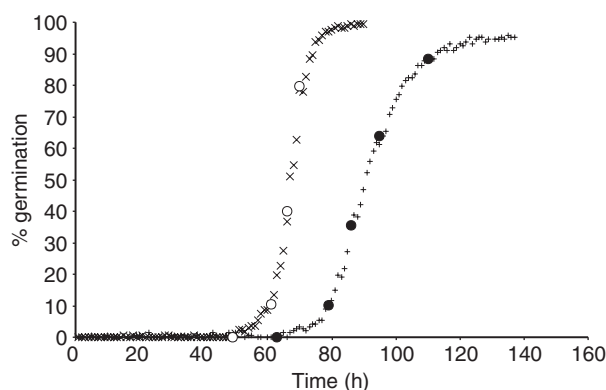
The completion of seed germination of *Arabidopsis* is a two-step process: first rupture of the testa, followed by the protrusion of the radicle through the micropylar endosperm (Liu *et al.*, 2005). The germination of a single seed was followed in time with high resolution imaging (see <http://www.pph.wur.nl/UK/seedlab/researchinterests/> for a time-lapse movie). The two steps of *Arabidopsis* germination are clearly distinguishable on these images: testa rupture after 35 h followed by endosperm rupture after 40 h (Figure 2a). The difference in the threshold area and threshold XY position in time are shown in Figure 2(b). This figure clearly shows that both the increase in area and the shift in XY position can serve as accurate indicators for germination *sensu stricto*.

To test the accuracy of the automatic germination scoring with lower resolution images that can be used to study seed batches we performed an interval experiment measuring the progression of germination of seed lots from two *Arabidopsis thaliana* accessions (Ler; 147 seeds and Col-0; 172 seeds) every hour. The automatic counts were verified at 9 time points by manual counting (Figure 3, Table 2, Table S1).

Measuring germination at 1-h intervals provides very accurate data, which enables precise curve-fitting. However, it is impossible to apply this without expensive robotics in large scale experiments. Therefore, we wanted to assess the



**Figure 2.** Analysis of germination. Comparison of the difference in either the area or XY position from a double color-threshold approach using a 10-min time lapse imaging series of a single seed. Dashed line = delta XY, line = delta area. The gray bar indicates the range for both area and XY position in which germination *sensu stricto* can be determined.



**Figure 3.** Comparison of manual and automated scoring of *Arabidopsis thaliana* Col-0 (x) and Ler (+) germination. Open circles, Col-0; filled circles, Ler.

effect of the number of measurements on the accuracy of the fitted curve using the data from the experiment depicted in Figure 3. Different time intervals were artificially created by removing data points from the 1-h interval dataset and curve fitting was tested to assay the minimum number of required data points during germination (Table S2).

From the example in Table 1 it is clear that our curve-fitting module is able to accurately predict the various parameters. The desired interval will be dependant on the

required accuracy and the level of difference between samples. Often, it is more convenient for practical reasons to use flexible intervals. Therefore, we tested with five replicates of an *Arabidopsis thaliana* Col-0 seed lot for which only six time points were acquired; care had been taken to obtain at least two measurements during the exponential phase of the curve (Figure 4, Table 2, Table S3).

This experiment shows that an accurate prediction for the various germination parameters can be obtained by as less as six data points with two data points in the exponential phase of the curve.

Salt tolerance during germination is an important but complex trait. Salt stress consists of an ionic and osmotic component which is influencing homeostasis signaling pathways, detoxification response pathways, and pathways for growth regulation (Zhu, 2002). Therefore, salt may influence the lag phase between testa and endosperm rupture, radicle growth and seedling establishment. We used germination on salt to test the accuracy of the measurement of germination *sensu stricto*. Figure 5 shows 25 seeds at different stages of germination during imbibition in 125 mM NaCl. Careful optimization of the threshold for both the area and XY difference between the double-color threshold enables scoring of germination which resembles manual scoring as close as possible. The most accurate threshold settings can be determined with the 'parameter screen' script that we included in the Germinator\_table file.

Both the germination rate and the maximum germination capacity are inhibited by sodium chloride (NaCl). We used a concentration range of NaCl to test the accuracy and flexibility of the automatic scoring on six replicates of an *Arabidopsis thaliana* Col-0 seed lot, here we have set the limit for  $\Delta$ area to 30 pixels and  $\Delta$ XY to 1.2 (Figure 6, Table S4).

Separation of the germination behavior in specific parameters can help to describe and compare many lines. Figure 7 shows a comparison of the four different parameters and their discriminative power for germination under salt stress based on the germination characteristics shown in Figure 6.

As shown in Figures 6 and 7 the cumulative germination is inhibited by NaCl. At lower concentrations only the  $t_{50}$  is reduced where at higher concentrations the maximum germination ( $G_{max}$ ) is affected as well.

### Natural variation for salt tolerance

To fully exploit the power of high throughput cumulative germination data we used the Germinator scripts to screen the core set (165 lines) of a Bay-0  $\times$  Sha recombinant inbred population (Loudet *et al.*, 2002) for salt tolerance. After-ripened seeds were germinated on water and 100 mM NaCl without prior stratification. We performed duplicate measurements of three different harvests resulting in a total of 1980 individual germination assays. Values obtained for

**Table 1** Comparison of germination curve fitting parameters  $t_{50}$ , uniformity ( $U_{7525}$ ) and area under the curve until 120 h ( $AUC_{120}$ ) with measurements at different time intervals (h)

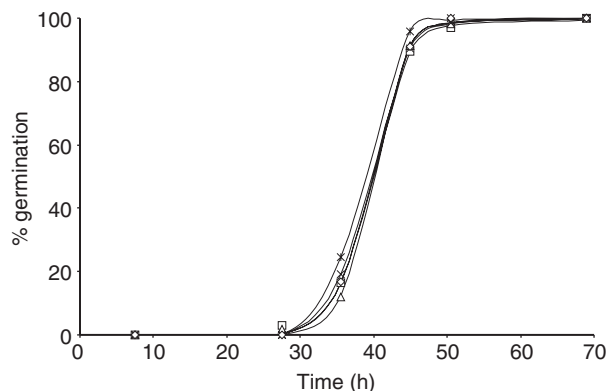
Interval (h)	Col-0				Ler			
	$t_{50}$	$U_{7525}$	$AUC_{120}$	$r^2$ fit	$t_{50}$	$U_{7525}$	$AUC_{120}$	$r^2$ fit
1	67.06	6.19	81.62	0.999	90.78	14.72	56.21	0.999
4	67.08	6.06	81.64	0.999	90.75	14.78	56.26	0.999
8	67.36	6.50	81.36	0.999	90.85	15.41	56.47	0.999
12	66.64	6.11	81.94	1.000	91.55	15.55	55.64	1.000
16	66.83	4.35	81.79	1.000	91.81	15.15	55.39	1.000

$r^2$  fit, determination coefficient;  $t_{50}$ , time to obtain 50% of germinated seeds;  $U_{7525}$ , time between 25% and 75% of germinated seeds;  $AUC_{120}$ , area under the curve until 120 h.

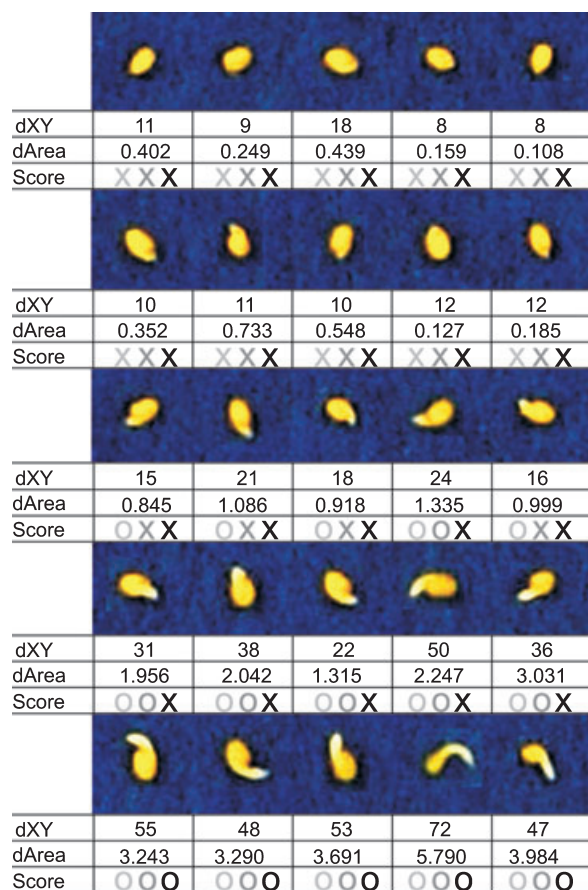
**Table 2** Parameters characterizing seed germination curves of five replicates of *Arabidopsis thaliana* Col-0 (Figure 4)

	1	2	3	4	5
Number of seeds	78	67	58	63	49
$r^2$ fit	1.000	0.999	1.000	1.000	1.000
$t_{50}$	39.1	39.3	39.5	38.9	37.8
$U_{7525}$	5.1	5.4	4.7	5.5	4.6
$AUC_{60}$	20.7	20.4	20.3	20.8	22.1

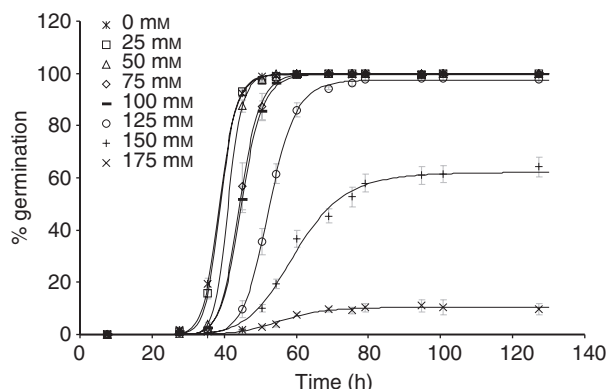
$r^2$  fit, determination coefficient;  $t_{50}$ , time to obtain 50% of germinated seeds;  $U_{7525}$ , time between 25% and 75% of germinated seeds;  $AUC_{60}$ , area under the curve until 60 h.

**Figure 4.** Germination curves of five replicates (indicated by the various symbols) of *Arabidopsis thaliana* Col-0. Different symbols indicate different replicates.

germination on 100 mM salt were subtracted from values derived from germination on water (Table S5). Figure 8(a–c) shows the frequency distribution of non-normalized data for  $G_{max}$ ,  $t_{50}$  and  $AUC$  in this population. Both parental lines are indicated with an arrow showing the large extent of transgression. After normalization (see Experimental procedures for details) of these trait data we detected multiple QTL for salt tolerance in six regions (Figure 8d). The QTL for both maximum germination and area under the curve could

**Figure 5.** *Arabidopsis* germination on 125 mM NaCl on blue filter paper. A compilation of images of different stages of germination derived from the original images used for automatic scoring is shown. Indicated are the differences in XY position and area between two color thresholds and scoring of germination based on different threshold settings (x = not germinated, 0 = germinated at thresholds: light grey = 20 area/0.8 XY, grey = 30 area/1.2 XY, black = 40 area/3.0 XY).

explain 49% of the total variance. The QTL for  $t_{50}$  could explain 39% of the total variance. No QTL for uniformity ( $U_{7525}$ ) was detected. The QTL on top of chromosome 1 is affecting germination capacity ( $G_{max}$ ) but not  $t_{50}$ . By



**Figure 6.** Germination of *Arabidopsis thaliana* (Col-0) seeds on different concentrations of NaCl. Results were analyzed with the Germinator curve fitting module. Error bars represent SEM ( $n = 6$ ).

contrast, we see QTL on chromosome 5 that affect rate of germination without affecting germination capacity.

### Scoring *Brassica* germination

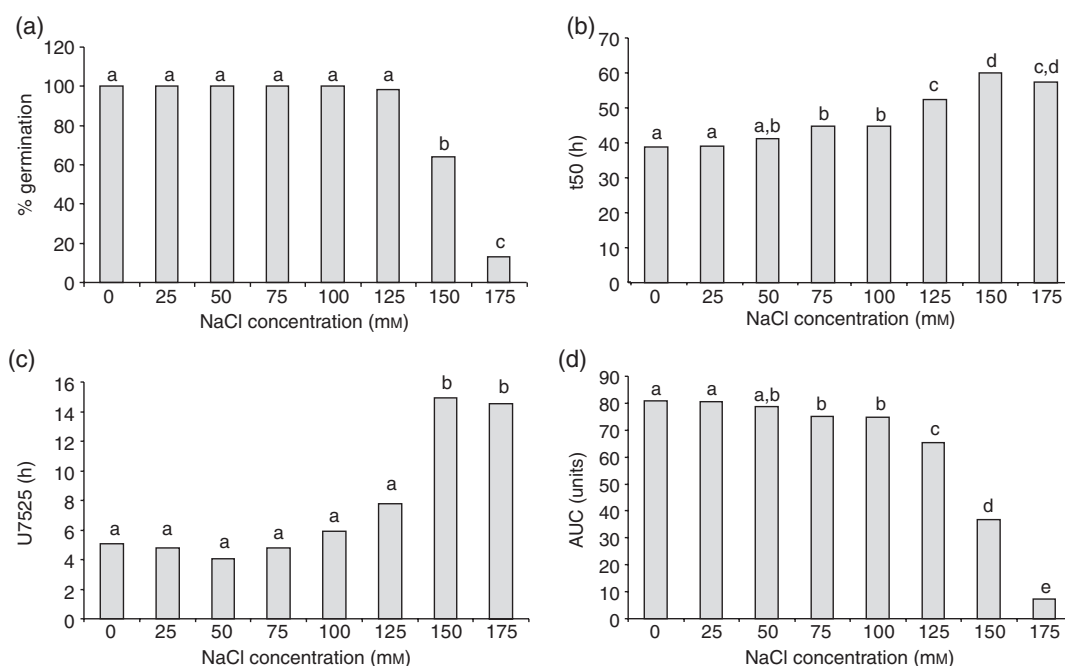
Currently the whole Germinator procedure is optimized for use with *Arabidopsis* but it is probably suitable to handle many other species as well. To test whether the same script is also applicable to other species we tested some lines from a *Brassica* doubled haploid population from which the seeds

strongly varied in color. Although some seeds are almost black while others are pale yellow it was possible to define two color thresholds which can distinguish between the protruding radicle and the rest of the seed (Figure 9). The differences between the two color thresholds can be used to automatically score germination, here we have set the limit for  $\Delta$ area to 100 pixels and  $\Delta$ XY to 1.4 (Figure 10, Table S6). Every seed with values below one of both limits will be scored as not germinated (e.g. Figure 9, seed III).

## DISCUSSION

### Features of Germinator

The procedure presented here represents a novel and efficient analysis tool for high-throughput monitoring of seed germination. A few other studies have used image analysis to score seed germination for other plant species such as cabbage, broccoli, cauliflower, sunflower, lentil, pepper, radish and tomato (Dell'Aquila *et al.*, 2000; Dell'Aquila, 2004, 2005, 2007, 2009; Ducournau *et al.*, 2005; Teixeira *et al.*, 2007). Basically all systems use a fixed imaging system that allows full automated scoring of germination. The advantage from such a setup is the possibility to precisely follow phenotypic properties of individual seeds and germination can be scored based on e.g. increase in seed size or changes in roundness of the individual seeds. As all these systems require a proper alignment between consecutive images



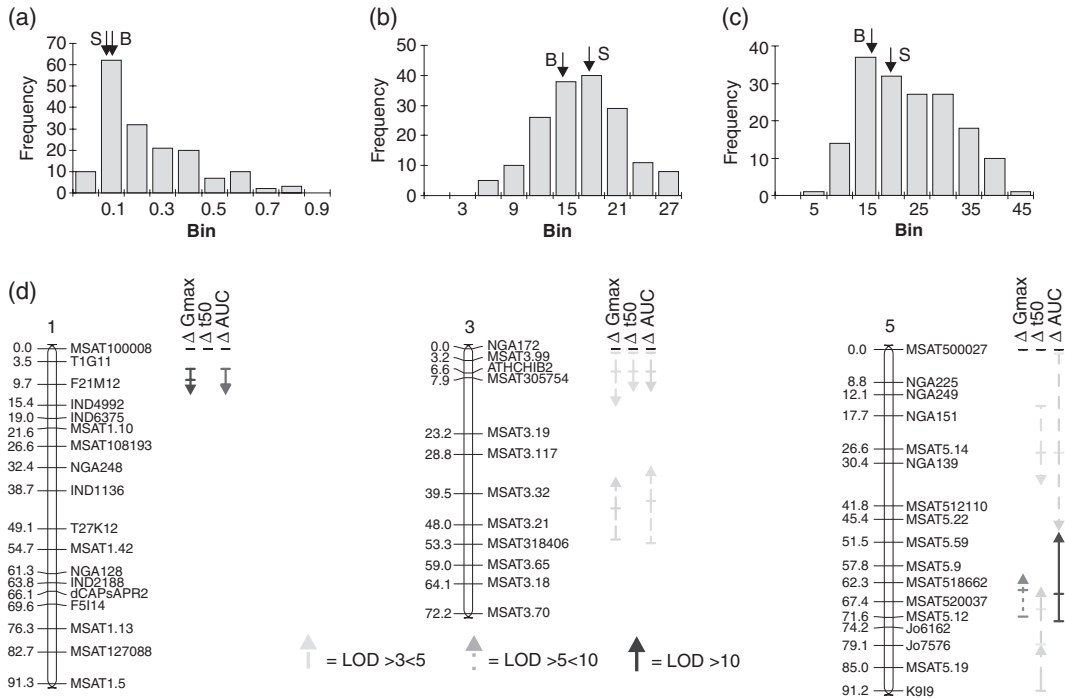
**Figure 7.** Germination of *Arabidopsis thaliana* (Col-0) seeds on different concentrations of NaCl.

(a) Maximum germination after 5 days.

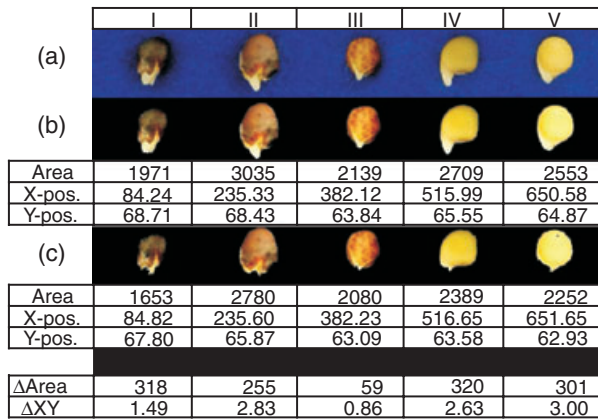
(b) Time to reach 50% germination, ( $t_{50}$ ).

(c) Uniformity ( $U_{7525}$ ).

(d) Area under the curve until 120 h (AUC<sub>120</sub>). Letters (a–e) represent statistical different subsets (Tukey HSD,  $P = 0.05$ ).

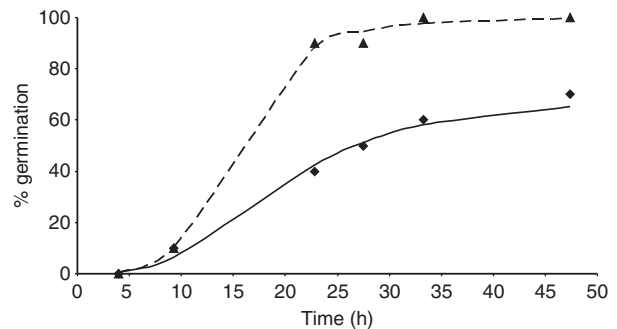


**Figure 8.** Quantitative trait locus analysis of salt tolerance of germination. Frequency distribution of non-normalized data for  $G_{max}$  (a),  $t_{50}$  (b) and area under the curve (c). (b)  $t_{50}$ ; and (c) area under the curve (AUC) in the Bay-0  $\times$  Sha recombinant inbred line (RIL) population for germination on 100 mm salt and corrected for germination on water:  $\Delta G_{max} = G_{max}(\text{water}) - G_{max}(\text{salt})$ ,  $\Delta t_{50} = t_{50}(\text{salt}) - t_{50}(\text{water})$ ,  $\Delta AUC = AUC(\text{water}) - AUC(\text{salt})$ . (d) The Bay-0  $\times$  Sha linkage map showing the genetic locations affecting germination on 100 mm salt. Mapped traits are indicated above each lane. Grayscales of the arrows indicate the LOD-score (darker = higher LOD scores). Arrows indicate the direction of the phenotypic effect; up: Sha increasing, Bay-0 decreasing; down: Bay-0 increasing, Sha decreasing. The length of the arrow depicts the 2-LOD support interval determined with restricted MQM mapping.



**Figure 9.** Five *Brassica* seeds that strongly vary in seed coat color were analyzed with the Germinator scripts. (a) Original image; (b) image after color threshold with settings:  $+Y_{0-255}U_{0-125}V_{135-255}$ ; (c) image after color threshold with settings:  $-Y_{0-255}U_{0-125}V_{120-255}$ . The difference in  $\Delta$  area and  $\Delta$  XY enables automatic scoring of germination.

they do not allow high-throughput analysis without huge investments in robotics. Therefore, we have chosen for a semi-automatic approach and have developed a system that can handle many samples which may be germinated at



**Figure 10.** Germinator scripts were used for automatic scoring of *Brassica* seed germination with strong variation in seed coat color. Two different mixtures of 24 lines that strongly vary in seed color were used.

different environmental conditions. The power of the presented procedure is that it does not score germination based on the difference between two consecutive images but instead uses the information from two different color threshold analyses on a single image. This process allows a much more flexible setup for screening large populations, but requires a good level of contrast between the radicle and seedcoat which may limit the usability for several species.



The high level of automation during experimental setup, image analysis and curve fitting provides a solid, reproducible and yet flexible system which can be implemented at very low costs. We have shown the accuracy of the automatic scoring and showed that only a limited number of measurements are needed for an accurate prediction of the germination curves. To achieve best accuracy care should be taken to obtain at least two measurements in the exponential phase of germination, although this procedure can be difficult when screening large populations with different times and rates of germination. Therefore, a critical assessment of the calculated germination curves cannot be skipped. Also, fungal contamination during the experiment should be prevented as much as possible because this may cause false positive scoring of germination. The flexibility of the curve fitting is shown by germinating *Arabidopsis* on different concentrations of NaCl, which caused reduced germination percentage, rate and uniformity. The curve fitting module was able to efficiently and accurately describe those curves.

### High-throughput phenotyping

The availability of large genetic and mutant populations offers great potential for seed science research but also demands a high-throughput procedure to score seed germination. Until now, scoring of *Arabidopsis* germination is a time consuming task, requiring manual inspection using binoculars. Especially in large scale experiments these human observations can easily lead to misjudgment and the number of experiments which can be handled is restricted by the desired interval between two inspections and the time it takes to count the individual experiments. This time is reduced dramatically in the procedure suggested here. Scoring six individual experiments (each 50–200 seeds) is reduced to the time it takes to take one photograph (approx. 5 sec). Registration of time intervals is automated and can therefore be corrected for each individual experiment. The short measuring time also contributes to the accuracy and reliability. Together, this enables the researcher to follow the cumulative germination in time and determine the germination curves in large scale experiments. Nevertheless, it should also be clear that automatic scoring of germination cannot be left unsupervised. If screening environmental perturbations like germination in the presence of NaCl one might expect effects on the lag time between testa and radicle protrusion and effects on radicle growth and seedling establishment. This requires accurate parameter thresholding which can only be achieved by manual counting of a small subset.

### Curve fitting for cumulative germination data

As described (by Brown and Mayer, 1988), fitted curves allow germination to be summarized in terms of a few curve coefficients, which offers a much better description of the

time course of germination than single value indices, such as the widely used maximum germination percentage. To optimize our analysis pipeline the Germinator curve-fitting script was developed. It provides a fast and easy tool for fitting germination curves from cumulative germination data. On a standard desktop computer (dual core 2.3, 4 Gb, Windows XP) the curve fitting and parameter extraction for 5000 germination tests was calculated in less than 15 min. The overall quality of the fit is of course strongly dependent on the quality and amount of data points but overall the coefficient of determination was close to 1.00. Here it should be noted that fewer data points automatically result in a higher coefficient of determination but that this might not always reflect the true germination curve. Therefore, the value attached to the coefficient of determination in experiments with only a few data points should be considered with care. The output from the Germinator curve fitting script contains all the parameters from the 4PHF function, total germination ( $G_{max}$ ), time to reach 50% germination ( $t_{50}$ ), uniformity ( $U_{b-a}$ ) and AUC. It offers the possibility to depict both the individual data points and the fitted curves and it can summarize the data by calculating averages and performing Student's *t*-test. The Germinator curve-fitting script enables analysis of general cumulative germination data and can be used for all plant species.

Gathering detailed germination data in an experiment on salt tolerance (Figure 6) clearly shows the added value of the cumulative germination curve compared to e.g. the total germination after 5 days. The latter is not discriminative until a concentration of 125 mM NaCl. The  $t_{50}$  is not discriminative until 75 mM NaCl and not between 150 and 175 mM. The uniformity ( $U_{7525}$ ) only shows a significant difference between the 125 and 150 mM points. The combined interpretation of the parameters  $G_{max}$ ,  $t_{50}$  and  $U_{7525}$  can accurately describe the cumulative germination curve. The AUC is summarizing these three parameters effectively and shows optimal discrimination among the different treatments (Figure 7). Calculating the difference of the AUC between germination on water and germination on a specific concentration of NaCl generates a value (stress index, SI), which can summarize the effect of the salt treatment based on maximum germination,  $t_{50}$  and  $U_{7525}$ . This approach can be used for any type of stress as well as for the release of dormancy by e.g. cold stratification. This parameter was introduced by El-Kassaby *et al.* (2008) as a useful index to describe dormancy (dormancy index, DI). We suggest using this parameter as well for normalized values of stress treatments.

### Natural variation for salt tolerance

The ability to handle large scale experiments was shown in a screen for allelic variation for salt tolerance in the *Arabidopsis* Bay-0 × Sha recombinant inbred population (165 lines). Repetitions and water control experiments raised the

number of individual germination experiments to 1980. It would have been impossible for one person to manually count this large number of experiments multiple times a day. The same experiment also clearly shows the large benefit of acquiring detailed germination curves (Figure 8d). The QTL on top of chromosome 1 is affecting germination capacity ( $G_{\max}$ ) but not rate of germination ( $t_{50}$ ). On the contrary, on chromosome 5 QTL are observed that affect rate of germination without affecting germination capacity. The AUC is summarizing both parameters and show QTL that are affected either in germination capacity or rate. Multiple QTL for germination on NaCl were identified in six regions. Distinct loci where either Bay-0 or Sha alleles improved germination were found, which could explain the observed transgression (Figure 8a). Comparing QTL for maximum germination capacity found in the Arabidopsis Ler  $\times$  Sha population (Clerkx *et al.*, 2004) revealed that the QTL on chromosome 1, 3 and lower arm of chromosome 5 could be in the same regions and show similar directions of the Shadara allelic effects. One of the apparent advantages of using cumulative germination data over endpoint germination is the ability to measure genetic variation for stress tolerance at lower concentrations (Figure 7). Furthermore, it is known that salt tolerance is realized via distinct pathways for high and low salt concentrations (Munnik *et al.*, 1999). Cumulative germination data might allow separate analysis of these pathways, whereas endpoint germination might be restricted to pathways for higher concentrations.

### Prospects

Although we optimized the Germinator scripts for *Arabidopsis thaliana* we were able to show that the same basic setup can also be employed for other seeds which have a good contrast between the seed coat and protruding radicle. The *Brassica* seeds we used for this test displayed considerable variation in seed color but, nevertheless, it was possible to define a color threshold setting that efficiently distinguished between seed coat and the protruding radicle.

We show that we have developed a package for high-throughput seed germination phenotyping. The Germinator pipeline offers a well defined and robust experimental setup but is very flexible in terms of numbers and treatments. The improved efficiency and absence of subjectivity are great advantages of computer aided assessment. The procedure presented in this paper offers great potential to perform high-throughput germination tests in large mutant or genetic populations. Automatic germination scoring is optimized for use with Arabidopsis and will most likely work for many other species as well. The curve fitting script enables analysis of general cumulative germination data and can be used for all plant species. Although we tried to optimize the package it is of crucial importance to set accurate thresholds by comparing the automated scoring with manual scoring. In conclusion, Germinator is a low-cost package that allows

the monitoring of several thousands of germination tests, several times a day by a single person.

## EXPERIMENTAL PROCEDURES

### Plant material and growth conditions

*Arabidopsis thaliana* plants from accessions Columbia and Landsberg erecta were grown on soil in a climate chamber (20°C day, 18°C night) with 16 h of light (35 W m<sup>-2</sup>) at a relative humidity of 70%. Seeds were bulk harvested and stored at 20°C under ambient relative humidity (around 40%) for 5 months. Seeds from the core population (165 lines) of the Arabidopsis Bayreuth-0  $\times$  Shadara recombinant inbred population (Loudet *et al.*, 2002) were obtained from the Versailles Biological Resource Centre for Arabidopsis (<http://dbsgap.versailles.inra.fr/vnat/>) and were grown in triplicate of 5 plants each in a fully randomized setup. Plants were grown on 4  $\times$  4 cm rockwool plugs (MM40/40, Grodan B.V., <http://www.grodan.com>) and watered with 1 g/l Hyponex (NPK = 7:6:19 <http://www.hyponex.co.jp>) fertilizer in a climate chamber (20°C day, 18°C night) with 16 h of light (35 W m<sup>-2</sup>) at a relative humidity of 70%. Seeds were bulk harvested and after ripened until they reached their maximum germination after 5 days of imbibition. Subsequently, the seeds were dried for 1 week at a relative humidity of 20% and stored at -80°C until further experimentation. To prevent fungus contamination during the experiment we surface sterilized the seeds by placing 50 mg of seeds per seed lot for 2 h in a dessicator jar above a solution of 100 ml 4% sodium hypochlorite + 3 ml concentrated HCl. *Brassica rapa* plants from a combined Double Haploid (DH) population containing plants from the DH38 population of Ping Lou *et al.* (2008) and plants from similar but reciprocal crossing were grown in the greenhouse. Seeds were harvested and stored at 20°C until use. Seeds from 24 lines representing the different classes of seed coat colors were used to test the Germinator.

### Germination assay

Germination experiments were performed in plastic (15  $\times$  21 cm) trays (DBP Plastics, <http://www.dbp.be>) containing 42 ml water or NaCl solution and two layers of blue filter paper (5.6'  $\times$  8' Blue Blotter Paper; Anchor Paper Company, <http://www.seedpaper.com>). Six samples of approximately 50–200 Arabidopsis seeds were dispersed on the filter paper using a mask to ensure an accurate and reproducible spacing. Clustering of seeds was prevented as much as possible. A maximum of 20 trays were piled with, on both the top and the bottom of the stack, two empty trays with 42 ml water and two layers of blue filter paper to prevent unequal evaporation and ensure equal distribution of light. The whole pile was wrapped in a closed transparent plastic bag and placed in an incubator. The incubator (type 5042; Seed Processing Holland, <http://www.seedprocessing.nl>) provides light from three sides and was set to a temperature of 20°C. For the interval experiment, the lower filter paper was used as a wedge inserted in a tray filled with water to prevent drying of the seeds and enable automatic hourly measurements. The experiment was carried out in an air-conditioned room (20°C). Experiment set up, automatic scoring and curve fitting was performed with the GERMINATOR package.

### Imaging

A digital camera (Nikon D80 with Nikkor AF-S 60 mm f/2.8 G Micro ED; Nikon, <http://www.nikon.com>) was fixed to a repro stand and connected to a computer, using Nikon camera control pro software version 2.0. Two vertically placed fluorescent tl-tubes (150 cm), 1.5 m left and right from the camera, were used as indirect light source; great care was taken to prevent any reflection. The camera

was set to full manual control (ISO400, F/18, 1/3 sec, manual focus). Image files are named following a strict convention: mmddyy-hhmm#seq, whereby the seq is an automatic sequential number indicating the tray number. A position mask is used to make sure that the trays are placed at the correct position under the camera.

### QTL analysis

For QTL analysis a genetic map consisting of 69 markers (provided by <http://dbsgap.versailles.inra.fr/vnat/>) with an average distance between the markers of 6.1 cM was used. To test and correct normality of the trait values we used the software package DISTRIBUTION ANALYZER v1.2 (<http://www.variation.com>). Multiple-QTL model mapping (MQM) was carried out by using the software package MAPQTL (version 5.0; Kyazma B.V., <http://www.kyazma.nl>). Cofactors were selected according to the program's reference manual and the 2LOD interval was determined with restricted MQM mapping. MAPCHART v2.2 (Plant Research International, <http://www.pri.wur.nl/UK>) was used to construct the linkage map shown in Figure 8.

### Downloading Germinator

The full GERMINATOR package (for Windows operating systems) is freely available for the scientific community. It can be easily downloaded from the website (<http://www.pph.wur.nl/UK/seedlab/resources/germinator/>). This website also contains a full manual and video demonstrations about the use of the various modules.

### ACKNOWLEDGEMENTS

This work was supported by the Technology Foundation STW, the Applied Science Division of NWO and the Technology Program of the Ministry of Economic Affairs. We would like to thank Marie Retiere for testing the *Brassica* seeds.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Table S1.** The Germinator curve-fitting module containing the raw cumulative germination data, fitted curves and extracted parameters from the interval experiment shown in Figure 3.

**Table S2.** The Germinator curve-fitting module containing the raw cumulative germination data, fitted curves and extracted parameters from the reduced interval experiment shown in Table 1.

**Table S3.** The Germinator curve-fitting module containing the raw cumulative germination data, fitted curves and extracted parameters from five replicates of Col-0 as shown in Figure 4 and Table 2.

**Table S4.** The Germinator curve-fitting module containing the raw cumulative germination data, fitted curves and extracted parameters from germination on various concentrations of NaCl as shown in Figures 5 and 6.

**Table S5.** Values for germination on NaCl for the *Arabidopsis* Bay × Sha recombinant inbred core population (Figure 8). All values are corrected for germination on water by subtraction.

**Table S6.** The Germinator curve-fitting module containing the raw cumulative germination data, fitted curves and extracted parameters from scoring *Brassica* germination as shown in Figure 10.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

### REFERENCES

- Baecker, P. and Travo, P. (2006) Cell image analyzer: a visual scripting interface for ImageJ and its usage at the microscopy facility Montpellier RIO. *Proc. ImageJ User Developer Conf.* **1**, 105–110.
- Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S.P.C. and Koornneef, M. (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiol.* **124**, 1595–1604.
- Bewley, J.D. (1997) Seed germination and dormancy. *Plant Cell*, **9**, 1055–1066.
- Brown, R.F. and Mayer, D.G. (1988) Representing cumulative germination. 2. The use of the Weibull function and other empirically derived curves. *Ann. Bot.* **61**, 127–138.
- Cadman, C.S.C., Toorop, P.E., Hilhorst, H.W.M. and Finch-Savage, W.E. (2006) Gene expression profiles of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. *Plant J.* **46**, 805–822.
- Clerkx, E.J.M., El-Lithy, M.E., Vierling, E., Ruys, G.J., Blankestijn-De Vries, H., Groot, S.P.C., Vreugdenhil, D. and Koornneef, M. (2004) Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions *Landsberg erecta* and Shakhara, using a new recombinant inbred line population. *Plant Physiol.* **135**, 432–443.
- Dell'Aquila, A. (2004) Cabbage, lentil, pepper and tomato seed germination monitored by an image analysis system. *Seed Sci. Technol.* **32**, 225–229.
- Dell'Aquila, A. (2005) The use of image analysis to monitor the germination of seeds of broccoli (*Brassica oleracea*) and radish (*Raphanus sativus*). *Ann. Appl. Biol.* **146**, 545–550.
- Dell'Aquila, A. (2007) Pepper seed germination assessed by combined X-radiography and computer-aided imaging analysis. *Biol. Plant.* **51**, 777–781.
- Dell'Aquila, A. (2009) Digital imaging information technology applied to seed germination testing. A review. *Agron. Sustain. Develop.* **29**, 213–221.
- Dell'Aquila, A., Van Eck, J.W. and Van Der Heijden, G.W.A.M. (2000) The application of image analysis in monitoring the imbibition process of white cabbage (*Brassica oleracea* L.) seeds. *Seed Sci. Res.* **10**, 163–169.
- Ducournau, S., Feutry, A., Plainchault, P., Revillon, P., Vigouroux, B. and Wagner, M.H. (2005) Using computer vision to monitor germination time course of sunflower (*Helianthus annuus* L.) seeds. *Seed Sci. Technol.* **33**, 329–340.
- Ei-Kassaby, Y.A., Moss, I., Kolotelo, D. and Stoehr, M. (2008) Seed germination: mathematical representation and parameters extraction. *Forest Sci.* **54**, 220–227.
- Espinosa-Ruiz, A., Bellés, J.M., Serrano, R. and Culiñán-Macià, F.A. (1999) *Arabidopsis thaliana* AtHAL3: a flavoprotein related to salt and osmotic tolerance and plant growth. *Plant J.* **20**, 529–539.
- Finch-Savage, W.E. and Leubner-Metzger, G. (2006) Seed dormancy and the control of germination. *New Phytol.* **171**, 501–523.
- Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2001) Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiol.* **126**, 835–848.
- Goda, H., Sasaki, E., Akiyama, K. et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* **55**, 526–542.
- Hobbs, D.H., Flintham, J.E. and Hills, M.J. (2004) Genetic control of storage oil synthesis in seeds of *Arabidopsis*. *Plant Physiol.* **136**, 3341–3349.
- Hong, S.W. and Vierling, E. (2000) Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proc. Natl Acad. Sci. USA*, **97**, 4392–4397.
- Khajeh-Hosseini, M., Lomholt, A. and Matthews, S. (2009) Mean germination time in the laboratory estimates the relative vigour and field performance of commercial seed lots of maize (*Zea mays* L.). *Seed Sci. Technol.* **37**, 446–456.
- Kim, Y.O., Kim, J.S. and Kang, H. (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. *Plant J.* **42**, 890–900.
- Kliebenstein, D.J., Gershenzon, J. and Mitchell-Olds, T. (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics*, **159**, 359–370.
- Leon-Kloosterziel, K.M., van de Bunt, G.A., Zeevaart, J.A. and Koornneef, M. (1996) *Arabidopsis* mutants with a reduced seed dormancy. *Plant Physiol.* **110**, 233–240.

- Lin, Y., Sun, L., Nguyen, L.V., Rachubinski, R.A. and Goodman, H.M. (1999) The Pex16p homolog SSE1 and storage organelle formation in *Arabidopsis* seeds. *Science*, **284**, 328–330.
- Liu, P.P., Koizuka, N., Homrichhausen, T.M., Hewitt, J.R., Martin, R.C. and Nonogaki, H. (2005) Large-scale screening of *Arabidopsis* enhancer-trap lines for seed germination-associated genes. *Plant J.* **41**, 936–944.
- Lou, P., Zhao, J., He, H., Hanhart, C., Del Carpio, D.P., Verkerk, R., Custers, J., Koornneef, M. and Bonnema, G. (2008) Quantitative trait loci for glucosinolate accumulation in *Brassica rapa* leaves. *New Phytol.* **179**, 1017–1032.
- Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D. and Daniel-Vedele, F. (2002) Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor. Appl. Genet.* **104**, 1173–1184.
- Muller, K., Tintelnot, S. and Leubner-Metzger, G. (2006) Endosperm-limited *Brassicaceae* seed germination: abscisic acid inhibits embryo-induced endosperm weakening of *Lepidium sativum* (cress) and endosperm rupture of cress and *Arabidopsis thaliana*. *Plant Cell Physiol.* **47**, 864–877.
- Munnik, T., Ligterink, W., Meskiene, I., Calderini, O., Beyerly, J., Musgrave, A. and Hirt, H. (1999) Distinct osmo-sensing protein kinase pathways are involved in signalling moderate and severe hyper-osmotic stress. *Plant J.* **20**, 381–388.
- Penfield, S. and King, J. (2009) Towards a systems biology approach to understanding seed dormancy and germination. *Proc. R. Soc. B: Biol. Sci.* **276**, 3561–3569.
- Quesada, V., García-Martínez, S., Piqueras, P., Ponce, M.R. and Micol, J.L. (2002) Genetic architecture of NaCl tolerance in *Arabidopsis*. *Plant Physiol.* **130**, 951–963.
- Routaboul, J.M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J. and Lepiniec, L. (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. *Planta*, **224**, 96–107.
- Schaar, W.V.D., Alonso Blanco, C., Leon Kloosterziel, K.M., Jansen, R.C., Ooijen, J.W.V. and Koornneef, M. (1997) QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. *Heredity*, **79**, 190–200.
- Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M. and Goodman, H.M. (1995) Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659–671.
- Teixeira, P.C.N., Coelho Neto, J.A., Rocha, H. and De Oliveira, J.M. (2007) An instrumental set up for seed germination studies with temperature control and automatic image recording. *Braz. J. Plant Physiol.* **19**, 99–108.
- Vreugdenhil, D., Aarts, M.G.M., Koornneef, M., Nelissen, H. and Ernst, W.H.O. (2004) Natural variation and QTL analysis for cationic mineral content in seeds of *Arabidopsis thaliana*. *Plant Cell Environ.* **27**, 828–839.
- Wehmeyer, N. and Vierling, E. (2000) The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiol.* **122**, 1099–1108.
- Zhu, J.K. (2002) Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53**, 247–273.