Automated scoring of *Arabidopsis* germination

Leo Willems¹, Ronny Joosen¹, Jan Kodde², Wilco Ligterink¹, Henk Hillhorst¹

**Introduction**

High throughput analysis of seed germination for phenotyping large genetic populations or mutant collections is very labor intensive and would highly benefit from an automated setup. We have developed a set of scripts that enables automatic scoring of germination, curve fitting and parameter extraction. Although very often used, the total germination after a nominated period of time is not very descriptive as it lacks information about start, rate and uniformity, which are highly indicative for many traits like dormancy, stress tolerance and longevity. To calculate a cumulative germination curve it is required to have information about germination at various time intervals.

Manual scoring of *Arabidopsis* germination is very laborious and therefore a fast and reliable automation procedure which fulfills the following criteria is needed.

- Flexible (number of samples, environments)
- High speed of analysis
- Optimized experimental design (randomization of samples)
- Solid administration and accurate time registration
- Automatic calculation of germination curves and parameters
- Visualize and summarize results with graphs and statistics

**Automated scoring procedure**

The setup that we developed is focused on large screens without the need for expensive robotics. We make use of germination trays which are kept in climatized cabinets. At flexible time intervals digital photographs are made and analyzed by image analysis software. 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. The non germinated seeds are scored based on similar area and XY positions.

**Table E: Threshold seeds only (C)**

<table>
<thead>
<tr>
<th>Seednr.</th>
<th>Area</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61</td>
<td>617.3723</td>
<td>247.6915</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>439.4508</td>
<td>184.2049</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table F: Threshold seed and roots (D)**

<table>
<thead>
<tr>
<th>Seednr.</th>
<th>Area</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81</td>
<td>617.4723</td>
<td>247.7915</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>439.4508</td>
<td>184.2049</td>
</tr>
<tr>
<td>3</td>
<td>84</td>
<td>491.8929</td>
<td>264.0595</td>
</tr>
</tbody>
</table>

**Figure 1**: A picture from a tray that contains 6 membranes (A) which will be analyzed with imageJ. The enlargements in B, C and D show the original picture (B), the first color-threshold which only selects the seed (C) and the second color-threshold which selects everything but the background (D). Table E shows the output from ImageJ. Only seednr 2 has similar size and XY coordinates and can therefore be labeled as “non germinated”.

**Curve fitting procedure**

Interpretation of germination performance can be accomplished by extracting different parameters from the germination curve. We have used the 4 parameter Hill function to mathematically describe the shape of the germination curve.

\[
y = \frac{ay}{cb + x^b}
\]

- \(y\) = cumulative germination percentage at time \(x\)
- \(a\) = time in hours
- \(b\) = intercept on the y axis
- \(c\) = shape and steepness of the curve
- \(d\) = time required for 50% of viable seeds to germinate (t50)

For the batch analysis of large experiments we developed a script called “Germinator” which automates the process of curve fitting and results in a clearly formatted output that summarizes the biological relevant parameters describing germination behavior.

**Figure 2**: A cumulative curve in which the biological relevant parameters for germination are indicated: 
- \(g_{\text{Max}}\): maximum germination
- \(t_{50}\): time required for 50% of viable seeds to germinate
- \(u75\%\): uniformity, time between 25% and 75% seed germination
- AUC: Area Under Curve

**Results**

To test the accuracy of the automatic germination scoring we performed an hourly interval measurement with two *Arabidopsis* accessions (Ler and Col). We verified the automatic scoring by comparing it to manual counting at 9 different time points (figure 3A). We also tested a concentration range of NaCl to assess the flexibility and accuracy of the automatic scoring and curve fitting (figure 3B).

**Figure 3**: Comparison of manual (filled red circles) and automatic scoring of *Arabidopsis* Col (x) and Ler (+) germination (A), germination of Col at different concentrations of NaCl and curve fitting with the Germinator (B)

The data interpretation with the Germinator (figure 4) clearly shows the added value of the cumulative germination curve compared to e.g., the total germination after 5 days which is not discriminative until a concentration of 125 mM NaCl. The combined interpretation of the parameters \(g_{\text{Max}}\) (figure 4A), \(t_{50}\) (figure 4B) and \(u75\%\) (figure 4C) can accurately describe the germination. The AUC (figure 4D) is summarizing these three parameters effectively and shows optimal discrimination between the different treatments.

**Figure 4**: Germination of *Arabidopsis* on different concentrations NaCl. A) germination after 5 days (\(g_{\text{Max}}\)). B) time required for 50% of viable seeds to germinate (\(t_{50}\)). C) uniformity, time between 25% and 75% of germination (\(u75\%\)). D) Area Under Curve after 120 hours (AUC). Statistical different subsets (Tukey HSD, p=0.05) are indicated by letters. Error bars represent SE (n=5)

**Conclusions**

The improved efficiency and absence of subjectivity are big advantages of computer aided assessment. The procedure presented offers great potential to perform high-throughput germination tests in large mutant or genetic populations. Although we did not test the automated scoring system for other plant species we expect that the same basic setup can be employed for many seeds which have a good contrast between the seed coat and protruding radicle. The Germinator is a script which enables batch analysis of general germination data and is not restricted to any plant species.

---

¹ Laboratory of Plant Physiology, WU ² Plant Research International

P.O. Box 9101, NL-6700 HB Wageningen, The Netherlands

Tel: +31 317 47 58 97

E-mail: Leo.Willems@WUR.nl