Whole-plant image processing for synthetic genetic circuits

Introduction

Synthetic Biology not only provide us a powerful toolbox to have a better understanding of the natural genetic circuits but also enable us to design and realize functionalities that are not found in nature (1). The emerging field of Synthetic Biology is characterized by the engineering oriented methodology, usually containing four iterative steps (2): 1) quantitatively characterize parts of genetic circuits; 2) design genetic circuits in silico using mathematical modeling; 3) assemble the designed genetic circuits and test their functionalities in quantitative experiments; 4) evaluate the performances of assembled genetic circuits using quantitative data generated in step 3). Therefore, quantitative data generated by quantitative experiments plays pivotal roles in Synthetic Biology.

In most Synthetic Biology studies using bacteria, yeast and mammalian cell culture, quantitative experiments can be carried out by expression of fluorescent proteins and quantitative data can be collected through fluorescence-aided cell sorting (FACS) or image processing routines (3). However, this traditional strategy is challenging in plant Synthetic Biology due to the high content of chlorophyll, making all the other fluorescent proteins hard to quantify due to the auto-fluorescence and wide emission spectrum of chlorophylls (3). One common solution in plant Synthetic Biology is to use luminescence instead of fluorescence, which can be collected in a high signal-to-noise ratio in a low-light setting. Two-channel imaging is also possible when a dual-reporter system is used (like Renilla luciferase and firely luciferase) that enables quantification of expression levels of two components in the genetic circuits. However, the low-light imaging condition poses additional challenges to quantify the luminescence data collected. The cameras used for low-light imaging are usually ultra-sensitive, which makes it hard to image the morphologies of plants simultaneously in an optimal lighting condition. Without the plant morphology, the quantification becomes challenging.

To quantify the expression levels of synthetic genes in transgenic plants and make comparable quantifications across different treatment conditions and time points, an imaging protocol and a corresponding image processing software were developed with the ability to distinguish the luminescence emitted from different tissues of leaves and roots.

Material and Methods

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia (Col-0) was used for all transformations. Plasmids were introduced into *Agrobacterium tumefaciens* GV3101 cells by electroporation. Agrobacterium transformed transgenic lines were generated by the common floral dipping method (4). Primary (T0) transgenic lines were selected on MS (5) media with 100µg/ml Kanamycin, 34µg/ml BASTA and 100µg/ml cefotaxime and screened initially for luciferase expression by spraying 500 µm d-luciferin solution (GOLDBIO) containing 0.01%Tween-20. Homozygous plants were isolated with no segregation for the respective resistance at T3 or T2 generations. For luciferase assay, thirteen days old homozygous lines were transferred to MS media supplemented with

corresponding inducers in 100mmx100mm square plates (Fisher Brand, Cat No FB0875711A) and grown vertically under short-day conditions (10 h light/14 h dark) at 22°C growth chamber.

Imaging

Luciferase activity was measured using low-light imaging with a XR/Mega-10 ICCD camera system (Stanford Photonics, Inc) and available Piper software (v. 2.6.17) by spraying 500 µm d-luciferin solution (GOLDBIO) containing 0.01%Tween-20 and after a minimum of 30 minutes dark adaptation. To obtain proper plant boundaries for quantification purpose, a colored image was taken by a regular digital camera and a bright-field image was taken by the XR/Mega-10 ICCD camera system right before measuring luciferase activities, so that the bright-field image contains the same localization of the plants as in the low-light images.

Image processing

A custom image processing software was written in Matlab, which combines a series of image-processing steps into a standardized workflow. Graphic user interfaces (GUI) and proper instructions were programmed wherever possible, so that a new user could start using the software to quantify images almost without training. A detailed step-by-step protocol was also prepared for users' further references. To ensure accuracy of the image processing and reduce possible artifacts caused by variable sources, each step in the workflow is fully supervised by the user.

In the initiation steps of the software, the classic GUI of selecting files is used to let the users select the input images (including a high-resolution color image of the plants taken under optimal lighting conditions by a digital camera and a bright-field image collected under the luciferase camera) and the input folder containing the low-light images of luciferase activities collected by the luciferase camera. The software first processes the high-resolution color image (**Figure 1 a)**. Region of interests (ROIs) of leaves and roots are identified in two separate steps through user-supervised color thresholding, where the user is able to adjust the threshold using a sliding bar till the thresholding level of satisfaction. An image registration step is then carried out between the high resolution and the bright-field images obtained by the luciferase camera by user defined control points ($n \geq 3$) (**Figure 1 b)**. More control points may improve the accuracy of this step but we found the improvement beyond 3 control points is trivial. This step provides the necessary information about the orientation, spatial shift and scaling to align the ROIs properly to the bright-field image and the low-light image. After the alignment, the ROIs are grouped by the user into individual plants. As shown in **Figure 1 c-d**, leaves ROIs 1, 2 and 3 were assigned to plant indexed as 1, 2 and 3 respectively. Similarly, roots ROIs 1 and 2 were assigned to plants 1, ROI 3 to plant 2 and ROIs 4 and 5 to plant 3. In the next step, the ROIs of leaves and roots of individual plants are applied to the low-light images of luciferase activities and several pixel-based measurements are carried out and recorded for further analysis (**Figure 1 e-f)**. These measurements of leaf and root ROIs include areas in number of pixels, total luminescence intensity levels and a complete list of intensity level per pixel.

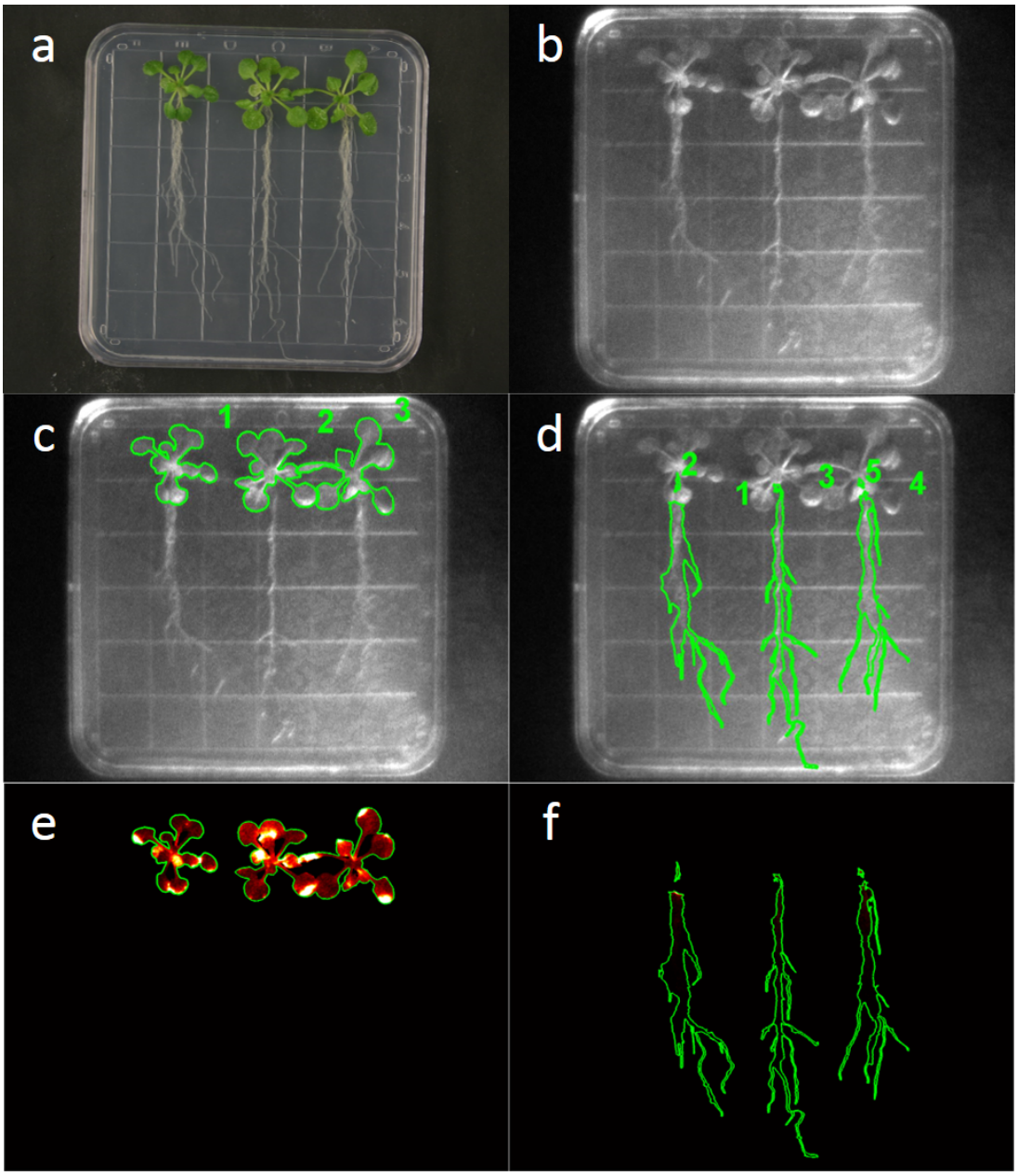

Figure 1. Whole plant image processing. **a**, High resolution color image collected using a regular digital camera under well controlled lighting conditions prior to the luciferase imaging, which was used for color thresholding and identifying the region of interests (ROIs) of leaves and roots for individual plants. **b**, Bright-field image collected *in situ* under the luciferase camera, which was used to align the ROIs obtained from **a**). It should be noted that there are differences in resolution, orientation and lighting conditions between **a)** and **b**). **c-d**, ROIs of leaves and roots aligned with the bright-field image with ROIs indexed spatially. **e-f**, ROIs of leaves and roots applied to the luminescence images with their luminescence data recorded. The two heatmaps were made using a same scale and color map.

Results

Differences between the luminescent levels across leaves and roots

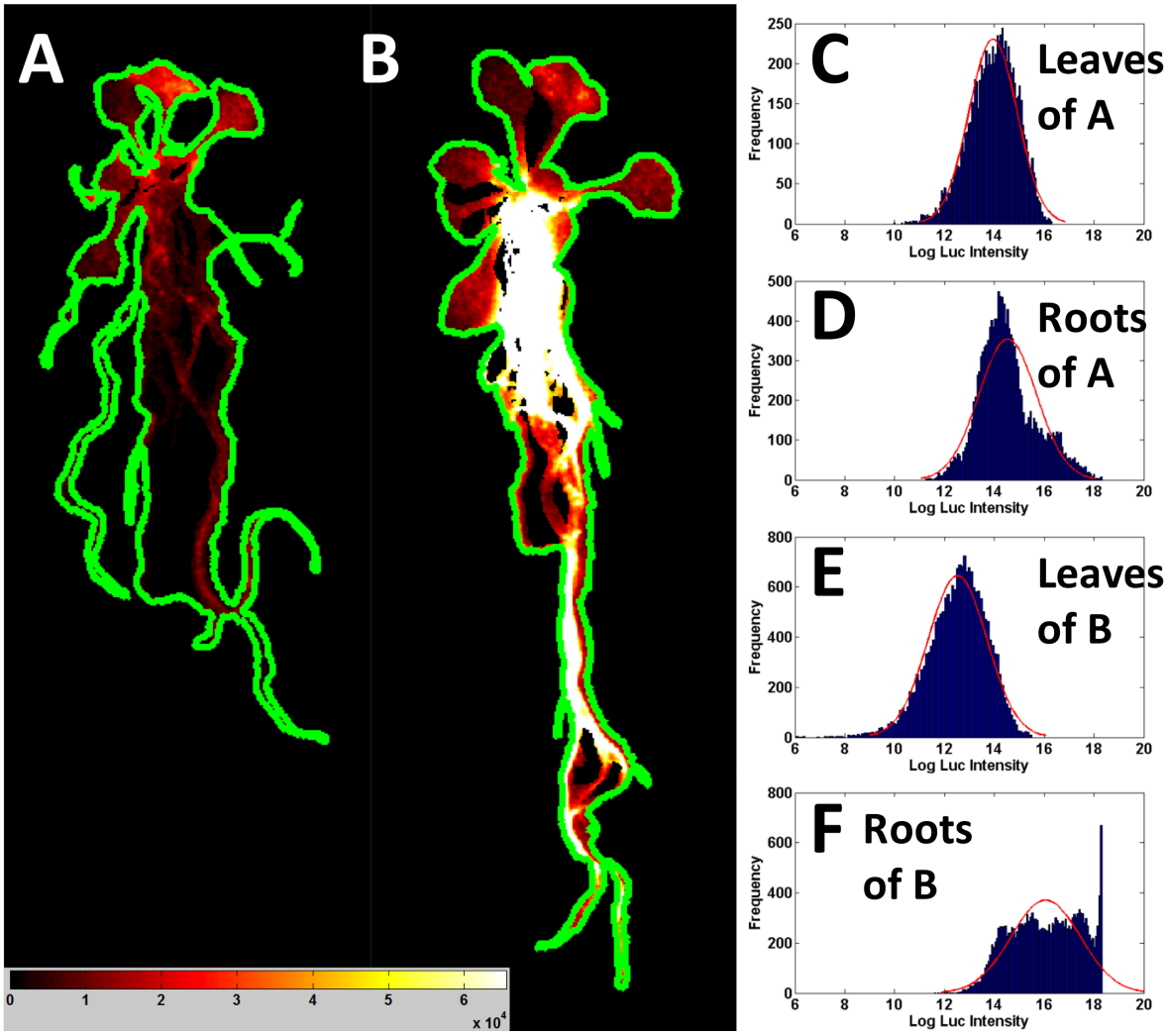

Figure 2. Different gene expression levels in difference tissue types in plants. A-B) Heatmaps of luciferase activities. Green lines mark out the boundaries of leaves and yellow lines the roots. Luciferase activities are represented by the brightness of the red color, with the brighter color indicating higher luciferase activities and the white color the saturation. C-F) Histograms of luminescence intensity per pixel. x-axis is the natural logarithmic of the luminescence intensity per pixel and the y-axis the frequency.

Similar to what was reported in a previous study (6), the distributions of luciferase luminescence per tissue of an individual plant was also found to be better described by lognormal distributions (*i.e.* a Gaussian distribution after a logarithmic transformation) than Gaussian distributions. This can be seen in the histograms in Figure 2 C-F). This helped us develop more accurate statistical summaries for the whole plant luminescence data. The exponentiated mean of logarithmic transformed data is more accurate than the direct mean, which tends to overestimate the mean due to the heavy tail of the distribution. It is the same for standard deviation. This knowledge also leads to more accurate calculation of important summarizing parameters, like the fold change between distributions before and after induction.

Furthermore, we can also determine a more accurate statistical test for the fold change as 2-sample t-test on the logarithmic scale, as the division of lognormally distributed variables is still lognormally distributed.

We also observed drastic differences in reporter expression levels of the same synthetic circuit between leaves and roots. Visually, the luciferase activities shown in the leaves are much dimmer than the roots, which are close to saturation in all the ROI (Figure 2 panel B)). Quantitatively, the histogram of luminescence intensity of the roots is significantly right-shifted compared to the leaves, and the mean expression level in roots is more than 20-times higher than the one of leaves (Figure 2 panel E-F)).

Quantify the growth rates

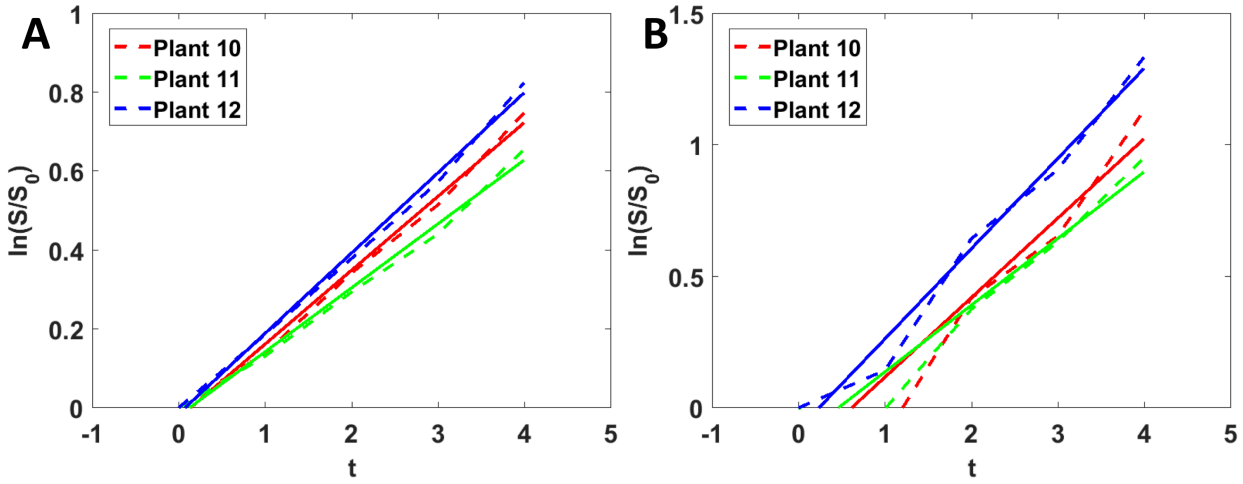

Figure 3. Growth rate of four plants measured separately in terms of the (A) leaves and (B) roots. x-axis indicates the time of measurements in unit of days. y-axis indicates the natural logarithmic of the normalized tissue sizes. The size of different tissues of each plant was firstly normalized by that of the first day and then taken the natural logarithmic. Dashed lines are experimental data and the solid lines are linear fits.

Using the information of the tissue ROIs collected in the image processing step, the plant growth can also be quantified in terms of pixel numbers. With the calibration information of length per pixel, physical units can be recovered.
We can model the growth of one tissue using a first order growth model as:

$$\frac{dS}{dt} = \gamma S$$

where the S indicates the size of a tissue (leaves or roots), t the time and γ the growth rate. The solution to this simple ODE is:

$$S = S_0 e^{\gamma t}$$

where the S_0 indicates the initial size of the tissue at t = 0. A straightforward algebra manipulation can be done as:

$$\ln \frac{S}{S_0} = \gamma t$$

Therefore, a linear curve is expected if the tissue sizes normalized by their initial sizes are plotted in a logarithmic scale against the time. This is exactly what we observed in Figure 3, where the growth curves are approximately linear. By fitting to a linear model with zero y-intercept, the growth rate constant can be estimated as the semi-logarithmic slope. This simple growth model applies relatively well for both leaves and roots (Table 1). The measured mean growth rate of roots (0.2999 ± 0.0448 day^{-1}, as mean ± standard deviation) is faster than that of leaves (0.1845 ± 0.0208 day^{-1}).

Table 1. Linear fitting results to the plant growth data shown in Figure 3. Fitting results are color-coded according to Figure 3.

Tissue	Leaves			Roots		
Plant Index	10	11	12	10	11	12
R2	0.9967	0.9960	0.9981	0.9488	0.9726	0.9892
Slope	0.1874	0.1623	0.2037	0.3026	0.2538	0.3433
y-intercept	-0.0259	-0.0203	-0.0149	-0.1861	-0.1170	-0.0812

References:

1. Nandagopal, N., and M.B. Elowitz. 2011. Synthetic Biology: Integrated Gene Circuits. Science (80-.). 333: 1244–1248.
2. Cheng, A.A., and T.K. Lu. 2012. Synthetic Biology: An Emerging Engineering Discipline. Annu. Rev. Biomed. Eng. 14: 155–178.
3. Medford, J.I., and A. Prasad. 2016. Towards programmable plant genetic circuits. Plant J. 87: 139–148.
4. Clough, S.J., and A.F. Bent. 1998. Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. 16: 735–743.
5. Murashige, T., and F. Skoog. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant. 15: 473–497.
6. Schaumberg, K.A., M.S. Antunes, T.K. Kassaw, W. Xu, C.S. Zalewski, J.I. Medford, and A. Prasad. 2016. Quantitative characterization of genetic parts and circuits for plant synthetic biology. Nat Meth. 13: 94–100.
7. Goldberg, R. 1988. Plants: novel developmental processes. Science (80-.). 240: 1460–1467.
8. Schaumberg, K.A., M.S. Antunes, T.K. Kassaw, W. Xu, C.S. Zalewski, J.I. Medford, and A. Prasad. 2015. Quantitative characterization of genetic parts and circuits for plant synthetic biology. Nat. Methods. .
9. Tan, C., P. Marguet, and L. You. 2009. Emergent bistability by a growth-modulating positive feedback circuit. Nat. Chem. Biol. 5: 842–848.