Versatile method for quantifying and analysing morphological differences in experimentally obtained images

Kristine S. Bagdassarian[1], Katherine A. Connor[1], Ian H Jermyn[2], J. Peter Etchells[1*]

[1]Department of Biosciences, Durham University, Stockton Road, Durham, DH1 3LE

[2]Department of Mathematical Sciences, Durham University, Lower Mountjoy, Stockton Road, Durham, DH1 3LE

*Corresponding author. Email: Peter.Etchells@Durham.ac.uk

Abstract:

Analysing high-resolution images to gain insight into anatomical properties is an essential tool for investigation in many scientific fields. In plant biology, studying plant phenotypes from micrographs is often used to build hypotheses on gene function. In this report, we discuss a bespoke method for inspecting the significance in the differences between the morphologies of several plant mutants at cellular level. By examining a specific example in the literature, we will detail the approach previously used to quantify the effects of two gene families on the vascular development of hypocotyls in *Arabidopsis thaliana*. The method incorporates a MATLAB algorithm and statistical tools which can be modified and enhanced to tailor to different research questions in future studies.

In many branches of science, an important step in investigating a mechanism or structure involves obtaining high-resolution images of the observed behaviours and/or morphology. In biology, the incorporation of image analysis tools has been increasing in popularity (Roeder et al., 2012; Wunderling et al., 2016). Several methods to analyse histology of roots and root architecture were introduced over the past decade (Burton et al., 2012; Chopin et al., 2015; Lartaud et al., 2015) and more recently, machine-learning based tools have begun to emerge (Hall et al., 2016; Sankar et al., 2014). Additionally, open-source platforms such as LithoGraphX, developed from MorphoGraphX (Barbier de Reuille et al., 2014) have also been established.

In the study of plant vascular tissue, one means of investigating genetic interactions includes the use of microscopy to image transverse or longitudinal cross-sections of plants and plant mutants. By studying the phenotypes generated through genome editing, it is possible to characterise the contribution of a specific gene to the plant growth and development. However, while intuitive understanding of the differences between genotypes is crucial for formulating hypotheses and prompting further tests, it is oftentimes difficult to quantify how significant certain variabilities are, especially between mutants with grossly irregular anatomy. Here, we describe a bespoke method for quantifying the vascular mutations between a range of genotypes using a MATLAB algorithm and appropriate statistical tools. This method can be applied for a range of similar research questions in various disciplines but was specifically employed in Wang et al. (2019) to quantify phenotypic variation in

Arabidopsis stems and hypocotyls. The discussion below may be viewed as an addendum to the Wang et al. (2019) publication, and as an illustrative example to provide clarity of the described approach.

In Wang et al. (2019), we aimed to address the question of how two gene families, the *PHLOEM INTERCALATED WITH XYLEM (PXY)* family of genes (*PX*f) (Fisher and Turner, 2007) and the *ERECTA (ER)* family of genes (*ER*f) (Shpak et al., 2004) function in concert to coordinate cell division and organisation. Here, we will detail the procedure we employed to quantify the morphological discrepancy we observed across different *px*f and *er*f genotypes.

The desired Arabidopsis mutant lines, including those with sextuple mutations across the gene families, were generated through crossing of previously described lines. A total of six mutant lines were analysed. The mutant lines were grown using standard protocols, and tissue was prepared for light microscopy imaging following fixation with FAA, embedding in JB4 resin, and sectioning (Wang et al., 2019).

To study the morphological differences, six images obtained through brightfield microscopy from each genotype were selected. The focus of our study were four of the main vascular cell types: xylem vessels, xylem fibers, phloem cells and parenchyma. From each image, a minimum of 10 cell representatives from each cell type were selected from a wedge of pre-defined size (60°). In order to account for the naturally occurring size variation from centre to border, all the cells along the length of the radius were included. The four cell types were assigned a unique colour and using the software GIMP, the cell interiors were manually coloured as appropriate (see: Supplemental Data S1).

Next, a MATLAB code was generated to study the properties of the individual cell types across the genotypes. The MATLAB Image Processing Toolbox was used, with the overall logic of the algorithm described below.

The manually manipulated images were separated into folders according to genotype. The programme then looped over the images in that folder. For each image in the folder, four entirely black images of the same size were generated. The original images were scanned, and the pre-defined cell colours were recognised. The entirely black images were then manipulated as follows: whenever a pixel of a 'known' colour was identified in a position (i,j), the pixel in position (i,j) in the new black image was coloured white. This was performed for each individual colour, isolating the different cell types for analysis, with one original cross-section image yielding four (new) binary images corresponding to the four cell types of interest (Fig. 1; Fig. 2).

The new images were read as binary images in the programme. The previously selected cells were now represented as white objects on a black background and their properties (area, perimeter, major and minor axis) could be quantified as follows. Each white object represents a connected component of pixels. Specifically, the binary images are scanned by the programme as a matrix of pixels, where each region of adjacent pixels with the same value (here, the colour white) is assigned a number identity by the programme. The various properties of that region can then be measured in pixels (Fig. 1b).

The correspondence between pixel size and actual size can be determined using an image of known size and calculating the micrometre per pixel ratio. Using this method, the data was transformed from pixels to microns/microns squared to calculate the correct perimeter and area of the sampled cells. The ellipticity of the objects (arguably, their level of deformity) were measured as the ratio of major to minor axis, which is a dimensionless parameter not requiring conversion.

In Wang et al. (2019), the converted measurements were then saved into an excel spreadsheet where each row corresponded to measurements taken from a different plant. This was done in order to implement a nested ANOVA analysis using R (packages multicomp, ggplot2, graphics).

The data from MATLAB was rearranged in an R-readable format of one column of measurements matched to a column of corresponding plant ID's and a column of corresponding plant genotypes. A nested ANOVA analysis and a post-hoc Tukey HSD test were performed to identify which genotypes were pairwise significantly different.

The method described above was employed due to its relative ease of design compared to other tools and ability to be readily adjusted and tailored to a specific research question. For instance, the method can be used to investigate the influence of different factors on organisms through changes in cell deformation and cell size using images of longitudinal and transverse sections. Other questions that can be addressed include extracting measurements of bubbles or polyhype structures in physics, engineering and chemistry as well as topological research (especially regarding objects of similar colour scheme) and applying statistical analysis.

The first step of the method represents its largest source of limitations. While in certain instances, where the objects of interest can be defined by a particular colour or colour range this step can be automated, in cases where the objects are not easily distinguished, the samples must be chosen manually which can be rather time-consuming. Notably, using digital tablets which are supplied with a pen, such as Wacom Drawing Tablets, could speed up the process and improve accuracy. In future studies, one might seek to refine the method for more speedy and automated results. The code used in Wang et al. (2019) can be obtained from Github (https://bit.ly/2Kht0BI). A standard operating procedure is available as Supplemental Data S1.

References:

Al-Tamimi, N., Brien, C., Oakey, H., Berger, B., Saade, S., Ho, Y.S., Schmöckel, S.M., Tester, M. and Negrao, S., 2016. Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. Nature Communications, 7, p.13342.

Burton, A.L., Williams, M., Lynch, J.P. and Brown, K.M., 2012. RootScan: software for high-throughput analysis of root anatomical traits. *Plant and Soil*, 357(1-2), pp.189-203.

Chopin, J., Laga, H., Huang, C.Y., Heuer, S. and Miklavcic, S.J., 2015. Rootanalyzer: a cross-section image analysis tool for automated characterization of root cells and tissues. PloS one, 10(9), p.e0137655.

Fisher, K. and Turner, S., 2007. PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. Current Biology, 17(12), pp.1061-1066.

Gehan, M.A., Fahlgren, N., Abbasi, A., Berry, J.C., Callen, S.T., Chavez, L., Doust, A.N., Feldman, M.J., Gilbert, K.B., Hodge, J.G. and Hoyer, J.S., 2017. PlantCV v2: Image analysis software for high-throughput plant phenotyping. PeerJ, 5, p.e4088.

Hall, H.C., Fakhrzadeh, A., Luengo Hendriks, C.L. and Fischer, U., 2016. Precision automation of cell type classification and sub-cellular fluorescence quantification from laser scanning confocal images. Frontiers in plant science, 7, p.119.

Lartaud, M., Perin, C., Courtois, B., Thomas, E., Henry, S., Bettembourg, M., Divol, F., Lanau, N., Artus, F., Bureau, C. and Verdeil, J.L., 2015. PHIV-RootCell: a supervised image analysis tool for rice root anatomical parameter quantification. Frontiers in plant science, 5, p.790.

Roeder, A.H., Cunha, A., Burl, M.C. and Meyerowitz, E.M., 2012. A computational image analysis glossary for biologists. *Development*, *139*(17), pp.3071-3080.

Shpak, E.D., Berthiaume, C.T., Hill, E.J. and Torii, K.U., 2004. Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development, 131(7), pp.1491-1501.

Tovar, J.C., Hoyer, J.S., Lin, A., Tielking, A., Callen, S.T., Elizabeth Castillo, S., Miller, M., Tessman, M., Fahlgren, N., Carrington, J.C. and Nusinow, D.A., 2018. Raspberry Pi–powered imaging for plant phenotyping. Applications in plant sciences, 6(3), p.e1031.

Wang, N., Bagdassarian, K.S., Doherty, R.E., Kroon, J.T., Connor, K.A., Wang, X.Y., Wang, W., Jermyn, I.H., Turner, S.R. and Etchells, J.P., 2019. Organ-specific genetic interactions between paralogues of the PXY and ER receptor kinases enforce radial patterning in Arabidopsis vascular tissue. *Development*, *146*(10), p.dev177105.

Wunderling, A., Ben Targem, M., Barbier de Reuille, P. and Ragni, L., 2016. Novel tools for quantifying secondary growth. *Journal of experimental botany*, 68(1), pp.89-95.

Githubcom. 2019. GitHub. [Online]. [5 August 2019]. Available from: https://github.com/KristineBagdassarian/PXY-ER_enforce_radial_growth

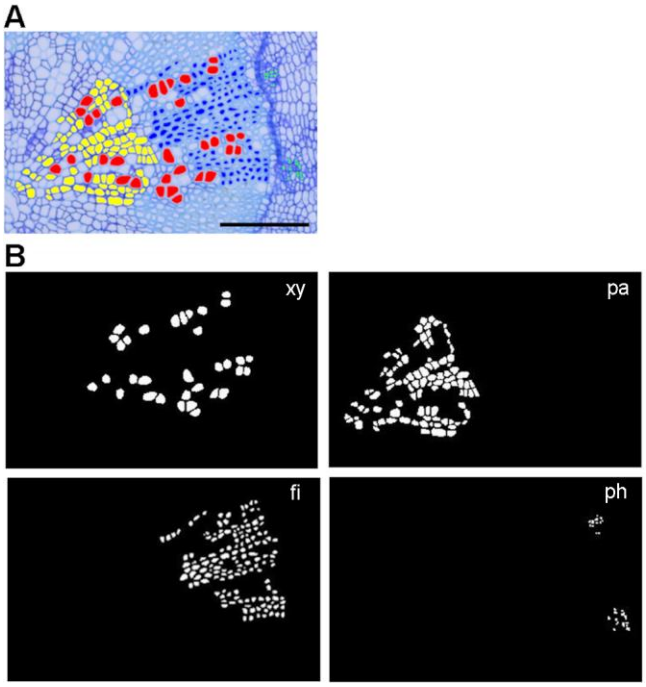

Figure 1. Example of transforming microscopic image (A) into binary images (B) with each image corresponding to four cell types: xylem vessels (xy), xylem parenchyma (pa), fibres (fi) and phloem cells (ph). Scale is 100µM in the first image (A). Each white object on the binary images (B) can be investigated as a connected component of white pixels and its properties such as area, perimeter and axis, measured.

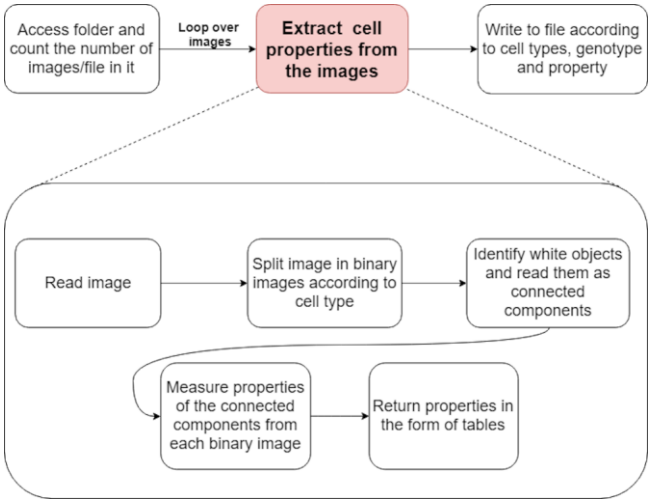

Figure 2. Block diagram, illustrating the main parts of the algorithm. The first row of blocks show the key stages of obtaining the data, with the steps comprising the extraction phase in the bubble below.