

A rapid method to quantify vein density in C₄ plants using starch staining

Conor J. C. Simpson¹, Pallavi Singh¹, Deedi E.O. Sogbohossou², M. Eric Schranz², Julian M. Hibberd^{1†}

¹University of Cambridge, Department of Plant Sciences, Downing Street, Cambridge, CB2 3EA, United Kingdom

²Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

[†]Correspondence: jmh65@cam.ac.uk

Keywords: C₄ photosynthesis, bundle sheath, starch, vein density, high-throughput phenotyping.

One sentence summary: Preferential accumulation of starch in bundle sheath cells of C₄ plants allows high-throughput phenotyping of vein density.

30 **Abstract**

31 C₄ photosynthesis has evolved multiple times in the angiosperms and typically involves
 32 alterations to the biochemistry, cell biology and development of leaves. One common
 33 modification found in C₄ plants compared with the ancestral C₃ state is an increase in vein
 34 density such that the leaf contains a larger proportion of bundle sheath cells. Recent findings
 35 indicate that there may be significant intra-specific variation in traits such as vein density in C₄
 36 plants but to use such natural variation for trait-mapping, rapid phenotyping would be
 37 required. Here we report a high-throughput method to quantify vein density that leverages the
 38 bundle sheath specific accumulation of starch found in C₄ species. Starch staining allowed
 39 high-contrast images to be acquired that permitted image analysis using a MATLAB-based
 40 program. The method works for the dicotyledon *Gynandropsis gynandra* where significant
 41 variation in vein density was detected between natural accessions, and the monocotyledon
 42 *Zea mays* where no variation was apparent in the genotypically diverse lines assessed. We
 43 anticipate this approach will be useful to map genes controlling vein density in C₄ species
 44 demonstrating natural variation for this trait.

Introduction

Photosynthesis is the basis of life on Earth and central to this process is the enzyme Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO) that operates in the Calvin-Benson-Bassham cycle to fix CO₂. However, oxygenase activity RuBisCO results in the energy-intensive photorespiratory pathway (Driever and Kromdijk, 2013) and in some environments costs of photorespiration are thought to have driven the evolution of carbon concentrating mechanisms such as C₄ photosynthesis and Crassulacean Acid Metabolism (CAM). Thus, C₃ photosynthesis is considered the ancestral state and despite its complexity, derived states involving carbon concentrating mechanisms are thought to have evolved repeatedly (Edwards, 2019). In the case of C₄ photosynthesis, current estimates are that it has evolved independently in more than sixty-six lineages (Sage *et al.*, 2011). The vast majority of C₄ plants evolved specialized anatomy to compartmentalise photosynthesis between two cell types. Typically, this involves the mesophyll being the site of initial HCO₃⁻ assimilation by Phosphoenolpyruvate Carboxylase (PEPC) and the bundle sheath being repurposed for the Calvin-Benson-Bassham cycle. Compared with the C₃ state, the partitioning of photosynthesis between these two cell types is associated with changes to transcriptional, post-transcriptional and also post-translational regulatory mechanisms (Hibberd and Covshoff, 2010; Reeves *et al.*, 2017). For some genes encoding components of the C₄ cycle a detailed understanding has emerged such that changes in *cis* (Nomura *et al.*, 2000, 2005a,b; Gowik *et al.*, 2004, 2017; Williams *et al.*, 2016) or *trans* (Brown *et al.*, 2011; Reyna-Llorens *et al.*, 2018) are considered the driving forces behind alterations to gene expression. In contrast, we have a more limited understanding of the genes underpinning the evolution of Kranz anatomy and how their expression has changed during the C₃ to C₄ transition (Sedelnikova *et al.*, 2018).

Reports of significant natural variation in traits including components of Kranz anatomy (Lundgren *et al.*, 2016; Yabiku and Ueno, 2017; Kolbe and Cousins, 2018; Reeves *et al.*, 2018) provide an opportunity to probe the basis of this trait using quantitative genetics. For example, if it were possible to screen vein density or bundle sheath size rapidly in populations made up of genetically diverse individuals such as Multi-Advanced Generation Inter-Crossing (MAGIC) populations or natural accessions designed for Genome Wide Association Study (GWAS) it should be possible to start to link genes with Kranz anatomy. This approach would be facilitated by rapid phenotyping of these traits. Significant advances have been made in phenotyping at scale for protein activity assays (Gibon *et al.*, 2004; Sulpice *et al.*, 2010) or the

use of robotics for data collection in the field (Virlet *et al.*, 2017). Notably, using a phenotyping facility Guo *et al.* (2018) rapidly analyzed fifty-one image-based traits in rice under drought stress. Correlating these traits with traditional drought resistance traits enabled novel genes linked to drought resistance to be identified via a GWAS. Such approaches may need support from non-invasive imaging techniques (Yang *et al.*, 2020) and consequently high-throughput image processing is key to alleviate the phenotyping bottleneck associated with laborious, error-prone manual image analysis (Moen *et al.*, 2019). Image analysis using Convolutional Neural Networks (CNNs) has been used to map Quantitative Trait Loci (QTL) associated with stomata size and density in maize (Xie *et al.*, 2021). Moreover, image segmentation tools have been developed to quantify vein density (Dhondt *et al.*, 2012; Parsons-Wingerter *et al.*, 2014; Bühler *et al.*, 2015), but such automated approaches have not yet replaced manual tracing (Perez-Harguindeguy *et al.*, 2016). Moreover, analysis of tissues such as cotyledons of *Arabidopsis thaliana* (Dhondt *et al.*, 2012; Parsons-Wingerter *et al.*, 2014) can require extensive manual supervision and may not be suited for tissues with higher vein density. Although, the powerful programme LeafVeinCNN (Xu *et al.*, 2021) enables high-throughput analysis of traits including vein order, areole number, vein width and vein density it currently requires manual real-time supervision of individual images and can be computationally slow due to the number of networks in use and the number of measurements being attained per image (Xu *et al.*, 2021). Further, to our knowledge, only one tool has been developed to assess vein structure in monocotyledons (Robil *et al.*, 2021). As C₄ photosynthesis is found in both monocotyledons and dicotyledons it would be useful to have a tool that permits analysis of vein density in both clades.

Here we aimed to test whether preferential accumulation of starch in the bundle sheath of C₄ plants (Lunn and Furbank, 1997; Figure 1A) could be used to generate high-contrast images so that a segmentation-based method could rapidly assess vein density in these plants. We show that such an approach can be used to measure vein density in both monocotyledons and dicotyledons. Moreover, this allowed us to develop a computational pipeline enabling high-contrast black and white images to be processed automatically using the image processing toolbox in MATLAB. Not only does the method allow rapid phenotyping but it also showed that whilst in genotypically diverse founders of a maize population (Dell'Acqua *et al.*, 2015) there was little variation in vein density, in the C₄ dicotyledonous model *Gynandropsis gynandra* there is significant natural variation in vein density.

Results and Discussion

Natural variation for vein density in *C₄ G. gynandra* and an automated image processing tool

Staining leaves of *G. gynandra* clearly indicated preferential accumulation of starch in the bundle sheath (Figure 1B). We therefore subjected an F_2 population of *G. gynandra* comprising 199 lines derived from a cross between parents differing in vein density (Reeves *et al.*, 2018) to starch staining. Initially, a random subset of seventeen plants were stained and this indicated significant natural variation in vein density amongst the F_2 plants ($p < 0.001$; Figure 1C,D&E). We reasoned that this variance in vein density between accessions provided a resource to develop and ground-truth a semi-automated procedure to quantify vein density.

Using MATLAB's Image Processing Toolbox (2019a) an automated pipeline, hereafter Starch4Kranz, was developed to reliably quantify bundle sheath density. This pipeline involved converting each image to greyscale (Figure 2A), enhancing the contrast to improve differences between bundle sheath strands and the background (Figure 2B) and then blurring such that a given pixel is converted to the mean pixel intensity in a sub-matrix of $2w+1$ where w is 20. Thus, for each pixel its value is converted to the average of the neighboring sub-matrix of 20×20 pixels. This smoothing resulted in pixels in each bundle sheath strand having similar values, and the same was true for background pixels (Figure 2C). Images are then binarised (Figure 2D) such that stained regions appear white (with a pixel value of 1) and the background is black (with a pixel value of 0). Skeletonisation converts all pixels with a value of 1 to curved lines a single pixel in width (Figure 2E). Lastly, these skeletonised images are trimmed such that superfluous pixels are removed (Figure 2F, Figure S1). The number of pixels trimmed is defined by the user as the `trim_factor`, but if the program detects a significantly larger number of superfluous pixels compared with expected vein density then the `trim_factor` can be calculated automatically based on the mode length of connected superfluous pixels. At this stage bundle sheath strands are represented by a line that is one pixel thick and so the total number of pixels provides an estimate of vein length. Vein density can then be determined from the total number of pixels and the image size.

Starch4Kranz processes all images in a given directory and saves results into a table containing image name and vein density. After each image is processed a window appears containing the vein trace superimposed on the original image in greyscale (Figure 2G). Compared with approximately eleven hours required to manually trace around 102 images

(Figure 1E) implementing the Starch4Kranz pipeline for the same samples took fifteen minutes.

Automatic quantification of high contrast images for vein density

The Starch4Kranz pipeline was next implemented on data from the seventeen F_2 plants for which vein density had been determined manually. From 102 images, the same pattern of variation in vein density was observed (Figure 3A) and estimates from manual and automated pipelines were highly correlated ($r = 0.92$; Figure 3B). When Starch4Kranz was applied to 1101 images, 98.5% passed the manual check. Of the seventeen images that did not, manual inspection indicated they were either poorly stained or images were of poor quality. When twenty-seven images were subjected to the LeafVeinCNN pipeline (Xu *et al.*, 2021) three and a half minutes were required to run the CNNs but repeated manual input was required to obtain optimal thresholding. Moreover, compared with LeafVeinCNN ($r = 0.74$; Figure S2B) the outputs from Starch4Kranz pipeline correlated more strongly with manual tracings ($r = 0.95$; Figure S2A). LeafVeinCNN performed less well when vein densities were low (Figure S2C), which may be because it was trained on data derived from leaves imaged at lower magnification.

We found that subdividing an image into multiple sections, and capturing slightly out-of-focus images could enhance contrast between veins and background. This was of particular use for samples with high vein densities, or when staining was less good. As we found that some images did not allow accurate vein density to be determined, a manual check is incorporated into the pipeline (Figure 3B). Compared with other methods that require continual manual checking of images during processing (Dhondt *et al.*, 2012; Parsons-Wingerter *et al.*, 2014; Bühler *et al.*, 2015; Xu *et al.*, 2021) Starch4Kranz saves time because it runs in the background across multiple images, and a manual check does not take place until the programme has finished. Machine learning techniques for image analysis are a useful tool for quantitative studies in plant sciences and have been employed to identify QTL associated with stomatal density in maize (Xie *et al.*, 2021) and estimate heritability of vein density in sunflower (Earley *et al.*, 2022). However, such approaches can require extensive computational understanding as well as power to build neural networks, and depending on the complexity of the phenotype may require manual annotation of thousands of images for network training (Moen *et al.*, 2019; van Dijk *et al.*, 2021). Thus, LeafVeinCNN is a powerful tool suitable for quantifying multiple traits associated with vein architecture, but Starch4Kranz

offers a rapid and robust tool allowing vein density to be determined in C_4 species. It also confirmed significant natural variation amongst F_2 lines of *G. gynandra* (Figure 3C).

Natural variation in vein density was not detectable in founder lines of a maize MAGIC population

Aside from Grasvick (Robil *et al.*, 2021) we noted a lack of automated methods for the assessment of vein density in monocotyledons. Because the preferential accumulation of starch in bundle sheath cells had first been reported in monocotyledons (Lunn and Furbank, 1997) we next assessed whether the Starch4Kranz algorithm could be used in maize. For this, an additional step was added to the pipeline to remove commissural veins (Figure S3), which are not considered important components of Kranz anatomy (Langdale, 2011). We also noted that they stained for starch inconsistently likely because they are considered a simplified vascular tissue not encircled by bundle sheath cells (Sakaguchi and Fukuda, 2008; Sakaguchi *et al.*, 2010). If commissural veins are of interest, and sufficient contrast obtained, then “auto” could be selected as an input variable instead of “monocot” (see Methods) and commissural veins will not be removed. Commissural veins were defined as the shortest branch stemming from a pixel with three branch points. Pattern recognition allowed these to be identified and removed (Figure S3). To ground-truth this approach, manual tracings that did not include these veins were undertaken. Images from starch staining of nine founder lines of a MAGIC population from maize (Dell’Acqua *et al.*, 2015) were subjected to Starch4Kranz. Estimates were highly correlated with manual analysis ($r = 0.91$; Figure 4A) but no statistically significant differences were observed between genotypes ($p > 0.05$; Figure 4B). Thus, Starch4Kranz can be used to quantify vein density in C_4 monocotyledons in addition to dicotyledons (Figure 4&S4). However, for the founder lines studied here, we detected no significant differences in vein density (Figure 4B).

This finding emphasises the need to generate specific mapping populations designed for the trait of interest. *G. gynandra* F_2 lines produced from a cross between two lines phenotypically distinct for vein density displayed abundant natural variation in this trait (Figures 3C). Although variation in vein density has been reported in maize (Kolbe and Cousins, 2018), equivalent variation was not present in the MAGIC maize founders (Figure 4B) indicating that the MAGIC maize population (Dell’Acqua *et al.*, 2015) is unlikely to be suitable for studying QTL underlying vein density. That said, the ease and speed with which Starch4Kranz permits quantification of vein density means that screening other C_4 mapping

209 populations such NAM populations of maize (Yu *et al.*, 2008; McMullen *et al.*, 2009) or
210 MAGIC populations of sorghum (Ongom and Ejeta, 2017) could be useful.

211 In summary, we report a simple protocol allowing vein density to be determined from starch
212 staining of vascular tissue in C₄ species. Automatation of this pipeline allowed rapid and
213 accurate quantification of vein density. As increased vein density is considered an important
214 trait allowing C₄ photosynthesis (Sage, 2001), where intra-specific natural variation for this
215 trait exists our approach could provide insight into mechanisms evolution has co-opted to
216 generate the complex C₄ phenotype.

217 **Methods**

218 **Plant Growth and leaf sampling**

219 *Gynandropsis gynandra* was grown under irrigated glasshouse conditions from April to July
 220 2019. Temperatures were maintained at 24°C in the day and 20°C during the night, with
 221 artificial lighting maintaining a minimum light intensity of 300 $\mu\text{mol m}^{-2}\text{s}^{-1}$ and a photoperiod
 222 of 16 hours light and 8 hours dark. Relative humidity was 60%. Maize was grown under field
 223 conditions from April to August 2022 at NIAB, Cambridge UK. Plants were treated with pre-
 224 emptive herbicide (Stomp® Aqua) one day after sowing, fertilizer (ammonium nitrate) five
 225 days after sowing, molluscicide (SluXX) one month after sowing, and herbicide (Leystar®) two
 226 months after sowing. During the growing season, average temperature was 19.8°C during the
 227 day and 15.2°C during the night. Maximum and minimum daytime temperature was 39.9°C
 228 and 8.9°C and that during the night was 34.2°C and 2.9°C. Total precipitation was 89mm and
 229 average relative humidity was 65%.

230 *G. gynandra* leaves were harvested over a three-day period from plants four weeks after
 231 germination. The youngest fully expanded leaf was selected. Tissue was harvested from the
 232 central leaflet of each leaf and immediately submerged in a plastic cuvette containing 3:1
 233 100% (v/v) ethanol:acetic acid solution for 4 hours before treatment with 70% (v/v) ethanol
 234 overnight at 37°C, being refreshed one hour into treatment. Samples were then cleared using
 235 5% (w/v) NaOH for 3 hours at 37°C before being washed with, and replaced in, 70% (v/v)
 236 ethanol until preparation for imaging. Maize leaves were harvested over a two-week period as
 237 part of a larger field trial. All disks were taken from the most recently fully expanded leaf.
 238 Leaves were harvested in the field and kept in water until leaf disks were taken. One disk was
 239 taken per leaf and immediately submerged in 3:1 100% (v/v) ethanol:acetic acid fixative
 240 solution for 4 hours before being transferred to 70% (v/v) ethanol solution and left overnight at
 241 37°C. Samples were then cleared using 5% (w/v) NaOH for 6 hours at 37°C before being
 242 washed with, and replaced in 70% (v/v) ethanol until preparation for imaging.

243 Prior to imaging, leaf disks were placed in Lugol's solution. Specifically, using forceps each
 244 disk was transferred to Lugol's solution and then removed immediately after submersion and
 245 placed into 70% (v/v) ethanol to minimize stain accumulation. If the stain did not accumulate
 246 rapidly the sample was re-placed in Lugol's solution until it stain had been taken up. Excess
 247 Lugol's solution was washed off in 70% (v/v) ethanol. Each leaf disk was then mounted on a
 248 slide using water and imaged at a magnification of 100x on an Olympus BX41 light

microscope with a mounted Micropublisher 3.3 RTV camera (Q Imaging). Images were captured with Q-Capture Pro 7 software.

Manual and automated analysis of vein density and statistical analysis

Veins were measured directly in ImageJ (Rasband, 2014) using the paint brush tool to quantify total length of veins in μm . Lengths were converted to density ($\mu m mm^{-2}$) by dividing by the image's area in mm^2 . Table 1 summarises the important MATLAB functions and their function. In addition to this, pattern recognition was used to identify commissural veins in maize (Figure S3). From a subset of 102 images, twenty-seven were also analysed using LeafVeinCNN (Xu *et al.*, 2021). These twenty-seven images comprised three images per plant from nine plants. LeafVeinCNN was run using default parameters consisting of 3 ensemble CNNs. Appropriate threshold values were determined for each image to attain optimal vein traces through segmentation. To achieve results comparable with Starch4Kranz the output value "VTotL" meaning total vein length in mm , was converted to $\mu m mm^{-2}$.

All statistical analysis were carried out in R Studio (V: 4.1.3). Figures were generated using the ggpubr (Kassambara, 2020) package. To compare manual and automated quantification methods in *G. gynandra* a subset of seventeen plants from a total of 199 were selected randomly. Six images were assessed per plant such that the dataset was made up of 102 images. To attain a vein density measurement for each of the seventeen lines, the mean of these six images was determined. We noted that using three images yielded no significant differences in plant vein densities when comparing manual (t-test, $p = 0.90$) and automated (t-test, $p = 0.85$) methods. Therefore for maize, six replicates were used for A632, F7, and W153R; five for B73, CML91, H99, HP301, and Mo17; and three for B96. The total dataset of 138 images was assessed both manually and automatically. For both treatments all lines were normally distributed (Shapiro-Wilk's test, $p > 0.05$) and Bartlett's test revealed homogeneity of variance ($p > 0.05$) meaning parametric analysis was carried by Analysis Of Variance (ANOVA). Correlation analysis was carried out using the Pearson correlation coefficient. In total a *G. gynandra* population of 199 plants, each with five or six image replicates meaning a total of 1102 images were assessed for variation.

Raw script availability and code implementation

The Starch4Kranz pipeline was designed to be simple to use and can run in the background while other tasks take place. While it requires a manual check, this can be done rapidly (~5-10 seconds per image) as the user checks each image within a whole directory once the script has stopped running. Only three scripts are needed to run the full pipeline, Starch4Kranz.m contains all executable code used in the production of our pipeline, the function, blur.m contains the blurring function, and running_Starch4Kranz.m contains a short script that runs Starch4Kranz.m while saving results to a table easily transferable to software for further data analysis. Script is available at <https://github.com/plycs5/Starch4Kranz>. To run the script the user must have the ImageProcessingToolbox activated and supply four inputs, filename, trim_factor (see Results), pixel_length_um (the image scale), y_pixels and x_pixels (image dimensions). A fifth input named split_mode must either be “auto”, “whole”, “split”, or “monocot”.

ACKNOWLEDGMENTS

We thank Angie Burnett, John Ferguson, and Johannes Kromdijk for provision of maize leaf material. The work was supported by a BBSRC DTP studentship to CJCS, European Research Council Grant 694733 Revolution to JMH, and Netherlands Organization for Scientific Research Grant W.08.270.350 to MES. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

Figure Legends

Figure 1: Starch staining in the bundle sheath of *C₄ Gynandropsis gynandra* demonstrates natural variation in vein density across a diversity panel. (A) Schematic of the *C₄* pathway. Arrows indicate carbon flux. Enzymes involved in carboxylation and decarboxylation shown in blue. Intracellular compartments represented schematically with mitochondria in pink and chloroplasts dark green. **(B, C, D)** Representative images of *Gynandropsis gynandra* leaves stained for starch at low magnification **(B)**, higher magnification with low vein density **(C)** and higher magnification with high vein density **(D)**. **(E)** Natural variation in vein density across seventeen lines. Abbreviations as follows: PEP = Phosphoenolpyruvate; OAA = Oxaloacetate; Asp = Aspartate; Mal = Malate; Pyr = Pyruvate; Ala = Alanine; 3-PGA = 3-Phosphoglycerate; G3P = Glyceraldehyde 3-phosphate; PEPC = Phosphoenolpyruvate carboxylase; NAD-ME = NAD-Malic Enzyme; RuBisCO = Ribulose Biphosphate Carboxylase/Oxygenase; CBB = Calvin Benson Bassham cycle. Scale bars represent 200µm. Six images per plant were assessed and p-value calculated by ANOVA.

Figure 2: Summary of critical automation steps. (A) A greyscale image is used as input and subjected to contrast enhancement **(B)**, followed by blurring **(C)**, binarization **(D)**, skeletonization **(E)**, and trimming **(F)**. An output is then produced for the user to check **(G)**. Scale bars represent 200µm. Skeletons have been thickened to ensure visibility.

Figure 3: Ground truthing of automated analysis. (A) Seventeen randomly selected lines from the 199 line population were assessed both automatically (red) and manually (blue) for vein density. **(B)** Correlation analysis between automatic and manual quantification of vein density. VD = Vein Density. Line of best fit shown with grey shading representing 95% confidence intervals. r = Pearson's correlation coefficient. **(C)** Distribution of vein density across 199 plants determined using the automated pipeline Starch4Kranz. **(D)** Example of a manually traced image. **(E)** Example of the automated output. Scale bars represent 200µm. Skeletons have been thickened to ensure visibility.

Figure 4: Starch staining of vein density in founder lines of a maize MAGIC population. (A) Vein density in maize determined automatically and manually after starch staining. **(B)** No significant differences between founder lines of maize were detected from either manual (red) or automatic (blue) methods. r = Pearson's correlation coefficient; VD = Vein Density; p-

values determined by ANOVA and student's t-test. **(C)** Representative images of maize leaves stained for starch. Scale bars represent 200 μ m.

Figure S1: Trimming of superfluous pixels. (A) A branch end is recognized as a line of pixels with a single branch point and a single terminus. The binary code represents background (0) and vein (1). Note values of one are no longer present after trimming **(B)**. So, if the length of this line is below the trim_factor threshold, i.e., it is small and therefore not considered a vein but an array of superfluous pixels, it is trimmed. Scale bars represent 200 μ m.

Figure S2: Comparison between Starch4Kranz and LeafVeinCNN. (A) Starch4Kranz correlated more strongly with manual traces compared with LeafVeinCNN **(B)**. Lower vein densities were detected less well by LeafVeinCNN **(B)** compared with Starch4Kranz **(D)**. r = Pearson's correlation coefficient; LVCNN – LeafVeinCNN. Scale bars represent 200 μ m. Skeletons have been thickened to ensure visibility.

Figure S3: Removal of commissural veins via pattern recognition. (A) A commissural vein is recognized as being a pixel with three branch points. The binary code in the circle represents background (0) and vein (1). Note values of one are no longer present once commissural vein are removed **(B)**. Scale bars represent 200 μ m.

Figure S4: Automated tracings for each accession of maize. Representative outputs from Starch4Kranz on each founder from the MAGIC maize population. Scale bars represent 200 μ m. Skeletons have been thickened to ensure visibility.

Table 1. Main MATLAB base and Image Processing Toolbox functions used in Starch4Kranz.

Function	Use
rgb2gray	Converts Red Blue Green (RGB) image to greyscale
adapthisteq	Improves contrast of image using contrast-limited adaptive histogram equalization (Zuiderveld, 1994)
blur	Blurs the image by converting each pixel to the mean of pixel value in a sub-matrix of 2w+1 (https://uk.mathworks.com/matlabcentral/answers/472573-write-a-function-called-blur-that-blurs-the-input-image?s_tid=mwa_osa_a)
imbinarize	Replaces all pixels above a threshold to a value of 1 and makes all others 0. Threshold is based on Otsu's method (Otsu, 1979)
imcomplement	Complements an image
bwskel	Reduces binary objects to a 1-pixel wide curved line while conserving image structure
bwmorph	Applies a particular morphological operation to a binary image. We used the 'branchpoints' and 'endpoints' operations to create two separate masks that showed the skeleton's branch points and end points, these could be used in conjunction with the find function to pinpoint their positions and lengths
bwdistgeodesic	Calculates the distance (in pixels) between defined points in a binary image. Used here to calculate the lengths of branches, i.e., length between a branch point and an end point. This value was important for the trimming step of the pipeline.

References

- Brown NJ, Newell S. CAS, Chen JE, Perrin AJ, Kajala K, Hibberd JM.** 2011. Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. *Science* **331**, 1436–1439.
- Bühler J, Rishmawi L, Pflugfelder D, Huber G, Hanno S, Hülkamp M, Koornneef M, Schurr U, Jahnke S.** 2015. phenoVein - A tool for leaf vein segmentation and analysis. *Plant Physiology* **169**, pp.00974.2015.
- Chen BL, Yang HK, Ma YN, Liu JR, Lv FJ, Chen J, Meng YL, Wang YH, Zhou ZG.** 2017. Effect of shading on yield, fiber quality and physiological characteristics of cotton subtending leaves on different fruiting positions. *Photosynthetica* **55**, 240–250.
- Dell'Acqua M, Gatti DM, Pea G, et al.** 2015. Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in Zea mays. *Genome Biology* **16**, 1–23.
- Dhondt S, Van Haerenborgh D, Van Cauwenbergh C, Merks RMH, Philips W, Beemster GTS, Inzé D.** 2012. Quantitative analysis of venation patterns of Arabidopsis leaves by supervised image analysis. *Plant Journal* **69**, 553–563.
- van Dijk ADJ, Kootstra G, Kruijer W, de Ridder D.** 2021. Machine learning in plant science and plant breeding. *iScience* **24**, 101890.
- Ding Z, Zhang Y, Xiao Y, et al.** 2016. Transcriptome response of cassava leaves under natural shade. *Scientific Reports* **6**, 1–14.
- Driever SM, Kromdijk J.** 2013. Will C3 crops enhanced with the C4 CO₂-concentrating mechanism live up to their full potential (yield)? *Journal of Experimental Botany* **64**, 3925–3935.
- Earley AM, Temme AA, Cotter CR, Burke JM.** 2022. Genomic regions associate with major axes of variation driven by gas exchange and leaf construction traits in cultivated sunflower (*Helianthus annuus* L.). *Plant Journal* **111**, 1425–1438.
- Edwards EJ.** 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytologist* **223**, 1742–1755.
- Gibon Y, Blaesing OE, Hannemann J, Carillo P, Höhne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M.** 2004. A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: Comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell* **16**, 3304–3325.

396 **Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P.** 2004.
397 cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria*
398 *trinervia*, the promoter of the C4 phosphoenolpyruvate carboxylase gene. *The Plant Cell* **16**,
399 1077–1090.

400 **Gowik U, Schulze S, Saladié M, Rolland V, Tanz SK, Westhoff P, Ludwig M.** 2017. A
401 MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4
402 carbonic anhydrase in *Flaveria*. *Journal of Experimental Botany* **68**, 311–320.

403 **Guo Z, Yang W, Chang Y, et al.** 2018. Genome-Wide Association Studies of Image Traits
404 Reveal Genetic Architecture of Drought Resistance in Rice. *Molecular Plant* **11**, 789–805.

405 **Hibberd JM, Covshoff S.** 2010. The Regulation of Gene Expression Required for C 4
406 Photosynthesis. *Annual Review of Plant Biology* **61**, 181–207.

407 **Kassambara A.** 2020. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version
408 0.4.0.

409 **Kolbe AR, Cousins AB.** 2018. Mesophyll conductance in *Zea mays* responds transiently to
410 CO₂ availability: implications for transpiration efficiency in C4 crops. *New Phytologist* **217**,
411 1463–1474.

412 **Langdale JA.** 2011. C4 Cycles: Past, Present, and Future Research on C4 Photosynthesis.
413 *The Plant Cell* **23**, 3879–3892.

414 **Lundgren MR, Christin PA, Escobar EG, Ripley BS, Besnard G, Long CM, Hattersley**
415 **PW, Ellis RP, Leegood RC, Osborne CP.** 2016. Evolutionary implications of C3–C4
416 intermediates in the grass *Alloteropsis semialata*. *Plant Cell and Environment* **39**, 1874–1885.

417 **Lunn JE, Furbank RT.** 1997. Localisation of sucrose-phosphate synthase and starch in
418 leaves of C4 plants. **61**, 106–111.

419 **McMullen MD, Kresovich S, Villeda HS, et al.** 2009. Genetic properties of the maize nested
420 association mapping population. *Science* **325**, 737–740.

421 **Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D.** 2019. Deep learning for
422 cellular image analysis. *Nature Methods* **16**, 1233–1246.

423 **Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M,**
424 **Matsuoka M, Tajima S.** 2005a. Differential Expression Pattern of C4 Bundle Sheath
425 Expression Genes in Rice, a C3 Plant. *Plant and Cell Physiology* **46**, 754–761.

426 **Nomura M, Higuchi T, Katayama K, Taniguchi M, Miyao-Tokutomi M, Matsuoka M,**
427 **Tajima S.** 2005b. The promoter for C4-type mitochondrial aspartate aminotransferase does
428 not direct bundle sheath-specific expression in transgenic rice plants. *Plant and Cell*

429 Physiology **46**, 743–753.

430 **Nomura M, Sentoku N, Nishimura A, et al.** 2000. The evolution of C4 plants: Acquisition of
 431 cis-regulatory sequences in the promoter of C4-type pyruvate, orthophosphate dikinase gene.
 432 Plant Journal **22**, 211–221.

433 **Ongom PO, Ejeta G.** 2017. Mating Design and Genetic Structure of a Multi-Parent Advanced
 434 Generation Intercross (MAGIC) Population of Sorghum (*Sorghum bicolor* L.) Moench). G3:
 435 Genes, Genomes, Genetics **8**, 331–341.

436 **Otsu N.** 1979. A threshold selection method from gray-level histograms. IEEE Transactions
 437 on Systems, Man, and Cybernetics **9**, 62–66.

438 **Parsons-Wingerter P, Vickerman MB, Paul A-L, Ferl RJ.** 2014. Mapping by VESGEN of
 439 Leaf Venation Patterning in *Arabidopsis thaliana* with Bioinformatic Dimensions of Gene
 440 Expression . Gravitational and Space Research **2**, 68–81.

441 **Perez-Harguindeguy N, Díaz S, Garnier E, et al.** 2016. Corrigendum: New handbook for
 442 standardised measurement of plant functional traits worldwide (Australian Journal of Botany
 443 (2013) 61:3 (167-234) DOI: 10.1071/BT12225). Australian Journal of Botany **64**, 715–716.

444 **Rasband WS.** 2014. ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA.
 445 <http://imagej.nih.gov/ij/>, 1997-2014.

446 **Reeves G, Grangé-Guermente MJ, Hibberd JM.** 2017. Regulatory gateways for cell-specific
 447 gene expression in C4 leaves with Kranz anatomy. Journal of Experimental Botany **68**, 107–
 448 116.

449 **Reeves G, Singh P, Rossberg TA, Sogbohossou EOD, Schranz ME, Hibberd JM.** 2018.
 450 Natural variation within a species for traits underpinning C 4 photosynthesis. Plant Physiology
 451 **177**, 504–512.

452 **Reyna-Llorens I, Burgess SJ, Reeves G, Singh P, Stevenson SR, Williams BP, Stanley
 453 S, Hibberd JM.** 2018. Ancient duons may underpin spatial patterning of gene expression in
 454 C4 leaves. Proc Natl Acad Sci U S A **115**, 1931–1936.

455 **Robil JM, Gao K, Neighbors CM, Boeding M, Carland FM, Bunyak F, McSteen P.** 2021.
 456 Grasviq: an Image Analysis Framework for Automatically Quantifying Vein Number and
 457 Morphology in Grass Leaves. Plant Journal **107**, 629–648.

458 **Sage RF.** 2001. Environmental and evolutionary preconditions for the origin and
 459 diversification of the C4 photosynthetic syndrome. Plant Biology **3**, 202–213.

460 **Sage RF, Christin P-A, Edwards EJ.** 2011. The C4 plant lineages of planet Earth. Journal of
 461 Experimental Botany **62**, 3155–3169.

462 **Sakaguchi J, Fukuda H.** 2008. Cell differentiation in the longitudinal veins and formation of
463 commissural veins in rice (*Oryza sativa*) and maize (*Zea mays*). *Journal of Plant Research*
464 **121**, 593–602.

465 **Sakaguchi J, Itoh JI, Ito Y, Nakamura A, Fukuda H, Sawa S.** 2010. COE1, an LRR-RLK
466 responsible for commissural vein pattern formation in rice. *Plant Journal* **63**, 405–416.

467 **Sedelnikova O V., Hughes TE, Langdale JA.** 2018. Understanding the Genetic Basis of C4
468 Kranz Anatomy with a View to Engineering C3 Crops. *Annual Review of Genetics* **52**, 249–
469 270.

470 **Sulpice R, Trenkamp S, Steinfath M, et al.** 2010. Network analysis of enzyme activities and
471 metabolite levels and their relationship to biomass in a large panel of *Arabidopsis* accessions.
472 *Plant Cell* **22**, 2872–2893.

473 **Virlet N, Sabermanesh K, Sadeghi-Tehran P, Hawkesford MJ.** 2017. Field Scanalyzer: An
474 automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant*
475 *Biology* **44**, 143–153.

476 **Williams BP, Burgess SJ, Reyna-Llorens I, Knerova J, Aubry S, Stanley S, Hibberd JM.**
477 2016. An Untranslated cis-Element Regulates the Accumulation of Multiple C4 Enzymes in
478 *Gynandropsis gynandra* Mesophyll Cells. *The Plant cell* **28**, 454–65.

479 **Williams BP, Johnston IG, Covshoff S, Hibberd JM.** 2013. Phenotypic landscape inference
480 reveals multiple evolutionary paths to C4 photosynthesis. *eLife* **2**, 1–19.

481 **Xie J, Fernandes SB, Mayfield-Jones D, Erice G, Choi M, E Lipka A, Leakey ADB.** 2021.
482 Optical topometry and machine learning to rapidly phenotype stomatal patterning traits for
483 maize QTL mapping. *Plant Physiology*, 1–19.

484 **Xu H, Blonder B, Jodra M, Malhi Y, Fricker M.** 2021. Automated and accurate
485 segmentation of leaf venation networks via deep learning. *New Phytologist* **229**, 631–648.

486 **Yabiku T, Ueno O.** 2017. Variations in physiological, biochemical, and structural traits of
487 photosynthesis and resource use efficiency in maize and teosintes (NADP-ME- type C4).
488 *Plant Production Science* **20**, 448–558.

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

492 **Yu J, Holland JB, McMullen MD, Buckler ES.** 2008. Genetic design and statistical power of
493 nested association mapping in maize. *Genetics* **178**, 539–551.

494 **Zuiderveld K.** 1994. Contrast limited adaptive histogram equalization. *Graphic Gems IV*. San

495 Diego: Academic Press Professional, 474–485.

496

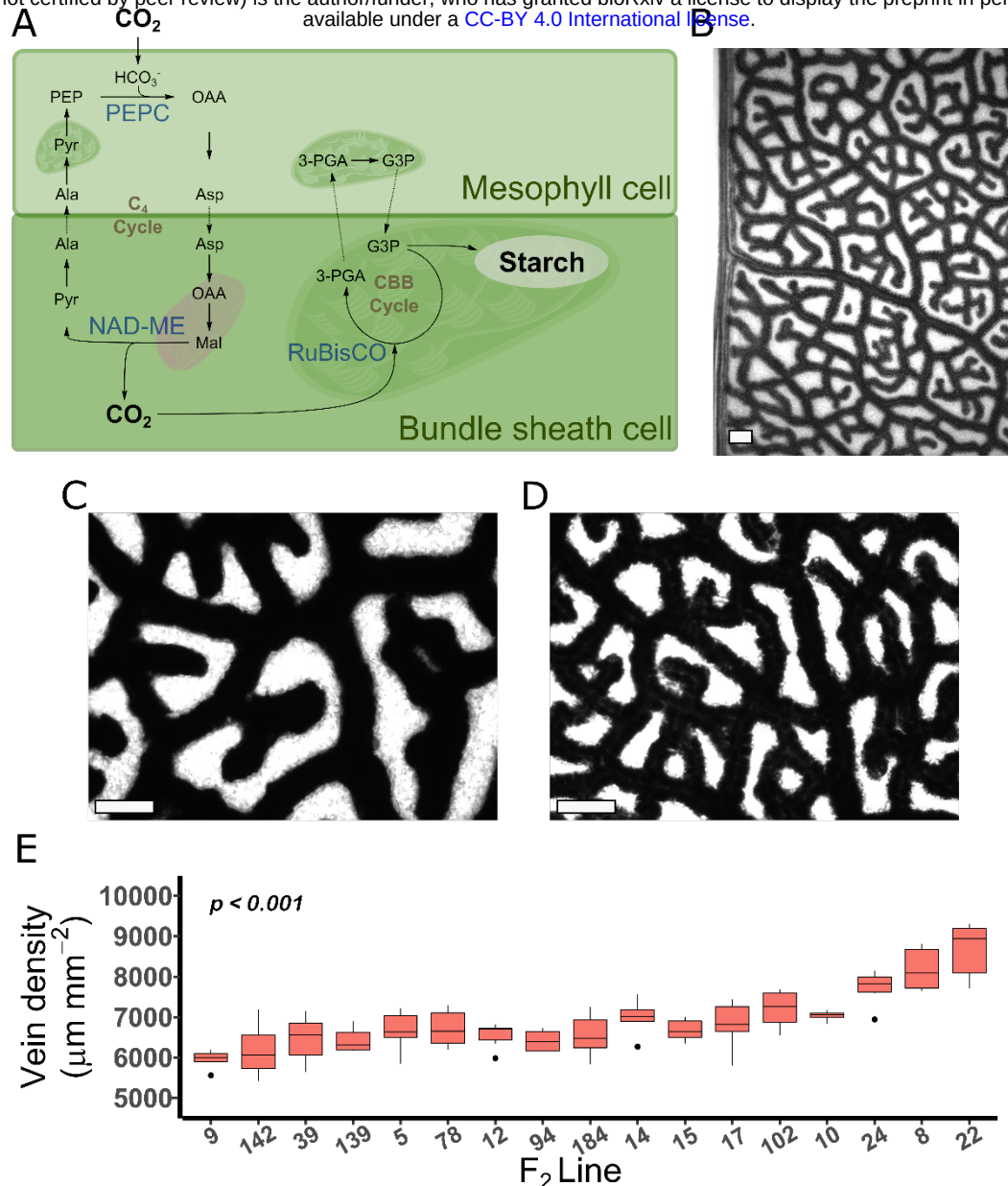


Figure 1: Starch staining in the bundle sheath of C₄ *Gynandropsis gynandra* demonstrates natural variation in vein density across a diversity panel. (A) Schematic of the C₄ pathway. Arrows indicate carbon flux. Enzymes involved in carboxylation and decarboxylation shown in blue. Intracellular compartments represented schematically with mitochondria in pink and chloroplasts dark green. (B, C, D) Representative images of *Gynandropsis gynandra* leaves stained for starch at low magnification (B), higher magnification with low vein density (C) and higher magnification with high vein density (D). (E) Natural variation in vein density across seventeen lines. Abbreviations as follows: PEP = Phosphoenolpyruvate; OAA = Oxaloacetate; Asp = Aspartate; Mal = Malate; Pyr = Pyruvate; Ala = Alanine; 3-PGA = 3-Phosphoglycerate; G3P = Glyceraldehyde 3-phosphate; PEPC = Phosphoenolpyruvate carboxylase; NAD-ME = NAD-Malic Enzyme; RuBisCO = Ribulose Bisphosphate Carboxylase/Oxygenase; CBB = Calvin Benson Bassham cycle. Scale bars represent 200μm. Six images per plant were assessed and p-value calculated by ANOVA.

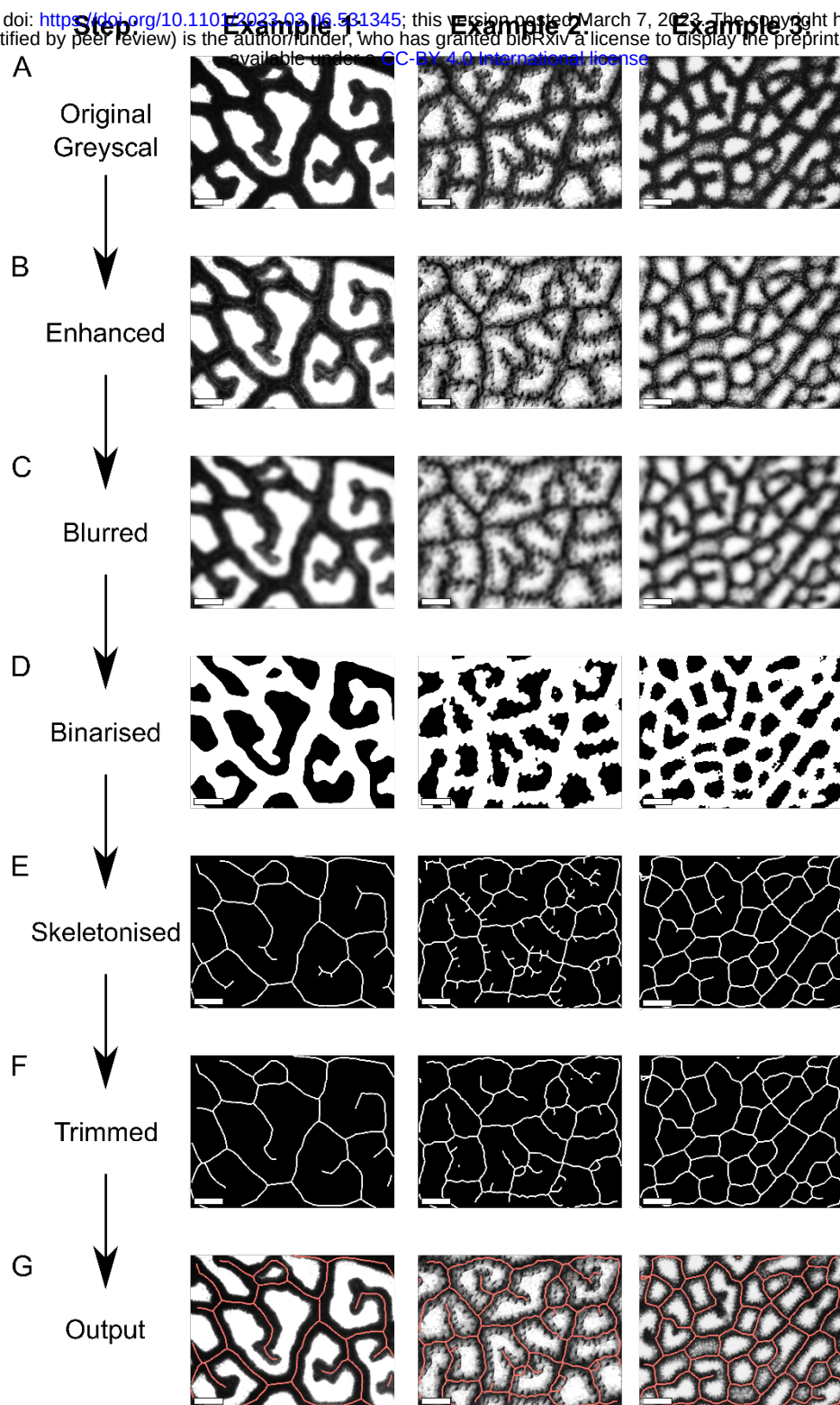


Figure 2: Summary of critical automation steps. (A) A greyscale image is used as input and subjected to contrast enhancement (B), followed by blurring (C), binarization (D), skeletonization (E), and trimming (F). An output is then produced for the user to check (G). Scale bars represent 200 μ m. Skeletons have been thickened to ensure visibility.

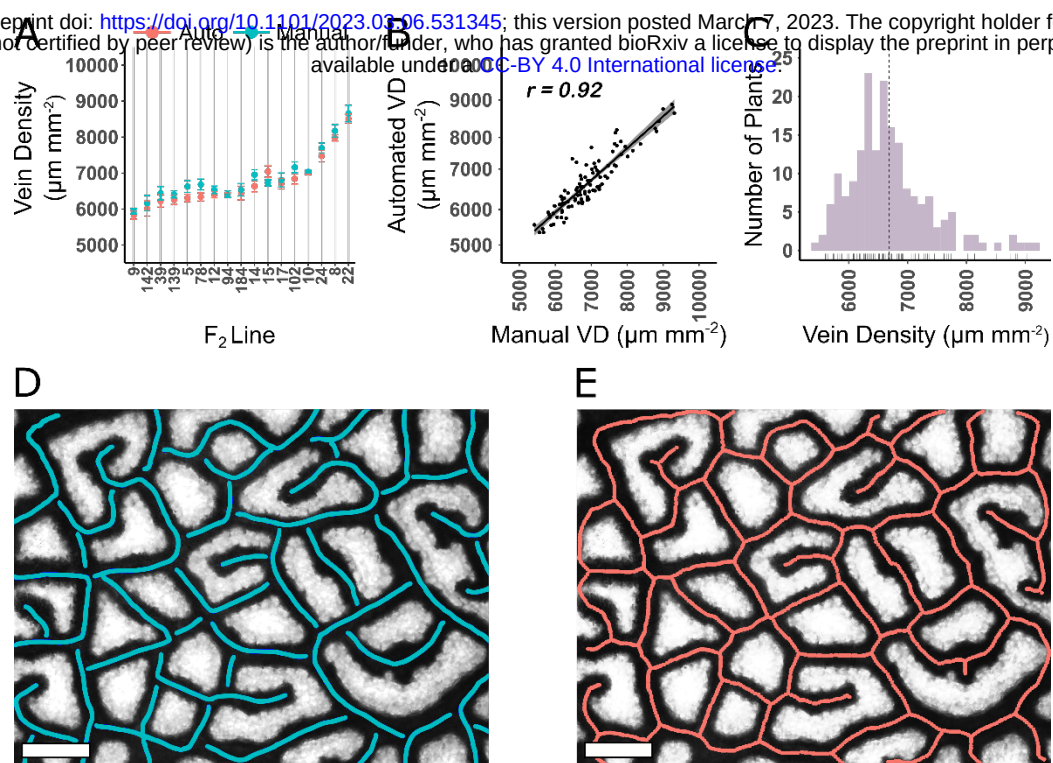


Figure 3: Ground truthing of automated analysis. (A) Seventeen randomly selected lines from the 199 line population were assessed both automatically (red) and manually (blue) for vein density. (B) Correlation analysis between automatic and manual quantification of vein density. VD = Vein Density. Line of best fit shown with grey shading representing 95% confidence intervals. r = Pearson's correlation coefficient. (C) Distribution of vein density across 199 plants determined using the automated pipeline Starch4Kranz. (D) Example of a manually traced image. (E) Example of the automated output. Scale bars represent $200\mu\text{m}$. Skeletons have been thickened to ensure visibility.

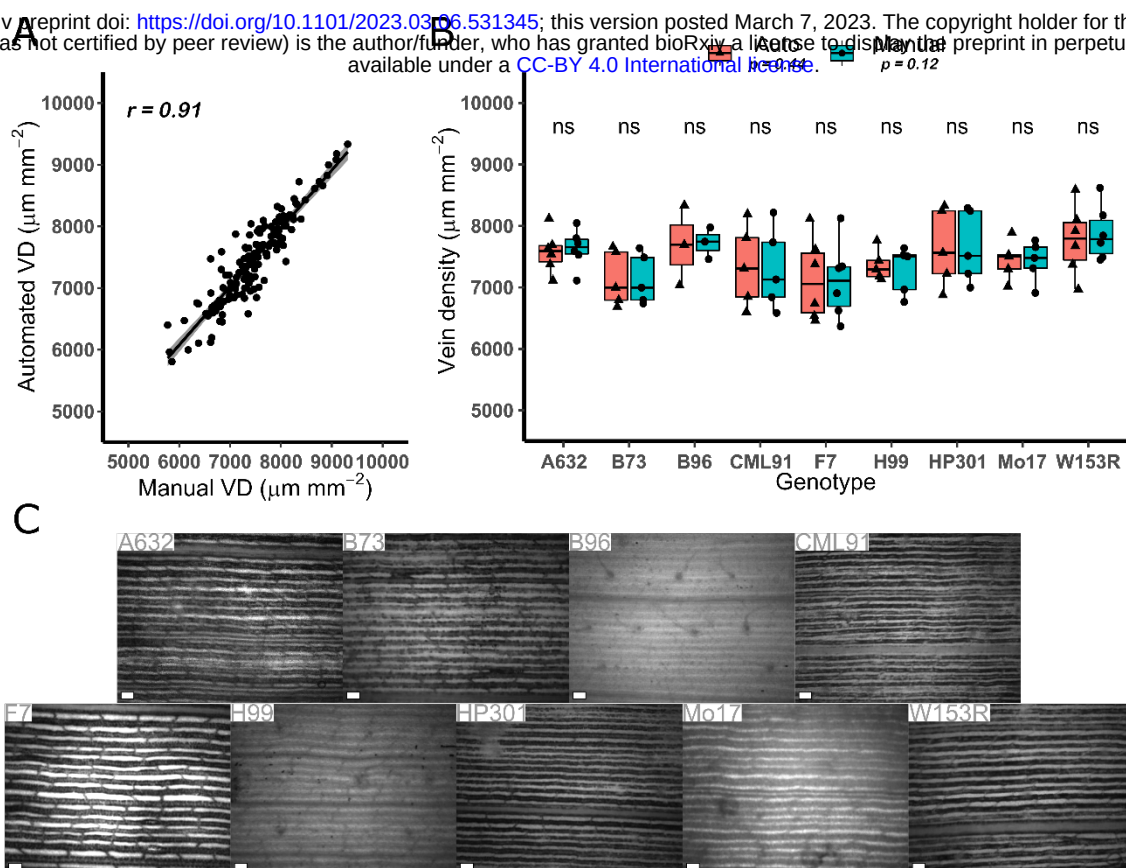


Figure 4: Starch staining of vein density in founder lines of a maize MAGIC population. (A) Vein density in maize determined automatically and manually after starch staining. **(B)** No significant differences between founder lines of maize were detected from either manual (red) or automatic (blue) methods. r = Pearson's correlation coefficient; VD = Vein Density; p -values determined by ANOVA and student's t -test. **(C)** Representative images of maize leaves stained for starch. Scale bars represent $200\mu\text{m}$.