




# Deep learning-based high-throughput phenotyping accelerates gene discovery for stomatal traits

Wei Zhang <sup>1,†</sup>, Bernarda Calla <sup>2,†</sup> and Dhineshkumar Thiruppathi <sup>3,†</sup>

<sup>1</sup> Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

<sup>2</sup> Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>3</sup> Donald Danforth Plant Science Center, Saint Louis, MO 63132, USA

<sup>†</sup>All authors contributed equally.

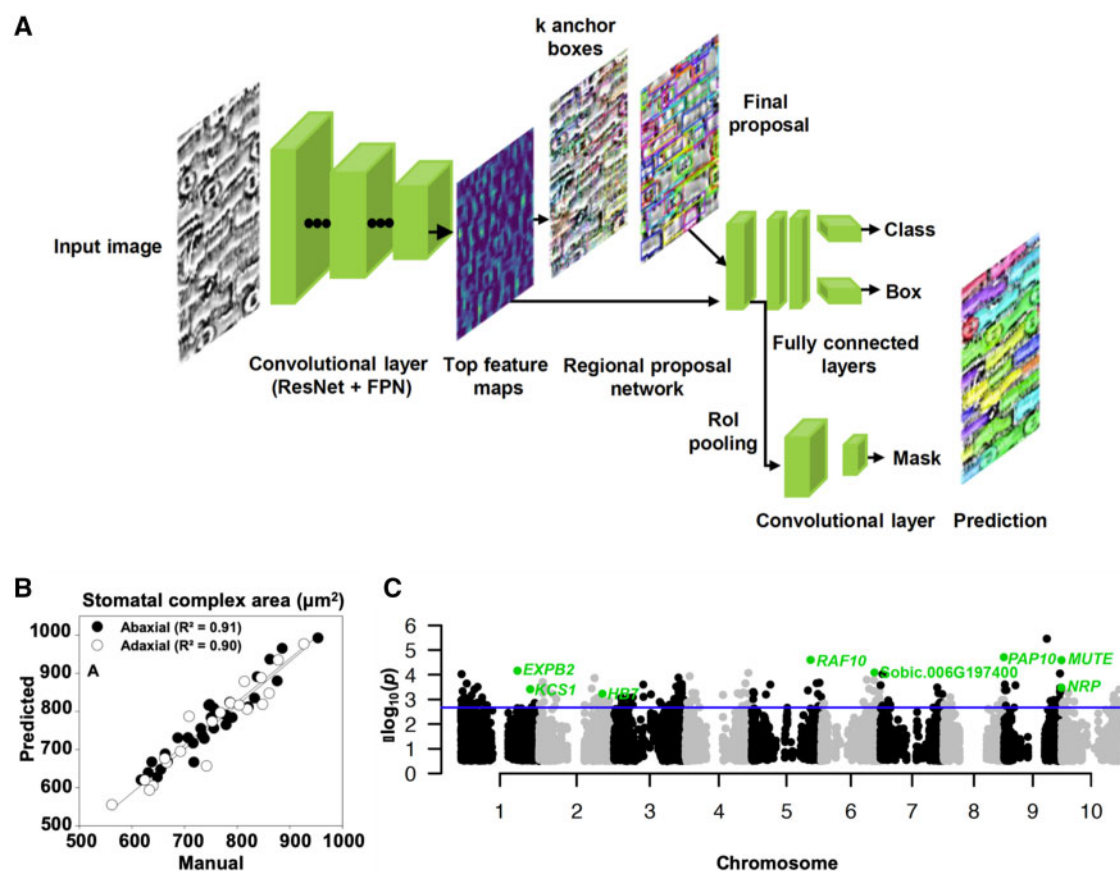
Future food security in the face of climate change requires rapid, efficient, and flexible plant genetic improvement. For this, an integrated understanding of developmental and physiological mechanisms from DNA sequences (genotypes) to terminal traits (phenotypes) under different environmental conditions is indispensable. Stomatal traits influencing photosynthesis, gas exchange, and water use are crucial targets for plant improvement programs involving major crop  $C_4$  plants sorghum (*Sorghum bicolor*) and maize (*Zea mays*; Leakey et al., 2019). However, the ability to select optimal plant genotypes is still challenged by the pace at which the acquisition and processing of stomatal phenotypic data can be accomplished. Artificial intelligence (AI) is revolutionizing the way in which problems are approached and solved across a wide range of disciplines. Machine learning, an AI subfield, is increasingly being used in agriculture to classify plants, identify pests, predict weather conditions, and track yield, among several other applications (van Dijk et al., 2021).

In this issue of *Plant Physiology*, Bheemanahalli et al. (2021), Ferguson et al. (2021), and Xie et al. (2021) introduce the use of AI-enabled high-throughput stomatal phenotyping platforms in combination with screening methods to identify specific genes and variations controlling stomatal-related traits in sorghum and maize. While considerable attempts have been made to address the bottlenecks associated with the phenotyping of stomatal traits through computer-aided image acquisition, previously developed methods suffered from issues of being time- and labor-intensive and of inaccurate stomata identification, classification, and quantification in  $C_4$  grass species (Furbank and Tester, 2011). By addressing those issues, the studies by the three groups present end-to-end pipelines that use a deep

learning algorithm to automatically identify, classify, and quantify stomatal traits associated with plant water use efficiency (WUE) and drought tolerance (Figure 1A). By integrating this pipeline with genomic studies, the authors further report the underlying genetic architecture of stomatal traits (Figure 1C).

Traditional stomatal phenotyping involves plant tissue collection and preparation for imaging, image data acquisition under microscope, and manual phenotyping of traits of interest. To relieve the phenotyping bottleneck, Ferguson et al. (2021) and Xie et al. (2021) used optical topometry, a rapid and nondestructive method for measuring surface characteristics at the nanometer scale, and acquired images of leaves to extract morphology-related stomata traits. The three-dimensional topographic layer of the raw images was first filtered to capture the points of interest and then flattened to two dimensions in grayscale with luminosity optimization and contrast enhancement.

After preprocessing, the authors trained a convolutional neural network model (Mask R-CNN) for automatic detection and counting of stomatal traits, such as number, density, and area (He et al., 2017). A typical deep learning framework for phenotyping starts with feeding and preprocessing of raw images, followed by several layers of automatic feature extraction during training, and ends up with the trained and validated model that can identify, classify, quantify, and predict the phenotypic traits of interest (Singh et al., 2018). Mask R-CNN detects and localizes objects of interest, such as stomata, using bounding boxes and generates precise segmentation masks (Figure 1A). The algorithm then uses several convolutional layers to identify and classify object region and then to predict object type. To train the Mask R-CNN model, the authors first labeled the input



**Figure 1** Developing and integrating deep learning-based high-throughput-phenotyping with genomic studies identifies genetic regions for stomatal traits in  $C_4$  grasses. A, A phenotypic pipeline showing rapid image acquisition by optical topometry and image analysis by a deep learning algorithm (Mask R-CNN) provides a powerful tool for identifying optimal stomatal traits. B, Correlation between manually measured and computationally predicted stomatal complex area in sorghum. C, Association of phenotypic data with genetic variants. Adapted from Figure 1B in Xie et al. (2021), Figure 3A in Bheemanahalli et al. (2021), and Figure 5D in Ferguson et al. (2021).

images as stomata and/or pavement cells and randomly split the entire image set into a training set for model training and validation set for model validation.

Stomatal traits predicted by the deep learning-based methods exhibited strong correlations with manual measurements across the three studies while dramatically reducing the required time and labor burden. For example, manual and computer-predicted measurements of stomatal complex area in sorghum leaves were significantly and positively correlated with each other ( $R^2 > 0.96$ , Figure 1B). In maize, the computer-predicted means of stomatal complex density and pavement cell density showed high significant correlations with those of the manually obtained values ( $R^2_{\text{SCD}} = 0.974$  and  $R^2_{\text{PD}} = 0.961$ , respectively). Bheemanahalli et al. (2021) showed a significant ( $P < 0.001$ ) strong relationship between predicted and manual observations of abaxial and adaxial stomata density, suggesting the reliability and accuracy of automated deep learning-based methods.

Next, the three sets of authors applied the deep learning-based pipelines to sets of maize and sorghum accessions and identified candidate genes with known and putative

roles for key WUE traits. Xie et al. (2021) utilized QTL mapping in a biparental mapping population of maize and identified high-confidence QTLs that were putatively pleiotropic and correlated with stomatal patterning and leaf gas exchange traits consistently in 2 years. They found these QTLs overlap with genetic position harboring not only known stomatal developmental and patterning genes, such as putative maize orthologs of *Arabidopsis thaliana* (*Arabidopsis*) *EPIDERMAL PATTERNING FACTOR 2*, *PANGLOSS1*, and *CYCLINA2;1*, but also genes previously not linked to stomatal traits.

A genome-wide association study (GWAS) on diverse grain sorghum accessions by Bheemanahalli et al. (2021) provided evidence for more than 71 genetic loci having significant association with stomatal traits, such as abaxial and adaxial stomatal density and stomatal complex area, and almost half as many overlapped with previously reported genomic regions. Further clarification of these regions revealed candidate putative genes including *ATP-binding cassette transporter*, *BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1*, *homeodomain-START transcription factor*, and *basic helix-loop-helix family transcription factor*, putative

orthologs of which are known to regulate leaf development, stomatal morphology, and stomatal lineage, respectively, in dicot and monocot models.

To improve the efficiency of GWAS and boost confidence in the identification of candidate genes, [Ferguson et al. \(2021\)](#) determined the transcriptomes of sorghum leaf tissues using a transcriptome-wide association study (TWAS). They used Fisher's combined test to integrate information from TWAS and GWAS and identified 394 unique candidate genes with high confidence of being associated with key stomatal and/or photosynthetic traits ([Figure 1C](#)). Included among these were two-thirds of genes that contained deleterious nonsynonymous/missense variants and had been previously identified as regulators of stomatal patterning and leaf development and anatomy in *Arabidopsis*. Examples include *BETA-KETOSYL-CoA SYNTHASE 1*, a cell wall *EXPANSIN-TYPE PROTEIN 2*, an ABA-sensitive *MAP KINASE*, *PURPLE ACID PHOSPHATASE 10*, and *EPIDERMAL PATTERNING FACTOR 2*.

Overall, the work by [Bheemanahalli et al. \(2021\)](#), [Ferguson et al. \(2021\)](#), and [Xie et al. \(2021\)](#) extends our understanding of stomatal biology and opens up the potential to engineer stomatal traits to enhance WUE and drought resistance without compromising yield in crops. Although the deep convolutional neural network platform showed potential for rapid and efficient WUE phenotyping, many questions arise, such as whether this platform is generalizable to diverse physiological traits across different crops and environments. In the coming years, it will be interesting to see how high-throughput phenotyping tools accelerate the rate at which

traits of interest are quantified and characterized, revealing gene-environment interaction modules of complex traits.

## References

- Bheemanahalli R, Wang C, Bashir E, Chilawal A, Pokharel M, Perumal R, Moghimi N, Ostmeyer T, Caragea D, Jagadish SVK** (2021) Classical phenotyping and deep learning concur on genetic control of stomatal density and area in sorghum. *Plant Physiol* **186**: 1562–1579
- Ferguson JN, Fernandes SB, Monier B, Miller ND, Allen D, Dmitrieva A, Schmuker P, Lozano R, Valluru R, Buckler ES, et al.** (2021) Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiol* **187**: 1481–1500
- Furbank RT, Tester M** (2011) Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* **16**: 635–644
- He K, Gkioxari G, Dollár P, Girshick R** (2017) Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp 2961–2969
- Leakey ADB, Ferguson JN, Pignon CP, Wu A, Jin Z, Hammer GL, Lobell DB** (2019) Water use efficiency as a constraint and target for improving the resilience and productivity of C3 and C4 crops. *Annu Rev Plant Biol* **70**: 781–808
- Singh AK, Ganapathysubramanian B, Sarkar S, Singh A** (2018) Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci* **23**: 883–898
- van Dijk ADJ, Kootstra G, Kruijer W, de Ridder D** (2021) Machine learning in plant science and plant breeding. *iScience* **24**: 101890
- Xie J, Fernandes SB, Mayfield-Jones D, Erice G, Choi M, Lipka AE, Leakey ADB** (2021) Optical topometry and machine learning to rapidly phenotype stomatal patterning traits for maize QTL mapping. *Plant Physiol* **187**: 1462–1480