

Adapting and Optimizing a Machine Learning Tool for Automated Cell Detection in *Setaria viridis*

Grace D. Tan^{1,2}, Jiayang Xie^{2,3}, Stephen Barretto², Andrew D.B. Leakey^{1,2,4}

¹Program in Ecology, Evolution, and Conservation Biology, ²Carl R. Woese Institute for Genomic Biology, ³College of Agriculture and Biotechnology, Zhejiang University, ⁴Departments of Plant Biology and Crop Sciences, University of Illinois at Urbana-Champaign

Pores in the leaf epidermis called stomata allow plants to take up carbon dioxide for photosynthesis, but are also pathways for water vapor loss. New image acquisition and analysis methods are allowing high-throughput phenotyping of stomatal patterning, which can be applied to better understand the genetic basis of variation in certain species. However, it takes considerable data and effort to train the models and their ability to accurately detect epidermal structures is constrained by the training data. This issue of context dependency, the inability to perform effectively in novel contexts, is the main hurdle preventing widespread adoption of machine learning in high-throughput phenotyping of intraspecific, interspecific, and environmental variation. Here we show the limited ability of a Mask-RCNN tool trained and successfully applied to *Zea mays*, to analyze images from a closely related grass called *Setaria viridis*. We then demonstrate successful retraining of the tool to cope with the novel amounts of diversity presented by this new species. The stomatal complexes in optical tomography images of mature *Setaria* leaves were accurately identified by comparison to expert raters ($R^2 = 0.84$). This study highlights the challenge of context dependency for widespread application of machine learning tools for phenotyping plant traits, even in closely related species. At the same time, it also provides a new tool that can be applied to leverage *Setaria* as a model C_4 species, and a roadmap for the translation of a machine learning tool to analyze stomatal patterning in diverse datasets of new plant species.