1 **Recognizing and overcoming context dependency in the application of a machine learning**
2 **tool for counting stomata in Setaria versus maize**
3
4 Grace D. Tan[1,2], Kevin Xie[2,3], Lauren J. Murphy[2], Andrew D.B. Leakey[4,2,3]
5
6 [1]Program in Ecology, Evolution and Conservation Biology, [2]Institute for Genomic Biology,
7 [3]Department of Crop Sciences, [4]Department of Plant Biology, University of Illinois at Urbana-
8 Champaign, Urbana, IL 61801
9
10 Stomata, microscopic pores on leaf surfaces, regulate the uptake of carbon dioxide and the
11 simultaneous loss of water vapor by leaves. New image acquisition and analysis methods are
12 allowing high-throughput phenotyping of stomatal patterning, which in turn have been applied
13 to better understand the genetic basis of variation in certain species. However, it takes
14 considerable data and effort to train the models, and their ability to accurately detect
15 epidermal structures is constrained to morphologies found within the training data. This issue
16 of context dependency, the inability to perform effectively in novel contexts, is the main hurdle
17 preventing widespread adoption of machine learning in high-throughput phenotyping of
18 intraspecific, interspecific, and environmental variation. Here we show the limited ability of a
19 Mask-RCNN tool, which was previously trained and successfully applied to *Zea mays*, to analyze
20 images from a closely related grass, *Setaria viridis*. We then demonstrate successful retraining
21 of the tool to cope with the novel diversity presented by this new species. The stomatal
22 complexes in optical tomography images of mature Setaria leaves were accurately identified by
23 comparison to expert raters (R^2 = 0.84). This study highlights the challenge of context
24 dependency for widespread application of machine learning tools for phenotyping plant traits,
25 even in closely related species. At the same time, it also provides a new tool that can be applied
26 to leverage Setaria as a model C_4 species, while also providing a roadmap for translation of a
27 machine learning to analyze stomatal patterning in new plant species.

Recognizing and overcoming context dependency in the application of a machine learning tool for counting stomata in Setaria versus maize

Grace Tan

University of Illinois Urbana-Champaign

Machine Learning

Stomata

High Throughput