Towards Automating Highly Heritable Phenotype Discovery For Plant Breeding

Ruhana N. Azam[1], Samuel B. Fernandes[2], Mohammed El-Kebir[1], Sanmi Koyejo[3], Alexander E. Lipka[1], and Andrew D.B. Leakey[1]

[1]Department of Computer Science, University of Illinois Urbana-Champaign
[2]Department of Crop, Soil, and Environmental Sciences, University of Arkansas
[3]Department of Computer Science, Stanford University
[4]Department of Crop Sciences, University of Illinois Urbana-Champaign
[4]Department of Plant Biology, University of Illinois Urbana-Champaign

Abstract

Genomic prediction uses genome-wide marker information to predict breeding values of agronomically important traits. This tool has become impactful because breeders can make selections earlier in the field season without waiting to observe traits that occur later in the field season. However, genomic prediction models are trained on biological traits that are labor-intensive to measure. One way to alleviate costs is to use highly heritable synthetic traits, i.e., functional combinations of observed low cost traits that correlate with the biological traits of interest. While heritability can be estimated, i.e., using variance component estimates from mixed models with genotype and environmental random effects, the search space for synthetic traits grows quickly with function complexity. Thus, current approaches that individually examine the heritability of each trait using brute-force are computationally intractable at scale. We propose the new problem of efficiently discovering highly-heritable synthetic traits. In this setting, the search algorithm can evaluate the heritability of a selected trait while being blind to the traits with the highest heritability. We investigate the characteristics of searching over heritability functions and the computational costs of search methods for discovering highly-heritable synthetic traits in large trait spaces.

References

[1] Ferguson, J. N., Fernandes, S. B., Monier, B., Miller, N. D., Allen, D., Dmitrieva, A., Schmuker, P., Lozano, R., Valluru, R., Buckler, E. S., Gore, M. A., Brown, P. J., Spalding, E. P., Leakey, A. D. B. (2021). Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. Plant Physiology, 187(3), 1481–1500. https://doi.org/10.1093/plphys/kiab346

[2] Fernandes, S. B., Azam, R. N., Paul, R. E., Yuan, M., El-Kebir, M., Koyejo, S., Lipka, A. E., Leakey, A. (2023). Including High-Throughput Phenotyping Derived Traits in Multi-Trait Genomic Analysis. Presented at the CSSA: Translational Genomics Workshop, Plant and Animal Genome XXI Conference.

[3] Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theoretical and Applied Genetics, 131(3), 747–755. https://doi.org/10.1007/s00122-017-3033-y