

# Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition

Sirjan Sapkota<sup>1,2\*</sup>, J. Lucas Boatwright<sup>1</sup>, Kathleen E. Jordan<sup>1</sup>, Richard E. Boyles<sup>2,3</sup>, Stephen Kresovich<sup>1,2</sup>

<sup>1</sup>Advanced Plant Technology Program, Clemson University, Clemson, SC, USA

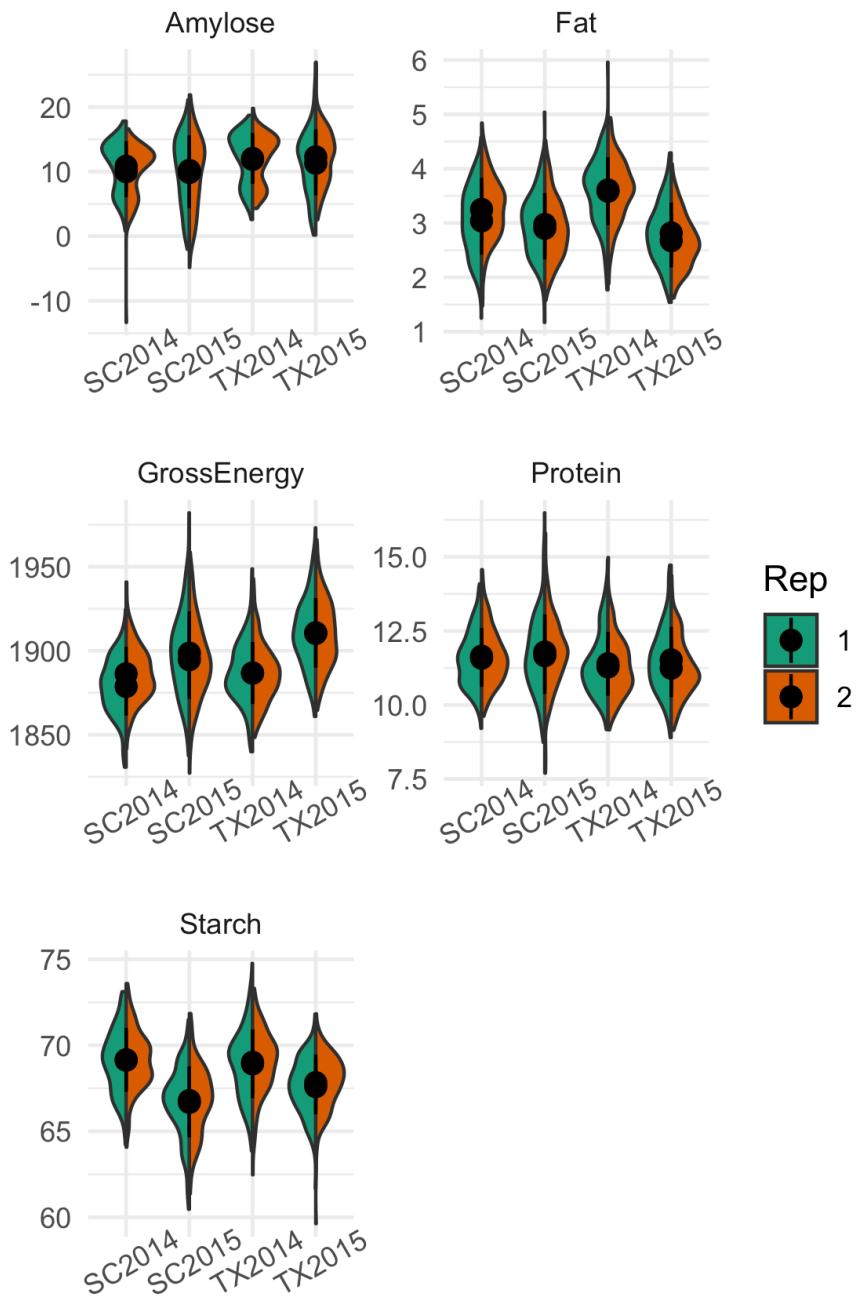
<sup>2</sup>Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA

<sup>3</sup>Pee Dee Research and Education Center, Clemson University, Florence, SC, USA

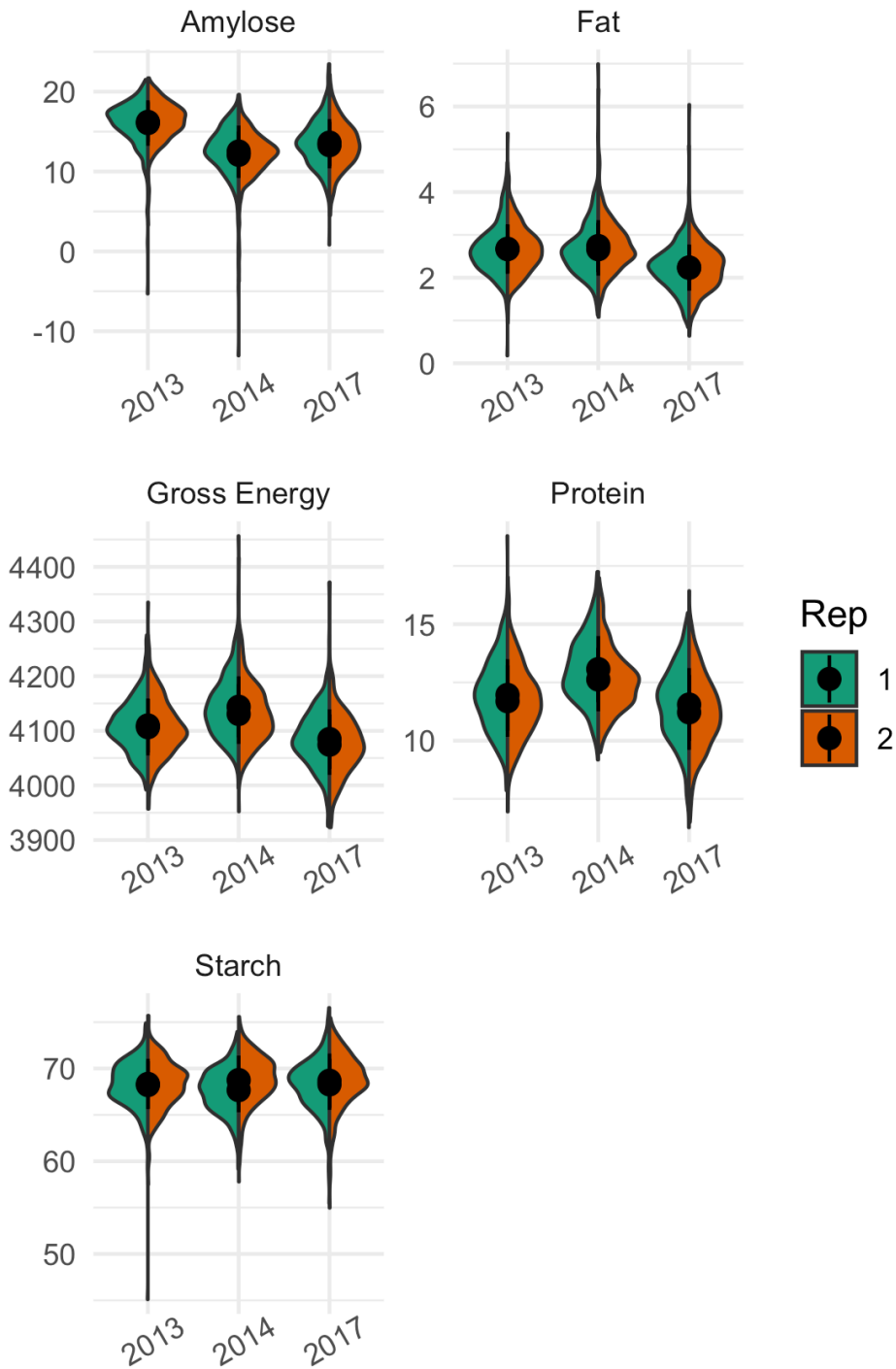
\* Corresponding author: [ssapkot@clemson.edu](mailto:ssapkot@clemson.edu)

Supplementary Information

**Figure S1. Phenotypic distribution of grain composition traits in the RILs.** In the x-axes, SC: South Carolina, TX: Texas, numbers represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in KCal/lb; and amylose is in percent of starch.

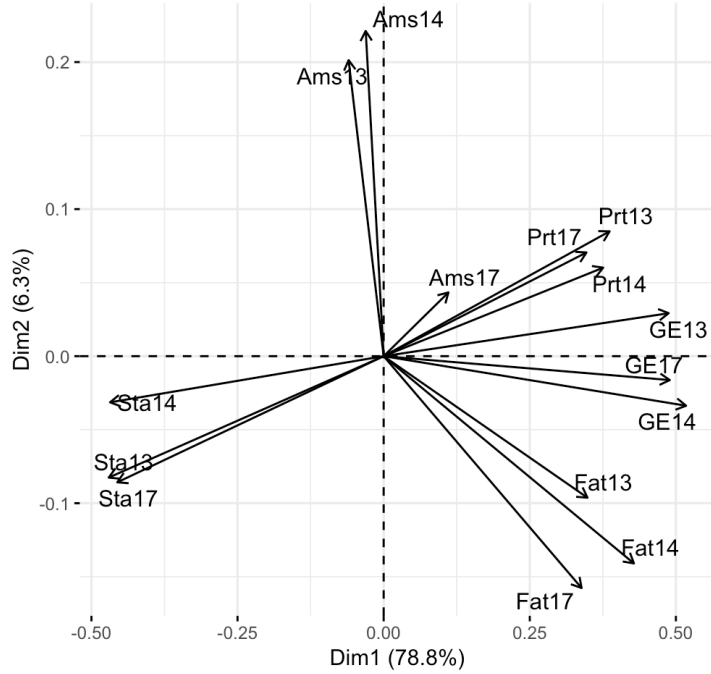


**Figure S2. Phenotypic distribution of grain composition traits in the GSDP.** Numbers in x-axes represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in Cal/g; and amylose is in percent of starch.

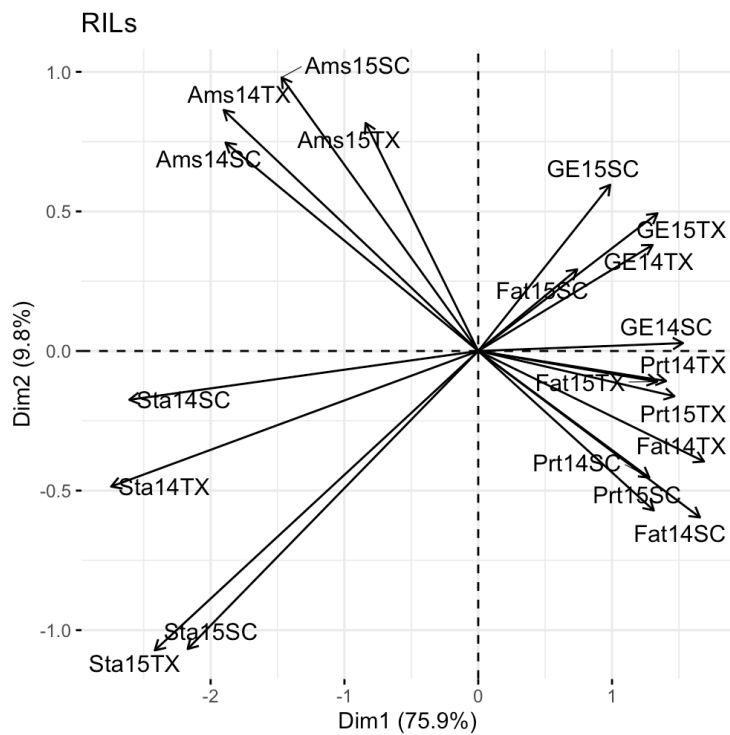


**Figure S3. PCA analysis of correlation matrix between traits. a. GSDP, and b. RILs.** Ams: amylose, GE: gross energy, Prt: protein, Sta: starch, SC: South Carolina, TX: Texas. The numbers in the text represent years of the environment.

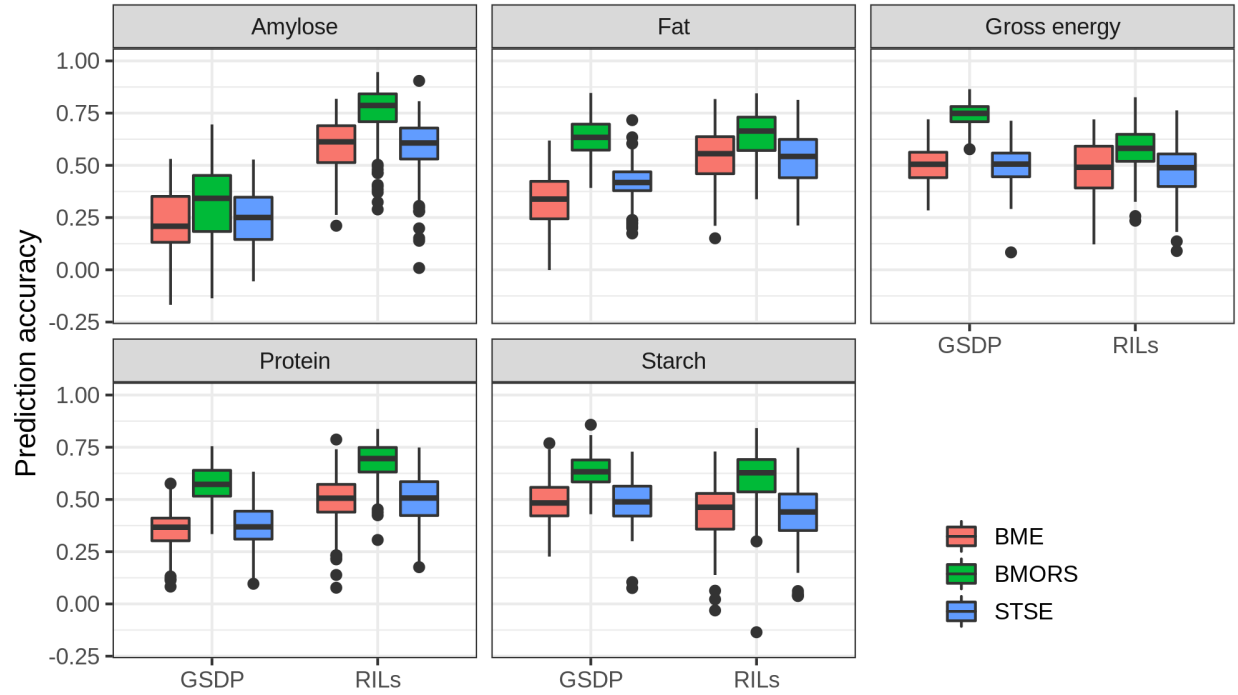
**a.**



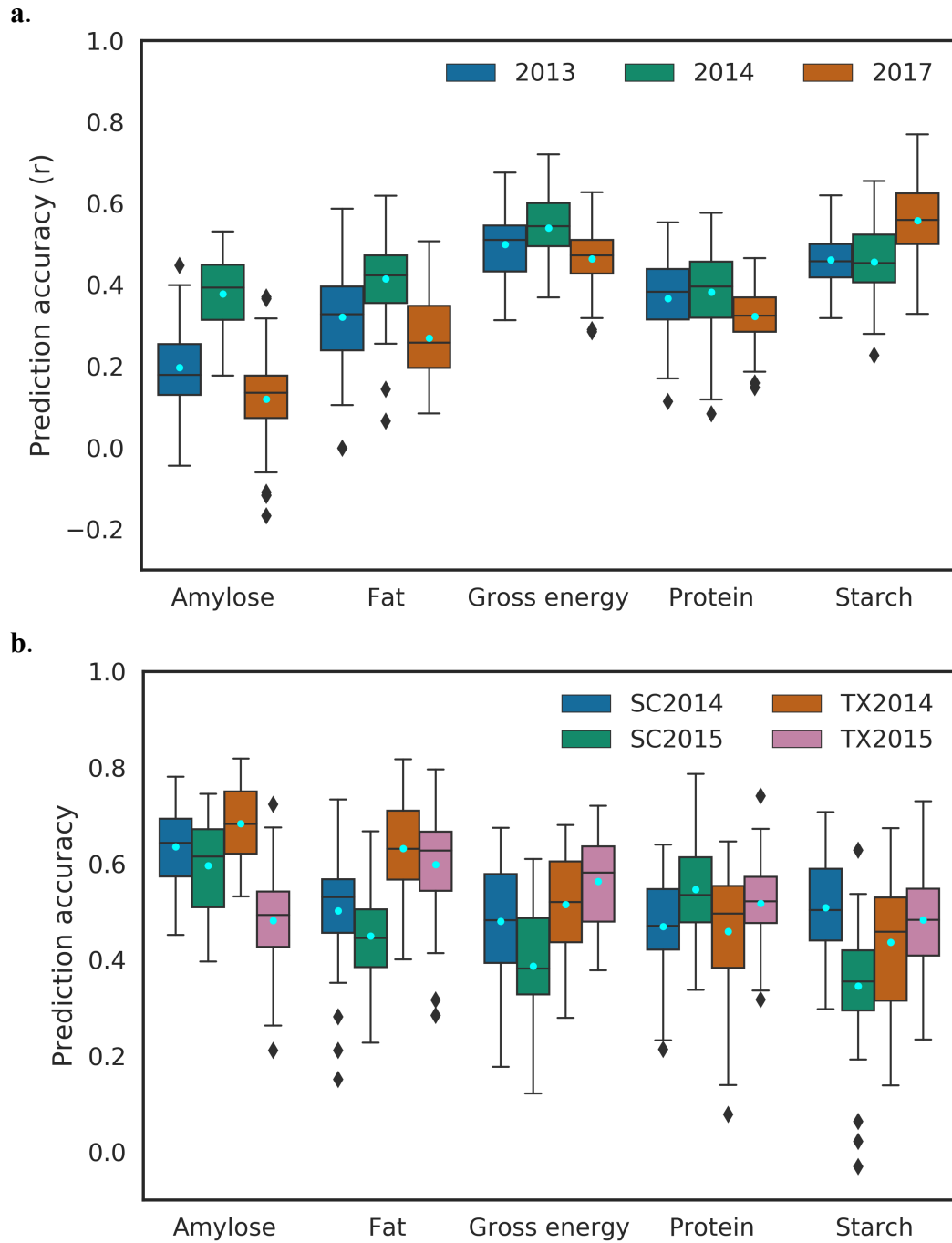
**b.**



**Figure S4. Overall prediction accuracy of traits across all the environment for the three prediction methods in the two populations.** The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Legend represents the environment/years. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, RILs: recombinant inbred lines.

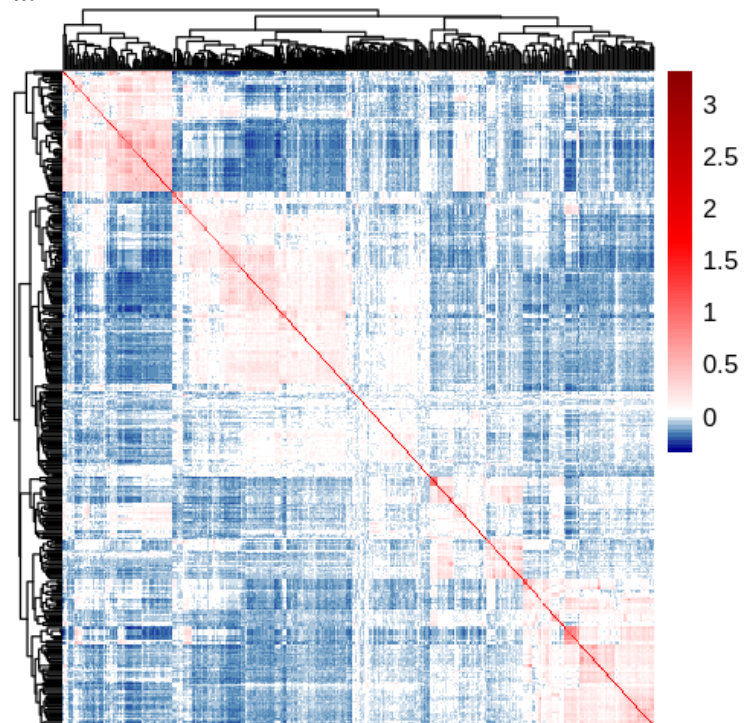


**Figure S5. Prediction accuracy using five-fold CV in Bayesian multi-environment (BME) model. a. GSDP, and b. RILs. Legend represents the environment/years. SC: South Carolina, TX: Texas. Pale blue dots represent the mean of prediction accuracy.**



**Figure S6. Heatmap for genomic relationship matrix calculated using vanRaden (2008). a. GSDP, b. RILs. Trees show hierarchical clustering using Euclidean distance.**

**a.**



**b.**

