BIOMASS YIELD PREDICTION IN PHOTOPERIOD-SENSITIVE SORGHUM

Samuel Bonfim Fernandes¹; Patrick James Brown²; Christopher Kaiser³; Payne Burks³; Elizabeth Hawkins⁴

¹PhD Candidate in Genetics and Plant breeding – UFLO/Lavras-MG/Brasil. Fellow FAPEMIG – email: samuelfernandes@agronomo.eng.br; ²Assistant Professor - Department of Crop Science, University of Illinois at Urbana-Champaign/Urbana-IL/United States. ³PhD Candidate in Genetics and Plant breeding – University of Illinois at Urbana-Champaign/Urbana-IL/United States. ⁴Undergraduate – University of Illinois at Urbana-Champaign/Urbana-IL/United States.

Phenotypic prediction using genotypic information and a trained model can accelerate genetic gain in plant breeding. We investigated the use of genomic selection (GS) and genome-wide association (GWAS) to predict biomass yield and its component traits (plant height, lodging, and moisture) in biomass sorghum, as well as the role of population structure in this prediction. A diverse panel of 430 photoperiod-sensitive sorghum lines obtained from Germplasm Resources Information Network (GRIN). Conducted in 2012, 2013 and 2014 in Urbana, IL, in an augmented block design with 24 common individuals across years. Biomass yield, moisture, and lodging were measured at harvest, and height was measured at 30, 60, 90 and 120 days after planting. 2012 used single row plots and 430 individuals, whereas 2013 and 2014 used four-row plots and 398 individuals. Repeatability for each trait in each year was calculated using the 24 individuals replicated in each year. Three datasets were used: 1) All years (2012, 2013 and 2014; n=430), 2) All years excluding individuals evaluated only in 2012 (n=398), and 3) years 2013 and 2014 only (n=398). Genotyping-by-sequencing was conducted using two pairs of restriction enzymes (PstI-HF/HinP1I and PstI-HF/BfaI), and SNPs were called using the TASSEL4 GBS pipeline. SNPs and individuals with more than 95% missing data were discarded as well as SNPs with MAF less than 5%. Imputation was conducted using BEAGLE4. Three models were used: A) regular rrBLUP (GS), B) GS using BLUPs from a model that included principal components 1-4 as covariates (GS-PC), C) Prediction using PCs 1-4 only (PC). Marker-trait associations were tested using an MLM implemented in GAPIT, and significance thresholds were determined using an FDR of 0.1. Five-fold cross-validation was repeated 60 times and prediction accuracy was calculated as the mean of Pearson’s correlations between genomic and phenotypically-estimated breeding values (GEBVs and PEBVs). For comparisons among traits the prediction accuracy was divided by the square root of the heritability of that trait. All sorghum races were represented in the top 5% of lines for biomass yield. Genetic correlations between biomass yield and other evaluated traits ranged from |0.31| to 0.57. Biomass yield and lodging had lower prediction accuracies than moisture and plant height traits. Prediction accuracy using population structure alone (PCs 1-4) is high in this diverse panel. A highly significant association with moisture was found near the sorghum Dry Stalk (D) locus in all analyses. Overall population structure assessed using PCs 1-4 accounts for much of the prediction accuracy in our diverse sorghum panel, even for traits like biomass yield for which top performers are scattered across all races of sorghum. Significant marker-trait associations were detected for traits genetically-correlated with biomass yield, but not for biomass yield itself.

Palavras-chave: Genomic Selection; GWAS; Bioenergy; rrBLUP.

Apoio Financeiro: DOE; BER; CAPES; FAPEMIG