GENOMIC SELECTION METHODS APPLIED TO HIGH BIOMASS SORGHUM FOR THE PRODUCTION OF SECOND GENERATION ETHANOL

Oliveira, AA[1]; Pastina, MM[2]; Souza, VF[2]; Parrella, RAC[2]; Noda, RW[2]; Magalhães, JV[2]; Schaffert, RE[2]; Damasceno, CMB[2]; Margarido, GRA[1].

1-Escola Superior de Agricultura Luiz de Queiroz, USP, São Paulo, SP.
2-Embrapa Milho e Sorgo, Sete Lagoas, MG.

avelar.a.o@gmail.com

Keywords: bioenergy; predictive models; accuracy; genotyping-by-sequencing; functional enrichment.

The increasing cost of energy and finite oil and gas reserves have created a need to develop alternative fuels from renewable sources. The use of plant biomass emerges as a viable alternative for biofuel production. Among the organic materials available, high biomass sorghum (*Sorghum bicolor* L. Moench) shows potential to be widely cultivated. Sorghum is a highly productive, drought-tolerant species with a history of improvement and production of lignocellulose, sugar and starch. Genomic selection is a useful tool for accelerating genetic gains and could, in the medium term, restructure plant breeding programs. This can promote greater genetic gains when compared to other methods and significantly reduce time required for the development of new cultivars through early selection. This work aimed at evaluating models of genomic selection and applying them to the prediction of breeding values for a diverse panel of high biomass sorghum from Embrapa Maize and Sorghum. The panel includes 200 genotypes from core collections from CIRAD and ICRISAT, as well as the germplasm bank and sorghum breeding program from Embrapa. The panel was divided in sub-panels 1 and 2, each comprised of 100 lines. Sub-panel 1 was evaluated phenotypically for two years (2011 and 2012), while sub-panel 2 was evaluated for only one (2011), both in the city of Sete Lagoas, Minas Gerais. We evaluated the following phenotypic traits related to biomass yield and quality: days to flowering, plant height, fresh and dry matter yield and fiber (neutral detergent and acid detergent fiber), cellulose, hemicellulose and lignin content. Subsequently, the whole panel was genotyped by the genotyping-by-sequencing technology. Using these genotypic and phenotypic data we fitted and compared the genomic selection models Bayes A, Bayes B, Bayes Cπ, Bayes Lasso, Bayes Ridge Regression and Random Regression BLUP (RRBLUP). The predictive capabilities obtained were high and varied little between the different models, ranging from 0.61 for days to flowering to 0.85 for acid detergent fiber, when the RRBLUP model was used for the combined analysis of the two sub-panels. Conversely, cross prediction between sub-panels resulted in lower predictive capability, never above 0.66, and in some scenarios virtually equal to zero, with greater variations between the fitted models. Simulations using subsets of molecular markers showed potential for achieving higher predictive capabilities. Functional enrichment analyses performed using the marker predicted effects suggested several interesting associations, such as larger effects of SNPs associated with the biosynthesis of aromatic amino acids for the lignin trait ($p = 0.0049$). These should be investigated in more detail in future studies, with potential for elucidating the genetic architecture of quantitative traits.

Financial Support: FAPESP, FAPEMIG, EMBRAPA.