

PREDICTIVE ABILITY BEHAVIOR ACROSS SITES AFTER DISCARD OF SNPS WITH UNSTABLE EFFECTS

João Filipi Rodrigues Guimarães¹; Janeo Eustáquio Almeida Filho¹; Marcio Fernando Resende Jr.²; Marcos Deon Vilela de Resende³; Fabyano Fonseca e Silva⁴; Patricio Muñoz⁵; Matias Kirst⁶

¹Doctorate student in Genetic and Breeding, UFV/Viçosa/MG/Brasil. * email: filipiguimaraes@yahoo.com.br; ² PhD in Genetic and Genomics, RAPiD Genomics LLC/Gainesville/FL/USA. ³Researcher at EMBRAPA Forestry and Dep. of Forest Engineering, UFV/Viçosa-MG/Brasil. Associate Professor, Dep. of Animal Science, UFV/Viçosa-MG/Brasil. ⁵ Assistant Professor, University of Florida/Gainesville/FL/USA, ⁶ Associate Professor, University of Florida/Gainesville/FL/USA.

Genomic Selection models typically take in account only additive genetic effects, leaving out of the predictive model non-additive effects and genotype by environment interactions. Modelling marker by environment interactions has the potential to create more stable prediction models and could potentially complement traditional genomic prediction methodologies. We hypothesized that by removing markers that have unstable effects across environments, we can increase the ability to predict phenotypes and select high-value genotypes across a broad range of conditions. This study aimed to verify the behavior of the predictive ability after the elimination of markers with unstable effect across sites. For the analysis, 700 individuals of *Pinus taeda* L. were genotyped with 4,722 SNPs and evaluated for the characteristic height in four sites across the United States: Palatka, Nassau, Cuthbert and BF Grant Forest. In the first step, the effects of all markers were estimated using a multivariate partial least squares methodology (MPLS), using an across-site 4-fold cross validation. The training population was composed by a combination of three sites and the validation was done with data from the fourth. This process was repeated four times until each site was used as a validation population. The number of latent components adopted for the MPLS was defined by cross-validation, and the estimate with the lowest mean square error was selected. The predictive ability was defined as the average of correlations across sites. The stability of the markers was evaluated by the methodology proposed by Huenh, where the stability across sites was tested using a chi-square test applied to statistics measures $Z1i$ $Z2i$. The measures $S1$, $S2$ and $S3$ were also obtained by Huenh's methodology, and refer to the specific stability of each marker (lower values indicate more stable markers). In the second step, we evaluated the effect of discarding 5%, 10%, 15% or 20% of the markers. The selection of markers to be removed was done using a rank index based in stability measures ($S1$, $S2$, $S3$) estimated in the first step. After removal of the markers, a new estimate of effects was made. The validation process was the same as previously described. There was a slight increase in predictive ability values after discard of markers, although it was not significant based on confidence intervals. We conclude that the discard of markers did not influence significantly the behavior of the prediction ability across environments.

Keywords: *Pinus taeda* L., GxE interaction, MPLS, stability.

Acknowledgments: The authors thank CAPES, CNPq, FAPEMIG, for support in student training. This work was supported by the US Department of Agriculture – National Institute of Food and Agriculture (awards no. 2013-67013-21159 and 2013-67009-21200 to MK and PM).