University of São Paulo
"Luiz de Queiroz" College of Agriculture

Genomic prediction for soybean segregating populations: selection strategies and training set establishment

Leandro de Freitas Mendonça

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2019

Leandro de Freitas Mendonça
Agronomist

Genomic prediction for soybean segregating populations: selection strategies and training set establishment

Advisor:
Prof. Dr. **ROBERTO FRITSCHE NETO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2019

RESUMO

Predição genômica para populações segregantes de soja: estratégias de seleção e estabelecimento da população de treinamento

Novas cultivares de soja são geradas a partir de cruzamentos bi-parentais, seguido de etapas de seleção e avanço de homozigose, cuja ordem de número de gerações varia de acordo com o método de melhoramento adotado. Nas etapas iniciais, a pouca quantidade de sementes por progênie, além da grande quantidade de indivíduos inviabiliza testes à campo com boa acurácia seletiva. Nesse contexto, a seleção genômica vem como método preditivo alternativo à simples amostragem aleatória nessas etapas. Sendo assim, o objetivo desta pesquisa for explorar aspectos relevantes ligados à aplicação de predição genômica nas etapas iniciais de um programa de melhoramento de soja. Os resultados mostram boa capacidade preditiva (acima 0.4) para os caracteres estudados (produtividade, altura de plantas e maturidade), mostrando ser possível aplicar seleção genômica já nas primeiras etapas do programa e obter ganhos de seleção. Além disso, demostrou-se que é possível obter capacidades preditivas equivalentes a um set de treinamento com irmãos completos, compondo-o apenas com linhagens homozigotas, possibilitando a criação de populações de treinamento performantes sem a necessidade de avaliação previa de progênies da mesma família, o que possibilita a criação de sets de treinamento estáveis ao longo ao longo dos anos e aplicáveis em distintas famílias.

Palavras-chave: Seleção genômica; Ganhos de seleção; Seleção precoce; Matriz de correlação genômica; *Glycine max*

ABSTRACT

Genomic prediction for soybean segregating populations: selection strategies and training set establishment

New soybean cultivars are generated from bi-parental crosses, followed by selection and homozygosis increasement stages, which the order of number of generations can vary according to the breeding method adopted. In the initial steps, the low quantities of seeds per progeny and the large number of individuals to be tested, makes it impossible to obtain a high-quality evaluation on field. In this context, genomic selection comes as an alternative predictive method, instead of simple random sampling. Therefore, the objective of this research is to explore relevant aspects related to the application of genomic prediction in the initial stages of a soybean breeding program. The results show good prediction ability (above 0.4) for traits tested evaluated (yield, plant height and maturity), showing that it is possible to apply genomic selection already in ealy steps of breding and obtain selection gains. In addition, it has been shown that it is possible to obtain predictive abilities equivalent to a full-sibs training set, establishing it only with pure lines, allowing the generation high predictive training populations without prior evaluation of within-family progenies, which allows the creation of stable training sets over the years and applicable in different families.

Keywords: Genomic selection; Selection gains; Early selection; Genomic relationship matrix; *Glycine max*

1. INTRODUCTION

One of the most important goals of breeding it to develop and release new varieties that should enables greater financial returns to the farmers. Breeders usually have some tools to make it easy and increase the efficiency of this process. One of these tools is call genomic selection (GS). The GS is a genetic-statistical approach capable to predict the performance of an individual, based in its molecular marker profile (Meuwissen et al., 2001). The first step is to obtain a training population, that must be genotyped, phenotype and related with the prediction set. Next, a linear or non-linear regression is usually used to estimate the alleles effect of each marker. Therefore, this vector can be used to predict the genomic estimated breeding value of a population (prediction set) that was just genotyped (Bernardo and Yu, 2007).

The use of GS is recommend in phases that the phenotypic evaluation and selection is inefficient (Jannink et al., 2010). The early steps of soybean breeding is a great example of it, once there are usually lots of new progenies to test and a reduced number of seeds (not enough for multi-location testing), what makes impossible a high-quality field evaluation. In this sense, GS could be applied to early select the best progenies, saving resources by non-evaluation of low potential ones in later steps.

The key point in the use of GS is the balance between cost of genotyping and prediction ability. Currently advances in the sequencing technology have reduced the cost of genotyping (Muir et al., 2016), what increase the advantages of applagin GS. In addition, the prediction ability of GS usually competes against low values of heritability in this phase for quantitative traits, such as yield. For that reason, in this research we investigated relevant aspects related to the use of GS in early steps of soybean breeding. In the first chapter, we focus on test different model's components and investigate the impact of selection intensity in the selection gains; and in the second, we explore different strategies to compose a stable and highly predictive training set.

2. CONCLUSIONS

In summary, we provided an interesting insight among the performance of GS, random sampling and selection based on traits measured in the field. It is important to remember that like GS, the field performance is also a prediction of the true genetic value of a variety, and both have a bias. Therefore, for each step of a breeding program, the prediction method with less error must be chosen. In this sense, our findings suggest GS is a useful prediction approach in the early stages of soybean breeding and obtain positive selection grains, early discarding those progenies with low potecial and avoiding the premature loss of progenies with high performance.

REFERENCES

Bernardo, R., and J. Yu. 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci. 47(3): 1082. doi: 10.2135/cropsci2006.11.0690.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9(2): 166–177. doi: 10.1093/bfgp/elq001.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–29. http://www.ncbi.nlm.nih.gov/pubmed/11290733 (accessed 22 September 2017).

Muir, P., S. Li, S. Lou, D. Wang, D.J. Spakowicz, L. Salichos, J. Zhang, G.M. Weinstock, F. Isaacs, J. Rozowsky, and M. Gerstein. 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 17(1): 53. doi: 10.1186/s13059-016-0917-0.