

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

New strategies for implementing genomic selection in breeding programs of
clonally propagated crops

Lorena Guimarães Batista

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

Piracicaba
2019

Lorena Guimarães Batista
Agronomist

New strategies for implementing genomic selection in breeding programs of clonally propagated crops

Advisor:

Prof. Dr. **GABRIEL RODRIGUES ALVES MARGARIDO**

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

Piracicaba
2019

RESUMO

Novas estratégias para a implementação de seleção genômica em programas de melhoramento de espécies de propagação vegetativa

A seleção genômica consiste no uso de efeitos preditos de marcadores genéticos para predizer os valores genéticos e/ou genotípicos de indivíduos genotipados. Desta forma, a seleção de genótipos superiores pode ser feita baseada apenas em valores genéticos preditos, reduzindo a necessidade de avaliações fenotípicas subsequentes. Isto representa um grande avanço em termos de custos e eficiência da seleção em programas de melhoramento de todos os tipos de culturas. No primeiro capítulo deste trabalho, nós exploramos uma das maneiras com que a seleção genômica pode ser utilizada para aumentar a eficiência no melhoramento simultâneo para múltiplos caracteres em espécies de propagação vegetativa. Utilizando simulações estocásticas, nós mostramos que um índice de seleção econômico deve ser utilizado no lugar da eliminação independente (*independent culling*). Os resultados mostram que o uso da seleção genômica pode tornar o custo-benefício da eliminação independente obsoleto se indivíduos em gerações iniciais forem genotipados e predições acuradas para todos os caracteres estiverem disponíveis desde o início. Apesar dos potenciais benefícios de realizar a seleção com base em valores genéticos preditos, para algumas espécies de propagação vegetativa a complexidade de seus genomas é um fator limitante para a efetiva implementação da seleção genômica em programas de melhoramento. Considerando que incluir a informação de dosagem alélica melhorou a performance de modelos de seleção genômica em espécies autotetraploides, nosso objetivo no segundo capítulo deste trabalho foi avaliar a acurácia da predição genômica com informação de dosagem alélica em cana-de-açúcar, que é uma complexa espécie poliploide. Neste capítulo, nós expandimos modelos GBLUP de seleção genômica desenvolvidos para autotetraploides para incluir níveis mais altos de ploidia. Dois modelos foram utilizados, um modelo com somente efeitos aditivos e um modelo com efeitos aditivos e efeitos de dominância digênica. Nós observamos uma modesta melhora na performance do modelo preditivo quando estimativas de ploidia e dosagem alélica foram incluídas, indicando que esta é uma possível maneira de aprimorar a seleção genômica em cana-de-açúcar. Os resultados obtidos nos dois estudos podem auxiliar pesquisadores e melhoristas de espécies de propagação vegetativa, abrindo portas para novas pesquisas e indicando as maneiras mais eficientes para implementação da seleção genômica.

Palavras-chave: Seleção genômica; Cana-de-açúcar; Dosagem alélica; Poliploides; Índice de seleção

ABSTRACT

New strategies for implementing of genomic selection in breeding programs of clonally propagated crops

Genomic selection consists of using predicted effects of genetic markers to predict breeding values and/or genotypic values of genotyped individuals. With this approach, selection can be carried based only on those predicted breeding values, reducing the need for further phenotypic evaluations. This represents a great advance in terms of cost and effectiveness of selection in breeding programs of all kinds of crops. In the first chapter of this work, we explore one of the ways genomic selection can be used to increase efficiency when breeding clonally propagated crops for multiple traits. Using stochastic simulations, we show that an economic selection index should be preferred over independent culling. Our results show that the use of genomic selection may render the cost-efficiency benefit of independent culling obsolete when all early generation individuals are genotyped and accurate prediction of all traits becomes available simultaneously. Despite the potential benefits of selecting based on predicted breeding values, for some clonally propagated species the complexity of their genomes limits the implementation of genomic selection in breeding programs. Since including allele dosage information has been shown to improve performance of genomic selection models in autotetraploid species, our objective in the second chapter of this work was to assess the accuracy of genome-wide prediction in the highly complex polyploid sugarcane when incorporating allele dosage information. In this chapter, we expanded GBLUP genomic selection models developed for autotetraploids to include higher levels of ploidy. Two types of model were used, one with additive effects only and one with additive and digenic dominance effects. We observed a modest improvement in the performance of the prediction model when ploidy and allele dosage estimates were included, indicating that this is a possible way of improving genomic selection in sugarcane. The results obtained in both studies can assist researchers and breeders of clonally propagated crops, opening new research opportunities and indicating the most efficient ways to implement genomic selection.

Keywords: Genomic selection; Sugarcane; Allele dosage; Polyploids; Selection Index

1. INTRODUCTION

Genomic selection is an approach that aims to increase the efficiency of selection in a breeding program, especially when selecting for complex traits, which usually requires evaluation of field trials in several locations and years, an expensive and time-consuming process (Heffner *et al.* 2009a). The method consists of using a training population that is both genotyped and phenotyped to predict the effect of genetic markers widely spread throughout the genome. The estimated effects can then be used to predict the phenotype of genotyped individuals in terms of estimated breeding values or estimated genotypic values (Meuwissen *et al.* 2001a). This allows selection to be carried without the need for further phenotypic evaluations, therefore shortening the time needed for selection of superior genotypes. Genomic selection can be implemented in any population of interest and has been successfully applied in several crop breeding programs (Bernardo and Yu 2007a; Heffner *et al.* 2009a; Crossa *et al.* 2010; Resende *et al.* 2012; Duhnen *et al.* 2017).

The breeding schemes for clonally propagated crops generally comprise several sequential steps that are carried over several years. The overall scheme can be simply summarized in generating genetic variation through crosses and subsequently selecting clones in the resulting F₁ progenies, which is done in several stages of selection, until the most promising clones can be released as cultivars, which often are also the candidate parents for the next breeding cycle (Simmonds 1979). Typically, the initial stages of selection include a large number of individuals to be evaluated and, to increase program efficiency, individuals are initially culled based on traits that can be phenotyped at a lower cost and, as the number of individuals decreases and higher-cost phenotyping becomes feasible, selection is performed for other traits in later stages of selection (Grüneberg *et al.* 2009a). In the context of genomic selection, the plant breeder may no longer be forced to cull if the individuals are genotyped. Because accurate prediction of breeding values of genotyped individuals can become available simultaneously for all traits, a selection index could be used instead of independent culling.

The selection index method involves selection for all traits simultaneously based on a linear or non-linear combination of individual traits weighted by their importance for the breeding objective (Hazel and Lush 1942). Theoretically, the selection index is the most efficient method of selection for multiple traits (Hazel and Lush 1942; Young 1961a). A major drawback of independent culling in comparison to selection index is that independent culling, if strictly applied, will not select individuals below the threshold for one single trait despite being exceptional for all other traits, while the use of a selection index makes it possible to retain those individuals (Bernardo 2010). Thus, especially for the selection of parents in breeding programs, the use of a selection index instead of independent culling might lead to higher genetic gains across cycles of selection, particularly when the correlation between traits is unfavorable. In this context, in the first chapter we investigated the gains over several generations of genomic selection in a recurrent selection breeding program using either a selection index or independent culling. We used simulations of recurrent breeding programs to evaluate and compare both strategies with the purpose of quantifying the magnitude of the difference between the different selection methods.

In the second chapter, we focus on the practical deployment of genomic selection in sugarcane breeding programs. Sugarcane cultivars are auto-allopolyploids, with 100 to 130 chromosomes and different number of chromosome copies between homology groups (i.e., aneuploid) (D'Hont *et al.* 1996, 1998). Due to this extremely complex genome structure, the majority of genetic studies in sugarcane use either dominant or single-dosage codominant markers (Wu *et al.* 1992; Huckett and Botha 1995; Besse *et al.* 1998; Nair *et al.* 2002; Gouy *et al.* 2013; Aitken *et al.* 2014; Racedo *et al.* 2016; Balsalobre *et al.* 2017), i.e., polymorphisms that were either detected in a

presence/absence fashion or that could only be detected in one chromosome per homology group, without considering information of other allele dosage levels. With the recent possibility of estimating the ploidy and allele dosage of markers (Serang *et al.* 2012; Garcia *et al.* 2013; Mollinari and Serang 2015), markers with higher dosages can be used in studies of polyploid species. Also, given that recent studies have shown that allele dosage information can improve the accuracy of genomic selection models in autotetraploid species (Slater *et al.* 2016, 2016; de Bem Oliveira *et al.* 2018; Hawkins and Yu 2018; Endelman *et al.* 2018), our objective in the second chapter was to assess the accuracy of genomic selection in sugarcane when incorporating allele dosage information.

Overall, we tackled possible ways to improve the implementation of genomic selection in breeding programs of clonally propagated crops in two levels. First, in terms of rearranging breeding schemes, by replacing the use of independent culling for an economic selection index; second, in terms of adapting genotyping techniques and genomic selection models to the complexity of the polyploid sugarcane genome, by estimating both ploidy and allele dosage of markers and incorporating this information in the prediction model.

2. CONCLUSION

2.1. First chapter

We evaluated and compared recurrent selection breeding programs using either independent culling or index selection for parent selection. The results show that, despite selection being carried out under unfavourable genetic correlations when using the selection index instead of independent culling, equivalent or higher genetic gains were achieved with index selection in all simulated scenarios. In terms of genetic diversity, the differences between methods in the studied system were driven mostly by differences in the generation of linkage disequilibrium between causal loci induced and not differences in allele frequencies. When linkage disequilibrium was not considered, both methods were equivalent in terms of loss of genetic diversity, and the differences between methods in terms of efficiency of converting genetic diversity into genetic gains mostly reflected the differences in the genetic gains obtained with each method. To obtain higher genetic gains, accurately assessing the economic importance of the traits is essential even when independent culling is performed, as optimal culling levels should be determined in order for maximum gain to be achieved. Given that optimal culling levels are complex to estimate, once the economic importance of each trait is known, maximum genetic gains are more easily achieved with index selection. Therefore, the best choice for plant breeding programs is to select parents using an economic selection index.

2.2. Second chapter

Overall, including estimates of ploidy and allele dosage of the SNPs led to a modest improvement of genomic selection models in sugarcane. The improvement we observed is likely to be more evident with larger training population sets that also display higher genetic variability, which would allow the models to have more precision to accurately estimate both the additive and the digenic dominance effects.

REFERENCES

- Aitken, K. S., M. D. McNeil, S. Hermann, P. C. Bundock, A. Kilian *et al.*, 2014 A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. *BMC genomics* 15: 152–152.
- Balsalobre, T. W. A., G. da Silva Pereira, G. R. A. Margarido, R. Gazaffi, F. Z. Barreto *et al.*, 2017 GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18:.
- de Bem Oliveira, I., M. F. Resende, F. Ferrao, R. Amadeu, J. Endelman *et al.*, 2018 Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *bioRxiv*.
- Bennett, G. L., and L. A. Swiger, 1980 Genetic variance and correlation after selection for two traits by index, independent culling levels and extreme selection. *Genetics* 94: 763–775.
- Bernardo, R., 2010 *Breeding for quantitative traits in plants*. Stemma Press, Woodbury.
- Bernardo, R., and J. Yu, 2007a Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47: 1082–1090.
- Bernardo, R., and J. Yu, 2007b Prospects for Genomewide Selection for Quantitative Traits in Maize All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any informati. *Crop Science* 47: 1082–1090.
- Besse, P., G. Taylor, B. Carroll, N. Berding, D. Burner *et al.*, 1998 Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. *Genetica* 104: 143–153.
- Bulmer, M. G., 1971 THE EFFECT OF SELECTION ON GENETIC VARIABILITY. 105: 1–88.
- Casler, M. D., and K. P. Vogel, 1999 Accomplishments and Impact from Breeding for Increased Forage Nutritional Value. *Crop Science* 39: 12–20.
- Chen, G. K., P. Marjoram, and J. D. Wall, 2009 Fast and flexible simulation of DNA sequence data. *Genome research* 19: 136–142.
- Clark, S. A., B. P. Kinghorn, J. M. Hickey, and J. H. J. Van Der Werf, 2013 The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genetics Selection Evolution* 45: 1–8.
- Cotterill, P. P., and J. W. James, 1981 Optimising two-stage independent culling selection in tree and animal breeding. *Theoretical and Applied Genetics* 59:.
- Cowling, W., and L. Li, 2018 *Turning the heat up on independent culling in crop breeding*.

- Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño *et al.*, 2010 Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186: 713.
- D'Hont, A., L. Grivet, P. Feldmann, J. C. Glaszmann, S. Rao *et al.*, 1996 Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics MGG* 250: 405–413.
- D'Hont, A., D. Ison, K. Alix, C. Roux, and J. C. Glaszmann, 1998 Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41: 221–225.
- Ducrocq, V., and J. J. Colleau, 1989 Optimum truncation points for independent culling level selection on a multivariate normal distribution, with an application to dairy cattle selection. *Genetics Selection Evolution* 21: 185.
- Duhnen, A., A. Gras, S. Teyssèdre, M. Romestant, B. Claustres *et al.*, 2017 Genomic Selection for Yield and Seed Protein Content in Soybean: A Study of Breeding Program Data and Assessment of Prediction Accuracy. *Crop Science* 57: 1325.
- Duvick, D. N., and K. G. Cassman, 1999 Post–Green Revolution Trends in Yield Potential of Temperate Maize in the North-Central United States. *Crop Science* 39: 1622–1630.
- Endelman, J. B., C. A. S. Carley, P. C. Bethke, J. J. Coombs, M. E. Clough *et al.*, 2018 Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. *Genetics* 209: 77.
- Erskine, W., P. C. Williams, and H. Nakkoul, 1985 Genetic and environmental variation in seed yield, seed size and cooking quality of lentil. *Field Crops Research* 12: 153–161.
- Falconer, D. S., T. F. Mackay, and R. Frankham, 1996 *Introduction to Quantitative Genetics* (4th edn). *Trends in Genetics* 12: 280.
- Garcia, A. A. F., M. Mollinari, T. G. Marconi, O. R. Serang, R. R. Silva *et al.*, 2013 SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Scientific Reports* 3: 3399.
- Gaynor, R. C., G. Gorjanc, A. R. Bentley, E. S. Ober, P. Howell *et al.*, 2017 A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. *Crop Science* 57: 2372–2386.
- Gaynor, R. C., G. Gorjanc, and D. L. Wilson AlphaSimR: An R Package for Breeding Program Simulations. *Manuscr Prep.*
- Goddard, M. E., 1983 Selection indices for non-linear profit functions. *Theoret. Appl. Genetics* 64: 339–344.
- Gorjanc, G., R. C. Gaynor, and J. M. Hickey, 2017 Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *bioRxiv*.

- Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal *et al.*, 2013 Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics* 126: 2575–2586.
- Grüneberg, W., R. Mwanga, M. Andrade, and J. Espinoza, 2009a Selection methods. Part 5: Breeding clonally propagated crops. *Plant Breeding and Farmer Participation* 275–322.
- Grüneberg, W., R. Mwanga, M. Andrade, and J. Espinoza, 2009b Selection methods. Part 5: Breeding clonally propagated crops. *Plant Breeding and Farmer Participation* 275–322.
- Hawkins, C., and L.-X. Yu, 2018 Recent progress in alfalfa (*Medicago sativa* L.) genomics and genomic selection. *The Crop Journal*.
- Hazel, L. N., and J. L. Lush, 1942 The Efficiency of Three Methods of Selection. *Journal of Heredity* 33: 393–399.
- Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009a Genomic Selection for Crop Improvement All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval syst. *Crop Science* 49: 1–12.
- Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009b Genomic Selection for Crop Improvement All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval syst. *Crop Science* 49: 1–12.
- Hickey, J. M., T. Chiurugwi, I. Mackay, W. Powell, and I. G. S. in C. B. P. W. Participants, 2017 Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics* 49: 1297.
- Huckett, B. I., and F. C. Botha, 1995 Stability and potential use of RAPD markers in a sugarcane genealogy. *Euphytica* 86: 117–125.
- Itoh, Y., 1991 Changes in genetic correlations by index selection. *Genetics selection evolution GSE* 23: 301–308.
- Kato, T., and K. Takeda, 1996 Associations among Characters Related to Yield Sink Capacity in Space-Planted Rice. *Crop Science* 36: 1135–1139.
- Kwon, S. H., and J. H. Torrie, 1964 Heritability of and interrelationships among traits of two soybean population.
- Meredith, W. R., and R. R. Bridge, 1971 Breakup of Linkage Blocks in Cotton, *Gossypium hirsutum* L.1. *Crop Science* 11: 695–698.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001a Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001b Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Mollinari, M., and O. Serang, 2015 Quantitative SNP Genotyping of Polyploids with MassARRAY and Other Platforms. Batley J. (eds) *Plant Genotyping. Methods in Molecular Biology (Methods and Protocols)* 1245:.
- Nair, N. V., A. Selvi, T. V. Sreenivasan, and K. N. Pushpalatha, 2002 Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms. *Euphytica* 127: 219–225.
- Namkoong, G., 1970 Optimum Allocation of Selection Intensity in Two Stages of Truncation Selection. *Biometrics* 26: 465.
- Racedo, J., L. Gutiérrez, M. F. Perera, S. Ostengo, E. M. Pardo *et al.*, 2016 Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC plant biology* 16: 142–142.
- Resende, M. D. V., M. F. R. Resende, C. P. Sansaloni, C. D. Petrolí, A. A. Missiaggia *et al.*, 2012 Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194: 116–128.
- Rharrabti, Y., S. Elhani, V. Martos-Núñez, and L. F. García del Moral, 2001 Protein and Lysine Content, Grain Yield, and Other Technological Traits in Durum Wheat under Mediterranean Conditions. *Journal of Agricultural and Food Chemistry* 49: 3802–3807.
- Rotundo, J. L., L. Borrás, M. E. Westgate, and J. H. Orf, 2009 Relationship between assimilate supply per seed during seed filling and soybean seed composition. *Field Crops Research* 112: 90–96.
- Saxton, A. M., 1989 INDCULL Version 3.0: Independent Culling for Two or More Traits. *Journal of Heredity* 80: 166–167.
- Serang, O., M. Mollinari, and A. A. F. Garcia, 2012 Efficient Exact Maximum a Posteriori Computation for Bayesian SNP Genotyping in Polyploids. *PLOS ONE* 7: e30906.
- Simmonds, N. W., 1979 *Principles of crop improvement*. Longman., London, K.
- Slater, A. T., N. O. I. Cogan, J. W. Forster, B. J. Hayes, and H. D. Daetwyler, 2016 Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. *The Plant Genome* 9:.
- Smedegaard-Petersen, V., and K. Tolstrup, 1985 The Limiting Effect of Disease Resistance on Yield. *Annual Review of Phytopathology* 23: 475–490.
- Smith, H. F., 1936 A Discriminant Function for Plant Selection. *Annals of Eugenics* 7: 240–250.
- Smith, S. P., and R. L. Quaas, 1982 Optimal Truncation Points for Independent Culling-Level Selection Involving Two Traits. *Biometrics* 38: 975.
- Tallis, G. M., 1987 Ancestral covariance and the Bulmer effect. *Theoretical and Applied Genetics* 73: 815–820.

- Triboi, E., P. Martre, C. Girousse, C. Ravel, and A. M. Triboi-Blondel, 2006 Unravelling environmental and genetic relationships between grain yield and nitrogen concentration for wheat. *European Journal of Agronomy* 25: 108–118.
- Venables, W. N., and B. D. Ripley, 2002 *Modern Applied Statistics with S*. Springer, New York.
- Woolliams, J. A., P. Berg, B. S. Dagnachew, and T. H. E. Meuwissen, 2015 Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics* 132: 89–99.
- Wu, K. K., W. Burnquist, M. E. Sorrells, T. L. Tew, P. H. Moore *et al.*, 1992 The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* 83: 294–300.
- Xu, S., and W. M. Muir, 1991 Multistage Selection for Genetic Gain by Orthogonal Transformation. *Genetics* 129: 963–974.
- Young, S., 1961a A further examination of the relative efficiency of three methods of selection for genetic gains under less-restricted conditions. *Genetical Research Cambridge* 2: 106–121.
- Young, S., 1961b A further examination of the relative efficiency of three methods of selection for genetic gains under less-restricted conditions. *Genetical Research Cambridge* 2: 106–121.
- Young, S. S. Y., and H. Weiler, 1960 Selection for two correlated traits by independent culling levels. *Journal of Genetics* 57: 329–338.