

**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

**Molecular pairwise relatedness in autopolyploids: a simulation study
considering linkage and many loci**

Rodrigo Rampazo Amadeu

Dissertation presented to obtain the degree of Master in
Science. Area: Genetics and Plant Breeding

**Piracicaba
2018**

Rodrigo Rampazo Amadeu
Agronomist

**Molecular pairwise relatedness in autopolyploids: a simulation study
considering linkage and many loci**

Advisor:

Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Dissertation presented to obtain the degree of Master in
Science. Area: Genetics and Plant Breeding

Piracicaba
2018

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Amadeu, Rodrigo Rampazo

Molecular pairwise relatedness in autopolyploids: a simulation study considering linkage and many loci/ Rodrigo Rampazo Amadeu. – Piracicaba, 2018 .

44 p.

Dissertação (Mestrado) – USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Parentesco molecular 2. Poliploides 3. Autotetraploides 4. Marcador molecular 5. SNP I. Título.

DEDICATORY

To Maria

ACKNOWLEDGMENTS

I'd like to acknowledge all those who support me to begin this scientific journey:

- My advisor Antonio Augusto Franco Garcia who taught me what is science.
- My other advisors Maria Marta Pastina and Patricio Munoz.
- All the friends of the eight years in the Statistical-Genetics Lab at ESALQ/USP.
- All the instructors at ESALQ/USP.
- The Department of Genetics at ESALQ/USP.
- The sponsors programs: Santander, CNPq, CAPES, and Science without Borders.
- My parents Marcia and Edson.
- My wife Maria.

*Quando eu te encarei frente a frente não vi o meu rosto
Chamei de mau gosto o que vi, de mau gosto, mau gosto
É que Narciso acha feio o que não é espelho
E à mente apavora o que ainda não é mesmo velho
Nada do que não era antes quando não somos Mutantes*

Sampa, Caetano Veloso

SUMMARY

Resumo	7
Abstract	8
1 Introduction	9
References	9
2 Conclusions	11

RESUMO

Parentesco molecular em autoploiploides: um estudo de simulação considerando ligação e muitos loci

Culturas agrícolas como cana-de-açúcar, batata, batata doce, diferentes forrageiras e frutas são autoploiploides. Em melhoramento de plantas, o conhecimento do parentesco entre cultivares é crítico. Essa informação pode ser incorporada em estudos de GWAS (genome-wide association studies) e GS (genomic selection) permitindo uma análise genética e predição de QTLs e valores genéticos de maior poder estatístico. No entanto, faltam estudos genéticos que considerem parentesco em um cenário autoploiploide. Nossos objetivos neste trabalho são: i) investigar, através de simulações, diferentes estimadores de parentesco entre indivíduos; ii) recomendar o melhor estimador baseado em cenários específicos considerando diferentes número de loci, número de alelos, ploidia, nível de dupla-redução e endogamia. Nossas recomendações poderão orientar estratégias de programas de melhoramento genético de espécies autoploiploides para escolha de estimadores baseados em seu cenário específico.

Palavras-chave: Parentesco molecular; Poliploides; Autotetraploides; Marcador molecular; SNP

ABSTRACT

Molecular pairwise relatedness in autopolyploids: a simulation study considering linkage and many loci

Crops, such as sugarcane, potato, and several forages and berries, are autopolyploids. In plant breeding studies, a key subject is the pairwise relatedness between cultivars. This information can be incorporated in GWAS (genome-wide association) and GS (genomic selection) studies allowing powerful genetic analysis. Despite the autopolyploid importance in agriculture, they lack genetic studies considering ploidy level, and, therefore, polysomic inheritance. Our objective in this work is: i) to investigate, through simulations, different molecular pairwise relatedness estimators; ii) to recommend which one is the best to be used in different scenarios regarding characteristic as number of loci, number of alleles, ploidy level, double-reduction, and inbreeding level. Our recommendations may guide autopolyploid breeding programs will be able to choose the best strategies and estimators based their reality.

Keywords: Molecular relationship; Polyploid; Autotetraploids; Molecular marker; SNP

1 INTRODUCTION

Estimation of pairwise relatedness is crucial to breeding programs. Cross planning alongside with modern molecular breeding tools as genomic selection and genome-wide association studies are some methods that are benefited from good relatedness estimation (DAETWYLER *et al.*, 2013). Based on pedigree and the works from Henderson in the 50's, relatedness has been used for prediction of BV (breeding values) based on BLUP (best linear unbiased prediction) models in animal (MRODE, 2014) and in plant breeding (PIEPHO *et al.*, 2008). Recently, relatedness has been estimated based on molecular information and incorporate in GBLUP (genomic best linear unbiased prediction) models (VANRADEN, 2008) for prediction of BV. However estimation based on molecular and pedigree for diploid are already long-termed established in the literature, considering polyploidy lack literature about how to achieve trustful pairwise relatedness.

Polyploids are individuals which had a whole genome duplication and can be divided in allopolyploids and autopolyploids. Allopolyploidy is the association of two distinct genomes into one, crops as wheat, cotton, and coffee tree are allopolyploids. Autopolyploidy is the whole genome duplication by itself, *e.g.*, an individual has its genome duplicated, crops as potato, sugarcane, and different forages and berries, are considered autopolyploids. Here, our focus are on autopolyploids. Different than allopolyploids where the inheritance pattern in like diploid, their inheritance is called polysomic: the homologous chromosomes can pair in different ways (bivalent, trivalent, quadrivalent etc) and may disrupt the expected loci segregations.

In autopolyploids, recent methods have been proposed to estimate pairwise relatedness based on genealogy (KERR *et al.*, 2012) and on modern molecular markers as SNPs (ASHRAF *et al.*, 2016) or even in SSRs (HUANG *et al.*, 2014, 2015). Although proposed, there is no study of how this methods works in the genomic era, *i.e.*, many loci and, therefore, linkage presence.

The objective of this work, written in the next chapter, is to present recommendational guidelines for breeding programs on molecular relatedness estimation with autopolyploidy. Here, we simulated several scenarios with different loci and alleles numbers and ploidy level. Based on these populations, we estimate relatedness with several available methods and evaluate their statistical power. At the end, we built a recommendational table with best methods (and their power) by scenario. Alongside with the recommendations, we discussed autopolyploidy genetics as Mendelian sampling behavior and the use of pseudo-diploid methodologies.

References

- ASHRAF, B. H., S. BYRNE, D. FÉ, A. CZABAN, T. ASP, M. G. PEDERSEN, I. LENK, N. ROULUND, T. DIDION, C. S. JENSEN, *et al.*, 2016 Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theor. Appl. Genet.* **129**: 45–52.
- DAETWYLER, H. D., M. P. CALUS, R. PONG-WONG, G. DE LOS CAMPOS, and J. M. HICKEY, 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* **193**: 347–365.

- HUANG, K., S. GUO, M. SHATTUCK, S. CHEN, X. QI, P. ZHANG, and B. LI, 2015 A maximum-likelihood estimation of pairwise relatedness for autopolyploids. *Heredity* **114**: 133–142.
- HUANG, K., K. RITLAND, S. GUO, M. SHATTUCK, and B. LI, 2014 A pairwise relatedness estimator for polyploids. *Mol. Ecol. Resour.* **14**: 734–744.
- KERR, R. J., L. LI, B. TIER, G. W. DUTKOWSKI, and T. A. MCRAE, 2012 Use of the numerator relationship matrix in genetic analysis of autopolyploid species. *Theor. Appl. Genet.* **124**: 1271–1282.
- MRODE, R. A., 2014 *Linear models for the prediction of animal breeding values*. Cabi.
- PIEPHO, H.-P., J. MÖHRING, A. E. MELCHINGER, and A. BÜCHSE, 2008 BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**: 209–228.
- VANRADEN, P. M., 2008 Efficient methods to compute genomic predictions. *J. Dairy. Sci.* **91**: 4414 – 4423.

2 CONCLUSIONS

- To consider polyploid dosage is crucial that a given statistical method achieve a high statistical power. In just few specific scenarios with highly unbalanced allele frequency pseudo-diploid methods may be satisfactory. Therefore, methods specifically developed for polyploids must be used.
- Relatedness estimated with the available multiallelic markers only have a high power with more than 100 loci and many alleles which. Nowadays, these may be unrealistic.
- To infer relatedness with high statistical power in a highly inbred autopolyploid population is harder than a population without inbred, available methods and markers are not satisfactory in this case.
- Nowadays, to have high statistical power in the relatedness estimators in autopolyploids, we recommend hundreds of effective biallelic SNP markers with reliable dosage inference.