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

Inferring heterotic patterns and the effect of incorporating dominance deviations for hybrid prediction: an example in tropical maize under nitrogen stress conditions

Júlia Silva Morosini

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

Piracicaba
2020

Júlia Silva Morosini
Agronomist

Inferring heterotic patterns and the effect of incorporating dominance deviations for hybrid prediction: an example in tropical maize under nitrogen stress conditions

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

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

Piracicaba
2020

To my beloved parents Carlos and Olivia,
and my sister Natália

ACKNOWLEDGMENTS

EPIGRAPH

Don't only practice your art,

But force your way into its secrets,

For it and knowledge can

Raise men to the Divine.

–Ludwig van Beethoven

RESUMO

Inferindo padrões heteróticos e o efeito da incorporação de desvios de dominância na predição de híbridos: um exemplo em milho tropical sob condições de estresse de nitrogênio

O melhoramento de híbridos explora a heterose, fenômeno melhor explorado quando há designação de grupos heteróticos (HG). A predição de híbridos incorpora informações de relacionamento nas análises estatístico-genéticas para aumentar a acurácia e melhorar a avaliação dos valores genéticos dos indivíduos. No entanto, tais matrizes de relacionamento geralmente negligenciam efeitos não aditivos. Além disso, um fator crítico em um programa de melhoramento híbrido é a previsão genômica ser realizada com base nos mesmos efeitos alélicos para dois grupos heteróticos. Finalmente, a deficiência de nitrogênio (N) é uma grande restrição para a produtividade do milho e pode confundir significativamente inferências de abordagens genéticas. Neste estudo, consideramos modelos que incluem efeito aditivo e a combinação de efeitos aditivo+dominância para estimar capacidades de combinação, determinar grupos heteróticos e o impacto da modelagem diferencial de efeitos de marcadores em grupos heteróticos em uma população de melhoramento de milho. Também investigamos o efeito do estresse N nesses parâmetros. Para isso, 906 híbridos simples obtidos de um dialelo de 49 linhagens de milho foram genotipados *in silico* usando 34,571 SNP e avaliados em quatro ambientes no Estado de São Paulo, Brasil, cada um com dois regimes N: ideal (IN) e estresse (LN). Três cenários de modelagem foram considerados: baseado em pedigree (I), onde nenhuma informação de relacionamento genômico foi considerada; aditivo (G_a), onde uma matriz de incidência aditiva foi atribuída às linhagens; e aditivo+dominância (G_{a+d}), onde os efeitos aditivo e de dominância foram considerados para linhagens e híbridos, respectivamente. Os HG foram definidos com base na capacidade específica de combinação (SCA) em cada cenário. As capacidades preditivas (PA) foram obtidas usando uma abordagem de validação cruzada de 5-fold. Os resultados indicam que a incorporação de ambos efeitos aditivos e desvios de dominância permite discriminar melhor as estimativas de capacidades geral e específica de combinação. A SCA designou agrupamentos distintos de linhagens genitoras para cada cenário. Ocorreram diferenças mais consideráveis na composição heterótica entre os cenários G_a e G_{a+d}, embora ainda houvesse uma sobreposição relevante. As estimativas genéticas foram significativamente menores para o LN em comparação ao IN, e a distância genética entre os grupos foi maior para o IN. As presentes descobertas revelam que agrupar HG com base em SCA do cenário G_{a+d} é uma abordagem útil para seleção recorrente recíproca quando não há informações anteriores disponíveis sobre a estrutura genética da população. Também foi observado que o estresse de N dificulta a resolução na determinação de padrões heteróticos, o que compromete a exploração da heterose. Por fim, a incorporação da dominância fornece informações relevantes para a determinação de grupos heteróticos, e a modelagem diferencial dos efeitos dos marcadores para cada HG é crucial para um programa de melhoramento sustentável.

Palavras-chave: *Zea mays*, Melhoramento, Capacidades de combinação, Estrutura de população, Grupos heteróticos

ABSTRACT

Inferring heterotic patterns and the effect of incorporating dominance deviations for hybrid prediction: an example in tropical maize under nitrogen stress conditions

Hybrid breeding exploits heterosis, a phenomenon that relies on the development of heterotic groups (HG). Hybrid prediction incorporates relationship information in genetic-statistical analyses to increase the accuracy and to improve the assessment of individuals' genetic values. Such relationship matrices, however, often neglect non-additive effects. Moreover, a critical factor in a hybrid breeding program is the genomic prediction being performed based on the same allelic effects for two heterotic groups. Finally, nitrogen (N) deficiency is a major constraint for maize productivity and may significantly confound inferences from genetic approaches. In this study, we considered models including additive and the combination of additive+dominance effects for the estimation of combining abilities, the determination of heterotic groups, and the impact of the differential modeling of marker effects on heterotic groups in a maize breeding population. We also investigated the N stress effect on these parameters. For that, 906 single crosses obtained from a diallel scheme of 49 inbred maize lines were genotyped in silico using 34,571 SNP and evaluated in four environments in State of São Paulo, Brazil, each with two N regimes: ideal (IN) and stress (LN). Three modeling scenarios were considered: pedigree-based (I), where no genomic relationship information was considered; additive (G_a), where an additive incidence matrix was assigned to lines; and additive+dominance (G_{a+d}), where the additive and the dominance effects were considered for lines and hybrids, respectively. HG were defined based on the specific combining ability (SCA) in each scenario. Prediction abilities (PA) were obtained using a 5-fold cross-validation approach. Our results indicate that the incorporation of both additive effects and dominance deviations allows us to discriminate the estimates of general and specific combining abilities better. SCA assigned distinct clustering of parents for each scenario. More considerable differences in heterotic pool composition occurred between scenarios G_a and G_{a+d}, although there was still a meaningful overlap. Incorporating HG into the prediction analysis provided a significantly increased SCA of the single crosses, and a noteworthy increase in PA. The genetic estimates were lower for LN compared to IN (significant at the 0.01 level), and the genetic distance between groups was higher for IN. Our findings reveal that additive effects are well assessed by using pedigree data when the population lacks structure. In addition, clustering HG based on SCA from the G_{a+d} scenario is a useful approach for reciprocal recurrent selection when no previous information is available about population genetic structure. We also observed that N stress hinders resolution in determining heterotic patterns, which compromises the exploration of heterosis. Ultimately, the incorporation of dominance provides relevant information for the determination of heterotic groups, and the differential modeling of marker effects for each HG is crucial to a sustainable breeding program.

Keywords: Hybrid breeding, Heterotic groups, Dominance, *Zea mays*, Population structure, Combining abilities, Nitrogen stress

1. INTRODUCTION

Hybrid breeding is based on the phenomenon called heterosis, which is the increased vigor and fitness of a crossbred genotype compared to corresponding parental inbred lines (Shull 1909). A key factor for the maximum exploitation of heterosis throughout the cycles is the establishment of heterotic groups (Boeven et al. 2016), which correspond to a set of genotypes genetically distant from another set (Larnkey et al. 1998; Technow et al. 2012).

The selection of a systematic method to developing progeny, called mating design, is a crucial step in hybrid breeding. Several types of mating design can be adopted to estimate additive and non-additive genetic variances, heritability, and breeding values (Falconer and Mackay 1996; Bernardo 2020). For hybrid breeding in special, the evaluation of all possible crosses is prohibitive in practice due to the high cost and the potentially enormous number of genotypes generated, especially in the early stages of a program. In this context, several genetic mating schemes have been proposed to identify elite parental lines and to evaluate hybrid performance in order to access heterosis. Among them, are the Griffing methods, Diallel, North Carolina Design II, and testcross (Hallauer et al. 2010).

Diallel designs have been widely used to estimate the effects of general combining ability (GCA) and specific combining ability (SCA), as well as other genetic parameters (Beyene et al. 2011; Fan et al. 2014). In maize, Griffing's methods have become popular to estimate the effects of GCA and SCA (Griffing 1956), and they define diallel crosses in terms of genotypic values: the sum of GCA for the two gametes is the genetic value to the cross, while SCA represents the value of dominance deviation in the simplest case, ignoring epistasis deviation (Olfati et al. 2012).

The reciprocal recurrent selection method (RRS) was first proposed by Comstock et al. (1949) and aims at the simultaneous improvement of two populations and their crossing, seeking the maximization of heterosis (Ordas et al. 2012). The genetic basis of heterosis is in the difference of allelic frequency between heterotic groups (Falconer and Mackay 1996), and heterotic grouping leads to a reduced SCA variance and a lower ratio of SCA to general combining ability GCA variance. Thus, early testing becomes more effective and superior hybrids can be identified and selected mainly based on their prediction from GCA effects.

Given the frequent practical impossibility of a full assessment of hybrid performance by the complete factorial mating design, only a small fraction of the genotypes is actually carried out in the field. For this reason, genomic models can be useful to predict the performance of the untested hybrids (Meuwissen et al. 2001; Schrag et al. 2010; Fan et al. 2014). Recent studies have

shown that genomic models can output considerably accurate predictions of the agronomic performance of hybrids (Kadam et al. 2016; Lyra et al. 2017). The choice of the mating design and its composition impact on the prediction accuracy. For instance, Fristche-Neto et al. (2018) reported the superiority of factorial and full diallel designs over testcrosses regarding the prediction of untested hybrids in maize, which also resulted in a better mating allocation.

Despite the essentiality of the principles of quantitative genetics for breeding studies and selection performance, some of their assumptions are rarely found in real life situations, which may lead to sensitive differences in estimations (Piepho et al. 2008). An example is considering parents in a breeding program as unrelated or uncorrelated. Thus, the incorporation of a kinship matrix based on the genealogy of the individuals has the potential to increase the prediction accuracy, the prediction of the genetic values, and the estimation of the components of variance (Holland et al. 2003).

In order to overcome the unrelated parent assumption, a pedigree matrix (A) that represents the expected additive genetic covariance between genotypes could be considered. However, this approach has some limitations, such as the non-capture of the realized kinship, and the non-detection of hidden co-ancestry and inbreeding. Additionally, this matrix ignores variation among family members (Hill and Weir 2011). Therefore, it overestimates either the individuals breeding values and the additive genetic variance (Gamal El-Dien et al. 2016). Conversely, through the use of molecular markers and genomic relationship matrices (G – Wright 1922; VanRaden 2008), it is possible to capture the realized resemblance between individuals, regardless of their genealogy, providing accurate estimates of variance and breeding values (Gamal El-Dien et al. 2016).

Genomic selection (GS), initially proposed by Meuwissen et al. (2001), is a method of marker-based selection that uses a large number of markers spread throughout the genome to predict the genetic value of selection candidates base on the genomic estimated breeding value (GEBV) predicted from the markers (Ward et al. 2019). GS combines molecular and phenotypic data in a training population to obtain estimates of GEBV of individuals in a testing population that has been genotyped but not phenotyped (Crossa et al. 2017). Since the production and testing of a substantial fraction of field hybrids are impractical, hybrid performance prediction is highly relevant for maize breeding (Technow et al. 2014).

Previous studies have shown that the use of molecular-based relationship matrices improves orthogonality and predictability of additive and non-additive effects (Vitezica et al. 2013; Dias et al. 2018). Nevertheless, most genetic analyses focus only on the estimation of additive or

total genetic effects, not considering dominance. The latter is crucial when it comes to hybrids performance due to its measurable effect on heterosis. As such, it is important to consider the dominant relationship among individuals to achieve more realistic combining abilities. Some studies have reported the positive impact of incorporating dominance deviations in the genetic models to achieve higher predictive abilities (Lyra et al. 2017; Alves et al. 2019).

Conducting GS requires estimation of the marker's effect, treated predominantly in practice as additive. Approaches including dominance (Technow et al. 2012), epistatic, and genotype-by-environment effects are being developed (Rice and Lipka 2019). Nevertheless, GS does not adjust consistently to the RRS method, since it considers only a vector of additive and dominant effects for the entire population and their respective groups. Based on quantitative genetics, what is expected in this case is that both groups are selected by the same allele effects, which would lead to a reduction in the genetic distance between them along the cycles, with consequent impairment of heterosis expression.

Response indices or selection indices are important tools in plant breeding because they allow to simultaneously summarize the performance of plants under different conditions or traits into a single value (Wu et al. 2011). The evaluation and selection of genotypes through this approach is highly advantageous for situations in which contrasting scenarios need to be evaluated, such as in nutritional stress breeding programs. For maize, nitrogen (N) deficiency represents a major constraint, negatively affecting plant development and yield. In this context, selection based on indices aids to improve the overall genotype performance based on several quantitative traits simultaneously (Lyra et al. 2017; Morosini et al. 2017).

In this context, the understand of the incorporation of genomic relationship matrices for the prediction of genetic values and its components in a diallel analyses context is of great interest. By comparing the methods regarding the testers elected and the heterotic group composition, the differential modeling of the marker effects for each heterotic group emerges as a potential verification approach, since it is in accordance with the principle of RSS. In addition, understading how the nitrogen stress affects the heterotic composition is crucial for the optimization of population clustering towards genetic gains. For that, we considered three scenarios: inbred lines parents as unrelated (I); inclusion of the genomic additive relationship matrix for parents (G_a); and inclusion of additive genomic information for parents and the genomic dominance relationship matrix for hybrids (G_{a+d}), for addressing these aspects.

2. CONCLUSION

In our study, we applied three modeling scenarios considering combinations of additive and dominance effects into a diallel analysis of a tropical maize population of 906 maize hybrids from 49 inbred lines to understand how the genetic relationship information impacts on the estimation of combining abilities, the assginment of heterotic groups, and genome wide prediction. The effect of nitrogen stress on pool clustering was also assessed. We generated the *in-silico* score GEV to compare I, G_a and G_{a+d}. From the SCA estimates, heterotic groups were assigned according to each scenario. Our results indicate that the incorporation of both additive effects and dominance deviations allows to discriminate better the GCA and SCA estimates. In addition, a population with poor structure benefits only marginally from the inclusion of additive genetic kernels in a diallel context, since the information provided by the crosses themselves is sufficient to estimate the effects. Heterotic patterns defined using SCA values estimated with dominance deviations provided a greater genetic distance between clusters and potential higher expression of heterosis compared to the groups defined using only pedigree/additive information. N stress hinders resolution in determining heterotic patterns, with consequent impact on heterosis assessment and exploration. In addition, we observed that a smaller set of crosses genetically well targeted by heterotic groups is as efficient as using all inter- and intra-crosses in terms of prediction ability. Finally, the differential modeling of marker effects for each heterotic group is crucial to a sustainable RRS breeding program.

REFERENCES

Alves FC, Granato ÍSC, Galli G, et al (2019) Bayesian analysis and prediction of hybrid performance. Plant Methods 15:14. doi: 10.1186/s13007-019-0388-x

Bernardo R (2020) Breeding for quantitative traits in plants, 3rd ed. Stemma Press, Woodbury, Minnesota

Beyene Y, Mugo S, Mutinda C, et al (2011) Genotype by environment interactions and yield stability of stemborer resistant maize hybrids in Kenya. African J Biotechnol 10:4752–4758

Boeven PHG, Longin CFH, Würschum T (2016) A unified framework for hybrid breeding and the establishment of heterotic groups in wheat. Theor Appl Genet 129:1231–1245. doi: 10.1007/s00122-016-2699-x

Comstock RE, Robinson HF, Harvey PH (1949) A Breeding Procedure Designed To Make Maximum Use of Both General and Specific Combining Ability1. Agron J 41:360. doi: 10.2134/agronj1949.00021962004100080006x

Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci 22:961–975. doi: 10.1016/j.tplants.2017.08.011

Dias KO das G, Alejandro Gezan S, Teixeira Guimarães C, et al (2018) Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity (Edinb) 121:24–37. doi: 10.1038/s41437-018-0053-6

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th edn. Pearson, Harlow, Essex, England

Fan XM, Zhang YD, Yao WH, et al (2014) Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. Crop Sci 54:89–97. doi: 10.2135/cropsci2013.06.0393

Fristche-Neto R, Akdemir D, Jannink JL (2018) Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. Theor Appl Genet 131:1153–1162. doi: 10.1007/s00122-018-3068-8

Gamal El-Dien O, Ratcliffe B, Klápště J, et al (2016) Implementation of the Realized Genomic Relationship Matrix to Open-Pollinated White Spruce Family Testing for Disentangling Additive from Nonadditive Genetic Effects. G3 Genes|Genomes|Genetics 6:743–753. doi: 10.1534/g3.115.025957

Griffing B (1956) Concept of General and Specific Combining Ability in Relation to Diallel Crossing Systems. Aust J Biol Sci 9:463–493. doi: 10.1071/BI9560463

Hallauer AC, Carena MJ, Miranda Filho JB (2010) Quantitative genetics in maize breeding. Springer, New York

Hill W., Weir B. (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet Res 93:47–64. doi: 10.1017/S0016672310000480.Variation

Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:9–112

Kadam DC, Potts SM, Bohn MO, et al (2016) Genomic Prediction of Single Crosses in the Early Stages of a Maize Hybrid Breeding Pipeline. G3 Genes|Genomes|Genetics 6:3443–3453. doi: 10.1534/g3.116.031286

Larnkey KR, Staub JE, Melchinger AE, Gumber RK (1998) Overview of Heterosis and Heterotic Groups in Agronomic Crops. In: Concepts and Breeding of Heterosis in Crop Plants. Crop Science Society of America, pp 29–44

Lyra DH, de Freitas Mendonça L, Galli G, et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed 37:. doi: 10.1007/s11032-017-0681-1

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–29

Morosini JS, Mendonça LDF, Vidotti MS, Fritsche-neto R (2017) Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. Plant Soil 421:453–463. doi: 10.1007/s11104-017-3479-3

Olfati JA, Samizadeh H, Rabiei B, Peyvast G (2012) Griffing's Methods Comparison for General and Specific Combining Ability in Cucumber. Sci World J 2012:1–4. doi: 10.1100/2012/524873

Ordas B, Butron A, Alvarez A, et al (2012) Comparison of two methods of reciprocal recurrent selection in maize (Zea mays L.). Theor Appl Genet 124:1183–1191. doi: 10.1007/s00122-011-1778-2

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. doi: 10.1007/s10681-007-9449-8

Rice B, Lipka AE (2019) Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum. Plant Genome 12:0. doi: 10.3835/plantgenome2018.07.0052

Schrag TA, Möhring J, Melchinger AE, et al (2010) Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor Appl Genet 120:451–461. doi: 10.1007/s00122-009-1208-x

Shull GH (1909) A pure-line method in corn breeding. J Hered. doi: 10.1093/jhered/os-5.1.51

Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor Appl Genet 125:1181–1194. doi: 10.1007/s00122-012-1905-8

Technow F, Schrag TA, Schipprack W, et al (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics. doi: 10.1534/genetics.114.165860

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91:4414–4423. doi: 10.3168/jds.2007-0980

Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195:1223–1230. doi: 10.1534/genetics.113.155176

Ward BP, Brown-Guedira G, Tyagi P, et al (2019) Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. Crop Sci 59:491. doi: 10.2135/cropsci2018.03.0189

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338

Wu Y, Liu W, Li X, et al (2011) Low-nitrogen stress tolerance and nitrogen agronomic efficiency among maize inbreds: Comparison of multiple indices and evaluation of genetic variation. Euphytica 180:281–290. doi: 10.1007/s10681-011-0409-y