

**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 FRITSCHÉ NETO**

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

Piracicaba
2020

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

Morosini, Júlia Silva

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. - - Piracicaba, 2020.

56 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. *Zea mays* 2. Melhoramento 3. Capacidades de combinação 4. Estrutura de população 5. Grupos heteróticos I. Título

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

ACKNOWLEDGMENTS

Foremost, I thank God and Our Lady for the blessings, strength, and good people They bestowed upon me during this journey. I extend my acknowledgments to the following:

To “Luiz de Queiroz” College of Agriculture from University of São Paulo – ESALQ/USP, which I very proudly have as my Alma Mater. I am grateful for the many outstanding opportunities I had here since I was an undergraduate.

To the Department of Genetics at ESALQ/USP, where I started my career as a geneticist. I am thankful to the faculty for the important knowledge transmitted, and staff, from the office to the field, for all the assistance and good moments we shared.

To the University of Wisconsin-Madison, where I integrated the Plant Breeding and Plant Genetics program for six months, for the extraordinary environment and experience, and for the good friends I made.

To Prof. Roberto Fritsche Neto for the guidance and opportunity to work as part of the Allogamous team at ESALQ/USP, where I met many talented colleagues and I had important lessons.

To Prof. Natalia de Leon, who kindly accepted me as part of her group at UW-Madison, for the encouragement, inspiration, and for taking my scientific/professional background to the next level.

To my parents, Carlos Alberto Morosini and Olivia Maria B. S. Morosini, for the unconditional love and care. We were together even when we were physically distant, and that made me through everything. Thank you for the amazing family you built for Natalia and I.

To my dear sister Natalia S. Morosini, such a strong and brave woman, for the constant talks, support, and for teaching me to always respect our own history as I navigated the PhD. Thank you for our beautiful friendship and for inspiring me.

To Fernando Garcia Espolador, a brilliant coworker and the love of my life, for the incredible teamwork and support. I am fortunate to share my days with you. To my father-in-law and mother-in-law, Carlos M. Espolador and Maria de Fátima G. Espolador, for being there for us every step of the way.

To my relatives and friends, who are always supportive and made this journey pleasant and fun.

To São Paulo Research Foundation (FAPESP – Grants 2017/25549-6 and 2019/06532-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 for the financial support.

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

TABLE OF CONTENTS

RESUMO	8
ABSTRACT.....	9
1. INTRODUCTION	11
2. MATERIALS AND METHODS.....	15
2.1. PHENOTYPIC DATASET.....	15
2.2. GENOTYPIC DATASET.....	15
2.3. PHENOTYPIC ANALYSES.....	16
2.4. DIALLEL ANALYSES.....	17
2.5. MODEL COMPARISONS	19
2.6. HETEROTIC GROUPING AND TESTERS	20
2.7. EFFECT OF NITROGEN STRESS ON HETEROTIC COMPOSITION	20
2.8. GENOMIC PREDICTION FOR SCENARIOS.....	20
3. RESULTS.....	23
3.1. PHENOTYPIC ANALYSIS.....	23
3.2. DIALLEL ANALYSES.....	25
3.3. MODEL COMPARISONS	27
3.4. HETEROTIC GROUPS.....	30
3.5. NITROGEN EFFECT ON HETEROTIC POOLS.....	30
3.6. GENOMIC PREDICTION	32
4. DISCUSSION	33
4.1. PHENOTYPIC DATA AND DIALLEL ANALYSIS.....	33
4.2. MODEL COMPARISON	34
4.3. HETEROTIC PATTERNS	36
5. CONCLUSION	39
REFERENCES	41
SUPPLEMENTARY MATERIAL.....	47

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; Olímpio Das Graças 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. MATERIALS AND METHODS

2.1. Phenotypic dataset

We used 906 maize single-crosses from a diallel mating design between 49 tropical inbred lines contrasting for N-use efficiency (Mendonca et al. 2016; Morosini et al. 2017). Field trials were carried out in Anhembi (22°50'51"S, 48°01'06"W, 466 m) and Piracicaba (22°42'23"S, 47°38'14"W, 535 m), in the State of São Paulo, Brazil, during the second growing season (January to May) of 2016 and 2017. In both sites, hybrids were evaluated under two N levels: low (LN) with 30 kg N ha⁻¹, and ideal (IN) with 100 kg N ha⁻¹. Each combination of season-location was considered an environment, totalizing four (1-Anhembi 2016, 1-Anhembi 2017, 3- Piracicaba 2016, and 4- Piracicaba 2017).

The experimental scheme was laid out as augmented blocks (unreplicated trial) consisting of 47 and 50 blocks (2016 and 2017, respectively), each with 16 unique hybrids and two checks. Plots of 7 m spaced at 0.50 m were used under conventional fertilization and weed and pest control. The traits evaluated included grain yield (GY, Mg ha⁻¹), plant height (PH, m), and ear height (EH, m). Plots were manually harvested, all ears in the plot were shelled and total GY was assessed. GY was corrected to 13% moisture. EH and PH were measured from the soil surface to the insertion of the main ear and to the flag leaf collar, respectively, both on five representative plants within each plot.

2.2. Genotypic dataset

The 49 tropical inbred lines were genotyped using the Affymetrix® Axiom® Maize Genotyping Array of 616,201 SNPs (Unterseer et al. 2014). Genotypic data was quality controlled by first removing all non-mapped SNPs, *loci* with at least one heterozygous, and the SNPs with call rate lower than 0.9. The remaining missing data were imputed based on the algorithms from Beagle 4.0 (Browning and Browning 2016) using the *codeGeno* function from Synbreed R package (Wimmer et al. 2012), restricting the imputation to homozygous genotypes. These high-quality polymorphic SNPs were used to build the genomic matrix for the 906 hybrids, deduced by combining the genotypes of the parental lines. Next, SNPs with minor allele frequency (MAF) < 0.05 were removed from the hybrid's genomic matrix. Finally, we performed a linkage disequilibrium (LD) pruning on the genotype matrix of hybrids, removing markers with pairwise linkage disequilibrium (r^2) greater than 0.9. This step was carried out using the *snpG- dsLDpruning* function of the

SNPRelate R package (Zheng et al. 2012). After these procedures, 34,571 high-quality SNPs were retained for further genomic analyses.

2.3. Phenotypic analyses

Phenotypes were first adjusted using a mixed model to obtain the best linear unbiased estimators (BLUEs) for the hybrids in each N condition within each environment, by using the *lmer4* package in R to fit the following model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is a vector of phenotypic values of hybrids; \mathbf{b} is the fixed effect of checks and hybrids; \mathbf{u} is the random effect of the block within site, where $\mathbf{u} \sim N(0, I\sigma_b^2)$; and $\boldsymbol{\varepsilon}$ is the vector of errors, where $\boldsymbol{\varepsilon} \sim N(0, I\sigma_e^2)$. \mathbf{X} and \mathbf{Z} are the respective incidence matrices.

Based on the adjusted data for IN and LN, the index Harmonic Mean (*HM* - Jafari *et al.*, 2009, Morosini *et al.* 2017, Mendonça *et al.* 2016) was calculated in each replication using the following equation:

$$HM_{ij} = \frac{2 \left(GY_{(IN)ij} \times GY_{(LN)ij} \right)}{GY_{(IN)ij} + GY_{(LN)ij}}$$

where *HM* is the harmonic mean of hybrid *i* in the environment *j*; $GY_{(IN)ij}$ is the adjusted mean of grain yield in the IN condition of hybrid *i* in the environment *j*; and $GY_{(LN)ij}$ is the adjusted mean of grain yield in the LN condition of hybrid *i* in the environment *j*.

Next, we estimated the Pearson's correlation coefficients between the four traits (GY, HM, EH, and PH). Also, for GY, EH, and PH, the significance of nitrogen level effect by each trait was assessed considering the following model in ASReml-R package (Gilmour *et al.* 2009):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{L}\mathbf{n} + \mathbf{V}\mathbf{s} + \mathbf{H}\mathbf{g} + \mathbf{I}\mathbf{gs} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is the vector of the adjusted phenotypic values of hybrids; \mathbf{b} is the vector of general mean; \mathbf{n} is the fixed effect of nitrogen stress; \mathbf{s} the random effect of environment, where $\mathbf{s} \sim N(0, \sigma_s^2 \mathbf{I})$; \mathbf{g} the random effect of hybrids, where $\mathbf{g} \sim N(0, \sigma_g^2 \mathbf{I})$; and \mathbf{gs} is the random effect of genotype \times environment interaction, where $\mathbf{gs} \sim N(0, \sigma_{gs}^2 \mathbf{I})$; and $\boldsymbol{\varepsilon}$ is the vector of errors, where $\boldsymbol{\varepsilon} \sim N(0, \sigma_e^2 \mathbf{I})$. \mathbf{X} , \mathbf{L} , \mathbf{V} , \mathbf{H} , and \mathbf{I} are the respective incidence matrices. Random effects were predicted as the best linear unbiased predictors (BLUPs), and their associated variance components

were obtained using the Maximum Restricted Likelihood (REML) method. Their significance was estimated by the likelihood-ratio test (LRT) at 5%, using the asremlPlus R package (Brien 2019). The Wald test implemented in ASReML-R tested the fixed effects.

2.4. Diallel analyses

The estimation of GCA and SCA was performed by fitting the Griffing's model II – method 4 (Griffing, 1956) for each trait in each N level, according to the equation:

$$\mathbf{y} = X\boldsymbol{\beta} + Z_a\mathbf{a} + Z_d\mathbf{d} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is a vector of hybrid adjusted data for the traits, $\boldsymbol{\beta}$ is the vector of general mean summed to environment effect, considered as fixed; \mathbf{a} is the vector of random effects of GCA, where $\mathbf{a} \sim N(0, \sigma_{GCA}^2 \mathbf{G}_a)$ and σ_{GCA}^2 is the associated variance component; \mathbf{d} is the vector of random effects of SCA, where $\mathbf{d} \sim N(0, \sigma_{SCA}^2 \mathbf{G}_d)$ and σ_{SCA}^2 is the associated variance component; and $\boldsymbol{\varepsilon}$ is the vector of errors, where $\boldsymbol{\varepsilon} \sim N(0, \sigma_e^2 \mathbf{I})$. X , Z_a and Z_d are the incidence matrices for $\boldsymbol{\beta}$, \mathbf{a} and \mathbf{d} , respectively.

Three scenarios for the incorporation of genetic information were considered:

1. Parents non-genetically related (I): the variance-covariance matrices among parents and hybrids are identities, where $\mathbf{a} \sim N(0, \sigma_{GCA}^2 \mathbf{I})$ e $\mathbf{d} \sim N(0, \sigma_{SCA}^2 \mathbf{I})$. In this case, $\mathbf{G}_a = \mathbf{I}_{49}$ and $\mathbf{G}_d = \mathbf{I}_{906}$;
2. *Genomic additive model* (G_a): the variance-covariance matrix for the parent lines accounts for the additive genomic relationship matrix, following the equation:

$$\mathbf{A}_g = \frac{\mathbf{W}_A \mathbf{W}_A'}{2 \sum_i^n p_i (1 - p_i)}$$

where \mathbf{W}_A is the matrix of markers incidence of the individuals, corrected for their averages in each locus ($2p_i$); p_i is the frequency of one allele of the locus i . In this scenario, $\mathbf{G}_a = \mathbf{A}_g$ and $\mathbf{G}_d = \mathbf{I}_{906}$.

3. *Genomic additive + dominance model* (G_{a+d}): both parent lines and hybrids are attributed to genomic relationship matrices. The dominance component is calculated as:

$$\mathbf{D}_g = \frac{\mathbf{W}_D \mathbf{W}_D'}{\sum_i^n \{(2p_i(1 - p_i))\}^2}$$

where W_D is the dominance deviation matrix, and p_i is the frequency of one allele of the locus i . Therefore, in this case, $G_a = A_g$ and $G_d = D_g$.

The incidence matrices W_A and W_D were designed following Vitezica *et al.* (2017), considering two alleles for a given marker locus:

$$W_A = \begin{cases} -(-p_{A_1A_2} - 2p_{A_2A_2}) \\ -(1 - p_{A_1A_2} - 2p_{A_2A_2}) \\ -(2 - p_{A_1A_2} - 2p_{A_2A_2}) \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases}$$

$$W_D = \begin{cases} -\frac{2p_{A_1A_2}p_{A_2A_2}}{p_{A_1A_1}+p_{A_2A_2}-(p_{A_1A_1}-p_{A_2A_2})^2} \\ \frac{4p_{A_1A_1}p_{A_2A_2}}{p_{A_1A_1}+p_{A_2A_2}-(p_{A_1A_1}-p_{A_2A_2})^2} \\ -\frac{2p_{A_1A_1}p_{A_1A_2}}{p_{A_1A_1}+p_{A_2A_2}-(p_{A_1A_1}-p_{A_2A_2})^2} \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases}$$

In order to build the W_A matrix, we used a genotypic incidence (S_A) matrix coded as 2 for homozygote A_1A_1 , 1 for heterozygote A_1A_2 and 0 for homozygote A_2A_2 . For W_D , the genotypic incidence matrix (S_D) was coded as 0 for both homozygotes and 1 to the heterozygote. The heatmaps of the additive and dominance kernels from the hybrid's genomic matrix were obtained using the superheat R package (Barter and Yu 2017) (Figure 1).

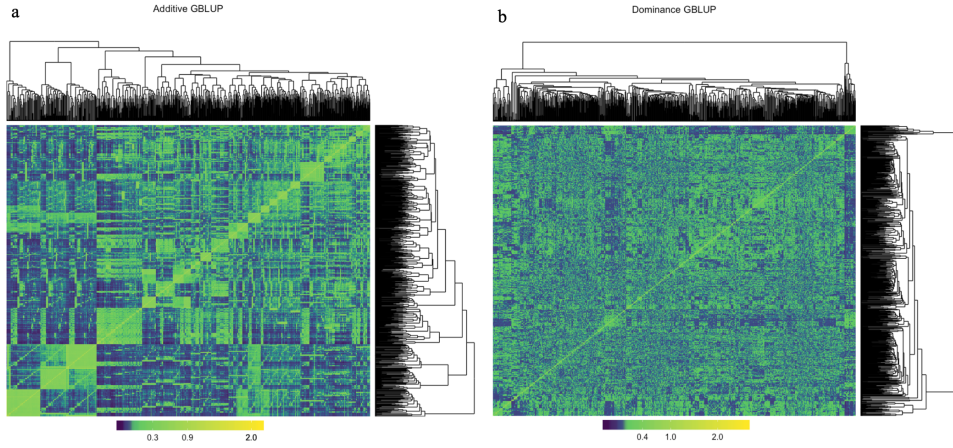


Figure 1. Heatmap of the genomic relationship matrix from 906 maize single crosses. (a) Additive genomic best linear unbiased prediction (GBLUP) matrix, (b) dominance GBLUP matrix.

Variance components, broad-sense genomic heritability (H^2), and narrow-sense genomic heritability (h^2) were obtained for GY, EH, and PH under IN and LN, and for HM, as follows:

$$H^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_a^2 + \sigma_d^2 + \sigma_\varepsilon^2}$$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_\varepsilon^2}$$

where σ_a^2 is the additive genetic variance, σ_d^2 is the dominance genetic variance, and σ_ε^2 is the residual variance. For scenario *I*, in which the coefficient of inbreeding (F) is assumed to be zero, $\sigma_a^2 = 4\sigma_{GCA}^2$ and $\sigma_d^2 = 4\sigma_{SCA}^2$. In all scenarios *I*, G_a , and G_{a+d} , the ponderation of GCA and SCA effects regarding the coefficient of inbreeding (F) is performed by their respective genomic relationship matrix, so the additive and dominance variance components are directly obtained as $\sigma_a^2 = \sigma_{GCA}^2$ and $\sigma_d^2 = \sigma_{SCA}^2$, where σ_{GCA}^2 and σ_{SCA}^2 are the GCA and SCA variances, respectively.

GCA and SCA estimates were obtained individually for each cross, each environment, and each nitrogen level – except for the HM index, which comprises both levels simultaneously. The associated variance components were obtained using the Maximum Restricted Likelihood (REML) method. The significance of the random effects was assessed by the likelihood ratio test (LRT), at 5% probability. Diallel analyses were performed using the ASReml-R package (Butler et al. 2009).

2.5. Model comparisons

In order to assess the reliability of each scenario to identify the best parent lines and then aid to discriminate heterotic groups and testers, we compared the rank of lines and hybrids in each scenario using Spearman's correlation. Since the main trait of interest is the GY and because we are seeking to evaluate the index usefulness, only GY and HM were considered for the comparisons. In this case, BLUPs for GY were obtained from an individual analysis for IN condition, that is, not considering N stress.

The genomic estimated breeding values (GEBV) of the lines were obtained directly from the GCA estimates from the diallel analysis in each scenario. For the hybrids, we constructed *in silico* the genomic estimated value (GEV) using the estimates of GCA and SCA according to the following strategy:

$$GEV_{hs} = GCA_{h1s} + GCA_{h2s} + SCA_{h1s \times h2s}$$

where GEV_{hs} is the genomic estimated value of hybrid h in scenario s , GCA_{h1s} is the general combining ability of hybrid h 's parent 1 in scenario s , GCA_{h2s} is the general combining ability of hybrid h 's parent 2 in scenario s , and $SCA_{h1s \times h2s}$ is the specific combining ability from the cross yielding hybrid h in scenario s .

Next, we selected the best hybrids according to their GEV in each scenario considering the intensity of selection of 10%. Finally, we identified the most frequent lines that appeared as parents of these best hybrids.

2.6. Heterotic grouping and testers

The determination of heterotic groups was performed based on the estimates of SCA for GY not considering N stress. From them, we calculated a Euclidean distance matrix for the 49 maize lines and performed a hierarchical agglomerative clustering by Ward's method (Ward 1963), which produces groups that minimize within-group dispersion at each binary fusion based on a classical sum-of-squares criterion.

Finally, according to the intersection between the hybrids identified as the 10% best for GEV in the three scenarios, the lines yielding higher GCA estimates, and the groups clustered by SCA scores, we evaluated the proportion of coincidence among scenarios and identified the potential testers for the best fit. For that, we considered the intersection of the lines with higher GCA and the most frequent ones as the parents of the best hybrids (according to their GEV).

2.7. Effect of nitrogen stress on heterotic composition

We investigated the effect of N deficiency on the determination of heterotic patterns. Only the G_{a+d} scenario was considered, where both additive and dominance genetic components are present. The clustering was performed based on the SCA estimates.

Based on composition of the pools, we calculated the potential mean SCA for IN and LN considering the crosses intra groups (lines within each group) and intergroups (groups crossed to each other). Finally, we compared the estimates to the mean SCA obtained when no heterotic groups were designated.

2.8. Genomic prediction for scenarios

We used the additive genomic best linear unbiased prediction (GBLUP) model to perform genomic prediction. Predictive ability (r) was calculated as the Pearson correlation between adjusted values and genomic estimated breeding values in 50 replications from independent validation

scenarios (Albrecht et al., 2014), randomly sampling 75% of the genotypes to form a training set, while the remaining 25% were used as a validation set. All prediction analyses were performed using the Bayesian Generalized Linear Regression (BGLR) R package (Perez and de los Campos, 2014).

3. RESULTS

3.1. Phenotypic analysis

There was significant effect of N level on the performance of all traits across environments. Average performance for GY, EH, and PH under LN across all hybrids was 5.09 Mg ha⁻¹, 1.02 m, and 1.96 m, respectively. For IN, significantly higher mean values were observed for these traits (GY 5.66 Mg ha⁻¹, EH 1.06 m, and PH 2.05 m). Considering the four environments, GY and HM outperformed in Anhembi in season 2017 (2AN), where the mean values were 6.96 Mg ha⁻¹ and 6.80, respectively (Figure 2). The traits PH and EH followed the same performance trend across environments, with higher mean values in Piracicaba in 2016.

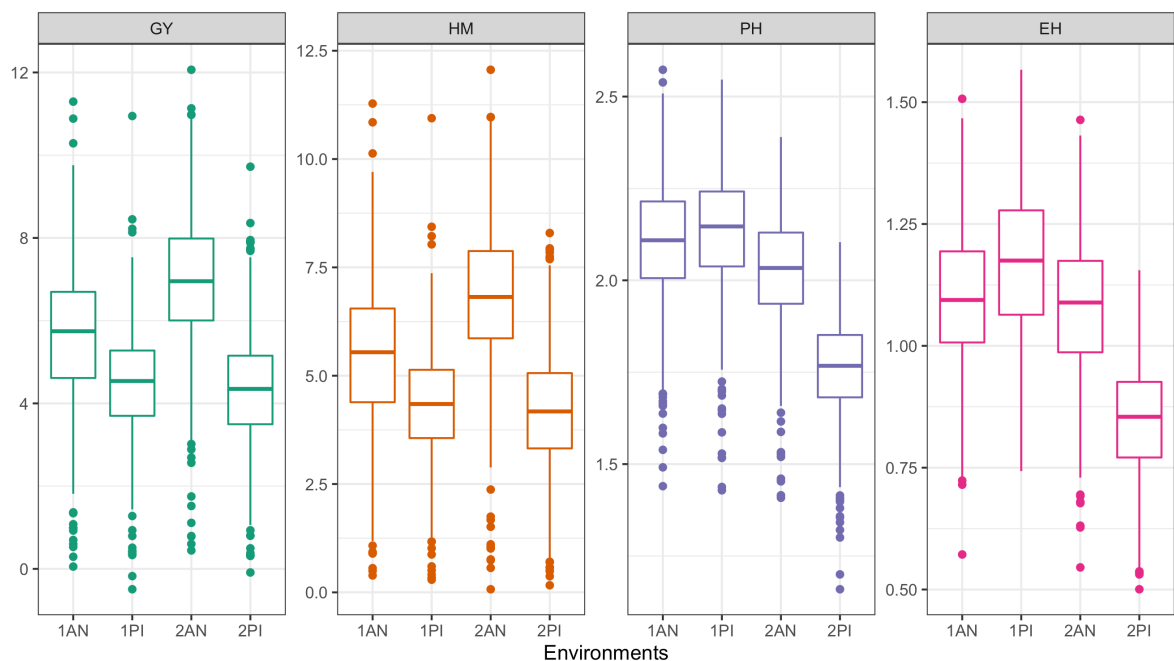


Figure 2. Boxplot of phenotypic traits. Adjusted mean values for grain yield (GY, Mg ha⁻¹), harmonic mean (HM, Mg ha⁻¹), plant height (PH, m), and ear height (EH, m) for 906 maize hybrids in four environments: Anhembi 2016 (1AN), Piracicaba 2016 (1PI), Anhembi 2017 (2AN), and Piracicaba 2017 (2PI).

The Pearson's coefficients were significant between all pairs of traits within and across the N conditions from a moderate positive correlation of 0.17 between GY in LN and EH in IN to a strong positive correlation of 0.90 for HM and GY in LN (Table 1). Interestingly, the index HM was more related to the GY in low nitrogen (0.90) than in the ideal N condition.

Nitrogen levels on the pre-adjusted data had a significant effect for GY, EH, and PH, indicating the differential impact of the N stress on performance (Table 2). Both hybrids and environments (season \times local) were also significantly different for all traits. Genotype \times environment interaction was significant for GY and PH, suggesting differential performance of hybrids between sites for these traits within each N level.

Table 1. Pearson's correlation coefficients between the traits for grain yield (GY), ear height (EH), plant height (PH), and harmonic mean index (HM) for 906 maize hybrids evaluated in four environments and two nitrogen treatments

	GY.IN	GY.LN	PH.IN	PH.LN	EH.IN	EH.LN	HM
GY.IN	-	0.60**	0.48**	0.38**	0.42**	0.30**	0.87**
GY.LN		-	0.19**	0.29**	0.17**	0.21**	0.90**
PH.IN			-	0.79**	0.85**	0.67**	0.35**
PH.LN				-	0.69**	0.82**	0.36**
EH.IN					-	0.77**	0.31**
EH.LN						-	0.28**
HM							-

Significant at ** $P = 0.01$

IN: ideal nitrogen; LN: low nitrogen

Even though the GXE was significant, rank (Spearman) correlation of genotypes between environments was significant, indicating that the interaction is due to scale and change in rank of genotypes within environments. For that reason, analyses were conducted across the four environments within each N level.

Table 2. Joint analysis of 906 tropical maize hybrids evaluated in ideal and low nitrogen levels across four environments for grain yield (GY), ear height (EH), and plant height (PH)

Variation source	GY	EH	PH
<i>Fixed</i>			
N level	300.06**	313.12**	1217.56**
<i>Random</i>			
Hybrid	1360.40**	3419.30**	3265.31**
Env	1931.96**	3085.34**	3389.15**
Hybrid/Env	56.23**	0.10 ^{NS}	23.64*

Significant at * $P = 0.05$; ** $P = 0.01$ by LRT (random effects) and Wald (fixed effects); ^{NS} non-significant

3.2. Diallel analyses

Aiming to determine the potential utility of dominance information to specify the best mechanism to establish heterotic patterns, the scenarios I , G_a , and G_{a+d} were considered. For all three, GCA and SCA components were significant for GY, PH, and EH in both N conditions and for HM (Table 3). The overall estimates were higher in ideal N than in low N for all traits, regardless of the scenario (Table S1, Figure 3, Figure 4). For example, in G_a for IN, scores of σ_{GCA}^2 were 0.42, 7.7×10^{-3} , and 5.4×10^{-3} for GY, PH, and EH, respectively, whereas in LN these values were 0.28, 5.9×10^{-3} , and 4.6×10^{-3} .

Table 3. Likelihood ratio test (LRT) of random effects estimated via REML/BLUP of 906 maize hybrids evaluated across four environments under two nitrogen treatments for grain yield (GY), ear height (EH), plant height (PH), and harmonic mean index (HM) in the genetic scenarios I , G_a , and G_{a+d}

Scenario	Effect	GY		EH		PH		HM
		Ideal N	Low N	Ideal N	Low N	Ideal N	Low N	
I	GCA	668.08**	434.62**	2336.48**	1803.42**	1695.68**	1134.06**	708.80**
	SCA	191.58**	131.49**	229.44**	121.30**	563.32**	279.71**	203.90**
G_a	GCA	692.57**	446.39**	2413.59**	1863.13**	1783.21**	1207.55**	703.50**
	SCA	167.25**	118.35**	173.42**	81.07**	482.57**	214.07**	206.82**
G_{a+d}	GCA	692.57**	446.33**	2413.59**	1863.13**	1783.21**	1207.55**	703.50**
	SCA	222.14**	117.23**	173.12**	65.12**	448.13**	215.58**	243.71**

Significant at * $P = 0.05$; ** $P = 0.01$; ^{NS} non-significant

GCA: general combining ability, SCA: specific combining ability

The additive effect was preponderant compared to the dominance for PH and EH. For instance, a 3.63-fold higher value was observed for the additive component EH in LN in scenario G_{a+d} (Figure 3a) compared to the dominance. Consequently, the broad and narrow-sense heritability estimates were higher for these traits. In G_{a+d} IN for instance, estimates of H^2 were 0.32, 0.58, and 0.50 for GY, PH, and EH, respectively, whereas h^2 output 0.18, 0.37, and 0.37 (Table 2, Figure 4a). Conversely, HM exhibited narrow-sense heritability values of 0.15 in scenario I , 0.16 in G_a , and 0.20 in G_{a+d} , which are very close to the mean ones observed for GY (Figure 3b).

The incorporation of genomic information to account for the relationship among individuals enabled more realistic genomic variance estimates. From scenario I to G_a , there was an increment of additive genetic variance for all traits excepted EH, and so from G_a to G_{a+d} (Figure 3). The lowest estimates of dominance effect were obtained in scenario G_{a+d} , following the order

$G_{a+d} < I \leq G_a$. That is, a more significant relative contribution of additive effect was observed from I to G_a and from G_a to G_{a+d} which led to an increase in narrow-sense heritability estimates along the scenarios, except for EH (Figure 4).

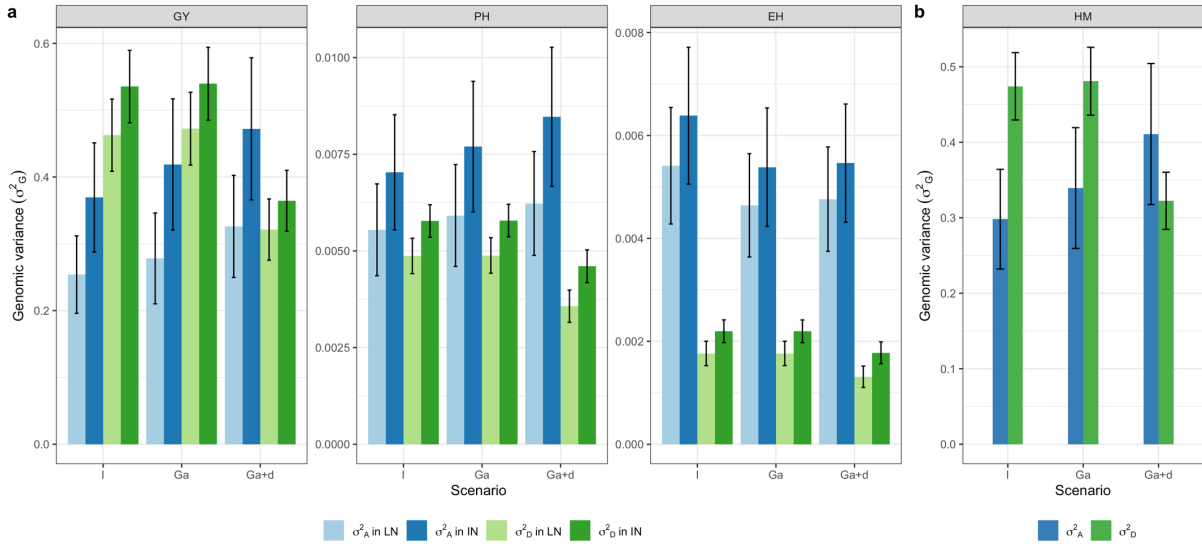


Figure 3. Genomic variance components from 906 maize hybrids. Traits a) grain yield (GY), plant height (PH), and ear height (EH) in ideal and low N condition, and b) harmonic mean (HM) evaluated in three genetic scenarios.

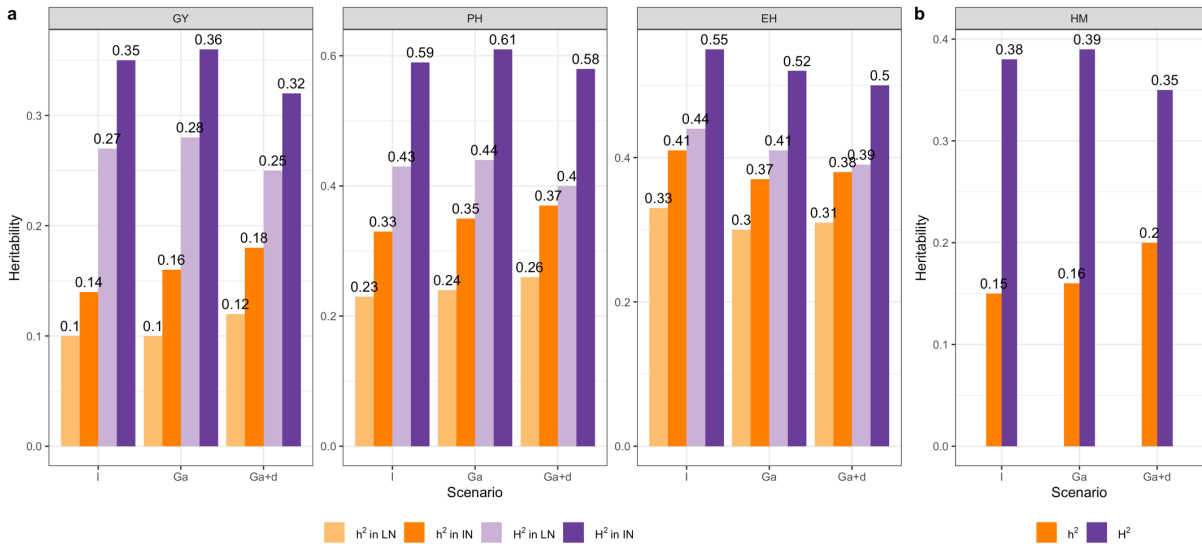


Figure 4. Narrow-sense and broad-sense heritability estimates in 906 maize hybrids. Traits a) grain yield (GY), plant height (PH), and ear height (EH) in ideal and low N condition, and b) harmonic mean (HM) evaluated in three genetic scenarios.

The distribution of the BLUP scores for GCA and SCA effects varied slightly among scenarios and between N condition within traits. Through the Pearson's correlation coefficients, it is possible to observe that the main differences come from the contrast of traits rather than across scenarios or N levels within each trait, especially for GCA. For instance, considering the SCA effects, while 0.51 was the lowest correlation within GY between G_a and G_{a+d} , the minimum coefficient contrasting GY and PH was 0.32.

3.3. Model comparisons

The ranking of lines based on their GEBVs (Table S2) in scenario *I* was statistically positively correlated with the ranking in scenario G_a for GY (0.991) and HM (0.985 - Figure 5a,b). The same pattern was observed contrasting the rank from G_a to G_{a+d} , which Spearman's correlation coefficient was 0.996 for GY and 0.986 for HM (Figure 5c,d).

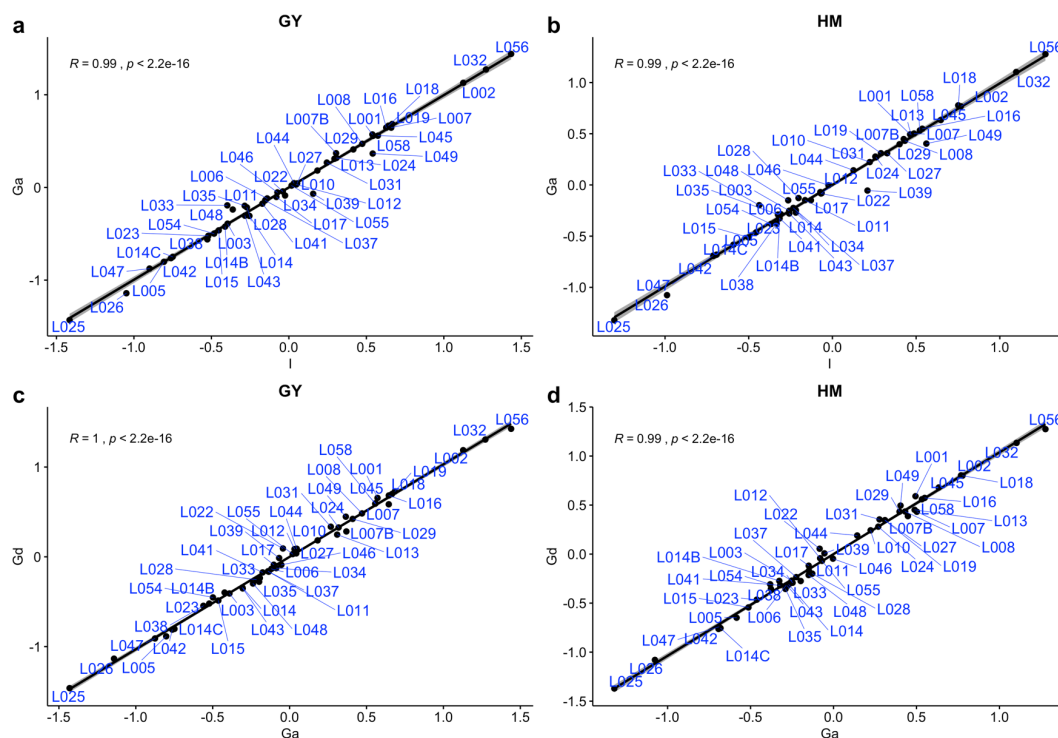


Figure 5. Spearman's correlation coefficients of general combining abilities (GCA) of 49 maize parent lines between scenarios $I-G_a$ and G_a-G_{a+d} for Grain Yield (GY – a, c) and the index Harmonic Mean (HM – b, d).

Seven of the best parental lines (L002, L016, L018, L032, L045, L056, and L058) appeared consistently across traits and scenarios (Figure 6). Lines L032 and L056 were the top-ranked genotypes across all scenarios.

The computed GEV was highly correlated between scenarios for both GY and HM. For instance, the lowest Spearman's correlation coefficient was 0.958 between G_a and G_{a+d} for HM. Nevertheless, we observed an interesting change of ranking when focusing on the hybrids with the lowest GEV estimates, mainly from G_a to G_{a+d} .

The group of hybrids selected as the best 10% (91 genotypes out of 906) according to the GEV had mean scores for GY of 1.760, 1.757, and 1.801 in I , G_a , and G_{a+d} , respectively (Table 4). They performed 397-fold higher in I , 499-fold higher in G_a , and 36-fold higher in G_{a+d} for GY comparing these estimates to the mean GEV of the whole panel of 906 hybrids. In total, the combination of the 91 best-ranked genotypes for each of the three scenario raised 103 hybrids as the best (Figure 7a). From them, 79 hybrids were common to all the three models, which indicates a reliable performance regardless of the scenario (Table S3). The lines that appeared more frequently as parents of the selected genotypes were L056, L032, and L002. Respectively, L056 and L032 are progenitor of 27, 22, and 17 hybrids for scenarios I , G_a , and G_{a+d} , while L002 is progenitor of 29, 21, and 19 hybrids (Table S4).

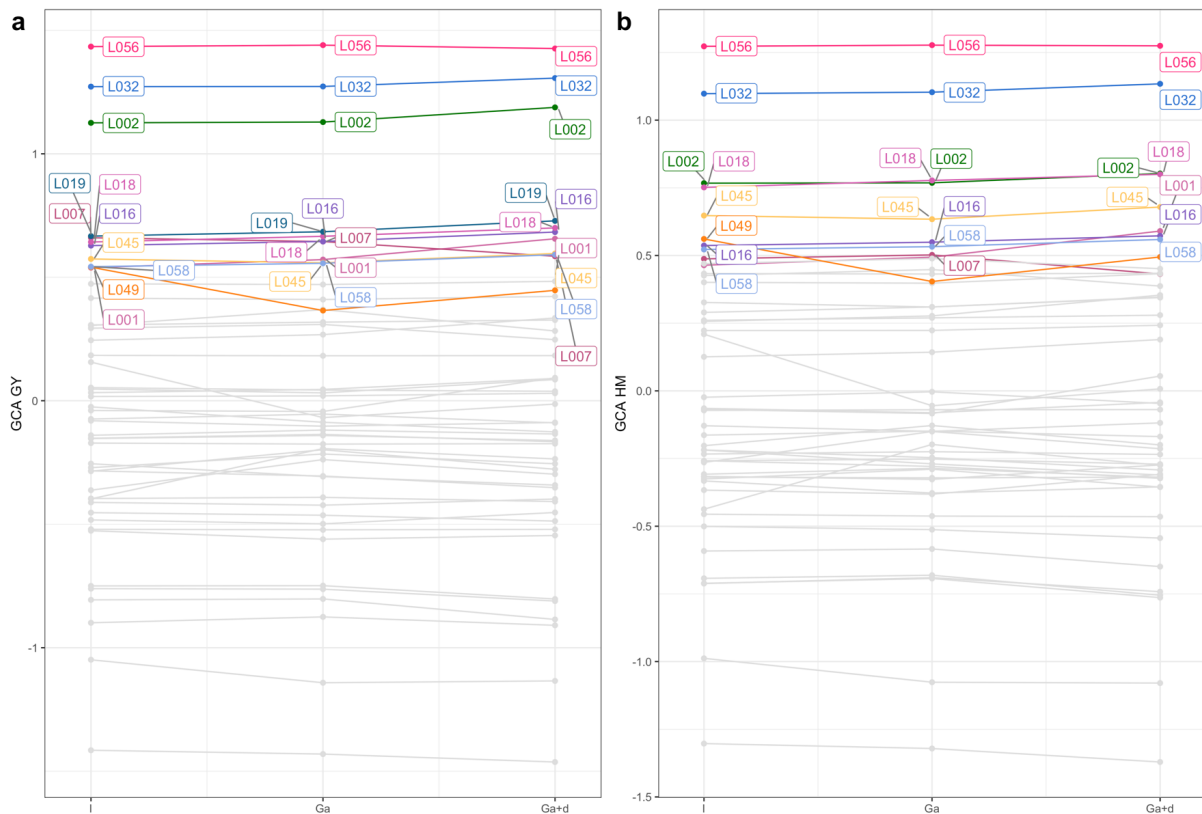


Figure 6. Ranking of 49 maize lines based on the genomic estimated breeding value (GEBV). Traits a) Grain yield (GY) and b) harmonic mean (HM) considered in three genetic scenarios.

Table 4. Statistics on the 10% (91) best hybrids from a population of 906 single crosses evaluated across four environments under two nitrogen treatments according to the genetic estimated value (GEV) scores for grain yield (GY) and a harmonic mean (HM) in the genetic scenarios I , G_a , and G_{a+d}

<i>Trait</i>	<i>Model</i>	<i>Mean</i>	<i>Var</i>	<i>CV (%)</i>
GY	I	1.760	0.200	11.4
	G_a	1.757	0.202	11.5
	G_{a+d}	1.810	0.177	9.78
HM	I	1.581	0.130	8.20
	G_a	1.581	0.130	8.20
	G_{a+d}	1.592	0.144	9.06

The same trend was observed for HM. The mean GEV scores considering the 10% best hybrids for the index were 1.581, 1.581, and 1.591 respectively for I , G_a , and G_{a+d} , respectively (Table 4). Contrasting these values to the mean GEV of the whole panel, the selected single-crosses were 516-fold, 665-fold, and 35-fold superior for I , G_a , and G_{a+d} , respectively. A total of 106 hybrids were identified among the 10% best for the three scenarios, being 77 common to all of them (Figure 7b). As we noted for GY, the lines present in the best crosses were L056, L032, and L002 (Table S5).

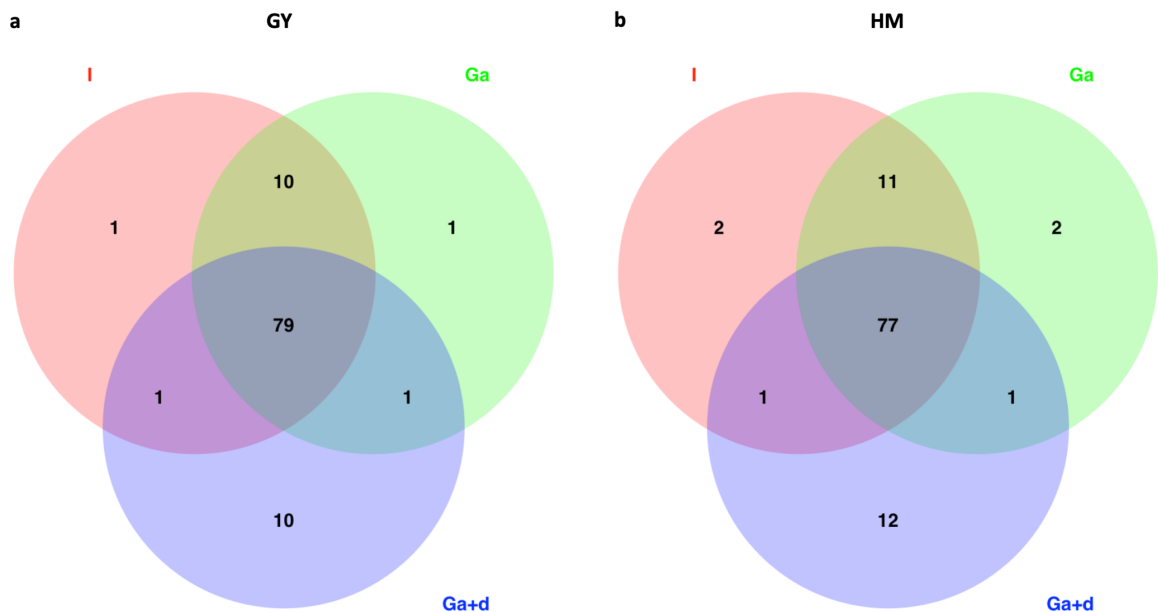


Figure 7. The coincidence of the selected hybrids in each genetic scenario (I , G_a , and G_{a+d}). Traits a) grain yield (GY) and b) harmonic mean index (HM).

3.4. Heterotic groups

The SCA estimates of GY in scenario I strongly correlated with the estimates in G_a . On the other hand, the ranking correlation based on SCA between G_a and G_{a+d} was 0.83, the lowest observed, although this coefficient is still very high. Accordingly, we observed different patterns of best crosses from G_a to G_{a+d} .

We identified two heterotic groups, G1 and G2, in each scenario (Fig. 8). G1 was composed of 30, 30, and 28 inbred lines in I , G_a and G_{a+d} , respectively, and G2 contained 19, 19, and 21 individuals (Figure 8). Twenty-five lines were in common in G1 between I and G_a and 16 between G_a and G_{a+d} , indicating a considerable overlap in the composition according to these scenarios. Contrasting G_a to G_{a+d} , 18 and 7 genotypes were in common for G1 and G2, respectively.

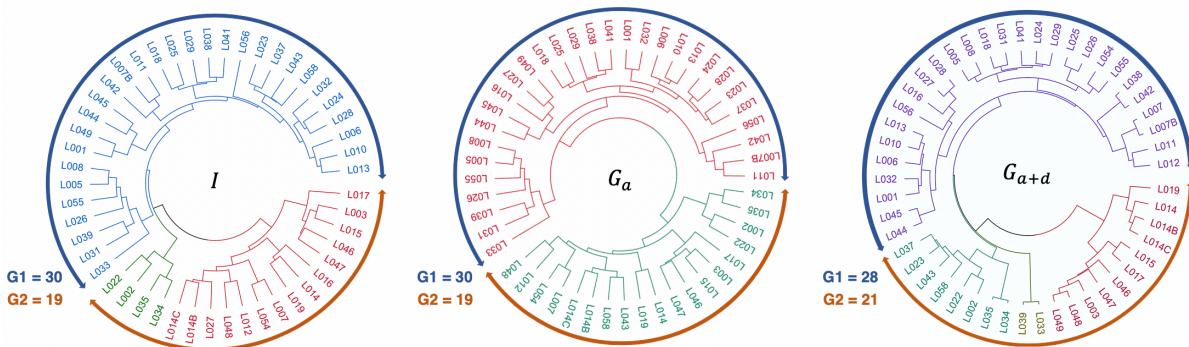


Figure 8. Dendrograms of 49 maize tropical lines for the genetic scenarios I , G_a , and G_{a+d} indicating the two heterotic groups designed (blue and orange circles) in each of them

In scenario G_{a+d} , the mean GEV from the hybrids of crosses produced between G1 and G2 was significantly superior to estimates in I and G_a . By combining information about the best lines identified by GCA, GEV, and their distribution on the suggested heterotic groups within the additive+dominance modeling scenario, we identified potential testers to be considered in this panel.

3.5. Nitrogen effect on heterotic pools

The composition of heterotic groups changed considerably between IN and LN (43% and 73% of coincidence for groups 1 and 2, respectively), and the proportion of lines in each

cluster remained similar (Figure 9). For IN, the best lines concerning GCA were designated to different groups, while for LN they were clustered in the same set.

We could observe that the progenitors of the best hybrid according to the GEV are clustered in different heterotic groups for both N conditions (Figure 9). On the other hand, parents of the lowest-GEV hybrid are assigned to the same heterotic group. This is an interesting evidence on how the accurate allocation of genotypes into heterotic pools plays an important role for exploring the potential genetic value towards heterosis.

Based on the pool's composition, the potential mean SCA considering crosses inter groups was 0.170 for LN and 0.173 for IN. When no heterotic group was assigned, the mean SCA accounting for all possible crosses was 0.039 for LN and 0.052 for IN. Interestingly, by considering the heterotic groups assigned in LN to calculate the mean SCA for inter groups crossing in IN, the output decreases drastically from 0.173 to 0.90, which indicates the important underestimation of genetic effects that stressful conditions may cause.

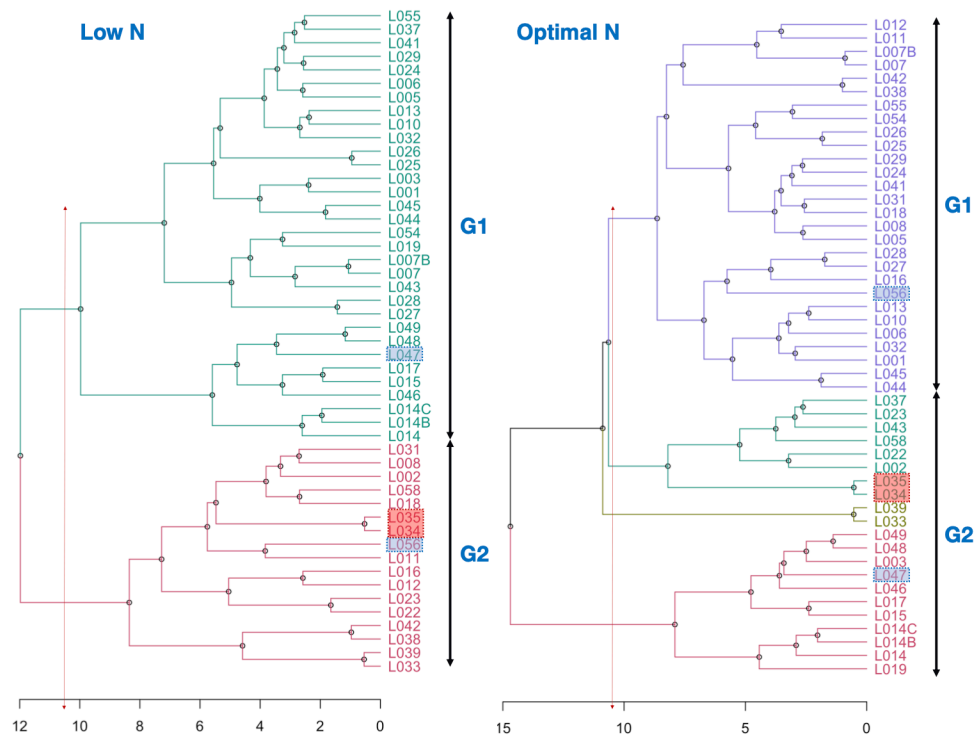


Figure 9. Dendrograms of 49 tropical maize lines constructed in the genetic scenario G_{a+d} for two nitrogen treatments, low (LN) and optimal supply, with indication of the two heterotic groups assigned (G1 and G2). Genotypes colored in blue box are the parents of the best hybrid according to their GEV; genotypes colored in red box are the parents of the worst hybrid according to their GEV.

3.6. Genomic prediction

Based on the assessment of the models described and assessed above, we selected the third one (additive+dominance) as the best for discriminating heterotic groups and exploring the genetic effects. Next, we considered three scenarios within a reciprocal recurrent selection (RRS) method context: additive using a single population (A), additive+dominance using a single population (A+D), and additive+dominance using a North Carolina II (NC) factorial design.

The heterotic groups used in the NC scenario were assigned based on the specific combining ability (SCA) estimates from the A+D model. The A model provided smaller prediction ability compared to A+D, which corroborates the importance of modeling the contribution of non-additive effects towards heterosis. NC provided a significantly increased SCA of the single crosses compared to A+D, which corroborates the importance of the genetic mating design towards the exploration of heterosis. The assignment of heterotic groups is determinant for exploring non-additive effects from a statistical standpoint, which allows to assess more accurate genetic values from both lines and hybrids. We also observed a noteworthy increase in PA (6%) from A+D model to NC.

4. DISCUSSION

4.1. Phenotypic data and diallel analysis

Phenotypic values varied significantly between low and ideal N conditions for all traits (Table 2), with reduced performance in the stress condition. These results are in accordance with previous findings (Betrán et al. 2003; Ning et al. 2018; LI et al. 2019). Nitrogen is a crucial nutrient required by plants for critical functions including protein synthesis, nucleic acids composition, photosynthesis, and carbohydrate production. Therefore, the amount of N assimilated and distributed has a large impact on plant metabolism and root-shoot growth (Bänziger et al. 2002; Morosini et al. 2017; Perchlik and Tegeder 2018).

As reported in the literature, EH and PH are highly positively correlated (Peiffer et al. 2014; Bennetzen et al. 2018), corroborating the Pearson's coefficients of 0.85 in IN and 0.82 in LN we found in this study (Table 1). The index HM is mathematically connected to GY ($HM = \frac{2(GY_{INx} * GY_{LNx})}{GY_{INx} + GY_{LNx}} = \frac{(GY_{INx} * GY_{LNx})}{GY_x}$), which explains the high correlation among them. We could also observe that the correlation between GY and EH decreased substantially in LN compared to IN, yielding a correlation of 0.21 (half of the coefficient obtained in IN). The same trend was found analyzing GY and PH.

In all three scenarios I , G_a , and G_{a+d} , GY presented substantial variation of additive effect from LN to IN and small variation of dominance effect between stressed and non-stressed conditions (Figure 3). On the other hand, PH and EH had a significant difference for the dominance variance between LN to IN, but not for the additive effect. These observations reinforce that the more complex the trait is, the greater the influence of environmental oscillations on genetic estimates tend to be, which are driven by N treatment in this case (Bennetzen et al. 2018; Bernardo 2020). It has been documented that GY presents lower heritability under abiotic stress conditions (Venuprasad et al. 2007; Bankole et al. 2017) and since it has a higher level of complexity compared to PH and EH, the impact of N stress on the additive component was higher than in the dominance one for this trait.

The estimates of narrow-sense genomic heritability increased from scenario I to G_a and G_a to G_{a+d} for all traits except EH (Figure 4, Table S1). Overall, the addition of only additive information in the model did not have a relevant impact on genetic estimates compared to a scenario that did not consider any genomic relationship among parents. On the other hand, when combined with a dominance matrix in scenario G_{a+d} , there was a significant reduction of the

dominance variance compared to G_a . Therefore, by including only the additive information, we cannot separate the additive and dominance genetic components entirely because it accounts for part of the dominance effect itself. That is, the dominance is lumped into the additive component when not included. Consequently, there is higher efficiency in capitalizing dominance variance by adding the dominance genomic relationship information into the model. In addition, it provides better discriminations of the residual. These results suggest that the incorporation of both additive and dominance effects increases the power for assessing the true breeding value of hybrids. It is important to highlight that the model proposed assumed no epistasis and since we have phenotypic information only for the hybrids, the additive component may be overestimated by the additive-by-additive interaction.

HM performed similarly to GY in IN. By its mathematical composition, this index identifies stress tolerant individuals as it favors genotypes that perform well in both conditions (stressed and non-stressed). The HM score can be useful for plant breeders as a tool for early selection and as an alternative to genome wide prediction. Since unfavorable conditions can affect genomic heritability and reduce estimation accuracy, performing predictions using phenotypic values that account for both conditions is an effective method of eliminating the effects of error on prediction and maximizing the amount of genetic information (Crossa et al. 2010; Ziyomo and Bernardo 2013; Wang et al. 2017).

4.2. Model comparison

The significant positive correlation coefficients for the ranking of lines indicate the order of parental maize lines barely changed among genetic scenarios for GY and HM (Figure 6). Furthermore, the same lines are reported as the best for both traits. The breeding value (BV) of an individual is the heritable portion of its genotypic value and it is due to additive effects (Bernardo 2020). The maintenance of the same lines as the best between scenario I and G_a suggest the pedigree deduced intrinsically in the diallel analysis was sufficient to capture their relationship. Therefore, the inclusion of the additive incidence matrix added little information regarding BV. This suggests that the population described here does not present a clear structure since the first-order relationship captured by the diallel scheme was sufficient to discriminate the additive effect. The incorporation of dominance deviations into the model (from G_a to G_{a+d}) did not provide changes in the ranking of lines since the GCA is due to the additive genetic component. These

findings strongly corroborate the importance of the additive effects as drivers for selecting inbred lines used in hybrids combinations and indicate the consistency of the analyses for GY and HM.

Interestingly, both GY and HM identified the same lines as the best according to GCA. Previous studies have reported the usefulness of HM as a discriminator of tolerant genotypes (Jafari et al. 2009; Mendonça et al. 2017; Zhao et al. 2019). Considering that GY and HM are highly correlated (0.99), a potential explanation for their similar ranking is that the majority of the genotypes present stability type 2, that is, they are responsive to the N increment response and their response is parallel to the mean response of all genotypes in the panel (Lin et al. 1986; Fasahat et al. 2015) (Figure 6). Furthermore, once HM discriminates the most tolerant and high yielding genotypes simultaneously, it is possible to infer the lines identified as the best are also the most tolerant ones for N stress. This evidence is in accordance with the origin of this panel, which was constructed through divergent selection regarding N use efficiency (Lanes et al. 2014; Morosini et al. 2017).

The ranking of hybrids by their GEV was highly similar within traits (comparing the scenarios) and between traits. However, we noticed an interesting change in the positions when focusing on the worst genotypes. This is a crucial feature to define the strategy of a breeding program. For instance, if the goal is to select the best hybrids, then the same individuals would be assigned regardless of the model. However, if the aim is to eliminate the worst genotypes, it must be done according to the specific genetic scenario adopted, otherwise the final results may be compromised. The strategy must also consider the genetic diversity the breeder wants to maintain in the population that remains after selection in each scenario.

The GEVs were higher when we incorporated the dominance information in the model (data not shown), and the increment in this parameter comparing the mean GEV value of the 10% selected individuals to the mean GEV of the total population (906 hybrids) was lower for G_{a+d} . For scenarios I and G_a , the mean GEV scores were lower than in G_{a+d} because we are assuming the hybrids are independent of each other from a genomic perspective since the only information on relatedness that can be captured is from the intrinsic pedigree deducted from the diallel analysis. Thus, the mean SCA in scenarios I and G_a tends to zero. On the other hand, considering the dominance matrix incidence for the hybrids in the model, the effects are no longer independent, which reflects directly in the SCA estimates. Since we did not model the scenarios I and G_a with dominance effects, the SCA estimates tend to be spurious in them, because the only source for effect discrimination is the genetic additive information. By that, we are not considering the mendelian sampling. In scenario G_{a+d} , the dominance incidence matrix allows us to assess

variation within crosses from the same parents due to SCA, therefore it is more accurate than the models without dominance.

The mean GEV of the 10% best hybrids were similar between scenarios within each trait, which corroborates the solid performance guided by the additive effects captured in the pedigree deduced in scenario *I*. The lines identified as the most frequent parents of the best hybrids based on GEV for GY and HM had substantial overlap. This result indicates the best parents are persistently identified throughout the scenarios and also report for tolerance to N stress, as previously mentioned. Among the best hybrids for GY and GEV, are L019xL056, L019xL032, L002xL056, L002xL013, L002xL045, and L001xL056.

4.3. Heterotic patterns

By using SCA estimates to identify structure in the panel, we clustered based on the difference, and potentially complementarity, of alleles between groups. Therefore, the genetic distance between them is maximized. The composition of heterotic groups changed slightly from scenarios *I* to G_a , and more substantially from G_a to G_{a+d} . Naturally, the determination of heterotic groups occurs as a function of the variance: since no significant increment was observed for the SCA from scenario *I* to G_a , both scenarios yielded similar heterotic patterns. The short initial branches of the phylograms indicated that the population is not divergent enough. Therefore, the best strategy in a first stage of allocation and selection of hybrids relies on SCA.

A well-established, long-term maize breeding program present distinct and genetically distant heterotic patterns and defined testers, which tend to provide a higher exploration of heterosis and maintenance of genetic variability through cycles of reciprocal recurrent selection. In this case, selection of hybrids based on GEV would be advantageous because the genetic distance between groups is high and SCA accounts for a greater proportion of the total genetic value. On the other hand, in initial maize breeding programs where heterotic groups are not well defined, SCA tends to represent just a small and constant portion of GEV, which mainly diverges from one individual to another due to the additive effects (GCAs of both parents). Therefore, the selection and the assignment of hybrids to heterotic groups should occur based on the SCA estimates towards genetic variability within the population and genetic distance between groups.

There were three potential clusters for scenarios *I* and G_a , and four in G_{a+d} . Nevertheless, we considered the determination of two heterotic groups in all scenarios. First, our panel is relatively small (49 lines) and does not present any clear structure, therefore a smaller number of

heterotic groups will be a more practical approach. Second, it is important to have heterotic groups with a similar number of individuals, especially in the initial stages and with limited number of parental lines. This strategy allows the breeder to deal with a greater population size and to avoid severe bottlenecks potentially induced by genetic drift.

We observed a considerable overlap among compositions of heterotic groups resulted from modeling scenario $I-G_a$ and G_{a+d} . Nevertheless, the dominance component in the last scenario allowed to reach higher heterosis, since the mean GEV from the hybrids of crosses produced between G1 and G2 was superior to estimates in I and G_a . Finally, the coincidence of best lines according to the parameters GCA, GEV, and SCA indicated the lines L056, L032, and L008 would be good testers.

N stress affected significantly the designation of heterotic groups (Figure 9). Since the environmental effects tend to be higher in stressful conditions, the genetic effects can be underestimated. We found an important decrease in the potential SCA when the clustering was performed based on LN data, which indicates that breeding trials conducted and evaluated in stressful conditions may lead to misleading results, negatively affecting mid-term and long-term steps in a breeding program. Therefore, the N stress condition seems to compromise the exploration of heterosis, the main goal in hybrid breeding, and must be appropriately addressed for studies involving N use efficiency.

5. 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 assignment 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 expression of heterosis compared to the groups defined using only pedigree/additive information, since the genetic distance between groups is increased. 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
- Bankole F, Menkir A, Olaoye G, et al (2017) Genetic Gains in Yield and Yield Related Traits under Drought Stress and Favorable Environments in a Maize Population Improved Using Marker Assisted Recurrent Selection. *Front Plant Sci* 8:808. doi: 10.3389/fpls.2017.00808
- Bänziger M, Edmeades GO, Lafitte HR (2002) Physiological mechanisms contributing to the increased N stress tolerance of tropical maize selected for drought tolerance. In: *Field Crops Research*. pp 223–233
- Barter R, Yu B (2017) superheat: A Graphical Tool for Exploring Complex Datasets Using Heatmaps. R Packag version 010 <https://cran.r-project.org/package=superheat>
- Bennetzen JL, Flint-Garcia S, Hirsch C, Tuberosa R (2018) *The maize genome*. Springer, Switzerland
- Bernardo R (2020) *Breeding for quantitative traits in plants*, 3rd ed. Stemma Press, Woodbury, Minnesota
- Betrán FJ, Beck D, Bänziger M, Edmeades GO (2003) Genetic analysis of inbred and hybrid grain yield under stress and nonstress environments in tropical maize. *Crop Sci* 43:807–817. doi: 10.2135/cropsci2003.0807
- 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
- Brien C (2019) asremlPlus: Augments “ASReML-R” in Fitting Mixed Models and Packages Generally in Exploring Prediction Differences. R Packag version 41-26
- Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. *Am J Hum Genet* 98:116–126. doi: 10.1016/j.ajhg.2015.11.020
- 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, De Los Campos G, Pérez P, et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724. doi: 10.1534/genetics.110.118521

- 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
- 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
- Fasahat P, Rajabi A, Bagher Mahmoudi S, et al (2015) An overview on the use of stability parameters in plant breeding. *Biometrics Biostat Int J* 2:149–159. doi: 10.15406/bbij.2015.02.00043
- 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
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) *ASReml user guide release 3.0*, VSN Intern. Hemel Hempstead, UK
- 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
- Jafari A, Paknejad F, Al-Ahmadi J (2009) Evaluation of selection indices for drought tolerance of corn (*Zea mays* L.) hybrids. *Int J Plant Prod* 3:1735–6814
- 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

- Lanes ECM, Viana JMS, Paes GP, et al (2014) Population structure and genetic diversity of maize inbreds derived from tropical hybrids. *Genet Mol Res* 13:7365–7376. doi: 10.4238/2014.September.12.2
- 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
- LI X ling, GUO L guo, ZHOU B yuan, et al (2019) Characterization of low-N responses in maize (*Zea mays* L.) cultivars with contrasting nitrogen use efficiency in the North China Plain. *J Integr Agric* 18:2141–2152. doi: 10.1016/S2095-3119(19)62597-9
- Lin CS, Binns MR, Lefkovitch LP (1986) Stability Analysis: Where Do We Stand?1. *Crop Sci* 26:894–900. doi: 10.2135/cropsci1986.0011183x002600050012x
- 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
- Mendonça LDF, Granato ÍSC, Alves FC, et al (2017) Accuracy and simultaneous selection gains for N-stress tolerance and N-use efficiency in maize tropical lines. *Sci Agric* 74:481–488. doi: 10.1590/1678-992x-2016-0313
- Mendonca LF, Fritsche-Neto R, Granato ISC, Alves FC (2016) Accuracy and simultaneous selection gains for grain yield and earliness in tropical maize lines. *Maydica* 1–7
- 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
- Ning P, Yang L, Li C, Fritschi FB (2018) Post-silking carbon partitioning under nitrogen deficiency revealed sink limitation of grain yield in maize. *J Exp Bot* 69:1707–1719. doi: 10.1093/jxb/erx496
- 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
- Olímpio Das Graças Dias K, 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

- 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
- Peiffer JA, Romay MC, Gore MA, et al (2014) The genetic architecture of maize height. *Genetics* 196:1337–1356. doi: 10.1534/genetics.113.159152
- Perchlik M, Tegeder M (2018) Leaf amino acid supply affects photosynthetic and plant nitrogen use efficiency under nitrogen stress. *Plant Physiol* 178:174–188. doi: 10.1104/pp.18.00597
- 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
- Unterseer S, Bauer E, Haberer G, et al (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15:823. doi: 10.1186/1471-2164-15-823
- VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *J Dairy Sci* 91:4414–4423. doi: 10.3168/jds.2007-0980
- Venuprasad R, Lafitte HR, Atlin GN (2007) Response to direct selection for grain yield under drought stress in rice. *Crop Sci* 47:285–293. doi: 10.2135/cropsci2006.03.0181
- Vitezica ZG, Legarra A, Toro MA, Varona L (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. *Genetics* 206:1297–1307. doi: 10.1534/genetics.116.199406
- 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
- Wang X, Li L, Yang Z, et al (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity (Edinb)* 118:302–310. doi: 10.1038/hdy.2016.87
- 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
- Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* 58:236–244. doi: 10.1080/01621459.1963.10500845
- Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28:2086–7. doi: 10.1093/bioinformatics/bts335
- 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
- Zhao Z, He K, Feng Z, et al (2019) Evaluation of Yield-Based Low Nitrogen Tolerance Indices for Screening Maize (*Zea mays* L.) Inbred Lines. *Agronomy* 9:240. doi: 10.3390/agronomy9050240
- Zheng X, Levine D, Shen J, et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328
- Ziyomo C, Bernardo R (2013) Drought tolerance in maize: Indirect selection through secondary traits versus genomewide selection. *Crop Sci* 53:1269–1275. doi: 10.2135/cropsci2012.11.0651

SUPPLEMENTARY MATERIAL

Table S1. Variance components of 906 maize hybrids evaluated across four environments under two nitrogen treatments for grain yield (GY), ear height (EH), plant height (PH), and harmonic mean index (HM) in the genetic scenarios I , G_a , and G_{a+d}

Scenario	Trait	N level	σ_{GCA}^2	σ_{SCA}^2	$\sigma_{SCA}^2/\sigma_{GCA}^2$	σ_{ϵ}^2	h^2	H^2	
I	GY	Ideal	0.3694	0.5353	1.45	1.7063	0.14	0.35	
		Low	0.2539	0.4626	1.82	1.9101	0.09	0.27	
	EH	Ideal	0.0064	0.0022	0.34	0.0071	0.41	0.55	
		Low	0.0054	0.0018	0.33	0.0092	0.33	0.44	
	PH	Ideal	0.0070	0.0058	0.82	0.0088	0.32	0.59	
		Low	0.0055	0.0049	0.88	0.0138	0.23	0.43	
	HM	-	0.2981	0.4741	1.59	1.2693	0.15	0.38	
	G_a	GY	Ideal	0.4188	0.5396	1.29	1.7071	0.16	0.36
			Low	0.2782	0.4723	1.70	1.9093	0.10	0.28
		EH	Ideal	0.0054	0.0022	0.41	0.0071	0.37	0.52
Low			0.0046	0.0018	0.38	0.0092	0.30	0.41	
PH		Ideal	0.0077	0.0058	0.75	0.0088	0.35	0.61	
		Low	0.0059	0.0049	0.83	0.0138	0.24	0.44	
HM		-	0.3394	0.4809	1.42	1.2696	0.16	0.39	
G_{a+d}		GY	Ideal	0.4721	0.3644	0.77	1.7689	0.18	0.32
			Low	0.3260	0.3213	0.99	1.9941	0.12	0.25
		EH	Ideal	0.0055	0.0018	0.33	0.0073	0.37	0.50
	Low		0.0048	0.0013	0.28	0.0095	0.30	0.39	
	PH	Ideal	0.0085	0.0046	0.54	0.0096	0.37	0.58	
		Low	0.0062	0.0036	0.57	0.0146	0.26	0.40	
	HM	-	0.4109	0.3225	0.78	1.3368	0.20	0.35	

Additive (σ_{GCA}^2), dominance (σ_{SCA}^2), error (σ_{ϵ}^2) variance, strict-sense heritability (h^2), and broad-sense heritability (H^2)

Table S2. Estimates of general combining ability (GCA) of 49 maize parent lines assessed via heir progeny of 906 hybrids, which were evaluated in four environments for grain yield (GY) and the index harmonic mean (HM) in the genetic scenarios I , G_a , and G_{a+d}

Line	GY			HM		
	I	G_a	G_{a+d}	I	G_a	G_{a+d}
L001	0.5390	0.5717	0.6560	0.4641	0.4940	0.5903
L002	1.1254	1.1286	1.1880	0.7673	0.7684	0.8029
L003	-0.3957	-0.3910	-0.4082	-0.2581	-0.2517	-0.2708
L005	-0.8063	-0.8024	-0.8855	-0.5916	-0.5840	-0.6492
L006	-0.1399	-0.1186	-0.1346	-0.3079	-0.2860	-0.3195
L007	0.6603	0.6447	0.5852	0.4876	0.5022	0.4318
L007B	0.3046	0.3691	0.2828	0.4259	0.4480	0.3871
L008	0.4735	0.4707	0.4838	0.4332	0.4317	0.4344
L010	0.1840	0.1821	0.1830	0.2230	0.2234	0.2424
L011	-0.1517	-0.1360	-0.1660	-0.1636	-0.1485	-0.1689
L012	0.0476	0.0320	0.0860	-0.0645	-0.0807	-0.0425
L013	0.2940	0.3092	0.2475	0.4700	0.4889	0.4511
L014	-0.2547	-0.3070	-0.3416	-0.2187	-0.2697	-0.3108
L014B	-0.4115	-0.4228	-0.3980	-0.3158	-0.3270	-0.2735
L014C	-0.7497	-0.7487	-0.8028	-0.6924	-0.6810	-0.7549
L015	-0.4530	-0.4635	-0.4872	-0.5006	-0.5126	-0.5438
L016	0.6283	0.6451	0.6832	0.5368	0.5492	0.5726
L017	-0.0808	-0.1028	-0.0880	-0.1287	-0.1495	-0.1182
L018	0.6429	0.6659	0.7001	0.7517	0.7772	0.7999
L019	0.6667	0.6845	0.7285	0.2898	0.3100	0.3446
L022	-0.0394	-0.0433	0.0929	-0.0682	-0.0836	0.0547
L023	-0.5204	-0.5226	-0.5207	-0.4555	-0.4624	-0.4647
L024	0.3070	0.3184	0.3269	0.2606	0.2697	0.2798
L025	-1.4152	-1.4306	-1.4628	-1.3029	-1.3210	-1.3710
L026	-1.0487	-1.1418	-1.1344	-0.9880	-1.0762	-1.0794
L027	0.0532	0.0433	0.0385	0.3268	0.3099	0.3464
L028	-0.2709	-0.2155	-0.2551	-0.2026	-0.1280	-0.1996
L029	0.4158	0.4100	0.4225	0.4010	0.3978	0.4328
L031	0.2450	0.2679	0.3353	0.2564	0.2769	0.3537
L032	1.2721	1.2728	1.3066	1.0979	1.1032	1.1342
L033	-0.3973	-0.1935	-0.2357	-0.4372	-0.1978	-0.2749
L034	-0.0246	-0.0871	-0.1262	-0.2182	-0.2467	-0.2924
L035	-0.2852	-0.1981	-0.2774	-0.3255	-0.2894	-0.3555
L037	-0.1529	-0.1404	-0.1608	-0.2334	-0.2246	-0.2347
L038	-0.5262	-0.5605	-0.5455	-0.3326	-0.3775	-0.3546
L039	0.1566	-0.0681	-0.0133	0.2094	-0.0549	0.0074
L041	-0.1713	-0.1755	-0.1738	-0.3181	-0.3214	-0.3183
L042	-0.7609	-0.7628	-0.8109	-0.7108	-0.6930	-0.7637
L043	-0.2838	-0.3048	-0.3518	-0.2579	-0.2791	-0.3231

L044	0.0329	0.0460	0.0904	0.1257	0.1429	0.1896
L045	0.5743	0.5586	0.5963	0.6475	0.6341	0.6792
L046	-0.0740	-0.0537	-0.0896	-0.0235	-0.0034	-0.0465
L047	-0.8990	-0.8751	-0.9093	-0.7118	-0.6898	-0.7419
L048	-0.3622	-0.2381	-0.2975	-0.2642	-0.1501	-0.2137
L049	0.5409	0.3654	0.4472	0.5612	0.4039	0.4946
L054	-0.4826	-0.4985	-0.4524	-0.3665	-0.3817	-0.3081
L055	0.0172	0.0199	0.0304	-0.0721	-0.0698	-0.0692
L056	1.4340	1.4402	1.4263	1.2727	1.2772	1.2746
L058	0.5429	0.5568	0.5915	0.5223	0.5326	0.5593

Table S3. The best 91 maize hybrids according to the genomic estimated value (GEV) scores for grain yield (GY) in the genetic scenarios I , G_a , and G_{a+d}

GY					
I		G_a		G_{a+d}	
Hybrid	GEV	Hybrid	GEV	Hybrid	GEV
L019xL056	3.20	L019xL056	3.22	L019xL056	2.96
L019xL032	3.16	L019xL032	3.18	L049xL056	2.94
L002xL056	2.80	L002xL056	2.81	L032xL056	2.89
L002xL013	2.66	L002xL013	2.68	L019xL032	2.83
L032xL056	2.61	L032xL056	2.61	L002xL056	2.78
L002xL045	2.52	L002xL045	2.52	L002xL013	2.71
L001xL056	2.40	L001xL056	2.42	L002xL045	2.60
L001xL002	2.36	L001xL002	2.37	L001xL002	2.48
L049xL056	2.35	L033xL056	2.34	L016xL032	2.46
L002xL007	2.32	L002xL007	2.32	L001xL056	2.45
L016xL056	2.30	L016xL056	2.31	L047xL056	2.39
L002xL019	2.30	L002xL019	2.31	L002xL019	2.39
L044xL056	2.29	L044xL056	2.31	L002xL018	2.36
L016xL032	2.27	L016xL018	2.29	L002xL007	2.27
L016xL018	2.26	L016xL032	2.28	L029xL032	2.20
L029xL032	2.26	L029xL032	2.26	L032xL058	2.15
L033xL056	2.25	L049xL056	2.25	L010xL056	2.15
L013xL056	2.22	L013xL056	2.24	L016xL056	2.13
L002xL018	2.13	L002xL018	2.15	L039xL056	2.09
L032xL058	2.13	L032xL058	2.14	L045xL056	2.09
L010xL056	2.13	L010xL056	2.13	L016xL018	2.05
L029xL056	2.12	L029xL056	2.13	L029xL056	2.04
L015xL056	2.09	L015xL056	2.09	L014xL056	2.01
L014xL056	2.09	L014xL056	2.07	L044xL056	2.00
L002xL049	2.05	L018xL056	1.98	L016xL045	2.00
L056xL049	1.98	L002xL049	1.95	L018xL056	1.99
L018xL056	1.97	L019xL045	1.89	L056xL049	1.96
L019xL045	1.89	L056xL049	1.87	L013xL056	1.95
L008xL032	1.86	L008xL032	1.86	L002xL031	1.90
L039xL049	1.85	L037xL056	1.85	L002xL044	1.89
L032xL045	1.84	L002xL031	1.84	L007xL032	1.89
L037xL056	1.84	L032xL045	1.83	L007xL017	1.87
L045xL056	1.83	L045xL056	1.83	L033xL056	1.87
L002xL031	1.82	L007xL032	1.79	L039xL019	1.87
L007xL032	1.80	L007xL056	1.77	L008xL032	1.85
L007xL056	1.78	L016xL045	1.77	L012xL032	1.84
L016xL045	1.77	L007xL017	1.74	L019xL045	1.84
L007xL017	1.75	L002xL044	1.73	L008xL056	1.84

L017xL032	1.72	L017xL032	1.71	L015xL056	1.83
L002xL044	1.72	L010xL058	1.69	L008xL045	1.78
L012xL032	1.69	L012xL032	1.68	L017xL032	1.77
L010xL058	1.68	L019xL031	1.67	L002xL029	1.72
L002xL029	1.66	L039xL049	1.67	L018xL019	1.71
L019xL031	1.65	L032xL055	1.66	L002xL010	1.70
L032xL055	1.65	L002xL029	1.65	L037xL056	1.70
L008xL056	1.64	L008xL056	1.64	L010xL058	1.65
L032xL039	1.61	L032xL007B	1.62	L002xL008	1.65
L039xL056	1.61	L016xL058	1.62	L048xL056	1.65
L034xL045	1.60	L002xL016	1.60	L056xL014B	1.64
L016xL058	1.60	L002xL010	1.59	L024xL032	1.62
L032xL007B	1.59	L034xL045	1.57	L032xL045	1.62
L002xL016	1.59	L008xL045	1.56	L032xL039	1.61
L002xL010	1.59	L002xL014	1.55	L002xL046	1.60
L002xL014	1.57	L024xL032	1.55	L008xL058	1.59
L008xL045	1.57	L011xL032	1.54	L032xL055	1.59
L024xL032	1.55	L032xL039	1.52	L007xL056	1.59
L011xL032	1.52	L008xL058	1.51	L002xL016	1.57
L008xL058	1.51	L039xL056	1.51	L019xL031	1.56
L002xL012	1.48	L018xL019	1.49	L027xL056	1.55
L032xL037	1.48	L032xL037	1.49	L003xL056	1.54
L032xL034	1.48	L002xL012	1.48	L008xL019	1.54
L002xL008	1.47	L002xL008	1.47	L017xL007B	1.53
L018xL019	1.46	L032xL034	1.45	L002xL049	1.52
L018xL032	1.42	L016xL019	1.44	L032xL007B	1.52
L016xL019	1.42	L001xL046	1.44	L007xL049	1.52
L001xL046	1.41	L018xL032	1.43	L032xL037	1.52
L013xL016	1.39	L013xL016	1.41	L017xL056	1.52
L007xL008	1.39	L032xL046	1.39	L034xL045	1.50
L032xL046	1.38	L048xL056	1.38	L002xL055	1.50
L039xL019	1.38	L016xL017	1.38	L016xL058	1.48
L016xL017	1.38	L007xL008	1.38	L002xL012	1.48
L008xL024	1.36	L008xL024	1.37	L007xL039	1.47
L056xL014C	1.34	L056xL014C	1.35	L011xL032	1.46
L055xL056	1.34	L055xL056	1.34	L032xL049	1.45
L032xL043	1.34	L018xL046	1.34	L032xL046	1.45
L008xL044	1.33	L008xL044	1.34	L007xL008	1.44
L048xL056	1.33	L008xL007B	1.34	L018xL032	1.42
L008xL019	1.33	L008xL019	1.34	L032xL034	1.42
L034xL056	1.32	L032xL043	1.33	L008xL044	1.41
L018xL046	1.32	L010xL016	1.32	L018xL049	1.40
L010xL016	1.31	L045xL046	1.30	L032xL033	1.39

L045xL046	1.30	L047xL056	1.30	L056xL014C	1.39
L008xL007B	1.30	L003xL056	1.29	L002xL011	1.36
L003xL056	1.28	L002xL046	1.28	L016xL019	1.35
L047xL056	1.28	L032xL033	1.28	L008xL007B	1.35
L002xL046	1.27	L034xL056	1.27	L013xL016	1.35
L018xL049	1.26	L039xL019	1.26	L055xL056	1.34
L014xL032	1.25	L032xL041	1.25	L002xL014	1.34
L032xL041	1.25	L027xL056	1.24	L035xL045	1.34
L007xL041	1.24	L007xL033	1.24	L001xL046	1.32
L027xL056	1.24	L007xL041	1.24	L056xL007B	1.32

Table S4. Maize lines progenitors of the 10% best hybrids based on their genomic estimated value (GEV) for grain yield (GY) and harmonic mean index (HM) in the genetic scenarios I , G_a , and G_{a+d} . Green color indicates the most frequent lines, Yellow represents the lines not present in all scenarios for GY, and Blue represents exclusive lines within HM

	GY						HM					
	I		G_a		G_{a+d}		I		G_a		G_{a+d}	
	Lines	Freq	Lines	Freq	Lines	Freq	Lines	Freq	Lines	Freq	Lines	Freq
1	L056	27	L056	27	L056	29	L056	26	L056	26	L056	29
2	L032	22	L032	22	L032	21	L032	22	L032	22	L032	20
3	L002	17	L002	17	L002	19	L002	14	L002	14	L002	12
4	L008	10	L008	10	L008	9	L018	9	L018	9	L018	10
5	L016	10	L016	10	L019	9	L008	8	L045	9	L045	10
6	L019	9	L019	9	L016	8	L016	8	L008	8	L008	9
7	L045	8	L045	8	L045	8	L045	8	L016	8	L016	9
8	L018	7	L007	7	L007	7	L007	6	L007	6	L019	8
9	L007	6	L018	6	L018	6	L019	6	L058	6	L007	6
10	L046	5	L046	5	L049	6	L029	6	L019	5	L049	6
11	L049	5	L010	4	L007B	4	L058	6	L029	5	L007B	5
12	L010	4	L039	4	L017	4	L039	5	L031	5	L017	5
13	L039	4	L049	4	L039	4	L049	5	L049	5	L044	5
14	L058	4	L058	4	L058	4	L010	4	L010	4	L058	5
15	L001	3	L001	3	L001	3	L013	4	L013	4	L010	4
16	L013	3	L013	3	L010	3	L014	4	L014	4	L013	4
17	L014	3	L017	3	L013	3	L044	4	L024	4	L029	4
18	L017	3	L029	3	L029	3	L001	3	L044	4	L031	4
19	L029	3	L033	3	L044	3	L007B	3	L007B	3	L014	3
20	L034	3	L034	3	L046	3	L017	3	L017	3	L039	3
21	L044	3	L044	3	L055	3	L024	3	L027	3	L001	2
22	L007B	2	L007B	2	L011	2	L031	3	L046	3	L012	2
23	L012	2	L012	2	L012	2	L046	3	L055	3	L024	2
24	L024	2	L014	2	L014	2	L055	3	L001	2	L027	2
25	L031	2	L024	2	L031	2	L012	2	L012	2	L037	2
26	L037	2	L031	2	L033	2	L027	2	L037	2	L046	2
27	L041	2	L037	2	L034	2	L037	2	L039	2	L055	2
28	L055	2	L041	2	L037	2	L003	1	L048	2	L003	1
29	L003	1	L055	2	L003	1	L014B	1	L003	1	L014B	1
30	L011	1	L003	1	L014B	1	L014C	1	L014B	1	L014C	1
31	L014C	1	L011	1	L014C	1	L015	1	L014C	1	L015	1
32	L015	1	L014C	1	L015	1	L033	1	L015	1	L041	1
33	L027	1	L015	1	L024	1	L034	1	L033	1	L047	1
34	L033	1	L027	1	L027	1	L038	1	L034	1	L048	1
35	L043	1	L043	1	L035	1	L041	1	L038	1	-	-
36	L047	1	L047	1	L047	1	L047	1	L041	1	-	-
37	L048	1	L048	1	L048	1	L048	1	L047	1	-	-

Table S5. The best 91 maize hybrids according to the genomic estimated value (GEV) scores for the index Harmonic mean (HM) in the genetic scenarios I , G_a , and G_{a+d}

HM					
I		G_a		G_{a+d}	
Hybrid	GEV	Hybrid	GEV	Hybrid	GEV
L019xL032	2.66	L019xL032	2.67	L049xL056	3.00
L001xL056	2.49	L001xL056	2.50	L032xL056	2.61
L016xL032	2.48	L016xL032	2.49	L016xL032	2.54
L032xL056	2.46	L032xL056	2.46	L001xL056	2.49
L049xL056	2.35	L010xL056	2.33	L019xL032	2.34
L010xL056	2.33	L002xL045	2.29	L010xL056	2.28
L002xL045	2.29	L049xL056	2.27	L047xL056	2.24
L002xL007	2.22	L002xL007	2.24	L002xL045	2.22
L002xL018	2.09	L002xL018	2.11	L002xL007	2.15
L008xL056	2.07	L008xL056	2.08	L002xL018	2.14
L045xL056	2.01	L045xL056	2.01	L056xL049	2.05
L019xL056	1.97	L019xL056	1.98	L008xL056	2.05
L056xL049	1.95	L013xL056	1.90	L045xL056	2.05
L002xL049	1.89	L016xL058	1.90	L027xL056	1.98
L019xL045	1.89	L019xL045	1.89	L019xL056	1.94
L013xL056	1.89	L033xL056	1.87	L029xL056	1.94
L016xL058	1.88	L056xL049	1.86	L002xL013	1.93
L007xL017	1.84	L016xL018	1.86	L048xL056	1.85
L027xL056	1.84	L013xL016	1.85	L003xL056	1.82
L016xL018	1.83	L007xL017	1.85	L008xL045	1.81
L013xL016	1.83	L027xL056	1.84	L002xL019	1.79
L015xL056	1.80	L002xL049	1.81	L056xL014B	1.78
L033xL056	1.77	L015xL056	1.80	L016xL018	1.77
L016xL056	1.76	L016xL056	1.77	L019xL045	1.75
L039xL049	1.76	L002xL013	1.74	L016xL058	1.74
L032xL039	1.76	L029xL056	1.73	L002xL049	1.71
L029xL056	1.73	L037xL056	1.72	L012xL032	1.71
L002xL013	1.72	L002xL019	1.72	L001xL002	1.71
L008xL045	1.72	L008xL045	1.71	L013xL016	1.70
L037xL056	1.72	L048xL056	1.68	L007xL017	1.65
L002xL019	1.71	L032xL039	1.66	L016xL056	1.65
L018xL027	1.64	L018xL027	1.65	L013xL056	1.64
L048xL056	1.63	L001xL002	1.63	L002xL056	1.61
L012xL032	1.62	L008xL007B	1.63	L017xL032	1.61
L003xL056	1.61	L003xL056	1.62	L014xL056	1.61
L001xL002	1.61	L012xL032	1.61	L015xL056	1.60
L008xL007B	1.61	L008xL058	1.61	L024xL032	1.59
L008xL058	1.60	L039xL049	1.60	L008xL032	1.58

L017xL032	1.59	L010xL058	1.59	L002xL031	1.57
L010xL058	1.58	L044xL056	1.59	L008xL058	1.57
L044xL056	1.58	L017xL032	1.58	L010xL058	1.57
L002xL056	1.58	L002xL056	1.57	L018xL027	1.57
L002xL031	1.56	L002xL031	1.57	L044xL056	1.54
L008xL032	1.56	L008xL032	1.56	L029xL032	1.54
L024xL032	1.53	L024xL032	1.54	L017xL007B	1.53
L014xL032	1.50	L007xL056	1.51	L037xL056	1.52
L007xL056	1.50	L032xL037	1.50	L018xL056	1.49
L032xL037	1.49	L014xL032	1.49	L032xL058	1.49
L056xL014B	1.49	L056xL014B	1.49	L039xL019	1.48
L032xL055	1.47	L007xL032	1.48	L032xL037	1.46
L018xL049	1.47	L032xL055	1.48	L008xL007B	1.46
L007xL032	1.47	L014xL056	1.44	L007xL032	1.46
L014xL056	1.46	L018xL056	1.43	L032xL055	1.44
L018xL056	1.42	L010xL016	1.42	L018xL032	1.43
L010xL016	1.41	L045xL046	1.41	L002xL010	1.43
L045xL046	1.40	L018xL049	1.40	L018xL049	1.42
L002xL012	1.39	L002xL012	1.38	L002xL044	1.42
L032xL034	1.37	L008xL044	1.38	L008xL044	1.41
L008xL044	1.37	L032xL034	1.37	L031xL045	1.41
L018xL032	1.35	L018xL032	1.36	L013xL045	1.37
L032xL045	1.35	L044xL058	1.36	L014xL032	1.37
L044xL058	1.35	L056xL014C	1.34	L017xL056	1.37
L029xL032	1.33	L032xL045	1.34	L032xL046	1.36
L056xL014C	1.33	L029xL032	1.33	L032xL045	1.36
L032xL058	1.32	L032xL058	1.32	L056xL014C	1.35
L013xL045	1.32	L013xL045	1.32	L029xL045	1.35
L016xL017	1.31	L047xL056	1.31	L016xL045	1.31
L032xL038	1.30	L016xL017	1.31	L039xL056	1.30
L055xL056	1.29	L055xL056	1.29	L055xL056	1.29
L047xL056	1.29	L002xL044	1.29	L056xL007B	1.29
L002xL044	1.28	L032xL038	1.29	L008xL019	1.28
L032xL046	1.28	L032xL046	1.29	L007xL056	1.26
L008xL019	1.27	L008xL019	1.29	L016xL017	1.26
L032xL041	1.26	L018xL046	1.28	L002xL012	1.26
L002xL010	1.26	L032xL041	1.27	L010xL016	1.25
L018xL046	1.26	L002xL010	1.26	L031xL032	1.25
L007xL031	1.24	L007xL031	1.26	L014xL018	1.24
L002xL014	1.24	L002xL014	1.22	L018xL019	1.24
L029xL045	1.22	L032xL007B	1.22	L032xL007B	1.24
L014xL018	1.22	L014xL018	1.21	L044xL058	1.22
L032xL007B	1.21	L007xL058	1.21	L007xL031	1.20

L007xL058	1.20	L029xL045	1.21	L018xL024	1.20
L016xL029	1.19	L018xL024	1.20	L019xL044	1.19
L039xL019	1.19	L016xL029	1.19	L007xL008	1.19
L018xL024	1.18	L031xL032	1.19	L032xL041	1.17
L031xL032	1.18	L029xL007B	1.15	L016xL029	1.16
L029xL007B	1.14	L027xL048	1.11	L008xL049	1.15
L029xL039	1.13	L024xL031	1.10	L046xL056	1.14
L001xL039	1.09	L024xL055	1.10	L032xL039	1.13
L024xL055	1.09	L031xL045	1.09	L049xL007B	1.12
L002xL008	1.09	L002xL008	1.08	L018xL045	1.12