Acta Scientiarum

http://www.periodicos.uem.br/ojs/
ISSN on-line: 1807-8621
Doi: 10.4025/actasciagron.v45i1.61550

Comparative study between phenotypic and genomic analyses aimed at choosing parents for hybridization purposes

Francyse Edite de Oliveira Chagas de Moraes[1*], Michele Jorge Silva Siqueira[1], Antônio Carlos da Silva Júnior[1], Renato Domiciano Silva Rosado[2] and Cosme Damião Cruz[1]

[1]Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. Peter Henry Rolfs, s/n, 36570-000, Campus Universitário, Viçosa, Minas Gerais, Brazil. [2]BASF – Pesquisador Digital e Inovação, Trindade, Goiás, Brazil. *Author for correspondence. E-mail: francysechagas70428@gmail.com

ABSTRACT. The development of superior cultivars involves parents with superiority for the traits of interest and wide genetic variability. Efficient plant breeding and selection strategies that allow for the identification of superior genotypes are essential in breeding programs. This work aims to carry out a comparative study between several strategies for choosing parents, for hybridization purposes, based on phenotypic analysis and molecular information. To obtain the phenotypic and genotypic information of the parents, data simulation was used. For genotyping, 2000 single nucleotide polymorphism markers were used, and from all possible gametes to be formed (2^{2000}), 5000 were randomly sampled to form each of the 100 individuals of the population of recombinant inbred strains. To obtain the phenotypic information, five characteristics with different levels of complexity were simulated. The comparative study was carried out using data referring to simulated genotypic values of hybrids and parents. Then, aiming to choose the parents destined for hybridization, different traditional selection strategies based on phenotypic analysis and the genome-wide selection methodology were approached. The genomic information resulted in the choice of the best lines and in obtaining superior hybrids when compared with traditional methodologies. The inclusion of the genomic genetic values of the parents in determining the crosses to be carried out increases the probability of generating phenotypically superior hybrids. Thus, the traditional methods of choosing parents for hybridization purposes are effective, but when incorporating the information from genome-wide selection, the choice of parents provides superior and promising results.

Keywords: genome-wide selection; genetic diversity; diallel analysis.

Received on November 15, 2021.
Accepted on June 10, 2022.

Introduction

Plant breeding has been used in agriculture since ancient times through domestication and selection; however, with advances in genetic engineering and biotechnology, breeding practices have improved significantly (Ahmar et al., 2020; Hill, 2012; Lenaerts, Collard, & Demont, 2018). Breeders have begun to interfere directly and intentionally in the DNA of plants and then define which desired characteristics should be passed on to their offspring, managing to select plants with greater potential based on more sophisticated data, such as genome information (Lyzenga, Pozniak, & Kagale, 2021). As the access to and quantity of genomic data increases, there are new opportunities for collaboration between classical improvement, which has been carried out for years, and biotechnology, bringing more precise improvements (Ahmar et al., 2020; Bohar et al., 2020; Garrido-Cardenas, Mesa-Valle, & Manzano-Agugliaro, 2017; Gupta, Kumar, Mir, & Kumar, 2010; Nadeem et al., 2018; Rasmussen, 2020).

In plant breeding programs, the presence of genetic variability is a basic and necessary condition to obtain genetic gains with selection (Khaki, Khalilzadeh, & Wang, 2020; Phuke et al., 2017). To develop a superior cultivar, it is necessary to carry out the crossing between parents properly. If selection involves parents with superiority for the traits of interest and wide genetic variability, there will be more superior segregating genotypes (Falconer, 1981). Therefore, the probability of success of a breeding program is closely linked to a careful selection of parents (Pimentel et al., 2013). Associated with this, efficient breeding and plant selection strategies that allow for the identification of superior genotypes are essential. Normally, crosses are carried out between parents with favorable phenotypes for complementary traits, to generate segregating populations with enough variability to carry out selections of superior lines for the characters of interest.

Among the existing methodologies for choosing parents, competition assays stand out, whether or not they are added to the predictive study of genetic diversity, or quantitative study of this diversity through experiments for diallel analysis purposes. Diallels consist of crossing all possible combinations of n parents (Hayman, 1954). In this crossing scheme, it is possible to identify the hybrids that manifest high heterosis, which are those with the greatest genetic diversity among the parents (Falconer, 1981). Therefore, segregating populations derived from these hybrids have greater potential to present superior individuals and should receive greater attention. Although this methodology allows selecting parents efficiently and still provides genetic information (Coelho et al., 2020; Kulka et al., 2018), there are some limitations—for example, in a situation with a high number of candidate parents, in which it is necessary to carry out several artificial crosses (Werle et al., 2014).

Another way to determine the best crosses is to aggregate information obtained by predictive procedures, which seek to estimate the genetic variance of the segregating population from estimates of the genetic distance between parents (Almeida et al., 2020; Bhandari et al., 2017; Müller et al., 2015; Swarup et al., 2020). Genetic diversity among parents is measured without the need for hybrid combinations, being established based on phenotypic (agronomic, morphological, and physiological differences) or genotypic (molecular markers) information of the parents themselves. Therefore, it can be predicted that the best hybrids will result from crossing parents with great diversity and desirable characteristics (Swarup et al., 2020) However, determining genetic diversity based on phenotypic characters is greatly influenced by the environment, especially for traits of agronomic interest (Bhandari et al., 2017).

With the improvement in molecular biology techniques and platforms for next-generation sequencing (NGS), a new approach has been proposed called genome-wide selection (GWS). This methodology associates phenotypic data with genotypic data and was originally applied to predict gains resulting from the selection (Meuwissen, Hayes, & Goddard, 2001). There is potential to use values predicted by the GWS approach in problems related to establishing a base population, which will be dealt with in this article. This technique has been explored through simulation studies and application in traditional breeding programs for several crops (Alkimim et al., 2020; Barbosa et al., 2021; Crossa et al., 2017; Sousa et al., 2021). GWS is not interested in identifying the function of each gene, but in establishing associations between markers throughout the entire genome with phenotypic characteristics of interest (Sousa et al., 2021). Although GWS is efficient, some challenges are still routinely faced by professionals in the field of biometrics, which include genetic issues inherent to the use of molecular markers, statistical questions about the use of different data analysis paradigms, and, especially, computational issues arising from the requirement analysis of large data sets (Cruz, Carneiro, & Bhering, 2021).

Given the above information, this work aims to carry out a comparative study between several strategies for choosing parents, to form a base population from which hybrids will be formed, based on traditional analysis and phenotypic information, and data analysis with aggregated molecular information.

Material and methods

Obtaining phenotypic and genotypic information from parents

This dataset served as the basis for the use of different approaches to parental choice to form a base population from which hybrids will be formed. For the experimental simulation of phenotypic and genotypic data, a diploid species was used (2n = 2x = 20). For genotyping, 2000 single nucleotide polymorphism (SNP) markers were used, and from all possible gametes to be formed (2^{2000}), 5000 were randomly sampled to form each of the 100 individuals of the population of recombinant inbred lines (RILs). The available gametes were originated from several generations of successive self-fertilization of F_1 individuals obtained by crossing contrasting homozygous parents. The controller genes were simulated and randomly distributed into 10 linkage groups, and the correlation between traits was given by genes present in the same linkage group. As discussed by Moura et al. (2018), this is one of the causes of genetic correlation, which may have positive, negative, or null values.

To obtain the phenotypic information, five characteristics with different levels of complexity were simulated. Complexity was considered directly proportional to the number of controlling genes and inversely proportional to heritability. The simulated characteristics had different dimensions, as evidenced by their different averages. The parameters adopted for the simulation of the characteristics are detailed in Table 1.

The information established is only reference values used to compare the performance of the methodologies used, without prejudice to the extrapolation of the results obtained.

Table 1. Parameters used to simulate the five characteristics with different levels of complexity.

Trait	NG	ADD	h^2	Average
X_1	12	0.4	0.8	5
X_2	21	0.7	0.7	30
X_3	32	1	0.5	80
X_4	73	0.5	0.4	100
X_5	100	0.3	0.2	150

NG: number of genes; ADD: average dominance of genes; h^2: heritability; Average: average value of the characteristic, in its respective unit of measure.

The total genetic value expressed by a given individual was estimated from the following expression:

$$G_i = \mu + a_i + d_i \quad (1)$$

where d_i represents the dominance effects in genotype i, that is, it is the effect given by the presence of two different alleles at the same locus in homologous chromosomes; a_i represents the additive effects present in genotype i, given by:

$$a_i = \sum_{j=1}^{g} p_j \alpha_j \quad (2)$$

Where α_j is the effect of the favorable allele in block j, considered equal to 1, 0, or -1 for the genotypic classes AA, Aa, and aa, respectively; and p_j is the contribution of locus j to the manifestation of the characters established in the work from weights generated by values of a uniform distribution. Thus, in the simulation process, it was assumed that all genes have the same effect on the trait.

The phenotypes of individuals (i) were generated according to the model:

$$F_i = G_i + M_i \quad (3)$$

Where F_i is the genetic effect given by the sum of the genetic effects of each locus and the dominance deviation, as shown in equation (1), and M_i is the environmental effect, generated according to a normal distribution with a mean of zero and variance compatible with the heritability of the simulated characteristic, given by the following expression:

$$h^2 = \frac{\sigma_g^2}{\sigma_f^2} \quad (4)$$

where σ^2_g refers to genetic variance associated with the simulated trait, and σ^2_f corresponds to phenotypic variance.

Obtaining phenotypic and genotypic information of hybrids

This information was used to attest to the effectiveness of the techniques for choosing parents to obtain superior hybrid populations. From the set of genotypic information of the parents, represented by RIL lineages, a new data set represented by the hybrid combinations was generated, with a total of 4950 ($C_{100,2}$) hybrids formed. Based on the established gene control, it was also possible to establish the expected phenotypic values of these hybrid combinations for the five phenotypic traits studied.

Strategies for selecting parents for hybridization purposes

From the original set of parents, the objective was to identify the 10 with the best attributes to be used in the formation of hybrids; for this endeavor, five different strategies were used.

Performance per se

For this approach, 10 of 100 available parents (RIL strains) were selected, having as the only reference the phenotypic value in relation to variable X_5. The choice for this methodology can be compared to the direct analysis of the phenotype in a given experimental evaluation.

Performance per se accompanied by predictive analysis of genetic diversity

In the first step, 20 parents were pre-selected, based exclusively on the phenotypic value in relation to variable X_5. In the second step, the 10 parents with the best potential for X_5 and genetic diversity revealed by

multivariate statistics were identified. Thus, for the analysis of genetic diversity, the Euclidean distances between each genotype pair were calculated, after data standardization, as follows:

$$d_{ii'} = \sqrt{\sum_j \quad (Y_{ij} - Y_{i'j})^2} \tag{5}$$

Where d_{ii} is the Euclidean distance between parents i and i' and Y_{ij} is the value of the ith parent in relation to the jth variable.

From the distance matrix, the Tocher grouping method was used, which consists of identifying the most similar pair of individuals to form the initial group (Cruz, Ferreira, & Pessoni, 2011). Subsequently, the possibility of including new individuals was evaluated, using as a criterion that the mean intragroup distance should be smaller than the mean intergroup distance.

Principal component analysis was used as a complementary analysis; it consists of a multivariate statistical technique that transforms a set of original variables into another set of variables of the same dimension, to redistribute the variation observed in the original axes to obtain a set of uncorrelated orthogonal axes. Thus, from a total of 20 parents considered to have good performance about the X_5 trait, the 10 most distant parents genetically regarding the 5 characteristics were chosen.

Performance per se accompanied by diallel analysis

This strategy also involved two steps. Initially, 20 parents were pre-selected taking into account the performance regarding the main characteristic X_5. In the second step, diallel design was used to select the final 10 parents based on the general combining ability (GCA) and the specific combining ability.

Performance per se accompanied by genetic and diallel diversity analysis

This approach involved three steps to identify good parents of high additive genetic value and with genetic complementarity: pre-selection, by competition assay, predictive study of genetic diversity, and, finally, quantitative study of genetic superiority and complementarity by diallel analysis. In this case, pre-selection (step 1) was carried out to identify the 40 best parents in relation to characteristic X_5. Subsequently, genetic diversity analysis (step 2) was performed using the Euclidean distance, the Tocher method, and principal component analysis. The 20 parents with good performance and divergence were selected. In step 3, diallel analysis was performed, and finally, 10 parents with good performance, divergent, and with greater additive genetic value were selected.

Selection of parents based on molecular information

Using the principles of the GWS methodology, the criterion for choosing the parents was based on the predicted genomic genetic value (VGG). This approach uses markers that have been analyzed previously for segregation.

The genomic selection methodology faces the problem that the number of markers is greater than the number of individuals in the population and the markers are highly correlated (Crossa et al., 2017). There are several models available to solve this problem. To model the effect of the markers, the ridge regression best linear unbiased prediction (RR-BLUP) was used, also called random regression. The RR-BLUP assumes that the effects of markers are random and have a normal distribution with constant variance, according to the following model:

$$y = Xg + e \tag{6}$$

Where y is the phenotype; X is an incidence matrix of dimensions N (number of individuals) × n (number of markers); g is the vector of regression coefficients, normally distributed (0, σ^2_g); and e is the random error with normal distribution (0, σ^2_e).

The effects of markers were estimated by the following equation:

$$g = (X'X + I\lambda)^{-1} X'y \tag{7}$$

where $\lambda = k = \sigma^2_e / \sigma^2_g$ is constant for all bookmarks. The variance of the markers is constant, and in general, k must be considered as a function of the additive genetic variance.

Having the predicted effect of each marker, it is possible to obtain the predicted VGG of the individuals that represented the population of possible parents:

$$\widehat{VGG} = \sum_i^n \quad X_i\, \hat{g}_i \tag{8}$$

where n is the number of markers arranged in the genome; X_i is the line of the incidence matrix that allocates the genotype of the i[th] marker for each individual; and $\hat{g}_i$ is the estimated effect of the i[th] marker.

In the prediction, the training and validation population were all strains genotyped for the markers and phenotyped for the traits of interest. The genomic genetic value of the lines was ordered in decreasing order and the 10 best lines were chosen. Pearson's simple correlation between the actual genetic value of individuals and the genomic genetic value obtained by GWS was used to measure the model's accuracy, as follows:

$$\rho_{x,y} = \frac{\sum_{i=1}^n \quad (x_i - \underline{x})(y_i - \underline{y})}{(n-1)\sigma_X \sigma_Y} \tag{9}$$

where $\rho_{x,y}$ denotes the Pearson coefficient; n is the number of terms; x_i and y_i are the values of the X and Y variables for the individual i, respectively; $\underline{x}$ and $\underline{y}$ are the means of the X and Y variables, respectively; and σ_X and σ_Y are the standard deviations of the X and Y variables, respectively.

Pearson's correlation between the genomic breeding value and the true breeding value was tested by the t-test at 1% probability, indicating the accuracy of the prediction model.

The simulations of phenotypic and genotypic data were performed using the GENES software (Cruz, Salgado, & Bhering, 2013). All analyzes were performed with the GENES software integrated in the R platform (Cruz, 2013; 2016; R Core Team, 2019).

Continuous and discrete real values of hybrids

For comparative purposes, data referring to the simulated genotypic values of the hybrids were used to organize the lines and hybrids into groups.

The GCA methodology was used to organize the information of the lineages into discrete groups, obtained through the type II Griffing diallel, in which reciprocal F_1 individuals are excluded. The genotypic values of the hybrids in relation to trait X_5, which has the highest number of controlling genes and the lowest heritability, were used to determine the GCA of each strain. The statistical model used in this diallel is represented in the following equation:

$$Y_{ij} = m + g_i + g_j + S_{ij} + \underline{e}_{ij} \tag{10}$$

where Y_{ij} is the average value of the hybrid combination ($i \neq j$) or the parent ($i = j$); m is the overall average; g_i and g_j the effects of the GCA of the i[th] and the j[th] parent, respectively; S_{ij} is the effect of specific combining ability for crosses between i and j order parents, respectively; and $\underline{e}_{ij}$ is the average experimental error.

The overall combining ability for each of the 100 parents was given by:

$$\hat{g}_i = \frac{1}{(p+2)} \left[Y_{ii} + Y_{i.} - \frac{2}{p} Y_{..}\right] \tag{11}$$

where $\hat{g}_i$ is the parent's GCA i; p is the number of parents analyzed; Y_{ii} is the total cross between parent i and parent i; $Y_{i.}$ is the total of the analyzed characteristic for the genotype i; and $Y_{..}$ is the total for the trait analyzed for all genotypes.

Based on the GCA result of each lineage, they were evenly organized into five groups: optimal, good, medium, bad, and very bad. The optimal group was further divided equally into two subgroups: suboptimal and superoptimal.

Subsequently, the transformation of data from hybrids with a continuous distribution into discrete values was performed, also within the seven groups mentioned.

Results and discussion

Performance per se

Based on performance per se, lines 2, 4, 5, 19, 22, 48, 49, 54, 55, and 76 were selected; among these, 10% belong to the group classified as superoptimal. After crossing these lines, out of a total of 45 ($C_{10,2}$) hybrids formed, it 24.4% also belong to the superoptimal group. This parent selection strategy depends only on information from competition trials and, compared with other strategies, it has a lower cost. The concept of adopting as a selection criterion only the individual performance of potential genotypes can be questioned for not considering the diversity that is essential in establishing the variability to be explored in future segregating generations.

Performance per se accompanied by predictive analysis of genetic diversity

In a new scenario, the strategy of pre-selecting 20 parents, among the 100 originally available, was adopted, based on performance per se. Afterward, genetic diversity analysis was carried out in the 20 selected parents, in which, after obtaining the similarity matrix, the grouping was carried out using the inverted Tocher method. Table 2 lists the selected parents, divided into nine groups.

Table 2. Grouping of parents selected according to performance per se accompanied by genetic diversity analysis.

Group*	Lines	Group	Lines
I	2+,4+,5+,12+,26+,37+,49+,54+,55, 74+ and 76+	VI	47
II	59 and 79	VII	22
III	19	VIII	69
IV	48	IX	62
V	33	-	-

*Each group aggregates parents with high dissimilarity; + parents selected only for good performance.

Friske, Schuster, Marcolin, and Silva (2018) also used the Tocher method for similarity pattern recognition in genetic diversity analysis, They evaluated the maturity characteristics and yield components in maize lines. They divided the analyzed the strains into eight groups, the first two groups being recommended for crossing between strains, as they belong to different heterotic groups and have good genetic complementation. Barros et al. (2019) reported the same strategy with open-pollinated maize varieties, noting that, by using the Tocher method, they allocated the varieties into five different groups, and that crosses between genetically distant varieties would produce hybrids with higher yields. As the inverted Tocher method was used in the work, the most divergent individuals were in one group, namely the selected group.

In addition to Tocher's method, principal component analysis was performed to recognize similarity patterns in scatter plots. In this study, the first three main components represented about 80% of the data variability; therefore, they were chosen for the analysis, with the three-dimensional graph being the most appropriate for the projection of the scores obtained by the components (Figure 1). The choice of these components is due to their involvement with a minimum of 80% of the available variation (Hongyu, Sandanielo, & Junior, 2016).

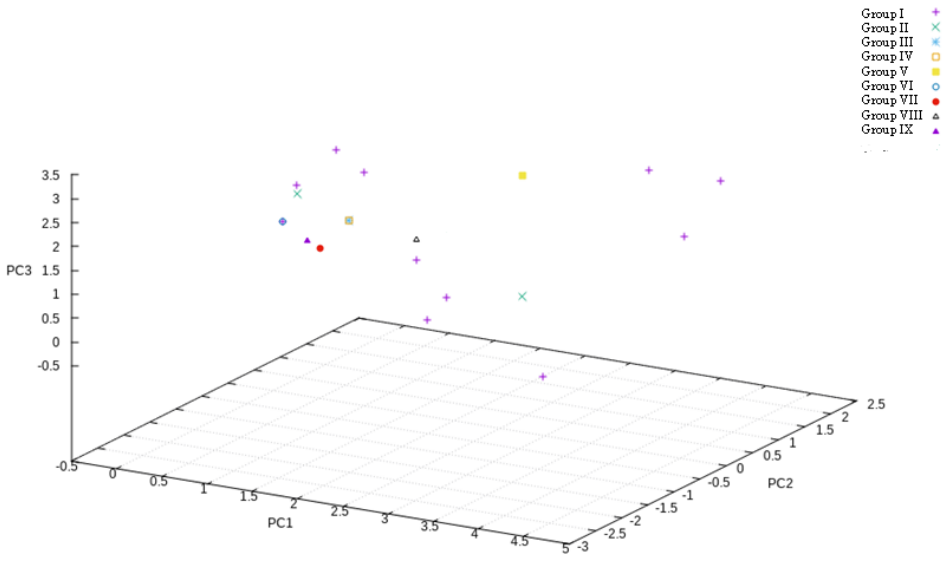

Figure 1. Graphical dispersion of selected parents according to performance per se and groups formed by the Tocher method.

Then, the 10 superior and divergent lines (2, 4, 5, 12, 26, 37, 49, 54, 74, and 76) were selected. When the 45 hybrids were formed, 20% of the lines and 37.8% of the hybrids belong to the group classified as superoptimal.

Performance per se accompanied by diallel analysis

By studying genetic diversity, it is possible to obtain genetically distant parents that should complement each other when they are crossed, but not parents with high additive genetic value, which could be a better alternative for breeding programs to obtain hybrids. Thus, one way to recognize the additive genetic potential

of parents is through Griffing diallel analysis. Initially, the 20 best lines were selected phenotypically to be crossed. From the results of the GCA of each strain, the 10 strains with the highest GCA were chosen. Among the chosen lines (2, 5, 12, 33, 48, 54, 62, 74, 76, and 79) and of the formed hybrids, 40 and 42.2%, respectively, belong to the superoptimal group, that is, in quantity greater than in the two strategies described above, showing that information about the GCA led to a better strategy for choosing parents and hybrids. Moura et al. (2018) used diallel analysis, an approach that allowed them to choose the best populations for the characteristics of grain yield and plant architecture in the common bean breeding program.

With diallel analysis, parents with a high concentration of favorable alleles are selected, generating a population with better agronomic performance. This performance can be improved by incorporating the analysis of genetic diversity, as the parents obtained have different favorable alleles, and there may be complementarity in the crosses.

Performance per se accompanied by genetic and diallel diversity analysis

For the analyses that addressed performance per se accompanied by genetic and diallel diversity analyses, initially, the 40 best parents were selected phenotypically in relation to the X_5 trait. Subsequently, the genetic diversity analysis was performed based on the similarity matrix, where 23 groups were formed, which can be seen in Table 3.

Table 3. Grouping of genotypes according to per se performance accompanied by genetic diversity analysis.

Groups*	Lines	Groups	Lines
I	1, 2, 5, 6, 13, 37, 49, 54, 72, 77, 83, 89, and 94	XIII	70
II	19, 21, and 41	XIV	9
III	4 and 11	XV	91
IV	66 and 76	XVI	22
V	55 and 84	XVII	26
VI	96	XVIII	47
VII	78	XIX	30
VIII	12	XX	48
IX	79	XXI	62
X	43	XXII	69
XI	74	XXIII	59
XII	33	-	-

*Each group aggregates parents with high dissimilarity.

Again, a few groups concentrated most of the genotypes and many groups were formed by only one genotype due to the global grouping criterion of the Tocher method (Vasconcelos, Cruz, Bhering, & Resende Junior, 2007). After applying the inverted Tocher method, the principal components analysis was performed, in which components 1, 2, and 3 together accounted more than 70% of the available variation. Hence, these were chosen for the analysis, which again provided the three-dimensional graph (Figure 2).

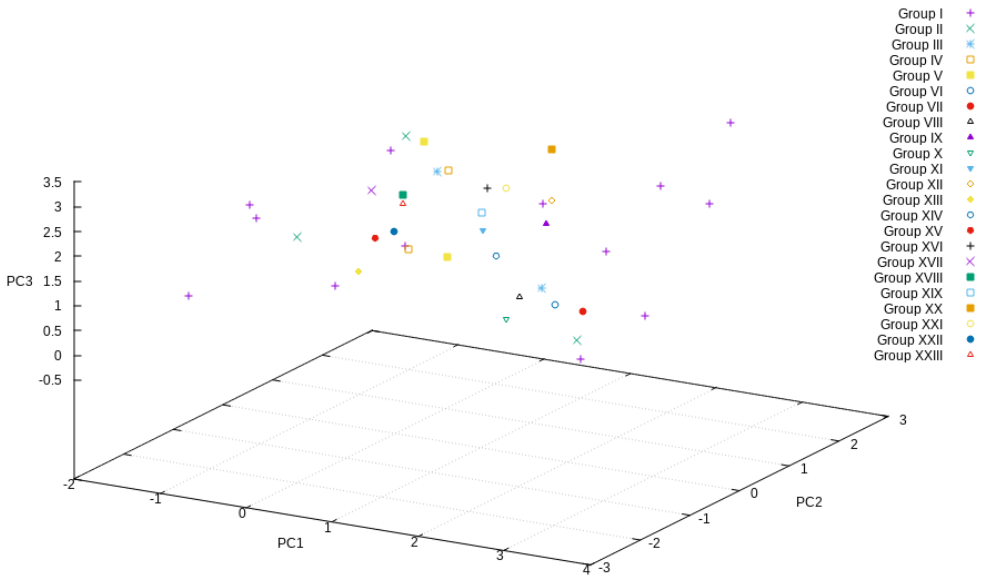

Figure 2. Graphical dispersion of selected parents according to performance per se and groups formed by the inverted Tocher method.

The 20 most distant parents were chosen, and after diallel analysis, the 10 most distant lines and with the highest GCA were chosen (1, 2, 4, 5, 37, 49, 54, 72, 76, and 89) and formed the 45 hybrids with, 10% and 24.4%, respectively, belonging to the superoptimal group. Differently from the three previous methodologies, there were fewer hybrids classified as superoptimal. This may have occurred because the genetic diversity analysis was performed before the diallel analysis. The chosen parents may be very distant, but they do not carry many favorable alleles, which could favor the formation of superior hybrids. Ferreira, Oliveira, Santos, and Ramalho (1995) reported a different result in maize, noting that performing a preliminary genetic diversity analysis resulted in varieties with the good combining ability and complementary.

Selection of parents based on molecular information

The criterion for choosing the parents for the formation of populations can be improved by using molecular information to predict the genomic genetic value of the parents. Thus, a new predicted phenotypic value (genomic genetic value) can be obtained from analyses using the broad genomic selection approach.

The markers used in the prediction were analyzed in relation to their quality, through the segregation behavior of the markers, noting that 0.3% presented skewed segregation in relation to the expected outcome (1:1), at 1% significance. This low frequency indicates that the markers are of good quality for genotyping. Whang et al. (2019) used genomic data to predict GCA and discarded markers that did not meet certain quality filters aimed at removing alleles with low frequency. From a total of 319,668 SNP markers in maize lines, only 61,468 were used for the subsequent analysis.

Among the lines obtained through the prediction model (2, 4, 9, 30, 48, 49, 51, 54, 74, and 78) 70% fall into the superoptimal group, indicating the capacity of GWS to select good parents. Then, from the cross between the lines, 80% of the hybrids formed belong to the superoptimal class. In addition, the Pearson correlation between the genomic breeding value and the real breeding value was significant by the t-test at 1% probability, with a value of 0.81. As noted previously, this correlation indicates satisfactory accuracy of the prediction model.

In this work, it can be noted that the genomic information resulted in the choice of the best lines and in obtaining superior hybrids when compared with traditional methodologies, as summarized in Table 4.

Table 4. Summary of the outcome and efficiency of the five parent selection strategies.

Strategy*	Selected Parents	Percent**
1	2, 4, 5, 19, 22, 48, 49, 54, 55, and 76	24.4%
2	2, 4, 5, 12, 26, 37, 49, 54, 74, and 76	37.8%
3	2, 5, 12, 33, 48, 54, 62, 74, 76, and 79	42.2%
4	1, 2, 4, 5, 37, 49, 54, 72, 76, and 89	24.4%
5	2, 4, 9, 30, 48, 49, 51, 54, 74, and 78	80%

*1 Per se performance; 2 per se performance accompanied by genetic diversity analysis; 3 per se performance accompanied by diallel analysis; 4 per se performance accompanied by genetic and diallel diversity analysis; 5 parent selection based on molecular information. **Corresponds to the percentage of hybrids present in the highest-ranked group: superoptimal.

Wang, Li, and Zhang (2020) reported similar findings when measuring the CGC of 266 maize lines by simulating hybrids originated from these lines. The predicted CGCs with the real phenotypic values of the hybrids obtained by the simulation and with the values predicted by the genomic selection were compared by Pearson correlation; considering the trait with 0.3 heritability and 2000 hybrid samples, they found a correlation of 0.80. This correlation, according to Fonseca and Martins (1996), can be classified as excellent. Thus, GWS is a useful and effective method to select parents and to form the best hybrids. The inclusion of the genomic genetic values of the parents in the determination of the crosses to be carried out increases the probability of generating phenotypically superior hybrids. It is emphasized that despite all its benefits, the GWS does not aim to replace the phenotypic analysis of the data, but rather to complement the process, to carry out the choice of parents more efficiently.

Conclusion

It was evident in this study that the traditional methods of choosing parents for hybridization purposes are effective, but when incorporating the information from GWS, the choice of parents presented superior and promising results.

Acknowledgements

The authors thank the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), the *Fundação de Pesquisa do Estado de Minas Gerais* (FAPEMIG), and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) for financial support and scholarships.

References

Ahmar, S., Gill, R. A., Jung, K. H., Faheem, A., Qasim, M. U., Mubeen, M., & Zhou, W. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *International Journal of Molecular Sciences*, *21*(7), 1-24. DOI: https://doi.org/10.3390/ijms21072590

Alkimim, E.R., Caixeta, E.T., Sousa, T.V., Resende, M. D. V., Silva, F. L., Sakiyama, N. S., & Zambolin, L. (2020). Selective efficiency of genome-wide selection in *Coffea canephora* breeding. *Tree Genetics & Genomes*, *16*(3), 1-11. DOI: https://doi.org/10.1007/s11295-020-01433-3

Almeida, C. P., Carvalho Paulino, J. F., Carbonell, S. A. M., Chiorato, A. F., Song, Q., Di Vittori, V., ... Benchimol-Reis, L. L. (2020). Genetic diversity, population structure, and Andean introgression in Brazilian common bean cultivars after half a century of genetic breeding. *Genes*, 11(11), 1-22. DOI: https://doi.org/10.3390/genes11111298

Barbosa, I. D. P., Silva, M. J., Costa, W. G., Castro Sant'Anna, I., Nascimento, M., & Cruz, C. D. (2021). Genome-enabled prediction through machine learning methods considering different levels of trait complexity. *Crop Science*, *61*(3),1890-1902. DOI: https://doi.org/10.1002/csc2.20488

Barros, L. M., Prochnow, D., Oliveira, V. F., Silva, Oliveira, A. C., & Maia, A. C. (2019). Characterization of open-pollinated maize varieties from Rio Grande do Sul State. *Journal of Crop Science and Biotechnology*, *22*(1), 31-36. DOI: https://doi.org/10.1007/s12892-018-0051-0

Bhandari, H. R., Bhanu, A. N., Srivastava, K., Singh, M. N., Shreya, & Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants - an overview. *Advances in Plants & Agriculture Research*, *7*(3), 279-286. DOI: https://doi.org/10.15406/apar.2017.07.00255

Bohar, R., Chitkineni, A., & Varshney, R. K. (2020). Genetic molecular markers to accelerate genetic gains in crops. *BioTechniques*, *69*(3), 158-160. DOI: https://doi.org/10.2144/btn-2020-0066

Coelho, I. F., Alves, R. S., Peixoto, M. A., Teodoro, L. P. R., Teodoro, P. E., Pinto, J. F. N., ... Bhering, L. L. (2020). Multi-trait multi-environment diallel analyses for maize breeding. *Euphytica*, *216*(9), 1-17. DOI: https://doi.org/10.1007/s10681-020-02677-9

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-Lopez, O., Jarquín, D., Los Campos, G., ... Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961-975. DOI: https://doi.org/10.1016/j.tplants.2017.08.011

Cruz, C.D. (2013). Genes Software para análise de dados em estatística experimental e em genética quantitativa. *Acta Scientiarum. Agronomy*, *35*(3), 271-276. DOI: https://doi.org/10.4025/actasciagron.v35i3.21251

Cruz, C. D. (2016). Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum*. *Agronomy*, *38*(4), 547-552. DOI: http://dx.doi.org/10.4025/actasciagron.v38i4.32629

Cruz, C. D., Carneiro, P. C. S., & Bhering, L. L. (2021). Biometry in plant breeding. *Crop Breeding and Applied Biotechnology*, *21*(S), 1-11. DOI: http://dx.doi.org/10.1590/1984- 70332021v21Sa18

Cruz, C. D., Ferreira, F. M., & Pessoni, L. A. (2011). *Biometria aplicada ao estudo da diversidade genética*. Viçosa, MG: UFV.

Cruz, C. D., Salgado, C. C., & Bhering, L. L. (2013). *Genômica aplicada*. Viçosa, MG: UFV.

Falconer, D. S. (1981). *Introdução à genética quantitativa*. Viçosa, MG: UFV.

Ferreira, D. F., Oliveira, A. C., Santos, M. X., & Ramalho, M. A. P. (1995). Métodos de avaliação da divergência genética em milho e suas relações com os cruzamentos dialélicos. *Pesquisa Agropecuária Brasileira*, *30*(9), 1189–1194.

Fonseca, J. S., & Martins, G. A. (1996). *Curso de estatística* (6. ed.). São Paulo, SP: Atlas.

Friske, É., Schuelter, A. R., Schuster, I., Marcolin, J., & Silva, M. F. (2018). Genetic diversity of maize lines for traits related to maturity and yield components. *Australian Journal of Crop Science*, *12*(12), 1820-1828. DOI: https://doi.org/10.21475/ajcs.18.12.12.p1005

Garrido-Cardenas, J. A., Mesa-Valle, C., & Manzano-Agugliaro, F. (2017). Trends in plant research using molecular markers. *Planta*, *247*, 543-557. DOI: https://doi.org/10.1007/s00425-017-2829-y

Gupta, P. K., Kumar, J., Mir, R. R., & Kumar, A. (2010). Marker-assisted selection as a component of conventional plant breeding. In J. Janick (Ed.), *Plant breeding reviews*. Hoboken, NJ: John Wiley & Sons. DOI: https://doi.org/10.1002/9780470535486.ch4

Hayman, B. I. (1954). The analysis of variance of diallel tables. *Biometrics*, *10*(2), 235-244. DOI: http://dx.doi.org/10.2307/3001877

Hill, W. G. (2012). Quantitative genetics in the genomics era. *Current Genomics*, *13*(3), 196-206. DOI: https://doi.org/10.2174/138920212800543110

Hongyu, K., Sandanielo, V. L. M., & Junior, G. J. O. (2016). Análise de Componentes Principais: resumo teórico, aplicação e interpretação. *Engineering and Science*, *5*(1), 83-90. DOI: https://doi.org/10.18607/ES201653398

Khaki, S., Khalilzadeh, Z., & Wang, L. (2020). Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach. *PLoS ONE*, *15*(15), 1-13. DOI: https://doi.org/10.1371/journal.pone.0233382

Kulka, V. P., Silva, T. A., Contreras-Soto, R. I., Maldonado, C., Mora, F., & Scapim, C. A. (2018). Diallel analysis and genetic differentiation of tropical and temperate maize inbred lines. *Crop Breeding and Applied Biotechnology*, *18*(1), 31-38. DOI: https://doi.org/10.1590/1984-70332018v18n1a5

Lenaerts, B., Collard, B. C., & Demont, M. (2018). Global survey of rice breeders to investigate characteristics and willingness to adopt alternative breeding methods. *Agriculture & Food Security*, *7*(40), 1-15. DOI: https://doi.org/10.1186/s40066-018-0191-3

Lyzenga, W. J., Pozniak, C. J., & Kagale, S. (2021). Advanced domestication: Harnessing the precision of gene editing in crop breeding. *Plant Biotechnology Journal*, *19*(4), 660-670. DOI: https://doi.org/10.1111/pbi.13576

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829. DOI: https://doi.org/10.1093/genetics/157.4.1819

Moura, L. M., Anjos, R. S. R., Batista, R. O., Vale, N. M., Cruz, C. D., Carneiro, J. E. S., ... Carneiro, P. C. S. (2018). Combining ability of common bean parents in different seasons, locations and generations. *Euphytica*, *214*(181), 1-13. DOI: https://doi.org/10.1007/s10681-018-2259-3

Müller, B. S. F., Pappas, G. J., Valdisser, P. A. M. R., Coelho, G. R. C., Menezes, I. P. P., Abreu, A. G., ... Vianello, R. P. (2015). An operational SNP panel integrated to SSR marker for the assessment of genetic diversity and population structure of the common bean. *Plant Molecular Biology Reporter*, *33*, 1697-1711. DOI: https://doi.org/10.1007s11105-015-0866-x

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., ... Labhane, N. (2018). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, *32*(2), 261-285. DOI: https://doi.org/10.1080/13102818.2017.1400401

Phuke, R. M., Anuradha, K., Radhika, K., Jabeen, F., Anuradha, G., Ramesh, T., ... Kumar, A. A. (2017). Genetic variability, genotype × environment interaction, correlation, and GGE biplot analysis for grain iron and zinc concentration and other agronomic traits in RIL population of sorghum (*Sorghum bicolor* L. Moench). *Frontiers in Plant Science*, *8*(712), 1-13. DOI: https://doi.org/10.3389/fpls.2017.00712

Pimentel, A. J. B., Ribeiro, G., Souza, M. A., Moura, L. M., Assis, J. C., & Machado, J. C. (2013). Comparação de métodos de seleção de genitores e populações segregantes aplicados ao melhoramento de trigo. *Bragantia*, *72*(2), 113-121. DOI: https://doi.org/10.1590/S0006-87052013005000026

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing. Retrieved on Aug. 10, 2021 from https://www.R-project.org/

Rasmussen, S. K. (2020). Molecular genetics, genomics, and biotechnology in crop plant breeding. *Agronomy*, *10*(3), 1-5. DOI: https://doi.org/10.3390/agronomy10030439

Sousa, I. C., Nascimento, M., Silva, G. N., Nascimento, A. C. C., Cruz, C. D., Silva, F. F. E, ... Caixeta, E. T. (2021). Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. *Scientia Agricola*, *78*(4), 1-8. DOI: https://doi.org/10.1590/1678-992X-2020-0021.

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K.C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science*, *61*(2), 839-852. DOI: https://doi.org/10.1002/csc2.20377

Vasconcelos, E. S., Cruz, C. D., Bhering, L. L., & Resende Junior, M. F. R. (2007). Método alternativo para análise de agrupamento. *Pesquisa Agropecuária Brasileira*, *42*(10), 1421-1428.

Wang, J., Li, H., & Zhang, L. (2020). *Genetic mapping and breeding design* (2nd ed.). Beijing, CH: Science Press.

Whang, X., Zhang, Z., Xu, Y., Li, P., Zhang, X., & Xu C. (2019). Using genomic data to improve the estimation of general combining ability based on sparse partial diallel cross designs in maize. *The Crop Journal*, *8*(5), 819-829. DOI: https://doi.org/10.1016/j.cj.2020.04.012

Werle, A. J. K., Ferreira, F. R. A., Pinto, R. J. B., Mangolin, C. A., Scapim, C. A., & Gonçalves, L. S. A. (2014). Diallel analysis of maize inbred lines for grain yield, oil and protein content. *Crop Breeding and Applied Biotechnology*, *14*(1), 23-28. DOI: https://doi.org/10.1590/S1984-70332014000100004