



Genetic variability analyses considering multi-environment trials in maize breeding

Igor Ferreira Coelho · Renan Garcia Malikouski · Jeniffer Santana Pinto Coelho Evangelista · Marco Antônio Peixoto · Rodrigo Silva Alves · Jefferson Fernando Naves Pinto · Edésio Fialho dos Reis · Leonardo Lopes Bhering

Received: 12 July 2020 / Accepted: 5 December 2021 / Published online: 13 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract Clustering techniques are widely adopted in genetic variability assessments. Aiming to understand the genotypes contrasting and complementarity, breeders have used those methodologies to guide them through crosses recommendation or heterotic group's formation. In multi-environment trials (MET) studies, the clustering analyses are under influence of the genotype by environment interaction (GEI) effect. Thus, the goal of this study was to compare clustering analyses dealing with MET data. For this purpose, eight traits were assessed from 84 maize genotypes, whereas Tocher and Unweighted Pair Group Method with Arithmetic Mean clustering analyses were applied. The variance components were estimated through restricted maximum likelihood and genetic values were predicted by best linear unbiased prediction. The significance of the random effects of the statistical model was tested by the likelihood ratio test,

attention was given to grain yield (GY) trait, that presented significant GEI effect. The variance components and genetic parameters varied among environments, considering the grain yield trait, for instance, the heritability varied from 21.51 to 47.65%, and in the joint analysis the heritability was 24.65%, evidencing the importance of joint analysis on MET studies. Finally, it was compared the number of clusters formed in the environments, individual and jointly, by both clustering methods. After these analyses, it was possible to conclude the importance of joint analysis in MET genetic variability study, recommending potential and complementary genetic materials, as the cross 11×65 indicated by both clustering methods.

Keywords Multivariate analysis · Genotype by environment interaction · Clustering analysis · Tocher · UPGMA

I. F. Coelho · R. G. Malikouski · J. S.
P. C. Evangelista · M. A. Peixoto · L. L. Bhering (✉)
Universidade Federal de Viçosa, Viçosa,
Minas Gerais 36570-900, Brazil
e-mail: leonardo.bhering@ufv.br

R. S. Alves
Instituto Nacional de Ciência e Tecnologia do Café/
Universidade Federal de Viçosa, Viçosa,
Minas Gerais 36570-900, Brazil

J. F. N. Pinto · E. F. dos Reis
Universidade Federal de Goiás, Jataí,
Goiás 75804-020, Brazil

Introduction

Maize (*Zea mays* L.) is considered the largest crop worldwide. In the 2018/2019 crop year its world yield production was 1.1 billion tons (USDA, 2019). An advantage of this crop is related to the capacity of being cultivated in a range of different environments and in different seasons. Such aspect leads to

differential responses by genotypes through the environments, and is known as phenotypic plasticity (Marchal et al. 2019) or genotype by environment interaction (GEI) (Smith and Cullis, 2018).

One of the main focus in maize breeding programs is the identification of contrasting genotypes, as the hybridization among them permits the exploration of heterosis, or hybrid vigor (Grigolo et al. 2018). The genetic variability evaluation among genotypes, through multivariate statistical methods, can enable a synthetic description of the genetic relationship between genetic materials or populations (Dias et al. 1997). Then, one of the most important parameters estimated by plant breeders is the genetic dissimilarity of the genotypes selected to be used in future crosses and improve knowledge about the genetic variability of the base germplasm.

Many breeders have been selecting commercial hybrids aiming to begin a breeding program with a population with high favorable alleles frequency (Oliboni et al. 2013; Coelho et al. 2020). Through crosses among these genotypes, it will be making genes shuffling, while generating inter-population hybrids. In this way, the studying of these inter-population hybrids is a key to the breeding program development. Since their variability and complementarity will be the basis of the heterotic groups, and allow the future hybrids to be developed in the maize breeding program.

The multivariate analyses techniques permits the study of the genetic variability and clustering genotypes in an efficient way, by genetically identifying the differences between pairs and among groups of genotypes (Godoy et al. 2006; Cruz et al. 2011). They also seek to minimize intra-group variation and maximize inter-group variation (Rodrigues et al. 2017). Morphological and/or molecular marks associated to the traits of interest can be used for genetic variability analyses (Silva and Neves 2011). The choice of the method for genetic variability evaluation depends on the dataset, its collection procedure and desired precision, since there is no defined parameter for the selection of the best method for genetic variability evaluation for a specific population (Cargnelutti et al. 2008; Cruz et al. 2012).

Among the available methods of genetic variability evaluation, the Mahalanobis distance is applied to measure genetic dissimilarity, which is crucial to hierarchical clustering methods, such as Unweighted

Pair Group Method with Arithmetic Mean (UPGMA), neighbor joining tree, principal component analysis and Tocher method of optimization (Silva et al. 2011; Azevedo et al. 2013). The method selected for the genetic variability study must guarantee the safety of the breeder for selecting the best parents for the crosses. By adopting more than one method for genetic variability evaluation, and if they are not concordant, the selection of the best parents will depend on the most efficient method applied (Cargnelutti et al. 2008).

There are some calculated parameters of the clustering for the determination of the clustering quality in the different methodologies adopted. For instance, the apparent error tax (TEA), by Anderson's discriminant analysis (Anderson, 1958), is applied to measure the clustering adequacy, by quantifying the numbers of different clusters between the predicted and the realized optimized method. Another one is the cophenetic correlation coefficient (CCC), which is regarded to the distortion of the clustering process. In addition, regarding plotting dissimilarity measures, there are the stress and distortion parameters, which use the dissimilarity distance to calculate the value (Cruz et al. 2011).

All the cited parameters provide, in each methodology, relevant information about the clustering quality. Several studies have applied different genetic variability methods (Cantelli et al. 2016, Rocha et al. 2017, Gonçalves et al., 2019, Pereira et al. 2019). However, it is still rare to find clustering studies dealing with GEI through mixed models (Steckling et al., 2017). Thus, this study aimed: (1) to contrast the Tocher and UPGMA clustering methods, using the best linear unbiased prediction (BLUP), considering GEI, in maize breeding; and (2) to indicate, under these perspectives, the better crosses to create potential heterotic groups.

Materials and Methods

Experimental data

The experiments were carried out between January and July 2018 in four sites, all located in the Southwestern Goiás, Brazil (Appendix – Table A1). The climate of the region is wet temperate, with dry

winters and hot summers (Cwa, according to Alvares et al. 2013).

The average annual temperature is around 21.5 °C and average rainfall is between 1400 and 2000 mm year⁻¹. The agricultural practices were based on those used for the maize crop in experimental and commercial production in Brazil (Cruz et al. 2008). The 78 inter-population hybrids (Appendix – Table A2) and the six commercial hybrids (F₁) most planted in the region were used in the experiment, which totaled 84 hybrids assessed.

In each environment (E1, E2, E3, and E4), a trial was conducted in a randomized complete block design with three replications and 44 plants per plot. The plots consisted of four 4-m rows, with a spacing of 0.40-m between plants and 0.45-m between rows. To eliminate the competition effect of each plot with its neighbor, only the two central rows were evaluated. Some cycle-related traits were analyzed: time for tasseling and time for flowering (TA and FL, the days were counted from seeding to 50% of tasseling – when more than half of the tassel releases pollen; and from seeding to 50% of flowering – when the ear starts silking and it is possible to see the silk outside of the husk), followed by the morphology traits: plant height and ear height (PH and EH, in meters – using a three-meter-high ruler and measuring from the ground to the flag leaf). Some yield related traits were also measured: ear length and ear diameter (EL and ED, in centimeters – length: mean of five unhusked ears in a row, mean of five unhusked ears; diameter: measured in the center side by side); and ear yield and grain yield (EY and GY, by the weight of the plot and conversion to hectares, kg ha⁻¹).

Statistical analyses

The statistical analyses followed three steps: (1) restricted maximum likelihood/best linear unbiased prediction (REML/BLUP) to estimate the genotypic values and evaluate the performance of the hybrids over the environments, individually and jointly; (2) multivariate analysis by the Mahalanobis distance, aiming to increase the understanding of the genetic variability among the hybrids; and (3) application of two clustering methodologies, to comparing their performances.

(1) Restricted maximum likelihood/best linear unbiased prediction (REML/BLUP)

The estimation of variance components and prediction of the genotypic values of the traits assessed were carried out through the REML/BLUP procedure, proposed by Patterson and Thompson (1971) and Henderson (1975), respectively. The statistical model associated with the evaluation of hybrids, with one observation per plot and in a single trial, was given by the following equation (Model 1):

$$y = Xr + Zg + e,$$

where y is the vector of phenotypes; r is the vector of fixed effects and comprises the replicate effect (fixed), added to the overall mean; g is the vector of genotypic effects [(assumed as random) ($g \sim N(0, \sigma_g^2)$), where σ_g^2 is the genotypic variance between hybrids; and, e is the vector of residual effects [(random), $e \sim N(0, \sigma_e^2)$, where σ_e^2 is the residual variance]. Uppercase letters refer to the incidence matrices for those effects.

The statistical model associated with the evaluation of hybrids, with one observation per plot and in a multi-environment trial, was given by the following equation (Model 2):

$$y = Xr + Zg + Wi + e,$$

where y is the vector of phenotypes; r is the vector of replication-environment combinations (fixed), added to the overall mean; g is the vector of genotypic effects (assumed as random) [$g \sim N(0, \sigma_g^2)$]; i is the vector of GEI effects (assumed to be random) [$i \sim N(0, \sigma_{gxe}^2)$, where, σ_{gxe}^2 is the GEI variance]; and e is the vector of residual variance [(random), $e \sim N(0, \sigma_e^2)$]. Uppercase letters refer to the incidence matrices for those effects.

For the random effects of the models, significance was tested by the likelihood ratio test (LRT), using chi-square statistics with 1 degree of freedom and 5% probability error type II (Rao 1952), as follows:

$$LRT = -2(\text{Log}L - \text{Log}L_R),$$

where $\text{Log}L$ is the logarithm of the maximum (L) of the restricted likelihood function and $\text{Log}L_R$ is the logarithm of the restricted likelihood function of the reduced models (without the genotypic or GEI effects).

In the model 1, the genotypic and residual variances ($\hat{\sigma}_{g1}^2$, and $\hat{\sigma}_{e1}^2$, respectively) allowed calculating the phenotypic variance ($\hat{\sigma}_{p1}^2$). In the model 2, the genotypic, GEI, and residual variances ($\hat{\sigma}_{g2}^2$, $\hat{\sigma}_{gxe}^2$, and $\hat{\sigma}_{e2}^2$, respectively) allowed calculating the phenotypic variance ($\hat{\sigma}_{p2}^2$). The broad-sense heritability (h_i^2), in percentage (%), coefficient of determination of GEI (c_{gxe}^2) effects, in percentage (%), genotypic correlation between performance in several environments (r_{gloc}), in percentage (%), and selective accuracy (r_{ggi}), in percentage (%), were calculated by the following expressions:

$$\hat{\sigma}_{p1}^2 = \hat{\sigma}_{g1}^2 + \hat{\sigma}_{e1}^2,$$

$$\hat{\sigma}_{p2}^2 = \hat{\sigma}_{g2}^2 + \hat{\sigma}_{gxe}^2 + \hat{\sigma}_{e2}^2,$$

$$h_i^2(\%) = (\hat{\sigma}_{gi}^2 / \hat{\sigma}_{pi}^2) 100,$$

$$c_{gxe}^2(\%) = (\hat{\sigma}_{gxe}^2 / \hat{\sigma}_{p2}^2) 100,$$

$$r_{gloc}(\%) = \left[\hat{\sigma}_{g2}^2 / (\hat{\sigma}_{g2}^2 + \hat{\sigma}_{gxe}^2) \right] 100, \text{ and,}$$

$$r_{ggi}(\%) = \left[\sqrt{1 - (PEV / \hat{\sigma}_{gi}^2)} \right] 100, \text{ where } i \text{ refers to the model (1 or 2), and } PEV \text{ is the prediction error variance extracted from the diagonal of the generalized inverse of the coefficient matrix of the mixed model equations.}$$

The agreement between selected hybrids by the two models was calculated using the Cohen's kappa coefficient (K) (Cohen 1960) given by:

$$K = [(A - C) / (D - C)] 100,$$

where: A is the number of coincident hybrids by the two models, D is the number of hybrids selected, and C is the number of hybrids coincident due to chance ($C = bD$, where: b is the selection intensity).

This analysis considered three different scenarios, with 3, 5 and 10 selected genotypes, which referred to selection intensity (b) of 3.57, 5.95 and 11.90%, respectively.

(2) Multivariate analysis

For multivariate analysis of genetic variability, the Mahalanobis genetic distance (Mahalanobis 1936) was used as a measure of dissimilarity, estimated from

the genetic values predicted by BLUP for each trait that presented significant genotypic effect. The Mahalanobis distance (D_M), between the individuals i and i' , was given by the following equation:

$D_M(i, i') = \left[(u_i - u_{i'})' \Sigma^{-1} (u_i - u_{i'}) \right]^{1/2}$, where i is individual i ; i' is the individual i' ; u_i is the vector of the means of the eight traits evaluated for individual i ; $u_{i'}$ is the vector of the means of the eight traits evaluated for individual i' ; and, Σ is the covariance matrix among traits.

The relative importance of the traits was determined via the interpretation of the eigenvectors associated with the eigenvalues, by principal component analysis (PCA), based on the predicted genotypic values (Rao 1952). It indicates the traits which contributed most for the expression of variation in the population, and the traits liable to be excluded in future works.

(3) Clustering analysis

Based on the Mahalanobis genetic distance calculated among the genotypes, it was studied the variability performance, by adopting two different clustering methodologies. It was adopted the UPGMA, a hierarchical, agglomerative and simple method, which uses the mean distances between genotypes as criteria for their allocation. The number of clusters was determined by the criteria proposed by Milligan and Cooper (1985). The analysis quality was verified based on the stability of the clustering, given by the CCC and distortion and stress levels, by comparing the cophenetic distances matrix with the original matrix.

The second method used was the Tocher method of optimization analysis. It considers, for the mutual exclusive cluster formation, one single criterion, that the intergroup distance is always higher than the intragroup distance. To evaluate the clustering quality, it was proceeded the Fisher discriminant analysis (Fisher 1938), which informed the result found by Tocher and obtained the apparent error tax (AET).

Statistical analyses were performed using the Selegen-REML/BLUP (Resende 2016) and R software (R Development Core Team 2020).

Table 1 Variance components and genetic and non-genetic parameters by individual and joint analyses for the traits: tasseling (TA), flowering (FL), ear height (EH), ear length (EL), ear diameter (ED), ear yield (EY), and grain yield (GY), evaluated in four environments (E)

Component / parameter	TA				FL					
	E1	E2	E3	E4	Joint	E1	E2	E3	E4	Joint
σ_g^2	0.6669*	1.2082*	0.7485*	0.2724*	0.6380*	0.2785*	0.2909 ns	0.6457*	0.2187 ns	0.3546*
σ_{gve}^2					0.0876 ns					0.0271 ns
σ_e^2	1.5779	1.2628	1.8278	1.4457	1.5271	1.2083	3.1012	2.2421	1.5330	1.9985
σ_p^2	2.2449	2.4709	2.5763	1.7181	2.2527	1.4868	3.3921	2.8877	1.7517	2.3802
h^2 (%)	29.71	48.90	29.05	15.85	28.32	18.73	8.58	22.36	12.48	14.90
c_{gve}^2 (%)					3.89					1.14
$\overline{r_{gg}}(\%)$	74.77	86.12	74.25	60.09	90.03	63.94	46.86	68.08	54.75	81.96
$r_{gloc}(\%)$					87.93					92.89
$CV_e(\%)$	2.07	1.83	2.20	1.82	1.98	1.81	2.85	2.43	1.88	2.26
μ	60.76	61.40	61.38	65.93	62.37	60.83	61.73	61.69	66.03	62.57
Component/parameter	PH	EH	E3	E4	Joint	E1	E2	E3	E4	Joint
σ_g^2	0.0080*	0.0078*	0.0080*	0.0036*	0.0068*	0.0050*	0.0042*	0.0040*	0.0036*	0.0042*
σ_{gve}^2					0.0001 ns					0.0001 ns
σ_e^2	0.0075	0.0048	0.0086	0.0098	0.0076	0.0056	0.0044	0.0069	0.0057	0.0056
σ_f^2	0.0155	0.0126	0.0166	0.0134	0.0145	0.0106	0.0085	0.0109	0.0094	0.0098
h^2 (%)	51.75	61.62	48.10	26.64	46.94	47.04	48.79	36.42	38.94	42.46
c_{gve}^2 (%)					83.02					86.80
$\overline{r_{gg}}(\%)$	87.34	91.00	85.76	72.21	95.47	85.27	86.07	79.50	81.04	94.65
$r_{gloc}(\%)$					98.26					98.00
$CV_e(\%)$	3.86	2.85	4.19	4.60	3.85	6.32	5.04	7.45	7.08	6.38
μ	2.24	2.44	2.22	2.16	2.26	1.19	1.31	1.12	1.07	1.17
Component/parameter	EL	ED	E3	E4	Joint	E1	E2	E3	E4	Joint
σ_g^2	5.0000*	10.9906*	5.9987*	5.1485 ns	6.8289*	0.3808*	0.2352	0.1722*	0.2454*	0.2727*
σ_{gve}^2					0.3864 ns					0.0088 ns
σ_e^2	26.4062	16.7364	56.5683	22.7931	30.2350	0.8294	0.8141	1.0347	0.8921	0.8714
σ_f^2	31.4062	27.7270	62.5670	27.9416	37.4503	1.2102	1.0493	1.2068	1.1375	1.1529

Table 1 continued

Component / parameter	TA		FL							
	E1	E2	E3	E4	Joint	E1	E2	E3	E4	Joint
h^2 (%)	15.92	39.64	9.59	18.43	18.23	31.47	22.41	14.26	21.58	23.66
c^2_{gve} (%)					1.03					0.76
$\overline{r_{gg}}$ (%)	60.19	81.44	49.13	63.55	85.03	76.12	68.14	57.70	67.24	88.59
r_{gloc} (%)					94.64					96.88
CV_e (%)	6.48	5.11	9.69	6.51	7.09	3.83	3.70	4.13	4.05	3.89
μ	79.34	80.06	77.58	73.36	77.58	23.77	24.36	24.61	23.34	24.02
Component/parameter	EY	GY								
	E1	E2	E3	E4	Joint	E1	E2	E3	E4	Joint
σ^2_g	1,152,525.72*	699,454.97*	456,344.81*	332,252.90*	512,234.50*	780,381.53*	534,926.72*	309,781.84*	233,962.91*	327,385.98*
σ^2_{gve}					153,363.65*					137,335.45*
σ^2_e	1,175,805.75	1,104,130.99	1,816,036.44	1,330,589.93	1,355,602.60	857,311.63	760,862.99	1,130,640.34	703,525.02	863,359.84
σ^2_f	2,328,331.46	1,803,585.96	2,272,381.25	1,662,842.83	2,021,200.75	1,637,693.16	1,295,789.71	1,440,422.18	937,487.93	1,328,081.27
h^2 (%)	49.50	38.78	20.08	19.98	25.34	47.65	41.28	21.51	24.96	24.65
c^2_{gve} (%)					7.59					10.34
$\overline{r_{gg}}$ (%)	86.38	80.95	65.56	65.44	87.86	85.55	82.36	67.17	70.67	86.89
r_{gloc} (%)					76.96					70.45
CV_e (%)	11.92	10.87	14.40	15.27	13.05	12.25	11.14	14.77	14.99	13.19
μ	9093.16	9670.50	9360.83	7556.03	8918.89	7559.58	7832.41	7198.91	5594.94	7045.01

ns and * : non-significant and significant at 5% probability by the chi-square test. The null hypothesis tested was that the complete model (n) and reduced model (n-1) did not differ from each other. σ^2_g : genotypic variance; σ^2_{gve} : genotype by environment interaction variance; σ^2_e : residual variance; h^2 : broad-sense heritability; c^2 : coefficient of determination genotype by environment interaction; $\overline{r_{gg}}$ (%): mean selective accuracy, in percentage; r_{gloc} (%): genotypic correlation between performances in several environments, in percentage; CV_e (%): coefficient of residual variation, in percentage; μ : phenotypic mean. The kappa coefficient reveals certain degree of agreement among the selected genotypes per environment, and with the jointly analysis (Fig. 1). The three selected genotypes varied from 0.31 to 0.65 of agreement among environments, and E4 presented the lowest agreement with other environments. The joint analysis presented agreement of 0.67 with all environments, except E4, which was 0.33 (Fig. 1a). The five selected genotypes presented agreement values from 0.15 to 0.57. The agreement values of E1 and E2 and the joint analysis and E1 figured out among the highest values (0.57), while the lowest agreement values (0.15) were found for E3 with E2 and E1, and E4 with all the other environments (Fig. 1b). The ten selected genotypes presented the lowest values (0.21) of agreement, between E3 and E4, for instance. The E1 and joint analysis exhibited the highest agreement value, 0.77 (Fig. 1c).

Results

Table 1 (and Appendix—Table A3) presents the variance components and genetic and non-genetic parameters estimates for all traits, in each environment and in the joint analysis. The LRT results were also presented in this table and shows which traits are significant or not in each analysis, individually and jointly. The FL trait had no significant genotypic effect ($p < 0.05$) on E1 and E4, which is also observed for the EL trait (E3), which has no significant genotypic effect ($p < 0.05$). In the joint analysis, all traits have significant genotypic effect, unlike the GEI effect, which was significant just for EY and GY traits ($p < 0.05$).

The GY trait is going to be the focus of this study, to demonstrate the variation over environments. For GY, the genotypic variance ranged from 233,962.91, in E4, to 780,380.53, in E1, whilst in the joint analysis, the genotypic variance was 327,385.9805. The residual variance presented values from 703,525.02 (E4) to 1,130,640.3391 (E3). In the joint analysis, this component was 863,359.84. The joint analysis includes the GEI effects in the model, which was 137,335.45. The variation of heritability along the environments doubled between E3 (21.51%) and E1 (41.28%). The heritability estimated by the joint analysis was 24.65%, similarly to E3 and E4. The mean selective accuracy ranged from 67.17 (E3) to 85.55% (E1) in the individual analyses, while in the joint analyses, it was 86.89%, thus overcoming the highest accuracy obtained in the individual analyses. The phenotypic means are also evidenced in Table 1, ranging from 5,594.94 (E4) to 7832.4135 (E2), while the overall mean was 7045.01.

Besides, the genetic material rankings (Appendix – Table A4) per environment and for the joint analysis evidence the high performance of some genetic materials developed in this study, compared to the commercial hybrids. At least five genetic materials figure out among the ten best genetic materials in all environments and in the joint analysis. In some cases, they appear among the five selected genetic materials.

The results of the principal component analysis are presented in Table 2. Among all environments and the joint analysis, the first eigenvalue (highlighted) presented at least 31% of accumulated importance. Their respective highest eigenvectors (highlighted) were the last, representing the GY trait, in E1, E2 and joint, or

the second last, regarding the EY trait, in both E3 and E4, all around 0.50.

The differences among the clustering by individual analyses and by the joint analysis, utilizing the two different clustering methodologies, are presented in Table 3 (and Appendix – Table A5). Both methodologies used the Mahalanobis distance based on the predicted genotypic values. For Tocher clustering, it is noted the variation of groups, from 9 (E4) to 18 groups formed (Joint analysis). The UPGMA method used the average distance to cluster the genetic materials and form the groups. Its number of groups varied around 10 groups, except the E3, containing 8 groups. This method presents the CCC, stress and distortion, in percentage. The CCC ranged from 62.69 to 74.53%, through the environments, and the result of the joint analysis was 69.98%. The stress was below 13.07% for all environments, individually and jointly. Distortion was around 35% for all results.

Discussion

Considering the individual analyses for the GY trait, it was observed that the residual variance in E3 was 67% higher than in E4, while the genotypic variance in E4 accounts for 29% of this component in E1. The heritability classification ranged from low, in E3 and E4, to moderate, in E1 and E2 (Resende and Duarte 2007). The heritability almost more than doubled from E3 (21.51%) to E1 (47.65%). In the joint analysis, it was observed a slightly higher value, 24.65%. The coefficients of residual variation were used to infer about the experimental quality. Variation is observed among the environments, as it raises almost 4% from E2 to E4, and the joint analysis is between these extreme values (11.14 and 14.99%), namely, 13.19%, which is inferred as high experimental quality (Coelho et al. 2020). This discrepancy among the coefficients of residual variation reinforces the advantage of the joint analysis, which maintained a low coefficient of residual variation.

The agreement index also demonstrates the importance of the joint analysis of multi-environment trials data, which presents low agreement values in all three scenarios, mainly when it is related to E4, reaching 0.20 of agreement when five hybrids are selected, in all other environments. Considering the agreement coefficients, it is important to highlight that the selection

Table 2 Eigenvalues and eigenvectors by the principal component analysis for the traits: tasseling (TA), flowering (FL), plant height (PH), ear height (EH), ear length (EL), ear diameter (ED), ear yield (EY), and grain yield (GY). The square root (SR), square root, in percentage (%) of importance, and accumulated eigenvalues importance (Ac) among all, in

percentage (%), are presented for each eigenvalue, per environment. The bold numbers represent, in Ac (%) column, the values until achieve near 80% of importance, and, in the eigenvectors, represent the highest trait value of the three first eigenvalues

	Eigenvalues			Eigenvectors							
	SR	SR (%)	Ac (%)	MF	FF	PH	EH	EL	ED	EY	GY
E1											
1	3.09	38.60	38.60	- 0.1131	- 0.0936	0.4189	0.3736	0.3141	0.2581	0.4971	0.5011
2	1.78	22.27	60.87	0.6813	0.6968	0.0446	0.0801	0.0330	- 0.1202	0.1225	0.1066
3	1.34	16.77	77.64	- 0.1490	0.0848	0.4966	0.6055	- 0.0949	- 0.3898	- 0.3220	- 0.3046
4	0.97	12.17	89.81	- 0.0053	- 0.1309	- 0.2441	- 0.0352	0.7225	- 0.6301	0.0470	0.0298
5	0.52	6.46	96.28	0.2388	- 0.0540	0.0744	0.0855	0.5652	0.5792	- 0.3562	- 0.3813
6	0.17	2.11	98.38	- 0.6567	0.6849	0.0113	- 0.1995	0.2128	0.1097	- 0.0355	- 0.0355
7	0.12	1.54	99.92	- 0.1108	0.0974	- 0.7146	0.6622	- 0.0641	0.1502	0.0122	0.0476
8	0.01	0.08	100.0	0.0168	0.0020	0.0094	- 0.0231	0.0211	0.0016	- 0.7098	0.7035
E2											
1	3.19	39.84	39.84	0.1668	- 0.0173	0.3833	0.3068	0.3914	0.2572	0.5102	0.5012
2	1.52	18.98	58.82	0.6596	0.6513	- 0.1729	- 0.2376	- 0.1111	0.2001	0.0289	0.0354
3	1.37	17.10	75.93	0.1272	0.3340	0.4820	0.6191	- 0.0606	- 0.2956	- 0.2704	- 0.3042
4	0.95	11.91	87.83	- 0.1143	- 0.0980	0.2861	0.0753	- 0.5696	0.7427	- 0.0745	- 0.0907
5	0.44	5.53	93.36	- 0.6588	0.6214	- 0.0010	- 0.0930	0.3336	0.2367	- 0.0255	- 0.0576
6	0.33	4.11	97.47	0.2711	- 0.2604	0.0386	- 0.0573	0.6286	0.3937	- 0.3934	- 0.3861
7	0.18	2.22	99.69	- 0.0153	0.0051	- 0.7117	0.6698	0.0014	0.2094	0.0027	0.0283
8	0.03	0.31	100.0	- 0.0163	0.0192	0.0307	0.0048	- 0.0007	0.0017	- 0.7105	0.7026
E3											
1	2.54	31.69	31.69	0.0386	- 0.0133	0.4209	0.3349	0.2656	0.2121	0.5454	0.5441
2	2.14	26.72	58.41	0.5767	0.6297	0.2764	0.2942	0.0762	- 0.0149	- 0.2227	- 0.2286
3	1.38	17.22	75.63	- 0.3690	- 0.2396	0.4434	0.5677	- 0.2716	- 0.3681	- 0.1984	- 0.1972
4	0.97	12.14	87.76	0.0496	- 0.0375	0.0953	0.0596	- 0.6826	0.7180	- 0.0397	- 0.0218
5	0.68	8.44	96.20	- 0.2621	- 0.1517	0.1635	0.0531	0.6129	0.5387	- 0.3302	- 0.3224
6	0.18	2.21	98.42	- 0.1062	0.1223	- 0.7094	0.6643	0.0596	0.1110	- 0.0328	0.1109
7	0.11	1.32	99.74	- 0.6659	0.7092	0.0801	- 0.1679	- 0.0693	0.0248	0.0147	0.1154
8	0.02	0.26	100.00	0.0622	- 0.0593	0.0660	- 0.0560	0.0050	- 0.0249	- 0.7083	0.6948
E4											
1	2.87	35.82	35.82	- 0.0403	- 0.0850	0.3289	0.2814	0.3903	0.3171	0.5241	0.5255
2	2.04	25.45	61.28	0.6670	0.6492	- 0.2230	- 0.1230	0.1842	0.1632	0.0727	0.0538
3	1.54	19.25	80.53	0.1745	0.2146	0.5623	0.6495	- 0.0486	- 0.0195	- 0.3067	- 0.2978
4	0.89	11.16	91.68	- 0.0005	- 0.0248	0.0541	- 0.0761	- 0.5931	0.7973	0.0160	- 0.0537
5	0.39	4.88	96.57	- 0.0816	- 0.2107	- 0.1151	- 0.0761	0.6706	0.4751	- 0.3466	- 0.3665
6	0.19	2.40	98.97	- 0.2412	0.1232	- 0.6921	0.6552	- 0.0520	0.0812	- 0.0201	0.0934
7	0.06	0.79	99.76	- 0.6712	0.6785	0.1721	- 0.2024	0.0819	0.0554	- 0.0817	0.0463
8	0.02	0.24	100.00	0.0876	- 0.0687	0.0345	- 0.0455	- 0.0209	0.0368	- 0.7060	0.6958
Joint											
1	3.17	39.62	39.62	0.1415	0.0501	0.4068	0.3272	0.3847	0.2435	0.4984	0.4992
2	1.89	23.58	63.19	0.6212	0.6976	0.0953	0.1661	0.0212	- 0.0882	- 0.2024	- 0.2039

Table 2 continued

	Eigenvalues			Eigenvectors							
	SR	SR (%)	Ac (%)	MF	FF	PH	EH	EL	ED	EY	GY
3	1.35	16.91	80.11	- 0.3311	- 0.1002	0.521	0.6401	- 0.1623	- 0.1902	- 0.2646	- 0.2582
4	1.08	13.50	93.61	0.0562	0.035	0.1546	- 0.0293	- 0.5823	0.7934	- 0.0208	- 0.0436
5	0.27	3.33	96.93	- 0.087	- 0.0848	- 0.0523	0.0169	0.6917	0.4997	- 0.3641	- 0.3487
6	0.13	1.62	98.56	- 0.4353	0.4454	- 0.5977	0.4548	- 0.0333	0.1306	0.0731	0.1566
7	0.11	1.34	99.90	- 0.5327	0.5419	0.4104	- 0.4963	0.08	- 0.0245	0.0284	0.0093
8	0.01	0.10	100.00	0.0225	- 0.0164	0.0415	- 0.0371	- 0.0134	- 0.0001	- 0.7081	0.7032

Table 3 Clustering analysis methodologies, based on the Mahalanobis distance of the predicted genotypic values of the evaluated traits. Number of clusters and the quantity of hybrids composing each one, for each environment individually (E1, E2, E3, E4), and all (ALL) simultaneously. The appearance

error tax (*AET*), in percentage, for Tocher analysis, and the cophenetic correlation coefficient (*CCC*), stress and distortion (*Dist.*), in percentage, for UPGMA analysis, are presented. The number of cluster (*N Clust*) is also presented for each methodology

Tocher	E1	E2	E3	E4	ALL	UPGMA	E1	E2	E3	E4	ALL
AET(%)	20.34	9.52	9.52	15.48	17.86	CCC (%)	65.23	62.69	74.53	61.47	69.98
						Stress(%)	12.19	13.07	11.38	11.87	11.13
						Dist.(%)	34.91	36.16	33.74	34.48	33.36
N Clust	17	12	15	9	18	NClust	10	10	8	10	10
1	53	70	63	71	48	1	69	61	72	59	63
2	5	2	2	3	9	2	3	2	1	8	7
3	2	2	2	2	4	3	3	8	2	4	5
4	2	2	2	3	4	4	1	1	2	3	2
5	4	1	2	1	2	5	1	2	2	3	1
6	3	1	2	1	2	6	2	3	2	1	2
7	3	1	2	1	2	7	2	3	2	1	1
8	2	1	2	1	2	8	1	2	1	2	1
9	2	1	1	1	2	9	1	1		2	1
10	1	1	1		1	10	1	1		1	1
11	1	1	1		1						
12	1	1	1		1						
13	1		1		1						
14	1		1		1						
15	1		1		1						
16	1				1						
17	1				1						
18					1						

by the joint analysis and the individual selection (per environment) led to the selection of different hybrids. It demonstrates the necessity of considering the joint

analysis for more accurate genetic selection and recommendation, since the joint analysis allows

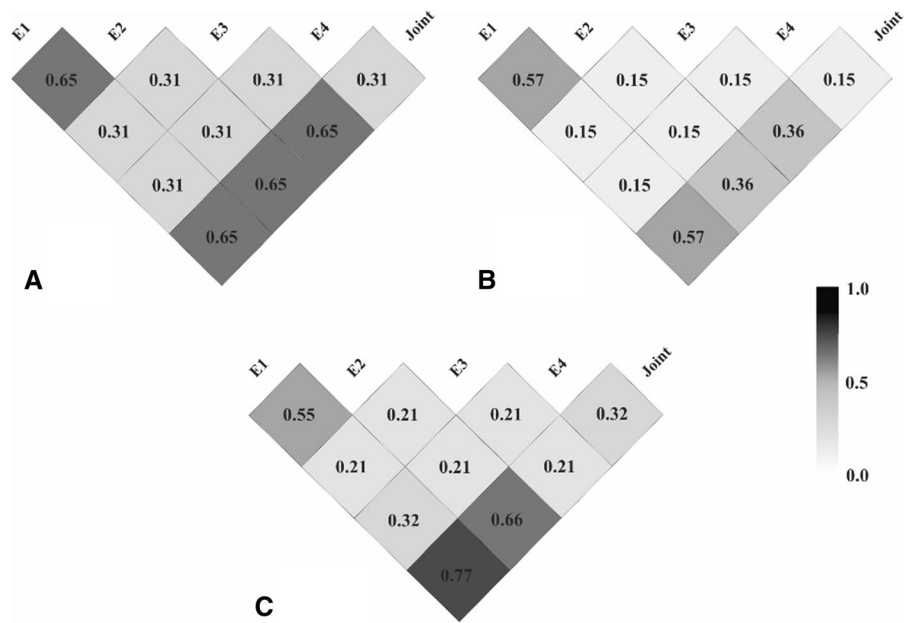


Fig. 1 Kappa coefficients for selection of 3 (A), 5 (B), and 10 (C) genotypes

evaluating the GEI. The GEI significance was confirmed by the LRT for EY and GY.

The different clustering methods provided distinct genetic variability results. Some studies also found differences among methods, when their clustering results were compared (Bhering et al. 2015; Oliveira et al. 2016; Silva et al. 2017). This is due to the multiple ways of determining the number of clusters by hierarchical methods and the Tocher methodology follows a pre-defined fixed methodology, making it impossible to establish a relationship between these two clustering methods. In the agglomerative methodologies, the first clusters present the highest number of genotypes.

The principal component analysis (PCA) allows genotype projection over a 2D-plot by obtaining coordinates and infers about the importance of the traits. Therefore, it indicates which traits can be excluded in the future analysis due to redundancy. The discussion of PCA projection in 2D or 3D-plots is not the aiming of this study, but is worth to remember that the minimum acceptable value of the total variation explained by the eigenvalues is 80% (Cruz et al. 2012).

The interpretation of eigenvalues, from the PCA, corroborates that EY is the most redundant trait and liable to be excluded in future works. The presence of the highest eigenvalues in the eigenvalues which

carry the lowest variance explains EY redundancy. It is also explained by the high correlation found in other morphological traits, such as PH.

The different clustering methods provided distinct results. Some studies evidence some differences in the results among the clustering methodologies (Bhering et al. 2015; Oliveira et al. 2016; Silva et al. 2017), since the variation of determination of the cluster numbers in the hierarchical methods and Tocher follows a pre-fixed methodology. This difference among the methodologies for clustering determination makes it impossible to establish a relationship between these clustering methods. The similarity between them is the fact of being agglomerative, which makes the first clusters contain the higher number of genotypes.

The environment E3 presented the best clustering results in both clustering methods, since it is indicated by the “quality parameters”, and because of its experimental quality. The high experimental quality of E3 conducted to the least experimental error, when compared to the other environments, which provided results that corroborated the clustering quality. Reduced stress values, distortion and AET, besides the high CCC, were the parameters used in this study. According to Cruz et al. (2011), stress and distortion values should be below 20%; AET, inferior to 5%; CCC, on the contrary, should be higher than 80%.

These referential values are close to those found in this study. E1 and E4 presented results opposite to those at E3 (stress, distortion and AET above, and CCC inferior), due to their experimental quality.

Even the joint analysis does not present the best results referring to the clustering quality. It is important to point out the importance of adopting the joint analysis in genetic variability studies, since it considers the GEI effects. It was observed, by the joint analysis results, that the environmental conditions are considered in the model and influence to encounter average values. The best and worst environments are considered, but the average among all is given by the joint analysis. Some studies considered the GEI effects and reinforces the importance of GEI over some traits, as confirmed in EY and GY traits, in this study. However, few studies have discussed and compared the impact of GEI effect on genetic variability studies (Bueno et al., 2013; Steckling et al., 2017), which highlights the importance of considering this effect on genetic variability studies in maize.

As defined by Melchinger and Gumber (2015): “a heterotic group denotes a group of related or unrelated genotypes from the same or different populations”. In this way, understanding the genetic diversity would allow the breeder to access some information regarding to the genotypes relationship. These authors also point out: “By comparison, the term heterotic pattern used herein refers to a specific pair of two heterotic groups, which express high heterosis and consequently high hybrid performance in their cross”. Where, it would be helpful to understand their relationship to make heterotic groups and as mentioned, to achieve: “high heterosis and consequently high hybrid performance in their cross”.

This study evidences the importance of the joint analysis instead of individual analyses. GEI must be considered in genetic variability evaluation, since it is frequently significant. GY trait, as observed, was the most important eigenvector in the first eigenvalue for E1, E4 and, mainly, in the joint analysis, thus confirming the importance of considering the GEI effect to develop the clustering analysis. The consideration of the GEI effects enriches the reliability of genetic variability evaluation and improves the clustering indexes, for Tocher or UPGMA analysis.

The recommended crosses are based on genetic distance, which is provided by the clusters containing

them, and on the genetic material ranking of the joint analysis, based on the GY trait. In other words, the farther their genetic distance, the more complementary they are when combined, which can result in offspring with more favorable genes. Considering the UPGMA, the 11 × 65 cross was indicated, where, both are present in the ten selected genetic materials and have good genetic divergence. 11 × 62 and 11 × 32 would be other cross possibilities, being in different clusters, and well ranked for the GY trait. Considering Tocher, 11 × 20, 11 × 7, 11 × 65, 11 × 45, 65 × 20, 65 × 7 and 65 × 45 are among the ten selected genetic materials and present good genetic divergence, as they figured out in different clusters.

To conclude, both methodologies presented similar crosses recommendations. The cross between the inter-population hybrids 11 and 65 have high potential to be good population founders, indicated by both methodologies. The next step could be the begin of two populations starting from these two inter-population hybrids indicated by the analyses as potential crosses.

Acknowledgements To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Financing Code 001), to Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and to Instituto Nacional de Ciência e Tecnologia do Café (INCT Café) for financial support.

Appendix

See Tables 4, 5, 6, 7 and 8

Table A1 Location of sites, with the corresponding geographic coordinates and altitudes

Site	Location	Coordinates	Altitude
E1	Jataí-Goiás	17°55'27,63" S 51°42'45,51" W	685 m
E2	Jataí-Goiás	17°50'04,70" S 51°29'53,29" W	865 m
E3	Caiapônia-Goiás	17°30'25,06" S 51°54'35,68" W	818 m
E4	Mineiros-Goiás	17°23'4,72" S 52°19'39,01" W	938 m

Table A2 Incomplete diallel scheme involving thirteen F₂ population (used as parents) and their respectively crosses (H), the bold letters refer to the F₂ population materials (used as parents) which were auto-pollinated, and are plotted in the diagonal of the scheme

1 ^a	2	3	4	5	6	7	8	9	10	11	12	13
1 ^a	H1,2 ^b	H1,3	H1,4	H1,5	H1,6	H1,7	H1,8	H1,9	H1,10	H1,11	H1,12	H1,13
2		H2,3	H2,4	H2,5	H2,6	H2,7	H2,8	H2,9	H2,10	H2,11	H2,12	H2,13
3			H3,4	H3,5	H3,6	H3,7	H3,8	H3,9	H3,10	H3,11	H3,12	H3,13
4				H4,5	H4,6	H4,7	H4,8	H4,9	H4,10	H4,11	H4,12	H4,13
5					H5,6	H5,7	H5,8	H5,9	H5,10	H5,11	H5,12	H5,13
6						H6,7	H6,8	H6,9	H6,10	H6,11	H6,12	H6,13
7							H7,8	H7,9	H7,10	H7,11	H7,12	H7,13
8								H8,9	H8,10	H8,11	H8,12	H8,13
9									H9,10	H9,11	H9,12	H9,13
10										H10,11	H10,12	H10,13
11											H11,12	H11,13
12												H12,13
13												

^aParents (F₂ population)—1: P4285; 2: 30F53; 3: P3646; 4: 30K75; 5: RB9110; 6: RB9210; 7: CD384; 8: AS1633; 9: AS1598; 10: 2B587; 11: BM709; 12: DKB390; 13: UFG Rustic Material

^bInter-population Hybrids

Table A3 Likelihood ratio test (LRT) for the genetic effects (individual analyses); and for the genetic and genotype by environment interaction effects (joint analysis) for the traits: tasseling (TA), flowering (FF), plant height (PH), ear height (EH), ear length (EL), ear diameter (ED), ear yield (EY), and grain yield (GY) evaluated in four environments (E)

Trait	Individual analyses				Joint analysis	
	σ_{gE1}^2	σ_{gE2}^2	σ_{gE3}^2	σ_{gE4}^2	σ_g^2	σ_{gxe}^2
TA	19.81*	- 54.83*	- 18.96*	- 5.79*	103.80*	2.34 ^{ns}
FL	8.02*	- 1.74 ^{ns}	- 11.33*	- 3.64 ^{ns}	47.50*	0.14 ^{ns}
PH	61.99*	- 92.29*	- 52.93*	- 15.99*	228.38*	0.19 ^{ns}
EH	50.47*	- 54.57*	- 29.76*	- 34.09*	199.50*	0.17 ^{ns}
EL	5.70*	- 35.35*	- 2.17 ^{ns}	- 7.55*	63.87*	0.13 ^{ns}
ED	22.11*	- 11.38*	- 4.71*	- 10.15*	94.14*	0.08 ^{ns}
EY	56.18*	- 33.50*	- 9.18*	- 8.72*	80.00*	7.97*
GY	50.79*	- 38.43*	- 10.43*	- 13.57*	71.81*	14.24*

σ_{gEi}^2 : genetic variance in the environment *i*; σ_g^2 : genetic variance considering all environments, and σ_{gxe}^2 : genetic by environment interaction variance. ^{ns} and * : non-significant and significant at 5% probability by the chi-square test. The null hypothesis tested was that the complete model (n) and reduced model (n - 1) did not differ from each other.

Table A4 Genetic values ranking by individual and joint analyses

Ranking	Environments				Joint
	E1	E2	E3	E4	
1	80	80	80	80	80
2	83	83	83	83	83
3	82	79	79	65	79
4	79	20	7	7	82
5	2	11	82	82	11
6	11	82	11	11	20
7	7	84	20	79	7
8	26	7	65	84	65
9	56	19	73	45	84
10	20	1	3	16	45
11	1	65	2	20	62
12	45	62	62	28	32
13	32	32	45	32	1
14	22	26	71	70	3
15	3	81	32	62	56
16	21	71	9	3	2
17	16	16	28	6	22
18	84	3	70	56	74
19	9	56	81	74	6
20	62	55	54	73	16
21	8	8	26	81	8
22	67	2	8	2	70
23	70	45	84	66	21
24	65	21	56	22	26
25	6	31	22	10	71
26	60	35	69	60	75
27	81	22	36	21	60
28	34	43	16	1	9
29	73	34	19	8	10
30	10	69	74	67	67
31	30	67	35	30	35
32	38	10	4	49	28
33	31	68	30	75	54
34	75	74	39	9	73
35	51	47	34	71	13
36	74	6	10	34	81
37	19	5	15	19	19
38	35	24	21	55	31
39	13	54	1	78	69
40	71	4	13	24	49
41	54	75	24	14	34
42	42	60	55	43	39
43	28	28	67	47	51

Table A4 continued

Ranking	Environments				Joint
	E1	E2	E3	E4	
44	37	73	5	37	30
45	5	49	60	26	24
46	55	9	51	13	4
47	49	39	66	4	14
48	58	51	31	41	55
49	76	37	23	17	43
50	14	76	75	77	5
51	43	70	49	5	38
52	78	13	38	31	47
53	66	38	61	72	58
54	77	57	6	36	36
55	68	23	47	38	15
56	15	42	17	61	66
57	39	15	58	27	27
58	17	30	42	15	17
59	23	27	27	23	42
60	47	58	43	76	37
61	69	52	48	54	23
62	61	14	76	52	68
63	24	44	14	35	57
64	52	12	37	48	29
65	64	36	46	68	76
66	4	50	50	29	48
67	41	48	57	58	61
68	27	29	52	51	44
69	36	61	59	39	52
70	44	59	44	42	46
71	46	66	29	57	64
72	48	64	12	63	59
73	59	78	68	50	77
74	29	41	78	69	78
75	50	72	63	59	50
76	57	77	77	64	41
77	12	63	41	46	12
78	72	46	53	44	72
79	33	17	72	40	53
80	63	40	64	12	63
81	53	53	40	53	40
82	18	18	18	33	33
83	40	33	33	18	18
84	25	25	25	25	25

The non-commercial hybrids are presented in bold. The dashed line bounds the top 10 genotypes

Table A5 Clustering analyses via Tocher (A) and UPGMA (B), by individual and joint analyses

Genetic Materials	(A)				
	Tocher				
	E1	E2	E3	E4	Joint
1	2	1	1	5	1
2	5	1	1	1	1
3	3	1	1	1	1
4	1	1	1	1	1
5	6	1	1	1	1
6	9	1	1	1	2
7	1	1	1	1	1
8	17	1	6	1	13
9	1	1	1	1	1
10	1	1	1	1	1
11	2	4	1	1	6
12	1	1	4	1	4
13	1	1	1	1	1
14	1	12	1	1	2
15	1	1	1	1	1
16	1	1	1	4	1
17	9	1	1	1	2
18	13	2	10	1	9
19	6	1	1	1	1
20	1	1	2	1	1
21	1	1	1	1	1
22	2	1	1	1	1
23	1	1	1	1	8
24	5	1	1	1	3
25	10	10	12	1	10
26	11	1	14	1	8
27	5	1	1	1	2
28	1	1	1	1	1
29	1	2	8	1	1
30	3	1	1	1	1
31	1	1	1	1	3
32	2	1	1	8	2
33	4	1	1	1	14
34	1	1	1	1	1
35	1	1	1	7	1
36	1	8	1	3	2
37	1	1	1	1	1
38	7	1	1	1	1
39	1	1	1	1	1
40	1	1	7	1	11
41	2	1	1	1	9

Table A5 continued

Genetic Materials	(A)				
	Tocher				
	E1	E2	E3	E4	Joint
42	6	1	1	1	1
43	1	1	1	1	1
44	1	1	1	3	3
45	1	1	1	1	1
46	1	1	1	1	1
47	1	1	1	1	1
48	1	1	1	1	1
49	1	1	15	1	1
50	1	5	1	1	7
51	5	7	1	1	3
52	1	1	1	1	1
53	1	1	5	2	7
54	1	1	1	1	1
55	1	1	1	1	1
56	7	1	1	1	6
57	1	1	1	1	1
58	1	1	1	1	1
59	1	1	1	1	1
60	1	1	1	1	1
61	8	1	5	1	2
62	1	1	9	2	2
63	1	1	1	1	2
64	1	1	1	1	1
65	1	1	8	1	12
66	1	1	1	2	1
67	1	1	1	1	1
68	1	1	4	1	1
69	1	1	1	1	1
70	1	1	1	6	5
71	1	4	1	1	1
72	1	1	13	9	4
73	15	1	1	1	1
74	7	11	1	1	1
75	1	1	1	1	1
76	1	1	7	1	1
77	1	6	1	1	4
78	1	1	11	1	4
79	8	1	3	1	16
80	16	3	3	4	17
81	4	1	1	4	1
82	1	1	1	1	15
83	12	3	2	1	18

Table A5 continued

(A)					
Genetic Materials	Tocher				
	E1	E2	E3	E4	Joint
84	14	9	6	1	5
(B)					
Genetic Materials	UPGMA				
	E1	E2	E3	E4	Joint
1	1	1	1	6	1
2	1	1	1	1	1
3	7	3	1	1	1
4	1	3	1	1	1
5	1	1	1	1	1
6	1	1	1	2	2
7	1	1	1	1	1
8	10	5	6	1	5
9	1	1	1	1	1
10	1	1	1	1	1
11	1	7	1	4	1
12	1	1	1	1	1
13	1	7	1	1	1
14	1	4	1	3	2
15	2	1	1	5	1
16	1	1	1	5	1
17	4	1	1	2	2
18	8	6	5	1	3
19	1	1	1	1	1
20	1	1	4	1	1
21	1	1	1	1	1
22	1	1	1	1	1
23	1	1	1	1	1
24	3	1	1	1	1
25	3	9	5	1	9
26	1	1	2	1	1
27	1	3	1	1	2
28	1	1	1	1	1
29	1	6	7	1	1
30	7	1	1	1	1
31	1	1	1	1	1
32	1	2	1	4	2
33	2	1	1	2	7
34	1	3	1	1	1
35	1	1	1	9	1
36	1	1	1	2	2
37	1	1	1	1	1

Table A5 continued

(B)					
Genetic Materials	UPGMA				
	E1	E2	E3	E4	Joint
38	1	1	1	1	1
39	1	1	1	1	1
40	1	1	1	1	1
41	1	1	1	1	3
42	1	1	1	1	1
43	1	3	1	1	1
44	1	1	1	7	1
45	1	1	1	1	1
46	1	1	1	1	1
47	1	1	1	1	1
48	1	3	1	1	1
49	1	1	6	1	1
50	1	3	1	1	3
51	3	10	1	1	1
52	1	1	1	1	1
53	1	1	1	8	3
54	1	1	1	1	1
55	1	1	1	1	1
56	1	1	1	1	1
57	1	1	1	1	1
58	1	1	1	1	1
59	1	1	1	2	1
60	1	1	1	1	1
61	1	1	1	2	1
62	1	1	1	2	2
63	1	1	1	2	1
64	1	6	1	1	3
65	1	3	7	1	6
66	1	1	1	8	1
67	1	1	1	1	1
68	1	1	1	1	1
69	1	1	1	1	1
70	1	1	1	3	4
71	1	7	1	1	1
72	1	1	8	10	1
73	1	1	1	1	1
74	6	2	1	1	1
75	1	1	1	1	1
76	1	1	1	3	1
77	1	1	1	1	1
78	1	1	1	1	1
79	1	1	3	3	6

Table A5 continued

Genetic Materials	UPGMA				
	E1	E2	E3	E4	Joint
80	9	8	3	5	10
81	2	1	1	9	1
82	1	1	1	1	1
83	6	8	4	4	8
84	5	5	1	1	4

The genetic material cells which are different of cluster one, are highlighted in bold

References

- Alvares CA, Stape JL, Sentelhas PC et al (2013) Köppen's climate classification map for Brazil. *Meteorol Zeitschrift* 22:711–728. <https://doi.org/10.1127/0941-2948/2013/0507>
- Anderson TW (1958) An introduction to multivariate statistical analysis. Wiley, New York
- Azevedo AM, Andrade Júnior VC de, Oliveira CM de, et al (2013) Seleção de genótipos de alface para cultivo protegido: divergência genética e importância de caracteres. *Hortic Bras* 31:260–265. <https://doi.org/10.1590/S0102-05362013000200014>
- Bhering LL, Peixoto LA, Ferreira Leite NLS, Laviola BG (2015) Molecular analysis reveals new strategy for data collection in order to explore variability in *Jatropha*. *Ind Crops Prod* 74:898–902. <https://doi.org/10.1016/j.indcrop.2015.06.004>
- Bueno RD, Borges LL, Arruda KMA et al (2013) Genetic parameters and genotype x environment interaction for productivity, oil and protein content in soybean. *African J Agric Res* 8:4853–4859. <https://doi.org/10.5897/AJAR2013.6924>
- Cantelli DAV, Hamawaki OT, Rocha MR et al (2016) Analysis of the genetic divergence of soybean lines through hierarchical and Tocher optimization methods. *Genet Mol Res* 15:1–13. <https://doi.org/10.4238/gmr.15048836>
- Cargnelutti AF, Ribeiro ND, Reis RCP et al (2008) Comparação de métodos de agrupamento para o estudo da divergência genética em cultivares de feijão. *Ciência Rural* 38:2138–2145
- Coelho IF, Alves RS, Rocha JRASC et al (2020) Multi-trait multi-environment diallel analyses for maize breeding. *Euphytica* 216:144. <https://doi.org/10.1007/s10681-020-02677-9>
- Cruz JC, Karam D, Monteiro MAR, Magalhães PC (2008) A cultura do milho Embrapa Milho e Sorgo, Sete Lagoas, MG
- Cruz CD, Ferreira FM, Pessoni LA (2011) *Biometria aplicada ao estudo da diversidade genética Visconde do Rio Branco Suprema*, p 620
- Cruz CD, Regazzi AJ, Carneiro PCS (2012) *Modelos biométricos aplicados ao melhoramento genético*, 3rd edn Editora UFV, Viçosa - MG
- Dias LAS, Kageyama PY, Castro GCT (1997) Divergência genética multivariada na preservação de germoplasma de cacau (*Theobroma cacao* L.). *Agrotropica* 9
- Fisher RA (1938) The statistical utilization of multiple measurements. *Ann Eugen* 8:376–386. <https://doi.org/10.1111/j.1469-1809.1938.tb02189.x>
- Godoy CV, Koga LJ, Canteri MG (2006) Diagrammatic scale for assessment of soybean rust severity. *Fitopatol Bras* 31:63–68. <https://doi.org/10.1590/S0100-41582006000100011>
- Gonçalves GMC, Gonçalves MMC, Medeiros AM et al (2019) Genetic dissimilarities between fava bean accessions using morphoagronomic characters. *Rev Caatinga* 32:1125–1132. <https://doi.org/10.1590/1983-21252019v32n430rc>
- Grigolo S, Fioreze ACCL, Denardi S, Vacari J (2018) Implicações da análise univariada e multivariada na dissimilaridade de acessos de feijão comum. *Rev Ciências Agroveterinárias* 17:351–360. <https://doi.org/10.5965/223811711732018351>
- Hallauer AR, Carena JM, Miranda Filho JB (2010) *Quantitative genetics in maize breeding*. Springer, New York, p 500
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447
- Mahalanobis PC (1936) On the generalised distance in statistics. *Proc Natl Inst Sci India* 12:49–55
- Marchal A, Schlichting CD, Gobin R et al (2019) Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. *Plant Genome* 13:281–288. <https://doi.org/10.1017/S0021859605005587>
- Melchinger A, Gumber R (2015) Overview of heterosis and heterotic groups in agronomic crops. In: *Concepts and breeding of heterosis in crop plants*, pp 29–44
- Milligan GW, Cooper MC (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159–179
- Oliboni R, Faria MV, Neumann M et al (2013) Análise dialélica na avaliação do potencial de híbridos de milho para a geração de populações- base para obtenção de linhagens. *Semin Agrar* 34:7–18. <https://doi.org/10.5433/1679-0359.2013v34n1p7>
- Oliveira RS, Queiróz MA, Romão RL et al (2016) Genetic diversity in accessions of *Stylosanthes* spp. using morphoagronomic descriptors. *Rev Caatinga* 29:101–112. <https://doi.org/10.1590/1983-21252016v29n112rc>
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554
- Pereira LD, Souza LKF, Valle KD et al (2019) Genetic diversity among red mombin fruits in the Southwest of Goiás. *Rev Ceres* 66:154–158. <https://doi.org/10.1590/0034-737x201966020010>
- R Development Core Team (2020) *R: a language and environment for statistical computing*
- Rao CR (1952) *Advanced statistical methods in biometric research* A division of Macmillan Publishing Co, Inc New York Collier-Macmillan

- Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed Appl Biotechnol* 16:330–339. <https://doi.org/10.1590/1984>
- Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui Agropecuária Trop* 37:182–194. <https://doi.org/10.5216/pat.v37i3.1867>
- Rocha VD, Tiago PV, Tiago AV et al (2017) Genetic diversity among *Hymenaea courbaril* L genotypes naturally occurring in the north of Mato Grosso State Brazil. *Genet Mol Res*. <https://doi.org/10.4238/gmr16039706>
- Rodrigues DL, Viana AP, Vieira HD et al (2017) Contribution of production and seed variables to the genetic divergence in passion fruit under different nutrient availabilities. *Pesqui Agropecuária Bras* 52:607–614. <https://doi.org/10.1590/s0100-204x2017000800006>
- Silva JAL, Neves JA (2011) Produção de feijão-caupi semi-prostrado em cultivos de sequeiro e irrigado. *Rev Bras Ciências Agrárias Brazilian J Agric Sci* 6:29–36
- Silva GC, Oliveira FJ, Anunciação Filho CJ et al (2011) Divergência genética entre genótipos de cana-de-açúcar. *Rev Bras Ciências Agrárias Brazilian J Agric Sci* 6:52–58
- Silva CA, Nascimento AL, Ferreira JP, Ferreira SO et al (2017) Genetic diversity among papaya accessions. *African J Agric Res* 12:2041–2048. <https://doi.org/10.5897/ajar2017.12387>
- Smith AB, Cullis BR (2018) Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica* 214:143. <https://doi.org/10.1007/s10681-018-2220-5>
- Steckling SM, Ribeiro ND, Arns FD et al (2017) Genetic diversity and selection of common bean lines based on technological quality and biofortification. *Genet Mol Res*. <https://doi.org/10.4238/gmr16019527>
- USDA (2019) Economic Research Service

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.