GYT Biplot Analysis: A New Approach for Cowpea Line Selection

Tâmara Rebecca Albuquerque de Oliveira[1*], Geraldo de Amaral Gravina[2],
Maurisrael de Moura Rocha[3], Francisco de Alcântara Neto[4],
Derivaldo Pureza da Cruz[2], Gustavo Hugo Ferreira de Oliveira[1],
Camila Queiroz da Silva Sanfim de Sant'Anna[2],
Mário Euclides Pechara da Costa Jaeggi[2] and Richardson Sales Rocha[2]

[1]*Universidade Federal de Sergipe, Rodovia Engenheiro Jorge Neto, Km 3 - Silos,*
Nossa Senhora da Glória, Sergipe, Brasil.
[2]*Universidade Estadual do Norte Fluminense Darcy Ribeiro, Avenida Alberto Lamego, 2000,*
Campos dos Goytacazes, Rio de Janeiro, Brasil.
[3]*Embrapa Meio-Norte, Avenida Duque de Caxias, 5650, Teresina, Piauí, Brasil.*
[4]*Universidade Federal do Piauí, Rua Cícero Duarte, 905, Picos, Piauí, Brasil.*

Authors' contributions

This work was carried out in collaboration among all authors. Authors GAG, MMR and GHFO designed the study and wrote the protocol. Authors TRAO and FAN performed the statistical analysis and wrote the first draft of the manuscript. Authors DPC, CQSSSA, MEPCJ and RSR managed the analyses of the study and the literature searches. All authors read and approved the final manuscript.

Article Information

Original Research Article

ABSTRACT

Cowpea beans is grown under different edaphoclimatic conditions throughout Brazilian regions causing them to perform differently due to the influence that environments have on genotypes. Thus, it is necessary to obtain lines adapted to the specific cultivation environments so that it can present high yield. The objective of this work was to select cowpea lines through the GYT biplot multivariate analysis. The experiment was carried out in Bom Jesus de Itabapoana, Rio de Janeiro

_____

Corresponding author: E-mail: tamara_rebecca@hotmail.com;

State, Brazil, in the 2016 and 2017 harvests. The randomized block design was performed with four replications and four lines per plot. Variance analyzes and biplot plots were applied for the number of days of flowering, final planting, harvest value, housing, pod yield, length, average number of beans per pod, average grain weight per pod and grain weight. The analysis of variance showed that there is genetic variability among the strains, requiring a detailed study to select those with the best agronomic performance. The first two major components of the biplot chart explained almost all of the variation between strains. All yield characteristics were negatively correlated with the set of productivity combinations with housing and number of days for flowering. Lines 3, 10, 4, 2, 6, 12, 7 and 11 showed better average performance for yield characteristics.

1. INTRODUCTION

The cowpea (*Vigna unguiculata* (L.) Walp.), as it is moderately resistant to drought, can be grown under different edaphoclimatic conditions. That vegetable is widely grown and consumed globally, being used as the main source of protein [1]. For the 2012-2014 period, the average global yield was 7.32 million tons [2]. As specified by Freire Filho [3], the Brazilian cowpea yield has still been considered low.

As this vegetable is grown in many Brazilian states, the development of adapted lines that provide good performance and agronomic traits of interest is a challenge. Therefore, knowing about the existing association between the traits of interest is essential when it is intended to make a simultaneous selection of characters and obtain progressive gains by means of indirect selection of easily measured characters [4,5].

The analysis of main components (MC) makes feasible to reduce a set of traits in linear combinations, enabling data interpretation. The biplot methodology, for its part, is an analysis that employs the first two MCs, describing data graphically on the basis of their pattern. According to Yan and Rajcan [6], Yan and Tinker [7] and Paramesh, et al. [8], the biplot analysis is superior to the other correlation and regression analyses, since it describes the correlation among all the variables simultaneously, provides a tool for visual comparison of genotypes based on the multiple traits, and permits its independent selection.

Yan and Frégeau-Reid [9] state that the GYT biplot analysis is a new methodology based on the genotype selection that combine high yield with other target traits, rather than by individual traits. They point out that the average tester coordination view (ATC) of the GYT biplot ranks the genotypes based on their general performance by means of the combinations of yield with the other variables, generating yield traits.

Considering that the breeding program aims to obtain high yielding genotypes and other agronomic traits of interest, the GYT biplot analysis appears as a solution select superior genotypes to reliably. According to Yan, et al. [10] the GYT index takes into account that yield is the most important trait, hindering the selection and recommendation of low yield genotypes, thus being superior to the traditional selection index. Thus, this analysis allows the breeder to choose genotypes that display a good performance for a trait of interest, being this connected to high-performance yield.

As a result, the purpose of this research was to select cowpea line by GYT biplot analysis and study the existing correlation among the yield traits.

2. MATERIALS AND METHODS

This work assessed 14 cowpea lines (Table 1), in which two cultivars were used as testimonies (*BRS Tumucumaque* and *BRS Imponente*), during the crops of 2015 and 2016, at Bom Jesus de Itabapoana Municipality, located in the Northeast of Rio de Janeiro State (latitude 21º08'02"S, c 41º40'47"W and altitude 88 m). The region has tropical climate according to the Köppen's climate classification [11], mean annual temperature of 23ºC and the soil classified as red-yellow latosoil + cambisoil.

The experimental design was based on randomized blocks, with four repetitions, with experimental plot constituted by four 5 m lines, 0.50-m spacing between lines, and 0.50 m spacing between plants; the two central lines were used as the useable area. Thinning was carried out 15 days after planting, leaving only one plant per hole, resulting in a stand of 40

plants per plot and a population of 160 thousand plants per hectare.

The traits assessed were number of days for flowering (NDF): count of total of days from planting to flowering; plant stand (PS): count of number of plants per unit of area; crop value (CV): product of percentage of pure plant by percentage of viable plant; lodging (LD): count of number of plants lodged per plot; pod yield (PY): quantification of total weight of pods of each plant in grams (g) from the regulated precision scale; number of grains per pods (NGP): determined by the count of the mean number of grains per plant; weight of grains per pod (WGP): total weight in grams (g) of grains/pod checked by regulated precision scale; grain yield (GY): quantification of the total weight in grams (g) of grains of each plant after pod threshing by precision scale; pod length (PL): measuring of pods in centimeters (cm) using a rule; weight of 100 grains (W100G): weighing of 100 grains from a sample selected at random in a regulated precision scale.

Table 1. Description of lines to be selected

N.	Line code	Commercial subclass
01	Bico-de-ouro 1-5-11	SV
02	Bico-de-ouro 1-5-15	SV
03	Bico-de-ouro 1-5-19	SV
04	Bico-de-ouro 1-5-24	ML
05	Pingo-de-ouro 1-5-26	ML
06	Pingo-de-ouro 1-5-4	ML
07	Pingo-de-ouro 1-5-5	ML
08	Pingo-de-ouro 1-5-7	ML
09	Pingo-de-ouro 1-5-8	ML
10	Pingo-de-ouro 1-5-10	ML
11	Pingo-de-ouro 1-5-11	ML
12	Pingo-de-ouro 1-5-14	ML
13	BRS Tumucumaque	BR
14	BRS Imponente	BC

BR - Branco; BC – Brancão; FR - Fradinho; ML- Mulato; SV - Sempre-Verde; CN – Canapu

Initially, an analysis of variance was performed for each environment aiming at verifying the homogeneity of the residue variances, using the model:

$$y_{ij} = \mu + b_j + g_i + \varepsilon_{ij}$$

where: y_{ij} is the observed value of the studied characteristic in genotype i and block j; μ is the overall mean of the experiment; t_i is the effect of treatment i; ε_{ij} is the error associated with y_{ij}.

Subsequently, a joint analysis of variance was conducted considering the effects of genotypes and environments to obtain the matrix GE, using the model:

$$y_{ijk} = \mu + g_i + a_j + ga_{ij} + b/a_{jk} + \varepsilon_{ijk}$$

where: y_{ijk} is the observed value of the studied characteristic in genotype i, environment j and block k; μ is the overall mean of the experiment; g_i is the effect of genotype i; a_j is the effect of environment j; ga_{ij} is the effect of the interaction of genotype i with environment j; b/a_{jk} is the effect of the interational of block k within environment j; ε_{ijk} is the error associated with y_{ijk}.

The GYT biplot multivariate analysis was based on information from phenotypic means, being the procedure for combining the means of the variables in accordance with the methodology proposed by Yan and Frégeau-Reid [9]. In this methodology, when there is a variable in which the focus is on reaching high values, it is multiplied by the mean of the yield (GY*CV), and those which interest is in achieving lower values, the means (GY/NDF) are divided. Being thus, the variable yield per se is not included in the biplot, provided that it is incorporated into each of the yield trait combinations.

The biplots were built applying the two first MCs, in such a way that the main component (MC1) was used in the horizontal axis and the MC2, in the vertical axis by means of the Singular Values Decomposition (SVD) on the basis of the equation proposed by Yan and Frégeau-Reid [9]:

$$P_{ij} = \left(d\lambda_1^\alpha \xi_{i1}\right)\left(\frac{\lambda_1^{1-\alpha} \tau_{1i}}{d}\right) + \left(d\lambda_2^\alpha \xi_{i2}\right)\left(\frac{\lambda_2^{1-\alpha} \tau_{i2}}{d}\right) + \varepsilon_{ij}$$

in which λ_1 and λ_2 are the singular values for the first and the second main components, respectively; α is the singular factor of value partitioning; ξ_{i1} and ξ_{i2} are the eigenvalues of the first and the second main components, respectively, for the genotype i; τ_{1j} and τ_{2j} are the eigenvalues for the first and the second main components, respectively, for the combination of yield trait j; and ε_{ij} is the residue of the first and the second main components for the genotype i in the combination of yield trait j.

The analysis of variance and the GYT biplots graphs were achieved by the R software with the

help of the ggplot2 package (R Development Core Team 2014).

3. RESULTS AND DISCUSSION

The significant effects for the variables GY, NDF, PS, NGP, and WGP revealed the lines displayed differences in performance caused by the genetic variability, which enables success in selection (Table 2). That result proves that the lines are altered in the rankings in accordance with the trait under study. The significance between the interaction genotypes x years indicates different behavior of the lines in the different years.

As a result of these alterations in the rankings, a behavior study of these lines is required to select the ones that present, together, agronomic traits of interest and good grain yield, taking into account that the genotype selection for a certain trait of agronomic interest is only valid when it is associated with high yield. Therefore, the combination levels between yield and traits are more significant than the levels of individual traits in the selection of superior cultivars [9].

The GYT biplot analysis demonstrated that the two first MCs explained almost the totality of the variation between the lines (89.73%), stating that there is high reliability of results provided by the multivariate analysis. Thus, the graphical representation was regarded as effective and could be used for a comprehensible data interpretation [12].

With regard to the yield performance of the genotypes, the position of the vertex shows the greatest distance with regard to the origin, ranking them as more responsive. Those genotypes should be used to identify possible yield patterns that help select the most responsive genotypes to specific traits. Then, the genotypes within the polygon are the least responsive to the traits that represent the group formed.

Under this context, the variables grouped within the polygon are considered similar in terms of the environmental influence generated on the genotypes. Four groups can be noticed. The performance of the grain formed the first group (G1), represented by the combination GY*W100G. The stand formed the second group (G2) by the combination GY*PS (Fig. 1). The third group clustered the traits of the performance of pod (G3), GY*PY, GY*NGP and GY*WGP. The fourth group, with the largest number of combinations, allocated the phenological traits of the plant and the pod length (G4), GY/NDF, GY/LD, GY*CV and GY*PL.

The line L8 was displayed at the vertex of the first group, which means that it was the best in the grain yield combination with weight of 100 grains. The best performance for G2 was from the lines L2 and L10. Regarding the group of combinations referring to the G3, they highlighted the line L7 as the best performance. In the G4, the cultivars *BRS Imponente* was higher for the combinations of grain yield with the number of days for flowering, lodging, crop value, and pod length.

It is worth noting that variable combinations, which are in the same group, enable to select more yielded lines and that present other traits of interest, as the line at the vertex stands out for all the sets of variables present in the group formed. Therefore, it is possible to designate a single combination and discard the other ones, reducing costs and time with evaluations.

The superiority ranking of the genotypes, on the basis of their yield trait combinations, was given in Fig. 2. In that graph, the ATC is based on the singular value partitioning focused on the genotype, so that the circle shaped represents the mean of the yield trait combination, defined by the coordinates of the set of yield trait combinations assessed in the biplot [9].

The vertical line separates the genotypes by the means; thus, those that are on the same side of the average tester axis (ATA), right side, present better mean performance for the yield trait combinations. The discriminance is represented by the length of the projection formed towards the ATA, in a way that the lines that present smaller projections tend to have a balanced yield trait set, while larger projections show higher discriminance for a combination set [9].

Consequently, the cultivar *BRS Imponente* and line L7 showed performance above overall average, and the lines L1, L2, L11 and cultivar *BRS Tumucumaque* were ranked as within overall average. On the other hand, the lines L3, L4, L6, L8, L9, L10 and L12 displayed mean values below the overall average, despite Fig. 1 illustrating that some of these values stood out in some trait combinations, such as weight of 100 grains and stand.

Table 2. Estimates of mean squares of the variables number of days for flowering (NDF); final plant stand (PS); crop value (CV); lodging (LD); pod yield (PY); pod length (PL); number of grains per pod (NGP); weight of grains per pod (WGP); grain yield (GY); and weight of 100 grains (W100G) of 14 cowpea lines, assessed at Bom Jesus de Itabapoana Municipality, Rio de Janeiro State, Brazil

VF	DF	QM									
		YLD	NDF	PS	CV	LD	PY	PL	NGP	WGP	W100G
Block (Year)	6	1.38	67.51	17.54	0.48	1.10	1.72	1.80	12.51	26.69	615.95
Years	1	0.66*	6.79	13622.70**	1.26	0.67	4.90*	300.87**	233.49**	47.69**	733.67
Genotypes	13	0.54**	116.79**	147.77**	0.88	0.65	1.73	4.46	17.21**	16.94**	672.70
Genotypes × Years	13	0.20	123.75**	60.56	0.72	0.57	1.25	2.82	4.99	9.48*	975.82
Error	77	0.16	41.44	57.45	0.59	0.96	1.07	3.74	3.51	4.61	830.66

*and ** indicate significance at 5% and 1%, respectively*

Table 3. The mean values of the different genotypes, of the variables number of days for flowering (NDF); final plant stand (PS); crop value (CV); lodging (LD); pod yield (PY); pod length (PL); number of grains per pod (NGP); weight of grains per pod (WGP); grain yield (GY); and weight of 100 grains (W100G) of 14 cowpea lines, assessed at Bom Jesus de Itabapoana Municipality, Rio de Janeiro State, Brazil

N.	Line Code	NDF	PS	CV	LD	PY	PL	NGP	WGP	W100G	GY
01	Bico-de-ouro 1-5-11	57.75a	1.75a	2.75b	1.75a	5.26a	9.00a	10.43a	1.51b	10.70a	117.12a
02	Bico-de-ouro 1-5-15	57.75a	1.50a	3.38a	1.50a	5.77a	9.17a	10.63a	4.13b	10.64a	13.13b
03	Bico-de-ouro 1-5-19	58.13a	1.25a	3.50a	1.25a	5.95a	9.31a	10.10a	4.19b	11.01a	141.08a
04	Bico-de-ouro 1-5-24	57.13a	1.63a	3.13a	1.63a	6.51a	9.79a	10.05a	4.32a	11.82a	148.16a
05	Pingo-de-ouro 1-5-26	57.38a	1.25a	3.25a	1.25a	6.51a	9.52a	10.33a	4.32a	11.36a	162.27a
06	Pingo-de-ouro 1-5-4	57.63a	1.88a	3.63a	1.88a	6.63a	9.79a	10.98a	4.44a	11.57a	164.96a
07	Pingo-de-ouro 1-5-5	57.63b	1.63a	3.38a	1.63a	6.45a	9.65a	10.38a	4.88a	10.32a	172.30a
08	Pingo-de-ouro 1-5-7	57.25a	1.38a	3.50a	1.38a	5.20a	8.52a	9.80a	4.94a	10.51a	174.59a
09	Pingo-de-ouro 1-5-8	36.75a	1.00b	1.75b	1.00a	1.70b	3.37b	6.05b	5.13a	3.13b	180.52a
10	Pingo-de-ouro 1-5-10	57.88a	1.88a	3.00b	1.88a	5.13a	8.95a	9.38a	5.26a	10.64a	184.78a
11	Pingo-de-ouro 1-5-11	56.88a	1.63a	3.38a	1.63a	6.38a	9.57a	10.55a	5.38a	11.64a	187.49a
12	Pingo-de-ouro 1-5-14	57.38a	1.38a	3.50a	1.38a	5.26a	8.37a	9.18a	5.38a	11.45a	191.34a
13	BRS Tumucumaque	58.50a	2.00a	3.50a	2.00a	5.26a	8.69a	10.45a	5.69a	9.95a	93.28a
14	BRS Imponente	58.00a	1.50a	3.63a	1.50a	6.13a	9.25a	10.48a	5.44a	12.39a	207.11a

In terms of profile, the lines that presented good means and that were shown balanced for the set of yield trait combinations were L5 and L1. L2, l7, and L11 were shown to be discriminant for the traits relating to G3 and the cultivars *BRS Imponente* and *BRS Tumucumaque*, for the traits concerning G4. All lines above the ATA tend to present the best levels for the set of traits existing in G4. Similarly, the lines below the ATA provide the best mean performance for the set of combinations in G3.

Being aware of the correlations between the variables makes it easier the decision-making for the breeding programs, which aim at selecting the genotypes of interest [13]. Considering that, the correlation among the yield traits can be verified in Fig. 3. This graph shows that the angles formed by the vectors denote the degree of association among the variables, in such a way that acute angles (<90º) designate positive correlations, right angles (=90º) indicate non-correlation, obtuse angles (>90º) reveal negative correlations, and vectors that form angles of 180º are strongly negatively correlated [14].

As such, the set of combinations concerning the phenology of the plant and length of the pod presented high positive correlation with the set of traits of pod performance, pointing out that lines with best performance for the combinations of yield with number of days for flowering, lodging, crop value, and pod length tend to have good performance for the combination of yield with pod yield, number of grains per pod, and weight of grains per pod.

Conversely, GY*W100G and GY*PS showed high negative correlation with the other sets of combinations, stating that lines that present the best means for the combination of yield with weight of 100 grains and stand tend to have the worst performance of pod and phenology of plant.

Machado, et al. [15] report that the existing positive correlation between number of days for flowering and weight of grains makes it difficult the selection of early and high yield lines. Nevertheless, the GYT biplot enabled to select line that provides those two traits together, like the cultivar *BRS Imponente*. Likewise, lines with a good combination of grain yield with number of grains per pod and weight of 100 grains were obtained, which was not possible in works performed with snap beans [14] and broad beans, respectively [16].
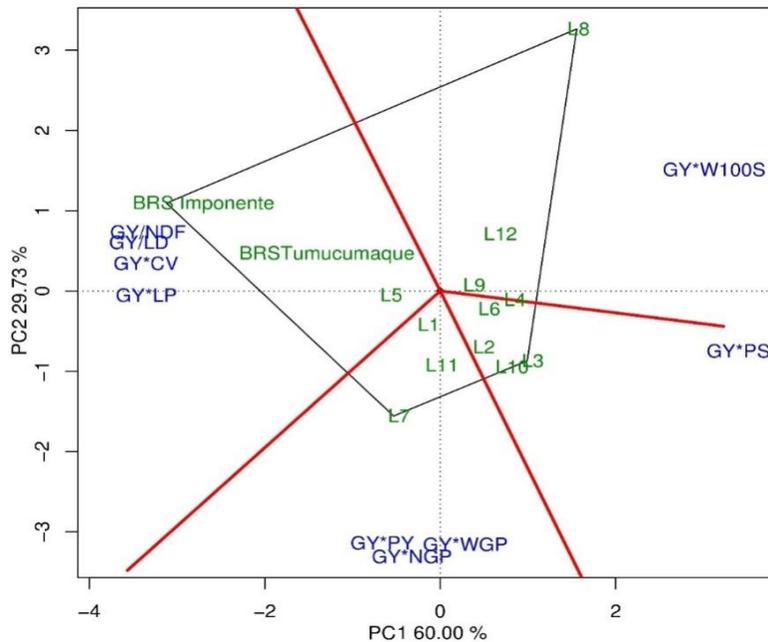

Fig. 1. GYT biplot representing "which-won-where" of the 14 cowpea lines, being NDF: number of days for flowering; PS: final plant stand; CV: crop value; LD: lodging; PY: pod yield; PL: pod length; NGP: number of grains per pod; WGP: weight of grains per pod; GY: grain yield; W100G: weight of 100 grains
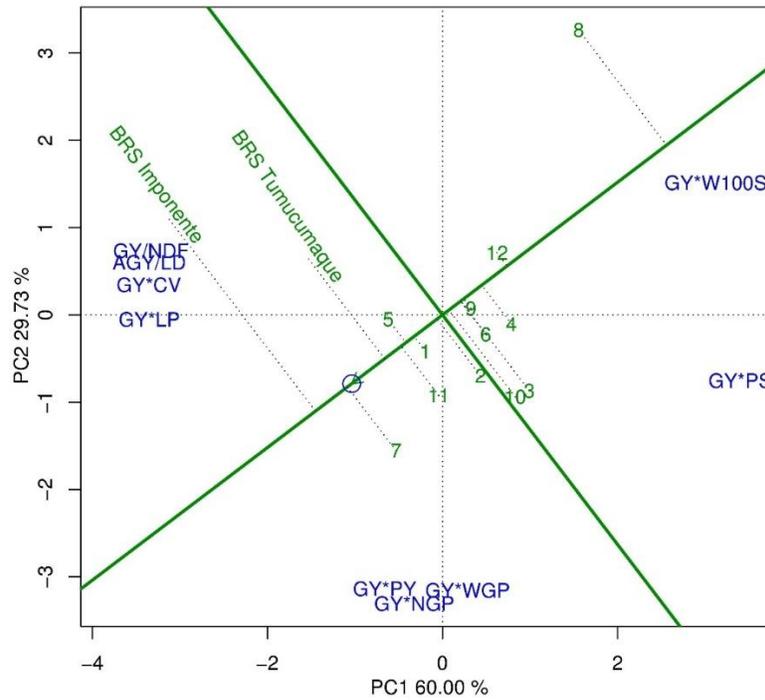
Fig. 2. GYT Biplot representing average × trait, indicating the ranking of the 14 cowpea lines for 10 traits, being number of days for flowering (NDF); final plant stand (PS); crop value (CV); lodging (LD); pod yield (PY); pod length (PL); number of grains per pods (NGP); weight of grains per pod (WGP); grain yield (GY); and weight of 100 grains (W100G)
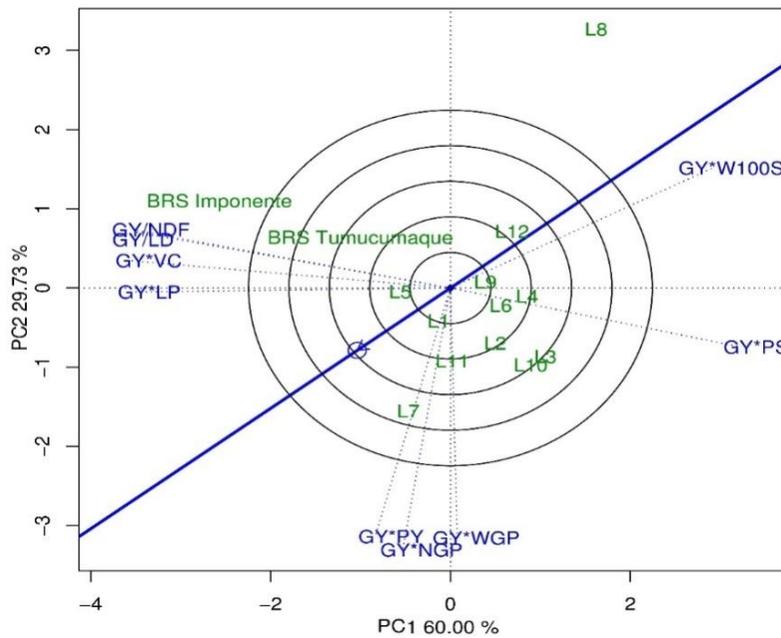

Fig. 3. GYT Biplot describing the 14 lines and representing the best ones for each trait, being number of days for flowering (NDF); final plant stand (PS); crop value (CV); lodging (LD); pod yield (PY); pod length (PL); number of grains per pods (NGP); weight of grains per pod (WGP); grain yield (GY); and weight of 100 grains (W100G)

Those results prove the superiority and ease of this analysis when it comes to selecting genotypes that respond well to the variable yield and that is connected to other wanted traits.

4. CONCLUSION

The GYT biplot proved to be a reliable and an easy-to-interpret analysis and visualization of the results.

Four groups that represented the performance of grains were formed, as follows: stand, grain performance, and phenology and pod length, respectively.

The lines with the best performance for each group were L8 (G1), L3 and L10 (G2), L7 (G3), and cultivar BRS Imponente (G4).

The combination between yield with weight of 100 grains and stand should not be used when it comes to select lines with good performance for the other groups of yield traits.

ACKNOWLEDGEMENT

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Freitas R, Dombroski JLD, Freitas FCLD, Nogueira NW, Pinto JRDS. Physiological responses of cowpea under water stress and rewatering in no-tillage and conventional tillage systems. Rev Caatinga. 2017;30(3):559-567.
Available:http://dx.doi.org/10.1590/1983-21252017v30n303rc
2. FAO. Food and agriculture organization of the united nations. Crops: cow peas, dry; 2017
3. Freire Filho FR. Feijão-caupi no Brasil: produção, melhoramento genético, avanços e desafios. Brasilia: Embrapa, Portuguese; 2011.
4. Carvalho FD, Lorencetti C, Benin G. Estimativas e implicações da correlação no melhoramento vegetal. Rio Grande do Sul: Ufpel, Portuguese; 2004.
5. Silva JALD, Neves JA. Production components and their correlations in caupi bean genotypes in rainfed and in irrigated cultivation. Rev Cienc Agron. 2011; 42(3):702-713.
Available:http://dx.doi.org/10.1590/S1806-66902011000300017
6. Yan W, Rajcan I. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Sci. 2002;42:11-20.
Available:http://dx.doi.org/10.2135/cropsci2002.1100
7. Yan W, Tinker NA. Biplot analysis of multi-environment trial data: Principles and applications. Can J Plant Sci. 2006; 86(3):623-645.
Available:http://dx.doi.org/10.4141/P05-169
8. Paramesh M, Reddy D, Priya MS, Sumathi P, Sudhakar P, Reddy K. GT biplot analysis for yield and drought related traits in mung bean (Vigna radiata L. Wilczek). Electron J Plant Breed. 2016;7(3): 538-543.
Available:http://dx.doi.org/10.5958/0975-928X.2016.00069.7
9. Yan W, Frégeau-Reid J. Genotype by yield* trait (GYT) Biplot: a Novel Approach for Genotype Selection based on Multiple Traits. Nature. 2018;8(8242):1-10.
Available:http://dx.doi.org/10.1038/s41598-018-26688-8
10. Yan W, Frégeau Reid Judith, Mountain N, Kobler J. Genotype and Management Evaluation Based on Genotype by Yield*Trait (GYT) Analysis. Crop Breed Genet Genom. 2019; 1:2-21.
Available: https://doi.org/10.20900/cbgg20190002
11. Alvares CA, Stape JL, Sentelhas PC, Moraes GJ, Sparovek G. Köppen's climate classification map for Brazil. Meteorol Zeitschrift. 2013;22(6):711-728.
Available: https://doi.org/10.1127/0941-2948/2013/0507
12. Yang RC, Crossa J, Cornelius PL, Burgueño J. Biplot analysis of multi-environment trial data: Principles and applications. Crop Sci. 2009;49:1564-1576.
Available:http://dx.doi.org/10.2135/cropsci2010.01.0001le
13. Mendonça MSD, Beber PM, Nascimento FSSD, Santos VBD, Marinho JT. Importance and correlations of characters

for cowpea diversity in traditional varieties. Rev Ciênc Agron. 2018;49(2):267-274. Available:http://dx.doi.org/10.5935/1806-6690.20180030

14. Oliveira TRA, Gravina GA, Oliveira GHF, Araújo KC, Araújo LC, Daher RF, et al. The GT biplot analysis of green bean traits. Cienc Rural. 2018:48(6):1678-4596. Available:http://dx.doi.org/10.1590/0103-8478cr20170757

15. Machado FCP, Teixeira NJP, Freire Filho FR, Rocha MMD, Gomes RLF.

Identificação de genótipos de feijão-caupi quanto à precocidade, arquitetura da planta e produtividade de grãos. Rev Cienc Agron. 2008;39(1): 114-123.

16. Gasim S, Hamad AS, Abdelmula A, Mohamed Ahmed IA. Yield and quality attributes of faba bean inbred lines grown under marginal environmental conditions of Sudan. Food Science & Nutrition. 2015;3:539-547. Available:https://doi.org/10.1002/fsn3.245