



# Genotype selection based on multiple traits in cotton crops: The application of genotype by yield\*trait biplot

Marco Antônio Peixoto<sup>1</sup>, Jeniffer Santana Pinto Coelho Evangelista<sup>1</sup>, Igor Ferreira Coelho<sup>1</sup>, Luiz Paulo Carvalho<sup>2</sup>, Francisco José Correa Farias<sup>2</sup>, Paulo Eduardo Teodoro<sup>3</sup> and Leonardo Lopes Bhering<sup>1\*</sup>

<sup>1</sup>Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. Peter Henry Rolfs, s/n., 36570-900, Viçosa, Minas Gerais, Brazil. <sup>2</sup>Empresa Brasileira de Pesquisa Agropecuária, Embrapa Algodão, Campina Grande, Paraíba, Brazil. <sup>3</sup>Departamento de Estatística e Genética, Universidade Federal de Mato Grosso do Sul, Chapadão do Sul, Mato Grosso do Sul, Brazil. \*Author for correspondence. E-mail: leonardo.bhering@ufv.br

**ABSTRACT.** In cotton crops, the cotton seed yield significantly contributes with the success of any cultivar. However, other traits are considered when an ideotype is pointed out in the selection, such as the fiber quality traits. The aim of this study was to applied genotype by yield\*trait (GYT) biplot to a multi-environment trial data of cotton genotypes and selected the best genotypes. For this end, thirteen genotypes from nineteen trials were assessed. Seven traits were evaluated [cotton seed yield (SY), fiber percentage (FP), fiber length (FL), fiber uniformity (FU), short fiber index (SFI), fiber strength (FS), and elongation (EL)] and residual error variances structures [identity variance (IDV) and diagonal (Diag)] were tested by bayesian information criterion. After, the REML/BLUP approach was applied to predict the genetic values of each trait and the selective accuracy were measured from the prediction. Then, the GYT-biplot were applied to the data. For SP and SFI traits, the model with Diag residual variance was indicated, whereas for SY FL, FU, FS, and EL traits the model with IDV residual variance demonstrated the best fit to the data. Values of accuracy were higher than 0.9 for all traits analyzed. In the GYT-biplot acute angles were find for all traits relations, which means high correlation between the yield\*traits combination. Besides that, the correlation still can be seen in the GYT-biplot, as shown by the magnitudes of the angles between the pairs Yield\*FU-Yield\*FS and Yield\*FS-Yield\*EL. Also, the GYT-biplot indicates the genotype G4 with the best performance for Yield\*FS, Yield\*SFI, Yield\*FU, Yield\*FL, and Yield\*FP combined. The genotypes G4, G1, G13, G8, and G9 represent those genotypes with yield advantage over the other cultivars. Then, the genotype G4 combines all desirable characteristics and demonstrate have large potential in the cotton breeding. The GYT approach were valuable and were highly recommended in cotton breeding programs for selection purpose in a multivariate scenario.

**Keywords:** biplot analysis; genotype by trait (GT) analysis; multi-environmental trial; residual error variance.

Received on June 5, 2020.  
Accepted on August 26, 2020.

## Introduction

Upland cotton (*Gossypium hirsutum* L.) is the most cultivated species worldwide for fiber production. It provides over 90% of the world's cotton and represents a crop with greatest industrial relevance. Its cultivation as an annual crop is widespread from south to north, from subtropical to temperate regions well over 30° (D'Eeckenbrugge & Lacape, 2014). In such crops, which encompasses large areas and different localities for evaluation of the best materials, the genotype × environment (G × E) interaction plays an essential role in genotypic expression and must be considered in the evaluation and selection of superior genotypes for cotton cultivation (Malosetti, Ribaut, & van Eeuwijk, 2013; van Eeuwijk, Bustos-Korts, & Malosetti, 2016; Li, Suontama, Burdon, & Dungey, 2017).

To obtain elite cultivars that are adapted to specific regions, it is essential to evaluate genotypes in a wide experimental network or the so-called multi-environmental trials (Smith, Cullis, & Thompson, 2005). In this sense, statistical methods have been proposed over the last few decades to deal with data emerging from the multi-environmental trials framework (van Eeuwijk et al., 2016; Li et al., 2017). In addition, an ideal genotype (ideotype) must present superior levels for many target traits, simultaneously. In this point, the challenge emerges once the correlation between pair of traits is not always positive or even large. Both aspects are crucial in a cotton breeding program, in which both yield and quality traits (such as fiber length) are desirable for improving the final value of the product (Teodoro et al., 2018) and the correlation between the yield and quality traits are lower or even negative (Ribeiro et al., 2018; Teodoro et al., 2019). Then, for cotton breeding

programs, methodologies that encompass multivariate analyses are core important in the breeding process toward the cotton ideotype.

One of the preeminent methods for dealing with multi-environmental trials is the biplot graphical display (Yan & Hunt, 2002; Yan & Tinker, 2006; Yan, Frégeau-Reid, Mountain, & Kobler, 2019), being a recent biplot approach the genotype by trait biplot (GT-biplot; Xu, Fok, Li, Yang, & Yan, 2017). This method emerges from multivariate methodologies, since it assessed the genotypes performance based on multiple traits and allows the identification of those superior combining all desirable traits (Xu et al., 2017; Oliveira et al., 2018). According to those authors, the GT-biplot, demonstrate the relations between the traits (antagonism or synergy) and clarify the traits profiles of the genotypes, pointing those genotypes that stands out with better performance, combining the desirable traits.

Another methodology indicated as more powerful than the GT-biplot for dealing with multivariate scenario is the genotype by yield-trait (GYT) biplot (Yan & Frégeau-Reid, 2018). It was proposed to tackle the problem of genotype evaluation of multiple traits. In several cultures, the most important trait is the yield, and secondary traits are desirable only when they combine with high yield. The GYT-biplot should reveal those superior genotypes that combine the yield with other target traits, rather than by the performance of individual traits (Kendal, 2019; Woyann et al., 2020). In this way, in the cotton crop, where more than one trait is the target for breeding programs, methodologies that encompass multivariate analyses in the selection of superior genotypes should be rather preferred.

The GYT arose as a linear combination between traits. In this approach, the established association between the main trait (generally yield) and all other traits occurs by multiplying yield with the trait when higher values of the trait are desirable and dividing the yield component by the trait that high values are not desirable. Results obtained from the GYT are visualized by GGE biplot (Yan, 2001), considering genotypes and yield-trait combination as fixed effects. While the application of the GYT-biplot is not documented for cotton crops, it has been implemented in other annual crops, such as soybean, wheat, and oat (Yan & Frégeau-Reid, 2018; Yan et al., 2019; Kendal, 2019; Merrick, Glover, Yabwalo, & Byamukama, 2020; Woyann et al., 2020). These studies demonstrate the usefulness of the GYT-biplot in dealing with the multivariate scenario. They highlight GYT as a method that: (i) measure the performance of the genotypes based on several traits and rank the genotypes, (ii) is based on the concept that yield is the most important trait and increase other traits combining them with yield level; (iii) avoid the low-yielding genotypes from being selected and recommended; and (iv) are easily to interpret, once a graphical dispersion are made through GYT-biplot, which facilitates the visualization and interpretation of the results, ranking the genotypes by its performance. Given the robustness of the GYT analyses in context of multivariate context, the study aims to (i) apply for the first time the GYT-biplot methodology in multi-environmental trial data from cotton genotypes and (ii) selected genotypes through the GYT-biplot toward the cotton ideotype, considering all traits combined.

## Material and methods

### Experimental data

The experiment was performed during the 2013/2014 and 2014/2015 cropping seasons in the Midwest region, Brazil. The environments consisted of the combinations between sites and cropping seasons of Brazilian Cerrado, whose edaphoclimatic characteristics are expressed in Table 1. Nineteen trials of cotton cultivars were performed in a randomized complete block design, with thirteen cotton genotypes with four replicates each. The experimental unit (plots) consisted of four 5.0 m rows, with 0.90 m between rows and 45 plants per row. The genotypes used in all the trials have a medium maturity (between 140 and 150 days) and are recommended for cultivation in the Brazilian upland region (Brazilian Cerrado). Seven traits were evaluated: cotton seed yield (SY, kg ha<sup>-1</sup>), fiber percentage (FP, %), fiber length (FL, mm), fiber uniformity (FU, %), short fiber index (SFI, %), fiber strength (FS, gf tex<sup>-1</sup>), and elongation (EL, mm).

### Statistical analyses

Variance components were estimated through restricted maximum likelihood (REML; Patterson & Thompson, 1971) and the prediction of genotypic values was made using best linear unbiased prediction (BLUP; Henderson, 1975) methods. For this end, the following model was used.

**Table 1.** Locality and edaphoclimatic characteristics of the 19 trials of cotton used in the analyses. m = meters. mm = millimeters. MG = Minas Gerais State. GO = Goiás State, MT = Mato Grosso State. PI = Piauí State. BA = Bahia State. MA = Maranhão State. MS = Mato Grosso do Sul State.

Locality	Crop year	Altitude (m)	Latitude (S)	Longitude (W)	Annual precipitation (mm)	Mean annual temperature (°C)	Climate <sup>1</sup>
Trindade (MG)	2013/2014	927	21° 06'	44° 10'	1,467	23.2	Aw
	2013/2014						
Santa Elena de Goiás (GO)	2013/2014	562	17° 48'	50° 35'	1,539	24.3	Aw
	2014/2015						
Primavera do Leste (MT)	2014/2015	465	15° 33'	54° 17'	1,784	22	Aw
	2014/2015						
Campo Verde (MT)	2013/2014	736	15° 32'	55° 10'	1,902	26.3	Af
	2014/2015						
Sinop (MT)	2013/2014	345	11° 51'	55° 30'	1,818	25	Aw
	2013/2014						
Pedra Preta (MT)	2013/2014	248	16° 37'	54° 28'	489	25.1	Bsh
	2014/2015						
Luiz Eduardo Magalhães (BA)	2013/2014	769	12° 5'	45° 47'	1,511	24.2	Aw
São Desidério (BA)	2013/2014	497	12° 21'	44° 58'	1,289	24.7	Aw
Montividiu (GO)	2013/2014	821	17° 26'	51° 10'	1,512	23	Aw
Magalhães de Almeida (MA)	2013/2014	36	03° 23'	42° 12'	1,430	27.2	Aw
Teresina (PI)	2013/2014	72	05° 05'	42° 48'	1,349	27.6	Aw
Chapadão do Sul (MS)	2014/2015	800	18° 47'	52° 37'	1,600	22.7	Aw
Sorriso (MT)	2014/2015	365	12° 32'	55° 42'	1,883	25	Aw

$$y = Xb + Zg + Wi + e$$

where:

$y$  is the vector of phenotypic data;

$b$  is the vector of replication-environment combinations (assumed to be fixed), which comprises the effects of environment and replication within the environment and is added to the overall mean;

$g$  is the vector of genotype effects [assumed to be random;  $g \sim N(0, \sigma_g^2)$ , where  $\sigma_g^2$  is the genotypic variance];  $i$  is the

vector of G×E interaction effects [assumed to be random;  $i \sim N(0, \sigma_i^2)$ , where  $\sigma_i^2$  is the G×E interaction variance]; and  $e$  is the vector of residuals [random;  $e \sim N(0, R)$ , where  $R$  represents a matrix of residual variances]. Capital

letters ( $X, Z, W$ ) represent the incidence matrices for  $b, g$ , and  $i$ , respectively.

### Residual structure and effects significance

Models with identity variance (IDV) and diagonal (Diag) residual variance structures were tested for all traits. The goodness-of-fit were measured by using the Bayesian Information Criterion (BIC; Schwarz, 1978) as follows:

$$BIC = -2\text{Log}L + p\text{Log}[n - r(x)],$$

where:

$\text{Log}L$  is the logarithm of the REML function,  $p$  is the number of estimated parameters,  $n$  is the number of observations, and  $r(x)$  is the rank of the incidence matrix of fixed effects. The significance of the random effects of the model were tested using the likelihood ratio test (LRT; Rao, 1952) as follows:

$$LRT = 2(\text{Log}L - \text{Log}L_R) \sim \chi^2$$

where:

$\text{Log}L_R$  is the logarithm of the REML function of the reduced models (without the genotype or G×E interaction effects).

### Selective accuracy

For the genotypic values predicted (BLUP means) from the phenotypic data, the mean selective accuracy ( $r_{\hat{g}_i}$ ) were calculated. This index measure how similar are the predicted values when compared with the real

genetic values (Resende, Silva, & Azevedo, 2014). The mean selective accuracy was obtained for each trait, by the following expressions:

$$r_{gg_i} = \sqrt{1 - \frac{PEV}{\sigma_g^2}};$$

where:

*PEV* is the prediction error variance extracted from the diagonal of the generalized inverse of the coefficient matrix of the mixed model equations.

### Genetic correlation and Path analysis

A correlation analysis and a path analysis were used to better estimated the relationship between each pair of traits. The genetic correlation coefficient between the genetic values predicted for each trait was calculated using the Pearson correlation and the t-test was used to test the significance of the correlations. For the path analysis, the SY was considered as the dependent variable and the genetic correlation matrix estimated was used for estimation of the direct and the indirect effects.

### Genotype by trait and genotype by yield\*trait combination table

The GYT table was obtained according (Yan & Fréreau-Reid, 2018): the traits FP, FL FU, FS, and EL (where large mean values are desirable in cotton breeding programs) were multiplied by the yield and the trait SFI (where large mean values are undesirable in cotton breeding programs) were divided by the yield.

### Data standardization

The GT table or the GYT table was standardized so that the mean for each trait or yield-trait combination becomes 0 and the variance becomes unit (Yan & Fréreau-Reid, 2018). The following formula was used:

$$P_{ij} = \frac{T_{ij} - \bar{T}_j}{S_j}$$

where:

$P_{ij}$  is the standardized value of the genotype  $i$  trait or yield combination  $j$  in the standardized table,  $T_{ij}$  is the genotypic value  $i$  for the trait or yield-trait combination  $j$  in the GT or GYT table,  $\bar{T}_j$  is the mean value over all genotypes for trait or yield-trait combination  $j$ ,  $S_j$  is the standard deviation for trait or yield-trait combination  $j$  among the genotype averages.

### Construction of GT and GYT biplot

The GT- and GYT-biplot analyses were applied to the data aiming to visualize the relation between the analyzed traits. The following equation was applied (Xu et al., 2017; Yan & Fréreau-Reid, 2018):

$$P_{ij} = (d\lambda_1^\sigma \zeta_{i1}) * (\lambda_1^{1-\sigma} \Gamma_{1j}/d) + (d\lambda_2^\sigma \zeta_{i2}) * (\lambda_2^{1-\sigma} \Gamma_{2j}/d) + e_{ij}$$

where:

$\zeta_{i1}$  and  $\zeta_{i2}$  are the eigenvalues for PC1 and PC2, respectively, for genotype  $i$ ;  $\Gamma_{1j}$  and  $\Gamma_{2j}$  are the eigenvalues for PC1 and PC2, respectively for yield-trait combination (or trait)  $j$ , and  $e_{ij}$  is the residual from fitting the PC1 and PC2 for genotype  $i$  on yield-trait combination (or trait)  $j$ ;  $\lambda_1^\sigma$  and  $\lambda_2^\sigma$  are the singular values for PC1 and PC2, respectively and  $\alpha$  is the singular value partitioning factor. When  $\alpha = 1$  (i.e., SVP = 1 in terms of GGE biplot), the biplot is said to be genotype-focused, and is suitable for comparing genotypes. When  $\alpha = 0$  (i.e., SVP = 2), the biplot is said to be yield\*trait combination-focused and is suitable for visualizing correlations among yield\*trait combination (or trait).

In GYT-biplot methodology only one trait can be considered as the main trait (Yan et al., 2019), even in the case that more than one trait could be considered as main trait (i.e. SY and FL for cotton breeding). In the analyses, the SY trait was considered as the main one. The GT- and GYT-biplot was constructed by plotting  $(d\lambda_1^\sigma \zeta_{i1})$  against  $(d\lambda_2^\sigma \zeta_{i2})$  for genotypes and plotting  $(\lambda_1^{1-\sigma} \Gamma_{1j}/d)$  against  $(\lambda_2^{1-\sigma} \Gamma_{2j}/d)$  for yield-trait combination (or traits) in the same plot (Yan & Fréreau-Reid, 2018).

All analyses were carried out in the Asreml (Gilmour, Gogel, Cullis, Welham, & Thompson, 2015), and GGEbiplotGUI package on R program (R Core Team, 2020).

## Results and discussion

The model selection criterion used (BIC) indicates models with different residual variance for the traits assessed in the analyses (Table 2). For SP and SFI traits the model with Diag residual variance was indicated, whereas for the remaining traits analyzed (FL, FU, FS, and EL) the model with IDV residual variance was assigned. However, the model with Diag residual variance for SY trait did not achieve convergence. In this case, the model accounting for IDV residual variance was considered.

**Table 2.** Mean selective accuracy and residual error variance for each trait assigned by the bayesian information criterion (BIC).

Model	Traits						
	SY	FP	FL	FU	SFI	FS	EL
IDV	<b>14331.65</b>	2487.506	<b>966.0651</b>	<b>1395.585</b>	613.4362	<b>2299.413</b>	<b>359.713</b>
Diag	NC	<b>2056.238</b>	999.1285	1480.909	<b>528.0298</b>	2316.296	430.5872
Accuracy	0.90	0.93	0.89	0.88	0.89	0.91	0.90

Bold letters indicated the selected model. Model with lower values of BIC were preferable. NC = not converged. SY = cotton seed yield (kg ha<sup>-1</sup>), FP = fiber percentage (%), FL = fiber length (mm), FU = fiber uniformity (%), SFI = short fiber index (%), FS = fiber strength (gf tex<sup>-1</sup>), and EL = elongation (mm).

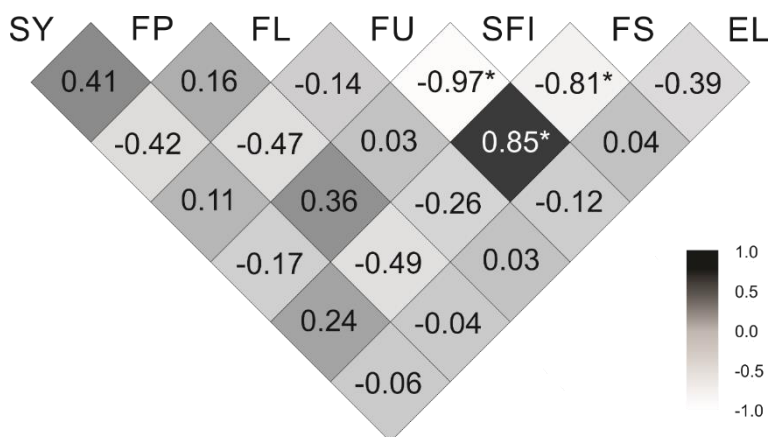
Then, in the subsequent analyses, the respective models selected by BIC were considered for the genotypic values prediction. The evaluation of models with different residual variance is a crucial step in data analyses from MET data (So & Edwards, 2011; Melo et al., 2020). The residual variance that best adjusts the data increment the reliability of genetic values prediction and impacts positively the subsequently analyses. In general, in annual crops, such as cotton, models with Diag residual variance are more acceptable, hence the residual variance is indicated for each environment individually. However, in MET data, there are cases that the goodness-of-fit are presented by models with IDV residual variance, similar with some traits analyzed here, demonstrating that only one estimated value for all environments is capable to represent the residual variance. Thus, to test the residual variance that best fits the data is an incipient step in any trustable study of MET data.

The results highlighted the significance for the genotypic and G×E effects for all traits analyzed, except for the SY trait, where the genotypic effect was presented as non-significant (Appendix S1). Probably, in this case, the G×E interaction exhaust the SY genotypic variability in the analyses. In the yield\*trait analyses, the aim was to evaluate the multivariate framework. In this sense, the SY trait was maintained in the analyses once it was the core trait in the analyses and could bring some important information under correlations. Further, the mean selective accuracy values were assigned as high ( $r_{gg_i} > 0.80$ ) for all traits analyzed (Table 2), which indicated reliability of the model for the BLUP prediction (Resende & Duarte, 2007).

The correlation between the BLUP values for each trait was presented in the Figure 1. Among all pairs of genetic correlations, only the pair of traits FU-SFI, FU-FS, and FS-SFI was significant under the t-test (5%). From the path analysis (Table 3), the coefficient of determination ( $r^2$ ) obtained indicate that the traits used to explain 63% of the variation obtained in the SY trait. The most preeminent values of direct effect in the SY were found for FP and FS traits. According to Cruz, Regazzi, and Carneiro (2012), traits that show favorable correlation but have direct effects in the opposite direction indicate the absence of cause and effect. For the traits FP, FL, SFO, and FS, there was a direct effect similar with the correlation between the trait and the SY, indicating the presence of cause and effect in this relation. Therefore, for the traits FU and EL, there are other trait that determine the changes in the variable of interest that will be more useful for selection not being clear the cause-and-effect relation with the main trait (SY).

The GT-biplot analysis represented a total of 66.81% of the variance, being 44.78 and 22.03% from the PCA1 and PCA2, respectively (Figure 2A), whereas the genotype vs. yield\*trait combination presents PCA1 (61.55%) and PCA2 (19.13%), summing 80.68% of the data variance explained by the two axes (Figure 1B). Both approaches presented values that were indicated as suitable to graphically display the data, where more than 70% should explain the data variation (Cruz et al., 2012). Based on the trait distribution, the GT-biplot demonstrate a high and positive correlation between FU-FS (acute angles), and positive moderate correlations (Resende, 2015) between the pair FP-SFI, SY-FP (Figure 2A). Negative correlation was found for the trait pairs (obtuse angles): FU-SFI, FP-FU, FP-FS, and SY-FL. However, the EL presented a small correlation with other traits, as demonstrate by its short vector (Yan & Frégeau-Reid, 2018). Besides that, the GT-biplot shows the trait profile of the genotype. The genotype G11 presented high FS content and the genotype G8 presented a high FP content. Further, the genotype G4 was highlighted by its performance in cotton seed yield. Other

authors also highlight the potential of the GT-biplot for demonstrate the correlations between traits (Akinwale, Fakorede, Badu-Apraku, & Oluwaranti, 2014; Oliveira et al., 2018). As a graphical display of a multivariate analyses, the GT-biplot represents a tool that combine several advantages from analyses that encompass relations between traits (such as correlation, path analyses, and joint regression; Akinwale et al., 2014), demonstrating the usefulness of this analysis.

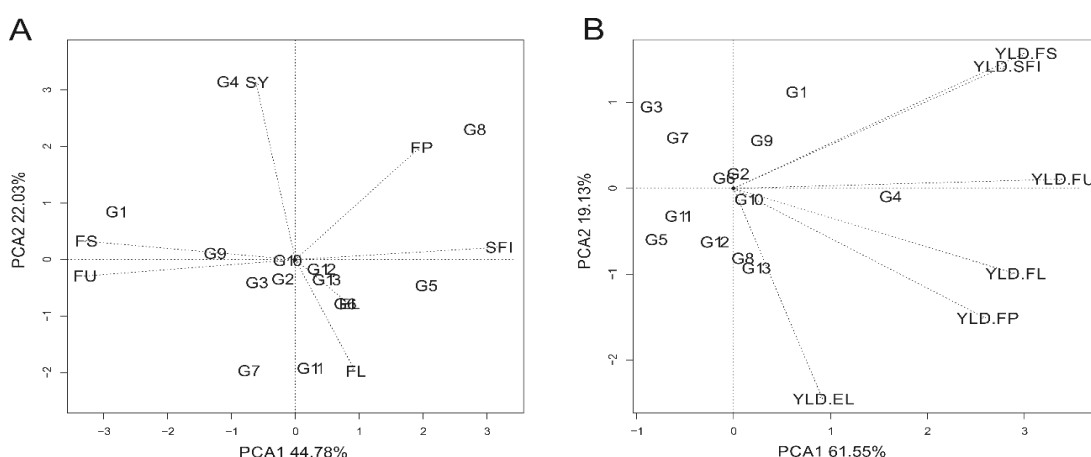


**Figure 1.** Correlation between the values of BLUP means of the traits evaluated in the 19 environments. SY = cotton seed yield (kg ha<sup>-1</sup>), FP = fiber percentage (%), FL = fiber length (mm), FU = fiber uniformity (%), SFI = short fiber index (%), FS = fiber strength (gf tex<sup>-1</sup>), and EL = elongation (mm). \*significant at 5% of probability by the t-test.

**Table 3.** Path coefficient estimates of direct (diagonal) and indirect effects (off – diagonal) among the studied traits: SY = cotton seed yield (kg ha<sup>-1</sup>), FP = fiber percentage (%), FL = fiber length (mm), FU = fiber uniformity (%), SFI = short fiber index (%), FS = fiber strength (gf tex<sup>-1</sup>), and EL = elongation (mm).

Traits	FP	FL	FU	SFI	FS	EL	Total
FP	0.74	-0.05	0.19	-0.05	-0.4	-0.01	0.41
FL	0.11	-0.37	0.05	-0.01	-0.21	0.01	-0.41
FU	-0.35	0.05	-0.41	0.14	0.71	-0.03	0.11
SFI	0.27	-0.01	0.40	-0.15	-0.68	0.01	-0.16
FS	-0.36	0.09	-0.35	0.12	0.83	-0.10	0.23
EL	-0.02	-0.01	0.04	-0.01	-0.32	0.26	-0.06
Residual				0.61			
r <sup>2</sup>				0.63			

r<sup>2</sup> = coefficient of determination provided by the analysis.



**Figure 2.** The tester vector view of genotype by trait (GT) biplot (A) and genotype by yield\*trait (GYT) biplot based on the original genotype by trait data. The biplot was based on singular value decomposition of trait-standardized data (‘Scaling = 1, Centering = 2’) and trait-focused singular value partition (‘SVP = 2’). SY = cotton seed yield (kg ha<sup>-1</sup>), FP = fiber percentage (%), FL = fiber length (mm), FU = fiber uniformity (%), SFI = short fiber index (%), FS = fiber strength (gf tex<sup>-1</sup>), and EL = elongation (mm).

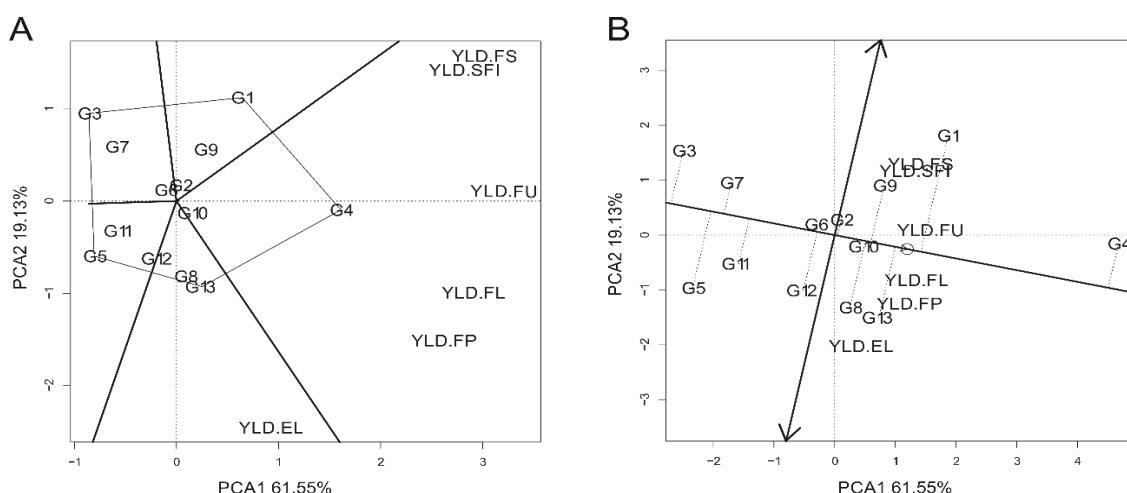
The results demonstrate that GT-biplot can be considered as a powerful tool for exploring the relation between traits, based on the correlation, presenting a graphical display of the genotypes and traits analyzed. However, the GYT-biplot have been pointed to overcome the GT-biplot analysis, once it combine information

of most traits and also information of yield, simultaneously (Kendal, 2019; Merrick et al., 2020). Then, GYT-biplot emerges as an alternative for dealing with multivariate analyses using a graphical biplot (Yan & Frégeau-Reid, 2018) and reliability.

Therefore, in the GYT-biplot (Figure 2B), the traits tend to be positive correlated since they presented the yield component, even if those traits per se are negatively correlated. Even though the GYT table presented the relation between traits, the GYT-biplot are more informative (Yan & Frégeau-Reid, 2018). This approach allows ranking the genotypes based on their levels of yield-trait combinations, showing the traits profiles and similarities/dissimilarities among the genotypes. In the results, the acute angles were then indicated in all traits relations. However, the correlation still can be seen in the GYT-biplots, as shown by the magnitudes of the angles between the Yield\*FU-Yield\*FS and Yield\*FS-Yield\*EL. Angles with high magnitude are more dissimilar from those genotypes in which angles are more acute. It is worth mentioning that the absence of significance for some genetic correlation between traits was an impediment for better understanding the relations displayed in the GYT-biplot.

Genotypes from the ‘which-won-where’ performance were displayed (Figure 3A). This view was useful to demonstrate the trait profile of the genotypes (Yan et al., 2019). An irregular polygon was formed connecting genotypes most distant from the origin in the GYT-biplot. Those genotypes were included in sectors formed by a line that emerges from the biplot origin. These lines divided the yield-trait combinations into five sectors (Figure 3A); corresponding to each sector there was a polygon vertex (Yan et al., 2019). The far is the genotype from the origin that determines the polygon vertex, the largest is the value for the yield-trait combinations placed within each corresponding sector. From the GYT-biplot, the genotype G4 represent the highest level of Yield\*FS, Yield\*SFI, Yield\*FU, Yield\*FL, and Yield\*FP, meaning that the genotype is indicated as the best in combining FS, SFI, FU, FL, FP, and yield. Similarly, the trait Yield\*EL was the best combination in the genotype G13.

Figure 3B presented the ATC view, based on the genotype-focused singular value partitioning. The focus of this analysis is in to compare the genotypes. The circle presented in the figure (near to the center) represents the placement of ‘average-trait-combination’. The line with two arrows is a good in a way to separates the genotype that presented values above and below the average genotype value. Following the ATC line, the genotypes G4, G1, G13, G8, and G9 represent those genotypes with yield advantage over the other cultivars. This figure also highlighted that the G13 was superior for FP and EL and G1 was superior for FS and SFI.



**Figure 3.** (A) The which-won-where view of the genotype by yield\*trait (GYT) biplot to highlight genotypes with outstanding profiles. The biplot was based on singular value decomposition of the standardized GYT table (‘Scaling = 1, Centering = 2’). The trait-focused singular value partition (‘SVP = 2’) was used. (B). The Average Tester Coordination view of the GYT-biplot to rank the genotypes based on their overall superiority and their strengths and weaknesses. The biplot was based on singular value decomposition of the standardized GYT table (‘Scaling = 1, Centering = 2’). The genotype- focused singular value partition (‘SVP = 1’) was used. YLD.FP = yield vs. fiber percentage, YLD.FL = yield vs. fiber length, YLD.FU = yield vs. fiber uniformity, YLD.SFI = yield divided by short fiber index, YLD.FS = yield vs. fiber strength, and YLD.EL = yield vs. elongation.

In cotton breeding programs, cultivars are not preferred by farmers solely when they present a large grain yield (Teodoro et al., 2018). Other traits, such FL and SFI are important to improve the quality and, consequently, the final value of the product. As a result, fiber quality traits are relevant in the analyses toward the cotton ideotype. However, the fiber quality traits are, generally negative correlated with SY, which implies

in difficult for breeders in the selection process (Ribeiro et al., 2018). The GYT-biplot combines the yield information with other traits and allowed the combined selection of the best cultivars. For instance, the genotype G4, G1, and G13 above mentioned were addressed by the desirable performance in several traits combined with yield and classified as stable for such traits were indicated as superior in the GYT-biplot traits. In the case that the aim of the breeding program is aiming to improve the performance of more than one trait, tools similar GYT-biplot are core important, once it allow the recommendation of genotypes with the best performance of several traits simultaneously.

Combining the GYT-biplot analyses with BLUP values represented a great advantage (Woyann et al., 2020). For instance, the mixed model methodology overcomes the usual methods used in plant breeding, such as ANOVA. It allows handling with unbalanced data, to add kinship information and consider the genotypes as a random effect, which implies in improvement of the reliability of the genotypic values prediction. On the other hand, the GYT-biplot is highlighted as a useful methodology that overcome the classic methodologies for deal with the multivariate scenario and for being a visual tool to describe genotypes strongness and ranking them. Ultimately, to combine such analyses improve the reliability of the results.

## Conclusion

The GYT-biplot methodology was successfully applied in the in multi-environment trial data from cotton genotypes. The GYT-biplot technique provides information regarding the genotype's performance based on the multivariate framework. The genotype G4 was selected based on the best performance for multiple traits. The genotypes G4, G1, and G13 combines yield\*trait superior performance when compared with the average genotypes. In breeding programs, the selection combining multivariate framework is core important and GYT-biplot presented an interesting solution for ideotype selection in cotton crops.

## Acknowledgements

We appreciate the financial support from the Brazilian Government offered by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq). This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (Capes) - Finance Code 001

## References

- Akinwale, R. O., Fakorede, M. A. B., Badu-Apraku, B., & Oluwaranti, A. (2014). Assessing the usefulness of GGE biplot as a statistical tool for plant breeders and agronomists. *Cereal Research Communications*, 42(3), 534-546. DOI: <https://doi.org/10.1556/crc.42.2014.3.16>
- Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, MG: UFV.
- D'Eeckenbrugge, G. C., & Lacape, J.-M. (2014). Distribution and differentiation of wild, feral, and cultivated populations of perennial upland cotton (*Gossypium hirsutum* L.) in Mesoamerica and the Caribbean. *PLoS One*, 9(9), e107458. DOI: <https://doi.org/10.1371/journal.pone.0107458>
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2015). *ASReml user guide release 4.1. Functional specification*. Hemel Hempstead, GB: VSN International Ltd.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2), 423-447. DOI: <https://doi.org/10.2307/2529430>
- Kendal, E. (2019). Comparing durum wheat cultivars by genotype × yield × trait and genotype × trait biplot method. *Chilean Journal of Agricultural Research*, 79(4), 512-522. DOI: <http://dx.doi.org/10.4067/S0718-58392019000400512>
- Li, Y., Suontama, M., Burdon, R. D., & Dungey, H. S. (2017). Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. *Tree Genetics & Genomes*, 13(60), 1-18. DOI: <https://doi.org/10.1007/s11295-017-1144-x>
- Malosetti, M., Ribaut, J.-M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, 4, 44. DOI: <https://doi.org/10.3389/fphys.2013.00044>



- Melo, V. L., Marçal, T. S., Rocha, J. R. A. S. C., Anjos, R. S. R., Carneiro, P. C. S., & Carneiro, J. E. S. (2020). Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. *Euphytica*, 216(5), 77. DOI: <https://doi.org/10.1007/s10681-020-02607-9>
- Merrick, L. F., Glover, K. D., Yabwalo, D., & Byamukama, E. (2020). Use of Genotype by Yield\*Trait (GYT) analysis to select hard red spring wheat with elevated performance for agronomic and disease resistance traits. *Crop Breeding, Genetics and Genomics*, 2(2), e200009. DOI: <https://doi.org/10.20900/cbgbg20200009>
- Oliveira, T. R. A., Gravina, G. A., Oliveira, G. H. F., Araújo, K. C., Araújo, L. C., Daher, R. F., ... Cruz, D. P. (2018). The GT biplot analysis of green bean traits. *Ciência Rural*, 48(6), e20170757. DOI: <https://doi.org/10.1590/0103-8478cr20170757>
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545-554. DOI: <https://doi.org/10.1093/biomet/58.3.545>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.
- Rao, C. R. (1952). *Advanced statistical methods in biometric research*. New York, NY: John Wiley & Sons.
- Resende, M. D. V. (2015). *Genética quantitativa e de populações*. Visconde do Rio Branco, MG: Suprema.
- Resende, M. D. V., & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical*, 37(3), 182-194.
- Resende, M. D. V., Silva, F. F., & Azevedo, C. F. (2014). *Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência*. Viçosa, MG: UFV.
- Ribeiro, L. P., Carvalho, L. P., Farias, F. J. C., Rodrigues, J. I. S., Teodoro, P. E., & Bhering, L. L. (2018). Genetic gains in agronomic and technological traits of elite cotton genotypes. *Bragantia*, 77(3), 466-475. DOI: <https://doi.org/10.1590/1678-4499.2017329>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464. DOI: <https://doi.org/10.1214/aos/1176344136>
- Smith, A. B., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *The Journal of Agricultural Science*, 143(6), 449-462. DOI: <https://doi.org/10.1017/S0021859605005587>
- So, Y.-S., & Edwards, J. (2011). Predictive ability assessment of linear mixed models in multi-environment trials in corn. *Crop Science*, 51(2), 542-552. DOI: <https://doi.org/10.2135/cropsci2010.06.0338>
- Teodoro, P. E., Carvalho, L. P., Rodrigues, J. I. S., Farias, F. J. C., Carneiro, P. C. S., & Bhering, L. L. (2018). Interrelations between agronomic and technological fiber traits in upland cotton. *Acta Scientiarum. Agronomy*, 40(1), e39364. DOI: <https://doi.org/10.4025/actasciagron.v40i1.39364>
- Teodoro, P. E., Farias, F. J. C., Carvalho, L. P., Ribeiro, L. P., Nascimento, M., Azevedo, C. F., ... Bhering, L. L. (2019). Adaptability and stability of cotton genotypes regarding fiber yield and quality traits. *Crop Science*, 59(2), 518-524. DOI: <https://doi.org/10.2135/cropsci2018.04.0250>
- van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype × Environment interactions? *Crop Science*, 56(5), 2119-2140. DOI: <https://doi.org/10.2135/cropsci2015.06.0375>
- Woyann, L. G., Meira, D., Matei, G., Zdziarski, A. D., Dallacorte, L. V., Madella, L. A., & Benin, G. (2020). Selection indexes based on linear-bilinear models applied to soybean breeding. *Agronomy Journal*, 112(1), 175-182. DOI: <https://doi.org/10.1002/agj2.20044>
- Xu, N., Fok, M., Li, J., Yang, X., & Yan, W. (2017). Optimization of cotton variety registration criteria aided with a genotype-by-trait biplot analysis. *Scientific Reports*, 7(1), 1-9. DOI: <https://doi.org/10.1038/s41598-017-17631-4>
- Yan, W. (2001). GGEbiplot - A windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agronomy Journal*, 93(5), 1111-1118. DOI: <https://doi.org/10.2134/agronj2001.9351111x>
- Yan, W., & Frégeau-Reid, J. (2018). Genotype by Yield\*Trait (GYT) biplot: a novel approach for genotype selection based on multiple traits. *Scientific Reports*, 8(1), 1-10. DOI: <https://doi.org/10.1038/s41598-018-26688-8>

Yan, W., & Hunt, L. A. (2002). Biplot analysis of diallel data. *Crop Science*, 42(1), 21-30.

DOI: <https://doi.org/10.2135/cropsci2002.0021>

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: principles and applications.

*Canadian Journal of Plant Science*, 86(3), 623-645. DOI: <https://doi.org/10.4141/P05-169>

Yan, W., Frégeau-Reid, J., Mountain, N., & Kobler, J. (2019). Genotype and management evaluation based on Genotype by Yield\*Trait (GYT) analysis. *Crop Breeding, Genetics and Genomics*, 1, e190002.

DOI: <https://doi.org/10.20900/cbgg20190002>