

## Correlations and path coefficient analysis for energy biomass production components in elephant grass (*Pennisetum purpureum* Schum)

Avelino S. Rocha<sup>1</sup>, Rogério F. Daher<sup>1</sup>, Camila Q.S.S. de Sant'Anna\*<sup>1</sup>, Erina Vítório Rodrigues<sup>2</sup>, Ana Késia F. Vidal<sup>1</sup>, Flávio D. Tardin<sup>3</sup>, Wanessa Francesconi Stida<sup>1</sup>, Rafael S. Freitas<sup>1</sup>, Eduardo P. Furlani<sup>4</sup>, Lília M. Gravina<sup>1</sup>

<sup>1</sup>Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego, nº2000, Parque California, CEP 28035-200, Campos dos Goytacazes, RJ/Brasil

<sup>2</sup>Universidade Federal de Viçosa, Avenida Peter Henry Rolfs, s/n - Campus Universitário, Viçosa - MG, 36570-900, Brasil

<sup>3</sup>Embrapa Milho e Sorgo, Rod. MG 424 Km 45, Zona Rural, CEP 35701-970, Sete Lagoas, MG/Brasil

<sup>4</sup>Universidade Federal de Juiz de Fora, Rua José Lourenço Kelmer, S/n - Martelos, Juiz de Fora - MG, 36100-040/Brasil

\*Corresponding author: agro.camilaqs@gmail.com

### Abstract

Elephant-grass (*Pennisetum purpureum* Schum) is known to have a great potential for biomass production, but studies on genotype responses to different sites and growing seasons are still scarce and necessary. This study aimed to estimate correlation and path coefficient analysis for biomass production trait in elephant grass under semi-annual cutting cycles. The experiment was conducted from 2012 to 2015 in a randomized block design with two replications. Seventy-three elephant grass cultivars (genotypes) were assessed during six cutting cycles. Cuts were manually performed when grasses reached a height of 1.5 m (semi-annually). The following variables were analyzed HGT, SD, NT, LW, %DM and DMP. Data were subjected to individual and joint variance analysis, in addition to analysis of genotypic, environmental, and phenotypic correlation and path analysis decomposition. Both genotypes and cuts had significant effects on all the variables. HGT, SD, and LW were significant, positive, and genotypically correlated with DMP. In terms of direct effects on DMP, NT (1.44) stood out with the highest effect and total correlation, thus showing a major contribution to DMP increases. Regarding indirect effects, %DM had a positive influence on DMP via NT and LW. The selection of plants with a high %DM included genotypes with lower SD, HGT, and LW and higher NT. The environmental component had a greater influence on the relationships involving both NT and SD.

**Keywords:** *Pennisetum purpureum*, variables, dry matter, plant breeding.

**Abbreviations:** HGT\_plant height, SD\_stem diameter, NT\_number of tillers, LW\_leaf blade width, %DM\_percentage of dry matter, DMP\_dry matter production.

### Introduction

Elephant grass (*Pennisetum purpureum* Schum) is a Poaceae species of tropical origin adapted to different environmental conditions (Kollet et al., 2006). This grass species has a high content of biomass (Lima et al., 2011), which makes it very attractive to the energy sector since it reaches a high productivity within a short-term. Added to this, other variables related to biomass quality make it a good choice for energy production such as its high fiber content, mainly related to high carbon content components (e.g. cellulose and lignin), as well as its high carbon/nitrogen ratio (C/N) (Zanetti, 2009).

The elephant-grass breeding program at the State University of Northern Rio de Janeiro Darcy Ribeiro (UENF) is responsible for studies on the feasibility of using elephant grass biomass, identification of genotypes with high productive potential, and biomass quality for energy

purposes (Souza et al., 2017). The study of relationships between traits of this species has a major importance for its breeding program (Cunha et al., 2011). However, most of these variables have a complex inheritance, i.e. having several genes with low heritability and high interaction between genotype and environment, what makes breeding a challenge (Cruz et al., 2012).

If compared to a direct selection, an indirect selection, by means of less complex traits with higher heritability and easy evaluation, could result in higher genetic progress. Thus, studies on correlated traits have received great attention in breeding programs (Yasin and Singh, 2010). According to Das and Taliaferro (2009), when one trait is correlated to another, it is possible to assess the changes in a trait using a plant selection process for its correlated one.

The quantification and interpretation of the magnitude of

correlation can result in misconceptions for selection strategy because high correlations may be derived from the effect of a third or a group of traits (Cruz et al., 2012). In this context, a path analysis can be used to understand the causes of associations between traits and to decompose the existing correlations into direct and indirect effects, using the main variable and explanatory variables. In short, according to Rao et al. (2013), a path analysis consists of formulating cause-effect relationships between two variables.

Our research has been innovative in assessing responses of genotypes to different cropping locations and seasons, what had not yet been done for this forage species. The main goal is to identify genotypes with high productive potential and biomass quality for use as energy source. Following this line, this study evaluated the phenotypic, genetic, and environmental correlations, as well as performing a path analysis of 73 elephant grass genotypes for energetic purposes.

## Results and Discussion

### Genetic estimates

The effects of genotypes and cuts were significant ( $p < 0.05$  and  $p < 0.01$ ) (Supplementary Table 2) for all the assessed variables. This result points to a genetic variability that allows identifying promising genotypes for breeding programs. Moreover, the interaction genotypes  $\times$  cuts was significant for all the variables ( $p < 0.05$ ), thus the responses of genotypes differed for the different cuts, which was already expected due to the heterogeneity of environmental conditions.

Maximum heritability estimates (Supplementary Table 2) had a satisfactory magnitude for LW (85.73), SD (85.31), and NT (85.34), which enables selection of superior genotypes. Knowing the maximum value of heritability is fundamental in case of perennial crops such as elephant grass. This is because of the time needed from the beginning of experimentation, after evaluations, and until the selection of promising genotypes to be used as bioenergy source (Shimoya et al., 2002).

### Correlation estimates

Table 1 shows the genotypic, phenotypic, and environmental correlation coefficients among the six studied traits. When considering the genotypic correlations, two out of 15 estimates did not differ from zero by the t-test ( $p > 0.05$ ). A significant correlation is an indicator of the viability of indirect selection to obtain gains in a given trait of great economic importance (Cruz et al., 2012).

DMP showed positive and significant genotypic and phenotypic correlations with NT, HGT, and SD at 1% and 5% probability levels. Conversely, dry matter production had no genetic correlation with %DM and LW, indicating no relationship between these variables.

In elephant-grass breeding programs for biomass production, genotypes with a higher number of tillers, larger leaf diameters, and high dry matter production are sought. In the case of elephant grass, genotypes with higher NT, SD, and DMP are ideal to select genotypes for biomass

production. Cunha et al. (2011) studied the association between morphological and productive variables in genotypes of *Pennisetum* sp. and associated plants with a higher DMP to those with a higher density of basal tillers, height, and lower number of green leaves per tiller. Likewise, Menezes et al. (2014) assessed the correlations of forty genotypes of elephant grass under two cutting regimes and the traits HGT, SD, and LW; these traits showed positive and significant genotypic correlation with dry matter production. The authors also observed that HGT showed the highest positive and direct effect on DMP during the second cut. Positive and significant genotypic and phenotypic correlations were found for %DM with NT, but negative with HGT, SD, and LW.

The genotypic correlation between NT and SD was significant and negative. According to Menezes et al. (2014), elephant grass genotypes with a higher NT tended to present a lower SD and vice versa due to competition for light and water, i.e. denser plantings (denser tillering) result in higher plants with lower DMP and SD.

The variables HGT, SD, and LW showed positive and significant genotypic correlation with DMP; therefore, by selecting genotypes with higher values of HGT, SD, and LW, plants showing larger dry matter production are directly selected. If two traits have significant correlation, gains in one could be achieved by indirect selection of the other. This is advantageous mainly when a high economic value trait has low heritability and/or is difficult to be evaluated in comparison to another with which it is associated. Thus, selection is made using a high heritability trait and/or an easy evaluation one, but aiming to improve the other (Rossi et al., 2014).

Thus, to select elephant grass genotypes with high heritability for DMP focusing on energy biomass production, plants could be indirectly selected by means of the following variables NT ( $h^2 = 85.34$ ) and SD ( $h^2 = 85.3$ ).

In cases of environmental influence, negative correlations between two traits indicate environmental advantages for one trait to the detriment of the other, while positive values signal that both traits are benefited or harmed by the same causes (Cruz et al., 2012). In general, neither positive nor negative values prevailed, which denotes a complex relationship regarding advantages and/or disadvantages caused by environmental conditions.

In some instances, environmental correlation was higher than its respective genotypic and phenotypic counterparts were; in other words, the environment had greater influence on the expression traits than did the genotype. Among them, DMP had positive environmental correlations with DM% and LW, while %DM had negative and positive correlations with NT and HGT, respectively. According to Falconer (1997), such differences involving genotypic and environmental correlations indicate that genetic variations influence traits in a rather different fashion than do the environmental changes.

### Effect estimates

Table 2 displays the direct and indirect effects of explanatory independent variables (%DM, HGT, SD, and NT) on DMP of elephant grass in six cuts.

The degree of multicollinearity among the explanatory

**Table 1.** Phenotypic (P), genotypic (G), and environmental (E) correlation coefficients for the six variables assessed in 73 elephant grass genotypes, based on the means of each cut.

	r	%DM	NT	HGT	SD	LW
DMP	P	0.24*	0.42**	0.51**	0.22*	0.07**
	G	0.19 <sup>ns</sup>	0.45**	0.58**	0.30*	0.02 <sup>ns</sup>
	E	0.87**	-0.60**	0.38**	-0.33*	0.53**
%DM	P		0.53**	0.25*	-0.44**	-0.28*
	G		0.69**	-0.36**	-0.55**	-0.31**
	E		-0.89**	0.76**	-0.09 <sup>ns</sup>	0.18 <sup>ns</sup>
NT	P			-0.27*	-0.66**	-0.61**
	G			-0.39**	-0.74**	-0.70**
	E			-0.83**	-0.20 <sup>ns</sup>	0.28*
HGT	P				0.56**	0.36**
	G				0.65**	0.42**
	E				0.01 <sup>ns</sup>	-0.16 <sup>ns</sup>
SD	P					0.75**
	G					0.80**
	E					-0.77**

DMP – dry matter production, %DM – percentage of dry matter, NT – number of tillers, HGT – plant height, SD – stem diameter, and LW – leaf blade width. <sup>ns</sup> non-significant, \*\*significant at 1%, and \*significant at 5% probability by the F-test.

**Table 2.** Path analysis: breakdown of genotypic correlations into components of direct and indirect effect involving the main dependent variable DMP (dry matter production, in Mg ha<sup>-1</sup>) and independent explanatory variables (%DM, HGT, NT, SD, and LW) assessed in six cuts of 73 elephant grass genotypes.

Variable	Estimation of correlation	Total
<b>%DM</b>		
Direct effect on DMP	-0.08	
Indirect effect of NT	1.00	
Indirect effect of HGT	-0.15	
Indirect effect of SD	-0.60	
Indirect effect of LW	0.02	
		0.19
<b>NT</b>		
Direct effect on DMP	1.44	
Indirect effect of %DM	-0.05	
Indirect effect of HGT	-0.16	
Indirect effect of SD	-0.81	
Indirect effect of LW	0.04	
		0.45
<b>HGT</b>		
Direct effect on DMP	0.42	
Indirect effect of %DM	0.03	
Indirect effect of NT	-0.56	
Indirect effect of SD	0.71	
Indirect effect of LW	-0.02	
		0.58
<b>SD</b>		
Direct effect on DMP	1.10	
Indirect effect of %DM	0.04	
Indirect effect of NT	-1.07	
Indirect effect of HGT	0.27	
Indirect effect of LW	-0.04	
		0.30
<b>LW</b>		
Direct effect on DMP	-0.05	
Indirect effect of %DM	0.02	
Indirect effect of NT	-1.01	
Indirect effect of HGT	0.18	
Indirect effect of SD	0.88	
		0.02
Determination coefficient (R <sup>2</sup> )	1.21	

%DM – percentage of dry matter; HGT – average plant height; SD – average stem diameter; NT – number of tillers per linear meter; PMS – dry matter production.

variables should be verified to obtain reliable results for path analysis (Cruz and Carneiro, 2014). According to Toebe and Cargnelutti Filho (2013), the lower the multicollinearity, the higher the determination coefficient (R<sup>2</sup>) and the lower the residual variable effects will be, so that a higher predictive power could be achieved using a path analysis. The explanatory model adopted was the most efficient in expressing cause and effect relationships between an explanatory variable and DMP since R<sup>2</sup> was 1.21 (Table 2). Among the variables directly influencing the DMP of elephant grass (included in Table 2), NT (1.44) showed the

highest effect and total correlation, thus contributing most to DMP increases, surpassing SD that also had a high direct effect (1.10). In contrast, HGT (0.42) had the lowest effect. These results may be partially explained by positive and significant genotypic correlations between explanatory variables and DMP (Table 1). Therefore, when selecting larger genotypes (higher NT, HGT, and DS), clones for higher dry matter production are directly selected. Furthermore, the variables %DM and LW also showed a direct influence on DMP, but negatively. Regarding indirect effects, an indirect and positive influence of %DM on DMP was obtained by

increasing NT and LW (Table 2). In short, to generate plants with a higher DMP for energy biomass production, plants with higher NT and LW could be selected, leading to a higher %DM and hence a higher DMP. However, the direct effect of LW on DMP was negative, whereas, except for NT, the indirect effects from other variables via LW were positive, although not high. The path analysis results showed that the wider the leaves, the lower the DMP of elephant grass plants (Table 2). Indirectly, the higher the HGT and the SD, the lower the LW will be, and thus the higher the DMP values.

## Materials and Methods

### Plant materials

Seventy-three elephant-grass genotypes (Supplementary Table 1) from Germplasm Active Bank at UENF (BAG-UENF) were assessed in this study.

### Experimental area location and characterization

The experiment was carried out at the Experimental Station of the Research Center on Agroenergy and Waste Utilization (PESAGRO-Rio), located at 21°19'23" S, 41°19'40" W, and an average altitude of 25 m, in the city of Campos dos Goytacazes - northern Rio de Janeiro State (Brazil). According to Köppen (1948), the local climate is classified as an Aw type, which stands for warm and tropical humid with annual rainfall of about 1,152 mm. The soil is classified as an Oxisol with a pH of 5.5, phosphorus of 18 mg dm<sup>-3</sup>, potassium of 83 mg dm<sup>-3</sup>, Ca of 4.6 cmolc dm<sup>-3</sup>, Mg of 3.0 cmolc dm<sup>-3</sup>, Al of 0.1 cmolc dm<sup>-3</sup>, H+Al of 4.5 cmolc dm<sup>-3</sup>, and C of 1.6%.

### Experimental design and genotypes

The experimental design was a completely randomized block with two replications. Each plot contained one row of 5.5 m, and plots were spaced 2 m apart, totaling 11 m<sup>2</sup>. Each replication contained 73 elephant grass (Supplementary Table 1). Six cuts were performed (June 2012, December 2012, August 2013, February 2014, August 2014, and February 2015), which were considered as the environments.

Planting fertilization consisted of 60 g single superphosphate evenly distributed along each row. After 50 days, topdressing was held spreading 70 g urea and 40 g KCl per row, which correspond to 28.6 kg N and 24 kg K<sub>2</sub>O per hectare. Topdressings were carried out after each cut. All the fertilizations were carried out based on soil analysis and recommendations proposed by Freire et al. (2013) for the Rio de Janeiro state.

After each cut, plant samples were collected for evaluation of the following variables: total dry matter production (DMP; t ha<sup>-1</sup> cut<sup>-1</sup>), number of tillers (NT; tiller per linear meter), plant height (HGT; meters), stem diameter (SD; mm), leaf blade width (LW; mm), and dry matter percentage per plant (%DM).

### Statistical analyses

GENES software version 1.0 (Cruz, 2013) was used for the genetic-statistical analysis. After measuring the variables, a

variance analysis was done using the mean of each plot, considering all random effects (random model) and using the genetic-statistical model below:

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

Wherein:  $Y_{ij}$  = phenotypic value of the ij-th observation relative to the i-th genotype in the j-th block;  $\mu$  = overall constant of the trait;  $g_i$  = effect of the i-th genotype;  $b_j$  = effect of the j-th block; and  $\varepsilon_{ij}$  = mean experimental error.

A split-plot model was used for the joint analysis as proposed by Steel and Torrie (1996), as described below:

$$Y_{ijk} = \mu + G_i + B_j + \varepsilon_a + C_k + \varepsilon_b + GC_{ik} + \varepsilon_c$$

Wherein:  $Y_{ijk}$  = observed value related to the i-th genotype in the j-th block in the k-th cut;  $\mu$  = overall constant of the experiment;  $G_i$  = random effect of the i-th genotype;  $B_j$  = effect of the j-th block;  $\varepsilon_a$  = effect of the error associated with the i-th genotype in the j-th block;  $C_k$  = random effect of k-th cut;  $\varepsilon_b$  = effect of the error b associated with the j-th block in k-th cut;  $GC_{ik}$  = effect of the interaction of the i-th genotype with the k-th cut; and  $\varepsilon_c$  = effect of the error c associated with the i-th genotype in the j-th block in the k-th cut.

### Phenotypic, genotypic, and environmental correlations

The estimators of phenotypic, genotypic, and environmental correlation coefficients were obtained through the equations proposed by Johanson et al. (1955), Miller et al. (1958), and Singh and Chandhury (1985), as follows:

a) Phenotypic correlations:

$$r_p = \frac{MGP}{\sqrt{MSG_x MSG_y}}$$

b) Genotypic correlations:

$$r_g = \frac{MGP_{xy} - MPR_{xy}/r}{\sqrt{\hat{\sigma}_{g(x)}\hat{\sigma}_{g(y)}}} = \frac{\hat{\sigma}_{g(xy)}}{\sqrt{\hat{\sigma}_{g(x)}\hat{\sigma}_{g(y)}}}$$

c) Environmental correlations:

$$r_e = \frac{MPR_{xy}}{\sqrt{MSR_x MSR_y}}$$

Wherein:  $MPG_{XY}$  = mean product deviation between genotypes and the traits X and Y;  $MPR_{XY}$  = mean product deviation between residuals for the traits X and Y;  $MSG_x$  = mean square deviation between genotypes for the trait X;  $MSG_y$  = mean square deviation between genotypes for the trait Y;  $MSR_x$  = mean square deviation between residuals for the trait X;  $MSR_y$  = mean square deviation between residuals for the trait Y; and  $\hat{\sigma}_{g(y)}$  = genotypic covariance estimator

represented by:

$$\hat{\sigma}_{g(y)} = \frac{MSG_y - MSR_y}{r}$$

Wherein:  $\hat{\sigma}_{g(x)}$  and  $\hat{\sigma}_{g(y)}$  = estimators of quadratic components associated with genotypic variability for the traits X and Y, respectively.

A multicollinearity diagnosis was performed involving the six traits assessed. Path analysis results were interpreted as proposed by Singh and Chandhury (1985), as follows: i) correlation coefficients and high direct effects (path coefficients) indicate that these independent variables explained much of the variation in the basic variable, and ii) positive (or negative) correlation coefficients, but with a

direct effect of different or negligible signal, indicates that the variables with the greatest indirect effects must be considered simultaneously to explain the variation of the basic variable.

## Conclusion

Dry matter production is correlated positively and directly with the number of tillers, plant height, and stem diameter. The variable that best explains directly the dry matter production is the number of tillers. Genotypes with a high percentage of dry matter can be selected by choosing plants with smaller stem diameter, height, and leaf width, but with a higher number of tillers. The environmental component had a greater influence on the relationships involving number of tillers and diameter of the stem with dry matter production.

## Acknowledgments

The authors thank the National Council for Scientific and Technological Development (CNPq) and the Rio de Janeiro Research Foundation (FAPERJ).

## References

- Cruz CD, Regazzi, AJ, Carneiro PCS (2012) Modelos Biométricos Aplicados ao Melhoramento Genético. Viçosa: UFV 1:514.
- Cruz CD (2013) GENES: a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum*, 35:271-276.
- Cruz CD, Carneiro PCS. (2014). Modelos Biométricos Aplicados ao Melhoramento Genético. Viçosa: UFV 1:668.
- Cunha MV, Lira MA, Santos MVF, Freitas EV, Dubeux Junior JCB, Mello, ACL, Martins, KGR (2011) Association between the morphological and productive traits in the selection of elephant grass. *Revista Brasileira de Zootecnia*, 40:482-488.
- Das MK, Taliaferro CM (2009) Genetic variability and interrelationships of seed yield and yield components in switchgrass. *Euphytica* 167:95-105.
- Falconer DS, Mackay TFC (1997) Introduction to quantitative genetics. Longman Press 459p.
- Freire LR, Balieiro FDC, Zonta E, Anjos LD, Pereira MG, Lima E, Polidoro JC (2013) Manual de Calagem e Adubação do Estado do Rio de Janeiro. Editora Universidade Rural, 179p.
- Kollet JL, Diogo JMS, Leite GG (2006) Rendimento forrageiro e composição bromatológica de variedades de milheto (*Pennisetum glaucum* L.). *Revista Brasileira de Zootecnia* 35:1308-1315.
- Lima RSN, Daher RF, Gonçalves LSA, Rossi DA, Amaral Junior AT, Pereira MG, Léo FJS (2006) Estimativa da diversidade genética entre genótipos de capim-elefante (*P. purpureum Schum*) baseada em marcadores de DNA (RAPD e ISSR). *Genetics and molecular research* 10:1304-131.
- Johnson HW, Robinson HF, Comstock RE (1955) Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47:314-318.
- Menezes BRF, Daher RF, Gravina GA, Amaral Junior AT, Oliveira AVDE, Schneider LSA, Silva VB (2014) Correlações e análise de trilha em capim-elefante para fins energéticos. *Revista Brasileira de Ciências Agrárias* 9:465-295.
- Miller PA, Williams JC, Comstock HF (1958) Estimation of genetic and environmental variance in upland cotton and their implications in selection. *Agronomy Journal* 50:26-131.
- Oliveira MLF, Daher RF, Gravina GA, Silva VB, Viana AP, Rodrigues EV, Shimoya A, Amaral Júnior AT, Menezes BRS, Rocha AS (2014) Pre-breeding of elephant grass for energy purposes and biomass analysis. *Africa Journal Agriculture* 9:2743-2758.
- Rao PJM, Malathi S, Reddy DVV, Upender M (2013) Genetic Studies of Association and Path Coefficient Analysis of Yield and its Component Traits in Pigeon Pea (*Cajanus cajan* L. Mill sp.). *International Journal of Scientific and Research Publications* 3:1-5.
- Rossi DA, Menezes BRS, Daher RF, Gravina GA, Lima RSN, Léo FJS, Gottardo RD, Campostrini E, Souza CLM (2014) Canonical correlations in elephant grass for energy purposes. *Africa Journal Biotechnology* 13:3666-3671.
- Shimoya A, Cruz CD, Ferreira RP, Pereira AV, Carneiro PCS (2002) Divergência genética entre acessos de um banco de germoplasma. *Pesquisa Agropecuária*, 37:971- 980.
- Singh RK, Chandhury BD (1985) Biometrical methods in quantitative genetic analysis. Kalyni Publishers. New Delhi. 318p.
- Steel RGD, Torrie JH (1996) Principles and Procedures of Statistics: a Biometrical Approach. 2 th ed. New York McGraw-Hill Kogakusha. 318p.
- Souza YP, Daher RF, Pereira AV, Silva VB, Freitas RS, Gravina GA (2017) Repeatability and minimum number of evaluations for morpho-agronomic traits of elephant-grass for energy purposes. *Revista Brasileira de Ciências Agrárias* 12:391-397
- Toebe M, Cargnelutti Filho A (2013) Não normalidade multivariada e multicolinearidade na análise de trilha em milho. *Pesquisa Agropecuária Brasileira* 48:466-477.
- Zanetti JB, Moraes RF, Leite JM, Soares LH, Jantalia CP, Alves BJR, Urquiaga, S, Boddey RM (2009) Produção e Qualidade da Biomassa de Genótipos de Capim-Elefante para uso como Fonte de Energia. *Revista Brasileira de Agroecologia* 4:100–102.
- Yasin AB, Singh S (2010) Correlation and path coefficient analyses in sunflower. *Journal of Plant Breed and Science* 2:129-133.