

Factor analysis and environmental stratification in the assessment of grain sorghum adaptability and stability

Pedro César de Oliveira Ribeiro¹, Felipe Vicentino Salvador¹, Isadora Cristina Martins Oliveira² and Cícero Beserra de Menezes^{2,*}

¹Departamento de Biologia, Universidade Federal de Viçosa (UFV), Viçosa, MG, Brazil. ²Embrapa Milho e Sorgo, Sete Lagoas, MG, Brazil. *Corresponding author: cicero.menezes@embrapa.br

ABSTRACT

The environmental stratification studies are crucial when releasing hybrids for different growing regions. An outstanding performance of a genotype in one environment does not qualify it for indication to all environments, due the occurrence of G×E interaction. Environmental stratification aim the breeders to form groups of environments that minimize G×E interaction. The purpose of this work was to evaluate the use of factor analysis in preliminary environmental stratification assisting at the recommendation of grain sorghum cultivars. Twenty-five hybrids were evaluated, using a randomized block design, in 12 locations during the 2015/16 season. Initially, the individual analysis of the experiments was carried out and later the joint analysis, aiming to examine the existence of G×E interaction. The means of the hybrids in the individual analyses were used to obtain the correlation matrix between pairs of environments. The factorization of this matrix was also carried out via factor analysis in order to group together the environments that most correlated with respect to the hybrids performance. Thus, differential performance between hybrids was observed through individual analyses for all the environments, with the exception of Sete Lagoas and Teresina. The joint analysis revealed the existence of a significant G×E interaction, that is, a differential behavior of the hybrids in relation to the evaluated environments. Based on the criterion of the analysis of the proportion of explained variance, it was found that six factors captured an accumulated variation of 86.29%, and the average communality observed was of 0.86. Considering the geographic and edaphoclimatic variables in the cultivation period, a pattern was not observed among the grouped places, but it is noteworthy that the grouping of places is a function of the performance of the evaluated genotypes, which can be similar even under different conditions. Given the results presented, factor analysis proved to be a tool with potential to perform environmental stratification and assist in the recommendation of grain sorghum cultivars for different regions.

Keywords: *Sorghum bicolor*, multi-environment trials, factor analytic models, animal feed, replacing corn, human food, flour for manufacture.

INTRODUCTION

For presenting characteristics such as drought tolerance and resistance to high temperatures, sorghum [*Sorghum bicolor* (L.) Moench] is one of the most cultivated cereals in the world. In Brazil, the planted area grows every year, becoming an alternative in succession to soybeans during the off-season (Rooney, 2007; Ribas, 2014; Companhia Nacional de Abastecimento [CONAB], 2018; Food and Agriculture Organization of the United Nations [FAO] 2018). Among the different types of sorghum, grain sorghum stands out, whose grain can be used in animal feed, replacing corn, and in human food, in the form of flour for the manufacture of baked goods and pasta (Rooney, 2007; Silva et al., 2013; Martino, Cardoso, Moraes, Sant'ana, & Queiroz, 2014).

Quantitative traits, such as grain yield, are greatly influenced by environmental factors under the phenotypic value of the evaluated individuals (Carvalho et al., 2021a; Carvalho et al., 2021b). In the case of the evaluation of a group of genotypes in various environments, the verification of the existence and magnitude of the genotypes × environments interaction (G×E) becomes crucial for breeding programs, as this one can affect the classification of the evaluated genotypes in each environment tested, making it difficult to select superior genotypes (Cruz, Regazzi, & Carneiro, 2012). Thus, it has been common practice, especially in the final phase of breeding programs, to evaluate potential cultivars in various locations, crops

and years, before they can be recommended to farmers (Bernardo, 2010).

Among the ways to circumvent the effects of the G×A interaction, Cruz et al. (2012) highlight environmental stratification, in which it is sought through biometric techniques, to identify patterns of similarity of genotype responses in the various environments considered; forming sub-regions of more homogeneous areas. In these sub-regions, the G×A interaction must be non-significant or of the simple type, which does not alter the classification of the genotypes, causing the recommendation for the set of environments to be facilitated (Cruz et al., 2012).

Murakami and Cruz (2004) proposed the use of the multivariate technique of the factor analysis in the context of environmental stratification. In this methodology, the various environments used for the evaluation of a particular characteristic are treated as variables and are grouped into latent variables called factors. Environments grouped in a same factor present a high correlation with each other and a low correlation with the environments of the other factors, thus allowing the stratification of environments into homogeneous groups (Murakami & Cruz, 2014). Environmental stratification allows for the disposal of some assessment sites with the objective of optimizing resources and can be replaced by new sites that represent edaphoclimatic regions not yet assessed in the final stage trials (Oliveira et al., 2020).

This study aimed to evaluate the use of factor analysis for a preliminary environmental stratification aiming at the adequate recommendation of grain sorghum cultivars and reduction of the number of environments evaluated in the final evaluation tests.

MATERIALS AND METHODS

Twenty-five grain sorghum hybrids were evaluated for determination of the value for cultivation and use (VCU), being 23 experimental hybrids (1-1167093, 2-1G244, 3-1421045, 4-1099034, 5-1167048, 6-1236043, 7-1099044, 8-1168092, 9-1170093, 10-1239020, 11-1096019, 12-0729033, 13-1167053, 14-0843009, 15-1168093, 16-1238020, 17-1236020, 18-1170017, 19-1105661, 20-1169092, 21-1167017, 22-1167092 and 23-1169054), belonging to the Embrapa Milho e Sorgo breeding program and two commercial checks, BRS 330 and DKB 550.

The experiments were conducted in 12 locations distributed throughout the Brazilian territory during the 2015/2016 off-season crop. In the evaluated locations, there are considerable geographic variations and distinct environmental conditions (Table 1). The experimental plots were composed of four rows of 5 m spaced 0.5 m considering that only the two central rows were considered as useful area. The design used was a randomized block design, with three replications. Two rows with commercial hybrids were also planted parallel to the blocks at the distance of 0.5 meters to serve as a border plants to the experimental plots of the blocks.

The planting and topdressing fertilizations were carried out based on soil analysis and cultural and phytosanitary treatments according to the needs of the crop and of each region. The trait grain yield (GY: grain weight of the useful area in t. ha⁻¹) was evaluated after the plots were threshed and the grain moisture was corrected to 13%.

Initially, the individual analysis of each experiment was carried out aiming to verify the existence of differential performance of the hybrids for grain yield in each environment, in addition to evaluating the experimental precision in each location given by the coefficient of variation (CV%).

Subsequently, the joint analysis of the 12 trials was conducted, aiming to ascertain the existence of variation due to genotypes × environments interaction (G×E). For this purpose, it was previously verified whether the ratio between the largest and smallest mean square of the residue were greater than seven, as proposed by Gomes (2009), in order to carry out the analysis with residue homogeneity.

The means of the 25 sorghum hybrids obtained in the individual analyses concerning to each location were used to obtain the matrix of correlations between pairs of environments (R). The factorization of this matrix of correlations by means of the multivariate technique of factor analysis was carried out in order to reduce it to a space with smaller dimensions, that is, into factors that can group environments that correlate with respect to the performance of the evaluated hybrids, as proposed by Murakami and Cruz (2004).

The factor analysis procedure was performed with the help of the GENES program (Cruz, 2016). In this methodology, the performance of hybrids for grain yield in each environment considered (x_j) could be described through a linear combination of factors, in addition to error, according to the following statistical model:

$$x_j = \sum_{k=1}^m \gamma_{jk} F_k + \varepsilon$$

where: m , corresponds to the number of factors used; γ_{jk} , is the factor loading for the j^{th} environment associated with the k^{th} common factor; ε , is the specific factor associated to the j^{th} environment.

To determine the number of factors to be used, the 12 eigenvalues (λ) of the R matrix were obtained and these were ordered in ascending order. Thus, the proportion of the total variation of the R matrix explained by each eigenvalue (PVE_k) was obtained, according to the formula:

$$PVE_k = \lambda_k / 12$$

Based on the PVE values obtained, the number of factors to be used was defined as that corresponding to the number of eigenvalues that explain at least 80% of the total variation. This step was performed with the help of the R software (R Core Team, 2018) according to the routine provided in anexo1.

After obtaining the initial factor loadings the factors were submitted to rotation via varimax (Kaiser, 1958), aiming to facilitate the interpretation of the factor loadings and the allocation of the environments, in which the performance of the hybrids is correlated within each factor and poorly correlated among the established factors.

Table 1. Geographical and edaphoclimatic attributes corresponding to each of the 12 sites evaluated (from January to July, 2016) aiming at the recommendation of grain sorghum hybrids.

Local	Estado	Long. ¹ °	Lat. °	Prec. mm	T°max. °C	T°min. °C	Alt. (m)
Sete Lagoas (SL)	MG	-44.28	-19.44	616.99	29.04	17.27	767
Goiânia (GOI)	GO	-49.25	-16.66	830.60	32.03	19.77	741
Montividiu (MON)	GO	-51.10	-17.26	843.40	34.60	22.05	857
Guaíra (GUA)	PR	-54.15	-24.04	823.60	30.66	20.50	220
Luís Eduardo Magalhães (LEM)	BA	-46.11	-12.24	690.42	32.27	20.18	720
Vilhena (VIL)	RO	-60.26	-11.98	1098.70	31.19	21.34	615
Teresina (TER)	PI	-42.83	-5.27	867.38	33.61	23.24	72
Sinop (SIN)	MT	-55.48	-11.78	1101.69	32.39	21.48	384
Rondonópolis (RON)	MT	-54.63	-16.45	732.29	31.90	20.83	227
Cáceres (CAC)	MT	-57.40	-16.04	-	31.19	24.96	176
Uberlândia (UBE)	MG	-47.97	-19.55	897.26	28.85	17.98	863
Rio Verde (RV)	GO	-51.02	-17.72	711.58	31.17	19.54	739

¹Long.: Longitude (degrees), Lat.: Latitude (degrees), Prec: Accumulated rainfall (mm), T° max.: Maximum Temperature (°C), T°min. Minimum temperature (°C), Alt.: Altitude (meters).

RESULTS AND DISCUSSION

Distinct performance for grain yield was observed among the sorghum hybrids by means of the individual analyses in most the environments tested, with the exception of Sete Lagoas (SL) and Teresina (TER) (Table 2). It is also found that, in general, the values of coefficients of variation were close to those reported in research studies on the sorghum crop, ranging from 11.04 to 30.84%, showing that the levels of experimental precision were in agreement with those reported in literature, 14.89 (Martins et al., 2016), 15.15 (Andrade et al., 2016) and 16.63 (Silva et al., 2016). On the other hand, the environments of Luís Eduardo Magalhães (LEM) and Rondonópolis (RON) presented a lower experimental precision (30.84 and 27.88%, respectively), which could be explained by the strong influence of the environment on the evaluation of grain yield.

Table 2. Summary of individual and joint analyzes referring to the evaluation of 25 grain sorghum hybrids for grain yield in 12 different environments, 2015/2016 crop.

Grain Yield (t.ha ⁻¹)																
Individual analyses													Joint analysis			
SV	FD	Mean Square											SV	FD	MS	
		SL ¹	GOI	MON	GUA	LEM	VIL	TER	SIN	RON	CAC	UBE				RV
Block	2	0.37	3.67	37.02	1.10	1.48	1.08	2.45	0.65	0.14	0.01	0.69	18.21	B/A	2	10.65
Treat.	24	0.80 ^{ns2}	2.43**	3.28**	2.23**	3.60**	2.27**	0.34 ^{ns}	1.70**	2.08**	5.34**	0.96**	2.70**	Amb.	11	112.16**
Res.	48	0.57	0.31	1.23	0.85	1.61	1.03	0.21	0.34	0.77	1.80	0.42	1.18	Trat.	24	7.33 *
Mean		4.35	4.35	5.01	6.83	4.12	4.92	4.14	3.55	3.15	5.80	3.06	6.52	G×A		1.86**
CV(%)		17.36	12.88	22.10	13.51	30.84	20.57	11.04	16.39	27.88	23.12	21.04	16.65	Res.	48	0.86

¹SL: Sete Lagoas; GOI: Goiânia; MON: Montividiu; GUA: Guaíba; LEM: Luís Eduardo Magalhães; VIL: Vilhena; TER: Teresina; SIN: Sinop; RON: Rondonópolis; CAC: Cáceres; UBE: Uberlândia; RV: Rio Verde. ²ns: non-significant (p>0.05); *, **: significant at 1% and 5% of probability (p<0.05) by the F test, respectively.

The joint analysis involving the 12 locations revealed the existence of a significant G×E interaction (Table 2), demonstrating the distinct behavior of the hybrids in relation to environmental variations. This lack of consistency in the performance of the hybrids evaluated across the considered environments makes it difficult to recommend a single hybrid for a set of locations. Thus, the use of biometric techniques in order to form groups of environments whose G×E interaction does not change the ranking of hybrids is justified, which can facilitate the recommendation of genotypes for different environments.

The similarity of performance between some pairs of environments can be observed by means of the phenotypic correlation estimates (Figure 1). The correlations ranged between -0.25 (Vilhena and Teresina) and 0.65 (Goiânia and Guaíra), with a mean of 0.24. The highest correlations (dark blue) were observed between Goiânia and Guaíra (0.65), Vilhena and Rio Verde (0.64), Sinop and Uberlândia (0.61) and Sete Lagoas and Goiânia (0.57), with Goiânia and Rio Verde being the places with the highest correlations between environments. It is also noticed that the lowest correlations (light pink) were observed for Luís Eduardo Magalhães, showing that the treatments behave in a more contrasting way in this location, when compared to the other trials. Another observation was the zero correlation found between the Sete Lagoas and Luís Eduardo Magalhães environments which reinforces the significance of the G×E interaction in this study.

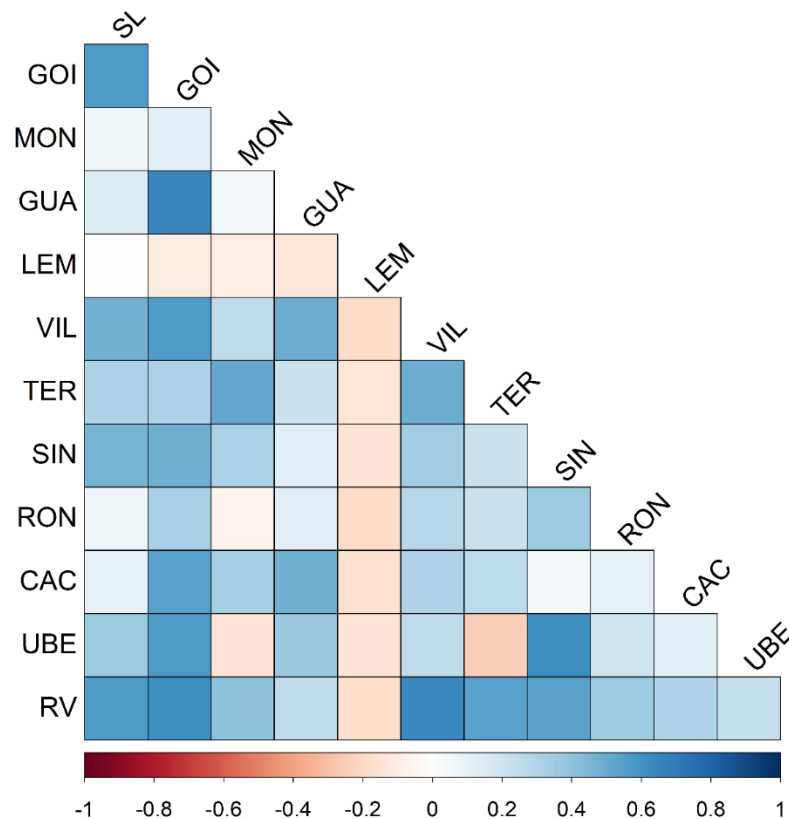


Figure 1. Phenotypic correlations between pairs of environments for grain sorghum yield. SL: Sete Lagoas; GOI: Goiânia; MON: Montividiu; GUA: Guaíra; LEM: Luís Eduardo Magalhães; VIL: Vilhena; TER: Teresina; SIN: Sinop; RON: Rondonópolis; CAC: Cáceres; UBE: Uberlândia; RV: Rio Verde.

In Table 3 are presented the eigenvalues of the correlation matrix and the proportion of variance explained by each eigenvalue. Based on the criterion of analysis of the proportion of explained variance, it was observed that six factors would be necessary to capture an accumulated variation of 86.29%, a value above the minimum of 80% used and proposed by Murakami and Cruz (2004) and treated in Cruz, Regazzi and Carneiro (2014). Therefore, six factors were used in the analysis aiming at grouping the tested environments.

The communality values for each environment are shown in Table 4. These ranged from 0.74 to 0.99 with a mean of 0.86. These results indicate that a large part of the variation observed among the hybrids in the tested environments was explained by six common factors, suggesting their potential to summarize the information and establish groups of correlated environments.

Table 3. Eigen values (λ) of the correlation matrix between pairs of environments and proportion of the variation explained by each eigenvalue (PVE_k) and accumulated (PVE) of the k factors under study.

k	λ_k	PVE_k (%)	PVE acumulated (%)
1	4.41	36.74	36.74
2	1.78	14.84	51.58
3	1.36	11.37	62.95
4	1.11	9.23	72.18
5	0.92	7.65	79.84
6	0.77	6.45	86.29
7	0.55	4.62	90.91
8	0.37	3.08	93.99
9	0.26	2.19	96.18
10	0.19	1.57	97.75
11	0.14	1.19	98.94
12	0.13	1.06	100.00

The factor loadings obtained for each environment after the rotation of the factors indicate the correlation of the environments with each common factor. According to the traditional factor analysis applied to multiple variables, as is the case presented in this work, environments with high loadings for a given factor are highly correlated within this factor and less correlated with the high correlation environments of the other factors (Cruz et al., 2014; Rocha et al., 2017). Thus, factor 1 grouped the environments of Sinop and Uberlândia, factor 2 comprised the environments Sete Lagoas, Vilhena and Rio Verde, and factor 3 grouped the environments Goiânia, Guaíra and Cáceres. The other factors, 4, 5 and 6, explained one environment each, namely: Luís Eduardo Magalhães, Montividiu and Rondonópolis, respectively (Table 4).

Table 4. Values of the factor loadings and communality after the rotation for the 12 environments under study.

Environments¹	Factors						Communality
	F1	F2	F3	F4	F5	F6	
SL	0.37	0.82	0.00	-0.08	-0.05	-0.16	0.85
GOI	0.43	0.50	0.65	-0.05	0.02	0.16	0.88
MON	0.02	0.13	0.09	0.04	0.94	-0.08	0.92
GUA	0.10	0.21	0.85	0.08	-0.11	0.04	0.80
LEM	-0.07	-0.05	-0.08	-0.98	-0.06	-0.08	0.99
VIL	0.04	0.73	0.38	0.14	0.08	0.17	0.74
TER	-0.33	0.61	0.13	0.08	0.51	0.23	0.82
SIN	0.79	0.32	-0.06	0.05	0.33	0.26	0.91
RON	0.15	0.12	0.08	0.08	-0.04	0.96	0.96
CAC	0.01	0.01	0.81	0.06	0.35	0.02	0.78
UBE	0.88	0.09	0.27	0.09	-0.23	0.04	0.92
RV	0.22	0.71	0.16	0.07	0.36	0.27	0.79

¹SL: Sete Lagoas; GOI: Goiânia; MON: Montividiu; GUA: Guaíra; LEM: Luís Eduardo Magalhães; VIL: Vilhena; TER: Teresina; SIN: Sinop; RON: Rondonópolis; CAC: Cáceres; UBE: Uberlândia; RV: Rio Verde.

Based on the groups of environments obtained, it is important to stress that the effect of the interaction genotypes x environments that is being mitigated, corresponds to the interaction of the hybrids tested with the different locations. Therefore, the grouping of locations obtained is only a previous result and indicates

the potential of the factor analysis to perform the environmental stratification which makes the recommendation of new cultivars easier. However, more decisive conclusions about that stratification for recommendation purposes could be obtained in future works which take into consideration also the evaluation of the hybrids in other years, crops and locations.

CONCLUSIONS

The factor analysis proved to be a tool with a high potential to perform environmental stratification, aiming to help the recommendation of grain sorghum cultivars for different regions.

REFERENCES

- Andrade, L. C., Menezes, C. B., Silva, K. J., Santos, C. V., Emygdio, B. M., & Tardin, F. D. (2016). Avaliação de produtividade, adaptabilidade e estabilidade genotípica de sorgo granífero em três ambientes. *Revista Agropecuária Técnica*, 37(0100–7467), 36–43.
- Bernardo, R. (2010). *Breeding for quantitative traits in plants*. Woodbury, MN, USA: Stemma Press.
- Carvalho, I. R., Eickhoff, F. G., Silva, T. S., Schulz, A. D., Ourique, R. S., Malheiros, T. F., ... Hutra, D. J. (2021a). Cultivation of maize in different environments and their effects on agronomic traits. *Agronomy Science and Biotechnology*, 7, 1–11. <https://doi.org/10.33158/asb.r125.v7.2021>
- Carvalho, I. R., Peter, M., Demari, G. H., Lautenchleger, F., Carlos, F. S., Pedó, T., ... Loro, M. V. (2021b). Efficiency in nitrogen management using conventional and transgenic technology in the cultivation of maize. *Agronomy Science and Biotechnology*, 7, 1–12. <https://doi.org/10.33158/asb.r124.v7.2021>
- CONAB - Companhia Nacional de Abastecimento. (2018). Sorgo. 11º Levantamento da Safra CONAB. Retrieved from http://www.conab.gov.br/index.../16482_45de496aac454b837dd01c66c884650
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético* (4th ed.). Viçosa, MG: Editora UFV.
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2014). *Modelos biométricos aplicados ao melhoramento genético* (3rd ed.). Viçosa, MG: Editora UFV.
- Cruz, C. D. (2016). Programa Genes – Ampliado e integrado aos aplicativos R, Matlab e Selegen. *Acta Scientiarum - Agronomy*, 38(4), 547–552. <https://doi.org/10.4025/actasciagron.v38i4.32629>
- FAO - Food and Agriculture Organization of the United Nations. (2018). *Global network on integrated soil management for sustainable use of salt-affected soils*. Rome, Italy. Retrieved from <http://www.fao.org/home/en/>
- Gomes, F. P. (2009). *Curso de estatística experimental* (15th ed.). Piracicaba, SP: Editora ESALQ.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200. <https://doi.org/10.1007/BF02289233>
- Martino, H. S. D., Cardoso, L. M., Moraes, E. A., Sant'ana, H. M. P., & Queiroz, V. A. V. (2014). *Por que utilizar o sorgo na alimentação humana?* Brasília, DF: Embrapa Milho e Sorgo. Retrieved from <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/993054/por-que-utilizar-o-sorgo-na-alimentacao-humana>
- Martins, L. S., Menezes, C. B., Simon, G. A., Silva, A. G., Tardin, F. D., & Gonçalves, F. H. (2016). Adaptability and stability of grain sorghum hybrids in Southwest of Goiás. *Revista Agrarian*, 9, 334–347. Retrieved from <http://ojs.ufgd.edu.br/.../3719>
- Murakami, D. M., & Cruz, C. D. (2004). Proposal of methodologies for environment stratification and analysis of genotype adaptability. *Cropps Breeding and Applied Biotechnology*, 4(1), 7–11. <https://doi.org/10.12702/1984-7033.v04n01a02>

- Oliveira, I. C. M., Guilhen, J. H. S., Ribeiro, P. C. de O., Gezan, S. A., Schaffert, R. E., Simeone, M. L. F., ... Pastina, M. M. (2020). Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. *Field Crops Research*, 257(June), 1–10. <https://doi.org/10.1016/j.fcr.2020.107929>
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Ribas, P. M. (2014). Origem e importância econômica. In R. A. C. Borém, A., Pimentel, L. D., Parrella (Ed.), *Sorgo: do plantio à colheita*. Viçosa, MG: Editora UFV.
- Rocha, J. R. A. S. C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy*, 10(1), 52–60. <https://doi.org/10.1111/gcbb.12443>
- Rooney, W. (2007). Sorghum breeding. In G. Acquaah (Ed.), *Principles of plant genetics and breeding*. Blackwell Publishing.
- Silva, K. J., Menezes, C. B., Tardin, F. D., Silva, A. R., Cardoso, M. J., Bastos, E. A., & Godinho, V. P. C. (2016). Selection of grain sorghum for grain productivity, adaptability and stability. *Revista Brasileira de Milho e Sorgo*, 15(1676–689X), 335–345. <https://doi.org/10.18512/1980-6477/rbms.v15n2p335-345>
- Silva, K. J., Menezes, C. B., Tardin, F. D., Emygdio, B. M., Souza, V. F., Carvalho, G. A., & Silva, M. J. (2013). Selection of sorghum hybrids cultivated in summer in three locations. *Revista Brasileira de Milho e Sorgo*, 12(1), 44–53.

Received: April 29, 2021.

Accepted: May 21, 2021.

Published: July 9, 2021.

English by: Eduardo Colombo, Lavras, MG, Brazil.