Genetic evaluation of grain sorghum hybrids in Brazilian environments using the REML/BLUP procedure

Janeo Eustáquio de Almeida Filho[1]*, Flávio Dessaune Tardin[2], Marcos Deon Vilela de Resende[1,3], Fabyano Fonseca e Silva[1], Ítalo Stefanine Correia Granato[1], Cicero Beserra de Menezes[2]

[1]Federal University of Viçosa – Dept. of Statistics, Av. P.H. Rolphs, s/n – 36570-000 – Viçosa, MG – Brazil.

[2]Embrapa Maize & Sorghum – Nucleus of Genetic Resources and Development of Cultivars, Rod. MG 424, km 45 – Sete Lagoas, MG – Brazil.

[3]Embrapa Forestry, Estrada da Ribeira, km 111 – Colombo, PR – Brazil.

*Corresponding author <janeo.filho@ufv.br>

ABSTRACT: When it comes to recommending sorghum (*Sorghum bicolor*) cultivars, it is essential to carry out a genetic evaluation of the agronomic traits of promising genotypes from several common environments where the crop is cultivated. This study consisted of a genetic evaluation of 52 experimental grain sorghum hybrids and eight commercial cultivars. Hybrids were evaluated in 19 experiments representing the most varied cultivation conditions in Brazil. Traits of agronomic interest such as grain yield, flowering and plant height were analysed. Genotypic evaluation was performed following the REML/BLUP (Restricted Maximum Likelihood/Best Linear Unbiased Predictor) procedure; the MHPRVG (Harmonic Mean of Relative Performance of Genotypic Values) method was also employed to study stability and adaptability. Hybrids which stood out in terms of highest grain yield based on genotypic values, stability and adaptability were 0306037, 1G150, DKB 599, 0306039, 1G282 and 0307671. Of these, only 1G282 showed restrictions as to plant height. For flowering, experimental hybrids showed shorter cycles than commercial cultivars, confirming the efficiency of genetic improvement for this trait. With the analysis of grain yield considering days to flowering and plant height as covariates, it was observed that most of the hybrids of greater performance, showed grain yield to be positively influenced by plant height and days to flowering.

Introduction

It is advisable that national competition assays between grain cultivars of sorghum (*Sorghum bicolor*) are carried out under the most diverse cultivation conditions to make it possible to select cultivars with genetic merit that are free of environmental effects, taking into account the interaction between genotypes and environments through their stability and adaptability, as well as capitalizing on adaptations specific to certain locations.

To select the best genotypes, the Brazilian Ministry of Agriculture, Livestock and Supply (MAPA) recommends the use of trials of value for cultivation and use (VCU). These trials, however, usually take the phenotypic mean into account, when the ideal would be to utilize genotypic means free of environmental effects.

The REML/BLUP (Restricted Maximum Likelihood/Best Linear Unbiased Predictor) procedure (mixed models method) is a very good tool for the genetic evaluation of individuals in a breeding program (Resende and Duarte 2007; Piepho et al., 2008), because it handles imbalanced data in an appropriate manner, making proper use of all the information contained in the experiments and giving precise estimates of genetic values. The use of the REML/BLUP procedure has been applied successfully to many crops (Resende and Thompson 2004; Viana et al., 2011; Barbosa et al., 2012; Ferreira et al., 2012; Mendes et al., 2012; Oliveira et al., 2012; Santos et al., 2012). For sorghum this methodology was used by Sabadin et al. (2012).

In this study, we evaluated the genetic value of grain sorghum hybrids focusing on the following traits: grain yield, flowering and plant height and compared their stability and adaptability in different cultivation environments.

Materials and Methods

Data from the evaluation of 19 experiments (Table 1) of single-cross grain-sorghum hybrids planted in several regions over three years of cultivation were evaluated. In all experiments, hybrids were arranged in a triple lattice design; some experiments utilized a 7 × 7 lattice and others a 5 × 5 lattice. Among the hybrids, 52 are

Table 1 – Locations and specific environmental conditions where the experiments were conducted.

Year	Locations
2008	Irrigated Janaúba* -15°48'S; 42°47' W[1/] (JAI)[2/] and Janaúba under water stress (JAS)
2009	Acreúna -17°23'S; 50°22' W (ACR), Itumbiara -18°24'S; 49°12' W (ITU), Patos de Minas -18°34'S; 46°31' W (PAT), Sete Lagoas -19°28'S; 44°14' W (SET), Sete Lagoas V20** (S20), Sete Lagoas V40** (S40), Teresina -5°5'S; 42°47' W (TER) and Vilhena -12°44'S; 60°7' W (VIL)
2011	Água Comprida -19°52'S; 48°25' W (AGU), Campos dos Goytacazes -21°45'S; 41°19' W (CAM), Guaíra -20°19'S; 48°18' W (GUA), Itumbiara, Montividiu -17°26'S; 51°10' W (MON), Rio Verde -17°47'S; 50°22' W (RIO), Sete Lagoas, Teresina and Vilhena.

*Irrigated Janaúba and Janaúba under water stress are experiments with plain irrigation during the entire cultivation cycle and induced water stress at 45 days after seeding, respectively; **Sete Lagoas V20 and V40 stand for cultivation in areas with aluminum saturation of 20 and 40 %, respectively. [1/]Geographic coordinates. [2/]Acronyms in parentheses represent the respective locations throughout the article.

experimental, originating from the breeding program of Embrapa (Brazilian Agricultural Research Corporation) Maize and Sorghum along with commercial cultivars.

The traits assessed were: i) grain yield with moisture corrected to 13 %; ii) plant height in the physiological maturation stage; and iii) days to flowering, assessed on the basis of number of days from seeding to the time when 50 % of the plants in the plots had the flowers of the middle part of the panicle at anthesis.

For statistical analysis, we adopted the mixed linear model: $y = Xl + Zg + Wgl + Qb + \varepsilon$, where: y = vector of phenotypic observations; l = vector with fixed-effect parameters corresponding to the location, added to the intercept; g, gl, b and ε = vectors with random-effect parameters corresponding to genotypes; interaction between genotypes and locations; block within repetition and location; and random error, respectively; X, Z, W and Q = incidence matrices of vectors l, g, gl and b, respectively.

For the analysis, the following was assumed:

$$\begin{bmatrix} y \\ g \\ gl \\ b \\ \varepsilon \end{bmatrix} \sim MVN \left(\begin{bmatrix} Xl \\ \phi \\ \phi \\ \phi \\ \phi \end{bmatrix}, \begin{bmatrix} V & Z\sigma_g^2 & W\sigma_{gl}^2 & Q\sigma_b^2 & I\sigma_\varepsilon^2 \\ Z'\sigma_g^2 & I\sigma_g^2 & \phi & \phi & \phi \\ W'\sigma_{gl}^2 & \phi & I\sigma_{gl}^2 & \phi & \phi \\ Q'\sigma_b^2 & \phi & \phi & I\sigma_b^2 & \phi \\ I\sigma_\varepsilon^2 & \phi & \phi & \phi & I\sigma_\varepsilon^2 \end{bmatrix} \right)$$

where:

$$V = ZZ'\sigma_g^2 + WW'\sigma_{gl}^2 + QQ'\sigma_b^2 + I\sigma_\varepsilon^2;$$

$$\phi \equiv 0$$

For the estimation of the BLUEs (best linear unbiased estimators) of the estimable functions of fixed effects and the BLUPs of random effect parameters, we adopted the equations of mixed models proposed by Henderson (1975), which, according to Resende (2007), are:

$$\begin{bmatrix} X'X & X'Z & X'W & X'Q \\ Z'X & Z'Z + I\sigma_\varepsilon^2/\sigma_g^2 & Z'W & Z'Q \\ W'X & W'Z & W'W + I\sigma_\varepsilon^2/\sigma_{gl}^2 & W'Q \\ Q'X & Q'Z & Q'W & Q'Q + I\sigma_\varepsilon^2/\sigma_b^2 \end{bmatrix} \begin{bmatrix} \hat{l} \\ \hat{g} \\ \widehat{gl} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ Q'y \end{bmatrix}$$

For the solution of the equations, the variance components were estimated by the Restricted Maximum Likelihood method (REML). Using the algorithm Expectation–Maximization (EM), the variance component estimators are:

$$\hat{\sigma}_\varepsilon^2 = (y'y - \hat{l}'X'y - \hat{g}'Z'y - \widehat{gl}'W'y - \hat{b}'Q'y)/[N - r(X)]$$

$$\hat{\sigma}_g^2 = [\hat{g}'\hat{g} + \hat{\sigma}_\varepsilon^2 tr(C^{22})]/q$$

$$\hat{\sigma}_{gl}^2 = [\widehat{gl}'\widehat{gl} + \hat{\sigma}_\varepsilon^2 tr(C^{33})]/p \qquad (4)$$

$$\hat{\sigma}_b^2 = [\hat{b}'\hat{b} + \hat{\sigma}_\varepsilon^2 tr(C^{44})]/s$$

where: N = number of observations; $r(X)$ = rank of X matrix; $tr(.)$ = function that returns the trace of a matrix; q, p and s are length of the vectors g, gl and b respectively; and C^{22}, C^{33}, C^{44} are obtained from:

$$C^{-1} = \begin{bmatrix} C^{11} & C^{21} & C^{31} & C^{41} \\ C^{21} & C^{22} & C^{32} & C^{42} \\ C^{31} & C^{32} & C^{33} & C^{43} \\ C^{41} & C^{42} & C^{34} & C^{44} \end{bmatrix}$$

C = matrix of the coefficients of the mixed model equations.

For evaluation of grain yield, two models were fitted, the first as previously reported, and the second similar to the first, but with the inclusion of days to flowering and plant height as covariates, in a similar way to that used by Sabadin et al. (2012).

For the variable grain yield, a stability and adaptability analysis using the Harmonic Mean of Relative Performance of Genotypic Values (MHPRVG) method was carried out as proposed by Resende (2007), obtained as follows:

$$MHPRVG_i = n_i / \sum_{j=1}^n \frac{\overline{GV}_{.j}}{GV_{ij}}$$

where: n_i = number of locations where genotype i was evaluated; $GV_{ij} = u_j + g_i + gl_{ij}$ = genotypic value of genotype i at a specific location j, in which u_j represents the average of location j, and g_i and gl_{ij} are the BLUPs of genotype i and the interaction between genotype i and location j, respectively; and $\overline{GV}_{.j}$ = average GV_{ij} in location j.

This method penalizes individuals that do not present a satisfactory performance in all locations, in relation to the average of the assessed environment, similar to the method of Lin and Bins (1988). Because the MHPRVG method considers additional genotypic values such as proportion of the average of each location, it is a measure of genotypic stability and adaptability.

All analyses were performed using Selegen-REML/BLUP software (Resende, 2007).

Results and Discussion

Based on the heritability values at the mean levels and on the level of accuracy, the experimental results and the fitted model for predictions of genetic values are appropriate for efficiently selecting hybrids based on the traits evaluated (Table 2). Resende and Duarte (2007) recommend accuracy values above 0.7 for intermediate stages of the breeding program and above 0.9 at the stage when cultivars are to be recommended.

Sabadin et al. (2012) fitted a multi-environment model, considering the results in each environment as a separate variable and the covariance matrix of the genotypic effects as unstructured. In our study, fitting a multivariate multi-environment model is prohibitive be-

cause of the high number of environments, which turns out to be computationally infeasible leading to numerical convergence problems for REML estimates of variance components.

The heritability of the variable grain yield was inferior to the other variables; this result can be explained by the polygenic nature of this trait, which is very much influenced by environmental variations. The genetic correlations between traits indicate a tendency of the more productive hybrids to have short cycles and also be greater in height (Table 2). Because grain sorghum is an off-season crop, its earliness is expected to foster an increase in productivity, since the off-season is a period of adverse precipitation conditions, so that hybrids with a shorter cycle are less susceptible to dry spells.

The positive association of plant height with grain yield is linked to vegetative development; the greater the vegetative growth, the larger the reserve of photo-assimilates to be used for the formation of grains. Although the number of days to flowering has a negative correlation with plant height, the ideal is to have early individuals with fast vegetative growth. Of the ten most productive hybrids, the majority presented genotypic values below the average for flowering and above the average for plant height (Table 3).

Given the correlations between the genotypic values over the locations, we can observe that the effects of the genotype × location interaction were more pronounced for grain yield than for the other traits, which indicates that the behavior of the hybrids was less predictable between locations in the case of this variable than for the others. The correlation between genotypic values over the locations for height was higher and indicates that this variable is less susceptible to environmental variations. These results are linked to the magnitude of the heritabilities, thereby revealing that traits of lower heritability interact more with their environments. Santos et al. (2012) observed a correlation between genotypic values over locations of 0.66 for grain yield in rice (*Oryza sativa*). For sugar cane (*Saccharum spp*), Oliveira et al. (2005) observed 0.62, for this correlation in terms of tons of stalk per hectare. Mendes et al. (2012) observed correlations of 0.6 and 0.25 in a group of favorable and unfavorable environments for grain yield of corn (*Zea mays*).

With regard to height, all hybrids showed genotypic values free of interaction $(u + g_i)$, and genotypic values considering the mean interaction $(u + g_i + glm_j)$ within the limits recommended for this crop. The height of a grain sorghum hybrid should be greater than 100 cm to favor mechanical harvest, and lower than 150 cm to avoid lodging. Observing the genotypic values for specific environments $(u_j + g_i + gl_{ij})$ we can see that the height of most hybrids was within the recommended range (Figure 1). Only Janaúba had an elevated number of hybrids shorter than the established limit; water deficiency was the main reason for such results.

For flowering, a large variation was observed in the results over the locations. In Brazil the time for sorghum to flower usually varies between 55 and 70 days. Most hybrids present genotypic values for flowering within this range. Only in the experiment conducted in Janaúba, with full irrigation, was a high number of hybrids present that exceeded 70 days; these results are
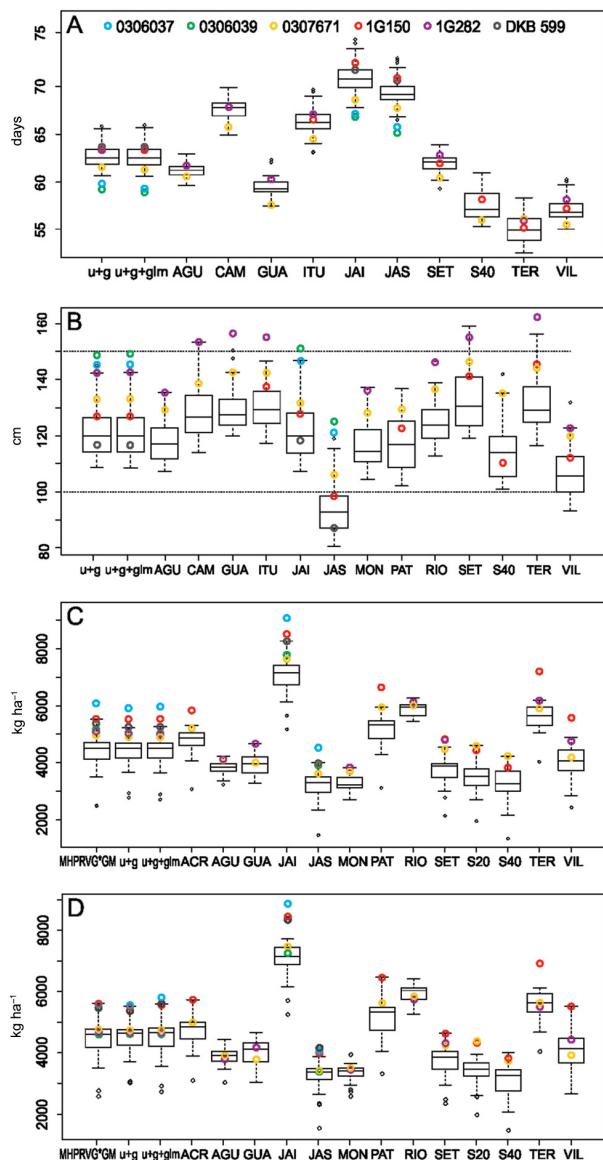
Table 2 – Estimates of genetic parameters of variables number of days between seeding and flowering (FL); plant height (HE); and grain yield (GY).

Parameters	FL	HE	GY
Heritability	0.81	0.88	0.76
Accuracy	0.90	0.94	0.87
r_{gl}[1]	0.74	0.84	0.69
r_{FL}[2]	1.00	-0.49	-0.35
r_{HE}	-0.49	1.00	0.30
r_{GY}	-0.35	0.30	1.00

[1]Correlation between genotypic values over the locations; [2]Genetic correlation of the variable in the column with the variable with prefix r.

Table 3 – Genotypic values of the 10 individuals with highest grain yield.

Genotype	Grain yield without covariates			Grain yield with covariates			Flowering			Height		
	GV	$CI_{95\%}$	Ac	GV	$CI_{95\%}$	Ac	GV	$CI_{95\%}$	Ac	GV	$CI_{95\%}$	Ac
0306037	5894.64	± 714.17	0.78	5557.89	± 697.76	0.77	59.42	± 1.61	0.83	145.16	± 8.69	0.89
1G150	5506.98	± 427.48	0.93	5507.89	± 417.09	0.92	63.16	± 1.05	0.93	126.93	± 5.21	0.96
DKB 599	5237.55	± 714.27	0.78	5383.79	± 691.88	0.77	65.57	± 1.61	0.83	116.73	± 8.69	0.89
0306039	5013.59	± 714.16	0.78	4627.19	± 699.33	0.77	59.04	± 1.61	0.83	148.78	± 8.69	0.89
1G282	4982.25	± 473.36	0.91	4649.89	± 468.64	0.9	61.72	± 1.61	0.83	142.3	± 5.07	0.97
0307671	4888.96	± 358.73	0.95	4741.79	± 352.80	0.95	61.33	± 0.86	0.96	132.98	± 4.28	0.98
0307167	4810.11	± 427.64	0.93	4491.11	± 424.73	0.92	60.66	± 1.03	0.93	145.01	± 5.22	0.96
0303001	4809.76	± 713.95	0.78	4731.09	± 692.08	0.77	61.38	± 1.61	0.83	130.43	± 8.68	0.89
0307423	4774.59	± 714.04	0.78	5005.59	± 692.27	0.77	61.09	± 1.61	0.83	113.00	± 8.68	0.89
0307107	4772.59	± 714.12	0.78	4951.89	± 691.68	0.77	62.58	± 1.61	0.83	115.74	± 8.69	0.89
Mean	4413.87			4413.87			62.6			121.43		
Maximum	5894.64			5557.89			65.86			148.78		
Minimum	2764.3			3014.89			59.04			108.64		

GV: genotypic value free of interaction; $CI_{95\%}$: asymptotic confidence interval (p < 0.05), according to Resende (2007); Ac: individual accuracy.

Figure 1 – Boxplot of the genotypic value free of the interaction ($u+g_i$), of the genotypic value considering the mean interaction ($u+g_i+glm$), of genotypic values considering the specific interaction with each location, for the trait days to flowering (A), plant height in cm (B), grain yield in kg ha^{-1} without covariables (C); grain yield in kg ha^{-1} using days to flowering and height as covariables (D). Bloxplot of the MHPRVG*GM values for the trait grain yield. Genotypes with highest grain yield had their genotypic values highlighted. The codification of the location is in accordance with Table 1.

attributable to the favorable water conditions. In the case of the lower limit of 55 days, we observed that only Teresina had a significant number of hybrids flowering before this age. These results can be explained by the elevated temperatures in this location.

Sorghum cultivars vary in their different requirements for flowering, because in some regions where the

interest lies in planting sorghum early, the producer can choose a cultivar with a longer cycle, since this cultivar is expected to be exceptionally productive as a result of this longer vegetative cycle. Moreover, the search for early cultivars is very appealing, because in addition to preventing the crop from suffering extended exposure in a condition susceptible to water stress, the producer can have more flexibility in the determination of their yield.

The results of genotypic values free of genotype-location interaction collate well with the MHPRVG method; the Spearman correlation between these statistics was 0.997, which shows that the more productive materials are also those of greatest stability and adaptability. Similar results were observed by Mendes et al. (2012) and Oliveira el al. (2005). Based on the results of MHPRVG, if hybrids 0306037, 1G150, DKB 599, 0306039, 1G282 and 0307671 are planted in one of the locations participating in the experiments, performance is expected to be at least 37, 25, 21, 16, 15 and 12 % above the average for each location, respectively. The MHPRVG method provides a measure of relative performance. When this result is multiplied by the overall mean (GM), we have a genotypic statistic value on the scale of the trait, as can be seen in Figure 1.

Overall, the most productive hybrids also stood out for their specific results for each environment (Figure 1). This is a result of the low genotype × environment interaction magnitude verified in the present study (correlation between genotypic values over locations equal to 0.69 for yield). When analyzing the genotypic value for the specific environments ($u_j + g_i + gl_{ij}$), cultivars 1G150 and 1G282 stood out: their performance was among the five most productive considering the locations where these hybrids were evaluated.

Although hybrids 0306037, DKB 599 and 030639 have shown high genotypic values for grain yield (Table 3), they need to be evaluated and proved in other sites because they were only assessed in the experiments of Janaúba. Of the hybrids that stood out most for grain yield, height was a limiting factor for cultivar 1G282, because this hybrid showed a genotypic value considering the specific interaction above 150 cm in several locations (Figure 1), indicating that the producer should be cautious when planting this hybrid in areas of strong winds.

Cultivar DKB 599 had a genotypic value for the specific location of Janaúba with water stress of 87 cm, which is well below the recommended lower limit. However, this hybrid reached this height in this location only.. Hybrids 0306039 and 0306037 were, in order, the earliest two, considering both their genotypic value free of interaction and the specific interactions of the locations where these hybrids were evaluated. Hybrid 0307671 also stood out as being among the earliest. In fact, it was one of the five earliest of the locations in which it was evaluated. Regarding the genotypic value free of interaction, this hybrid was also observed to be early. The genotypic values of cultivars 1G282, 1G150 and DKB 599 were from medium to late, considering

the specific interactions and the genotypic value free of interaction (Table 3). The divergent numbers of days to flowering between productive hybrids favors the establishment of strategies for the selection of planting time.

Due to associations between variables, a model was fitted for yield considering plant height and days to flowering as covariates. Consequently, it was possible to predict the genotypic value for yield free from effects of the divergence between plant heights and days to flowering. The predicted genotypic values of the hybrids DKB 599, 0307423 and 0307107 in the model that uses the covariates were higher than in the model without covariates, indicating that the plant height and flowering have negatively influenced the productivity of these hybrids. However, the opposite effect was observed for hybrids 0306037, 0306039, 1G282, 0307671, 0307167 and 0303001. In most places it was also observed that plant height and days to flowering have positively influenced the grain yield of cultivar 1G150 (Figure 1, Table 2).

As most of these hybrids are in the final stages of the breeding program, it is not interesting to carry out the recommendation by genotypic value of grain yield free from the effects of other important variables such as plant height and days to flowering. However, the genotypic value of grain yield free from the effects of these variables may indicate the potential for individuals to be used as parents in subsequent stages of the breeding program, as the use of these covariates helps to discriminate the effects of genes that are directly related to yield (Sabadin et al., 2012).

Acknowledgements

References

Barbosa, M.H.P.; Resende, M.D.V.; Dias, L.A.S.; Barbosa, G.V.S.; Oliveira, R.A.; Peternelli, L.A.; Daros, E. 2012. Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. Crop Breeding and Applied Biotechnology 12: 87-98.

Ferreira, R.T.; Viana, A.P.; Barroso, D.G.; Resende, M.D.V.; Amaral Júnior, A.T. 2012. Toona ciliata genotype selection with the use of individual BLUP with repeated measures. Scientia Agricola 69: 210-216.

Lin, C.S.; Binns, M.R.A. 1988. Superiority measure of cultivar performance for cultivar x location data. Canadian Journal of Plant Science 68: 193-198.

Mendes, F.F.; Guimarães, L.J.M.; Souza, J.C.; Guimarães, P.E.O.; Pacheco, C.A.P.; Machado, J.R.A.; Meirelles, W.F.; da Silva, A.R.; Parentoni, S.N. 2012. Adaptability and stability of maize varieties using mixed model methodology. Crop Breeding and Applied Biotechnology 12: 111-117.

Oliveira, A.R.; Resende, M.D.V.; Daros, E.; Bespalhok Filho, J.C.; Zambon, J.L.C.; Ido, O.T.; Weber, H.; Koehler, H.S. 2005. Genotypic evaluation and selection of sugarcane clones in three environments in the state of Paraná. Crop Breeding and Applied Biotechnology 5: 426-434.

Oliveira, E.J.; Fraife Filho, G.A.; Freitas, J.P.X.; Dantas, J.L.L.; Resende, M.D.V. 2012. Plant selection in F2 segregating populations of papaya from commercial hybrids. Crop Breeding and Applied Biotechnology 12: 191-198.

Piepho, H.P.; Möhring, J.; Melchinger, A.E.; Büchse, A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161: 209-228

Resende, M.D.V.; Duarte, J.B. 2007. Precision and quality control in variety trials. Pesquisa Agropecuária Tropical 37: 182-194 (in Portuguese, with abstract in English).

Resende, M.D.V. 2007. Selegen-Reml/Blup: statistical system and computerized genetic selection through linear mixed models = Selegen-Reml/Blup: Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos. Embrapa Florestas, Colombo, PR, Brazil (in Portuguese).

Resende, M.D.V.; Thompson, R. 2004. Factor analytic multiplicative mixed models in the analysis of multiple experiments. Revista de Matemática e Estatística 22: 1-22.

Santos, V.B.; Soares, A.A.; Reis, M.S.; Resende, M.D.V.; Cornelio, V.M.O. 2012. Value for cultivation and use of upland rice cultivars tested in multi-environments and analyzed by Reml/ Blup methodology. Crop Breeding and Applied Biotechnology 12: 25-33.

Sabadin, P.K.; Tardin, F.D.; Eeuwijk, F.A.; Malosetti, M.; Albuquerque, P.E.P.; Garcia, A.A.F.; Caniato, F.F.; Schaffert, R.E.; Gomide, R.L.; Guimarães, C.T.; Andrade, C.L.T.; Mollinari, M.; Oliveira, B.F.; Santos, F.G.; Margarido, G.R.A.; Boer, M.P.; Magalhães, J.V. 2012. Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. Theoretical and Applied Genetics 124: 1389-1402.

Viana, J.M.S.; Faria V.R.; Silva, F.F.; Resende, M.D.V. 2011. Best linear unbiased prediction and family selection in crop species. Crop Science 51: 2371-2381.