

## Minimum number of measurements for evaluating *Bertholletia excelsa*

A.B. Baldoni<sup>1</sup>, H. Tonini<sup>1</sup>, F.D. Tardin<sup>1</sup>, S.C.C. Botelho<sup>1</sup> and P.E. Teodoro<sup>2</sup>

<sup>1</sup>Embrapa Agrossilvipastoril, Sinop, MT, Brasil

<sup>2</sup>Laboratório de Biometria, Departamento de Biologia Geral,  
Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: F.D. Tardin

E-mail: flavio.tardin@embrapa.br

Genet. Mol. Res. 16 (3): gmr16039783

Received July 24, 2017

Accepted August 28, 2017

Published September 27, 2017

DOI <http://dx.doi.org/10.4238/gmr16039783>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Repeatability studies on fruit species are of great importance to identify the minimum number of measurements necessary to accurately select superior genotypes. This study aimed to identify the most efficient method to estimate the repeatability coefficient ( $r$ ) and predict the minimum number of measurements needed for a more accurate evaluation of Brazil nut tree (*Bertholletia excelsa*) genotypes based on fruit yield. For this, we assessed the number of fruits and dry mass of seeds of 75 Brazil nut genotypes, from native forest, located in the municipality of Itaúba, MT, for 5 years. To better estimate  $r$ , four procedures were used: analysis of variance (ANOVA), principal component analysis based on the correlation matrix (CPCOR), principal component analysis based on the phenotypic variance and covariance matrix (CPCOV), and structural analysis based on the correlation matrix (mean  $r$  - AECOR). There was a significant effect of genotypes and measurements, which reveals the need to study the minimum number of measurements for selecting superior Brazil nut genotypes for a production increase. Estimates of  $r$  by ANOVA were lower than those observed with the principal component methodology and close to

AECOR. The CPCOV methodology provided the highest estimate of  $r$ , which resulted in a lower number of measurements needed to identify superior Brazil nut genotypes for the number of fruits and dry mass of seeds. Based on this methodology, three measurements are necessary to predict the true value of the Brazil nut genotypes with a minimum accuracy of 85%.

**Key words:** Brazil nut; Repeatability; Number of fruits; Dry mass of seeds

## INTRODUCTION

Amazon region is known for the richness of plant species, which have a high potential for use in agriculture and breeding. Among the species, it is worth mentioning the Brazil nut tree (*Bertholletia excelsa* Bonpl.), belonging to the Lecythidaceae family, which occurs in Brazil in the States of Acre, Amazonas, Pará, Roraima, Rondônia, Mato Grosso, Maranhão, and Tocantins (Mori and Prance, 1990; Lorenzi, 2002). Brazil nut is one of the most important species of extractive exploration, being considered one of the main non-timber forest products of the Amazon Forest, of great economic importance, mainly for the North and Mid-West Regions of Brazil (Wadt et al., 2008).

The production of Brazil nut fruits oscillates annually, both at the population level and at the individual level. The variation in fruit and seed production in native forest trees can be attributed to exogenous and endogenous factors (diameter at breast height, crown shape, the presence of lianas, diseases, herbivory, among others) being the climate (precipitation and temperature) one of the main factors responsible for this variation (Kainer et al., 2007).

Tonini and Pedrozo (2014), studying the annual variations in fruit and seed production in native Brazil nut populations in Roraima, observed that the production varied between populations and individuals and most of the plants did not produce fruits every year. In the year of greatest production, a planting of Brazil nut tree produced 52 times more than in the low years, and total production was concentrated in a small number of individuals. Already Kainer et al. (2007) observed in the State of Acre that when evaluating the annual variation of the Brazil nut fruit production, at the individual level, it was relatively high, and at the general level, it was deficient, showing that in the 5 years of evaluation, in the long term, Brazil nut populations presented relatively constant fruit yield.

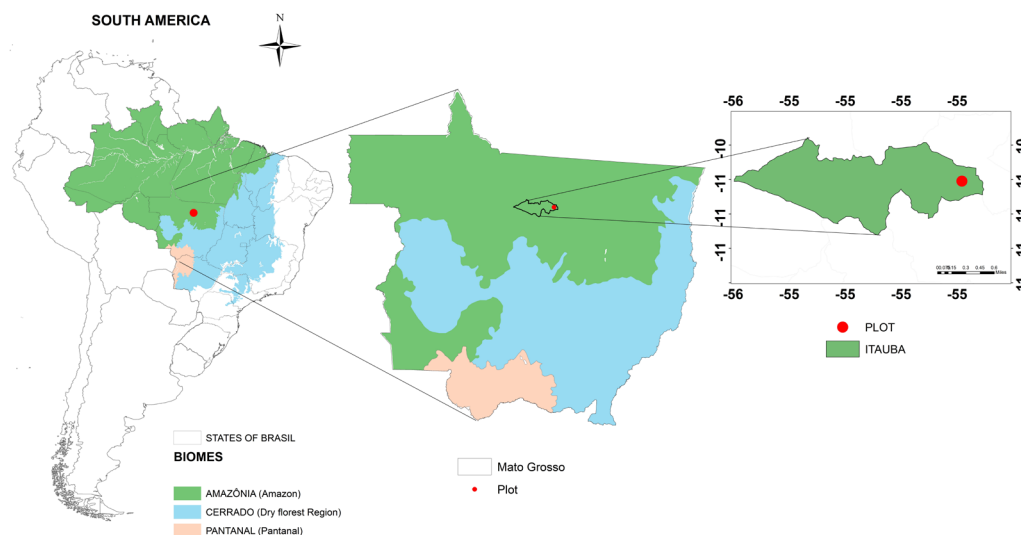
In research on perennial species such as Brazil nut, whether in breeding or conservation programs, several assessments are required over its cycle to identify superior individuals regarding the trait of interest. Thus, the dimensioning of the number of measurements has been performed using repeatability coefficients that are fundamental for improving perennial species, since they provide an approximation of the maximum value that the trait heritability in the broad sense needs. Besides, this information is also important to determine the least possible selection cycle, which has a direct impact on the human and financial research resources (Cruz et al., 2012).

The dimensioning of the number of measurements has been conducted for several perennial crops such as acerola (Lopes et al., 2001), guava (Degenhardt et al., 2002), araçazeiro and pitangueira (Danner et al., 2010), peach (Bruna et al., 2012), sweet orange (Negreiros et al., 2014), banana (Lessa et al., 2014), and Jatropha (Teodoro et al., 2016). The

repeatability coefficient in the Brazil nut was estimated by Pedrozo et al. (2015) and Assis (2016) from the methodology of mixed models. However, there is still little research that evaluates the number of measurements needed to select superior genotypes of the Brazil nut, using multivariate techniques. Therefore, we performed this study to estimate the minimum number of measurements needed for more accurate evaluation of Brazil nut genotypes based on the number of fruits and dry mass of seeds, in addition to identifying the most efficient method to estimate the repeatability coefficients.

## MATERIAL AND METHODS

Data were collected in a permanent parcel of nine hectares of native forest, located in the protected reserve area of Fazenda Santo Ângelo (geographical coordinates 11°05'54.0" 55°02'11.0"W, with an altitude of approximately 380 m), in the municipality of Itaúba, Mato Grosso (Figure 1). The forest is Ombrophilous Dense type, with flat to undulating relief. The climate is Am tropical humid or sub-humid type, with average annual rainfall around 2000 mm and dry period well marked, ranging from 2 to 3 months (Alvares et al., 2013).



**Figure 1.** Location of the study area, in the municipality of Itaúba, Mato Grosso, Brazil.

Fifty-five Brazil nut genotypes were evaluated in five harvests from 2013 to 2017. Number of fruits and seed dry mass was measured, being counted the fruits fallen on the ground per tree, under the crown projection, throughout the fruit fall period and, from these fruits, we measured the dry mass of seeds per tree, after subjecting them to a forced circulation oven at 65°C, until constant mass. To better estimate  $r$ , four procedures were used: analysis of variance (ANOVA), principal component analysis based on the correlation matrix (CPCOR), principal component analysis based on the phenotypic variance and covariance matrix (CPCOV), and structural analysis based on the correlation matrix (mean  $r$  - AECOR).

Data were analyzed using ANOVA, where the statistical model adopted considered two factors of variation according to Equation 1:

$$Y_{ij} = \mu + g_i + a_j + \varepsilon_{ij} \quad (\text{Equation 1})$$

where  $Y_{ij}$  is the observation for the  $i$ -th genotype for the  $j$ -th measurement;  $\mu$  is the overall mean;  $g_i$  is the fixed effect of the  $i$ -th genotype in the permanent environment;  $a_j$  is the fixed effect of the temporary environment for the  $j$ -th measurement; and  $\varepsilon_{ij}$  is the experimental error established by the temporary effects of the environment at the  $j$ -th measurement of the  $i$ -th genotype. The genotypes were clustered using the Scott-Knott test at a 5% probability according to their respective average fruit yield over the 16 years of evaluation, as well as based on the fruit yield obtained each year.

Based on ANOVA, estimates of the mean squared of the measurements, mean squared of genotypes ( $MS_G$ ), mean squared error ( $MS_E$ ), and the F-test value for the genotype were obtained. Then, the  $r$  was estimated following Equation 2:

$$r = \frac{MS_G - MS_E}{MS_G + (\eta - 1)MS_E} \quad (\text{Equation 2})$$

where  $\eta$  is the number of measurements performed.

The CPCOR method consists of obtaining a correlation matrix between the progenies for each pair of measurements. From this correlation matrix, the eigenvalues and their associated normalized eigenvectors are determined. The eigenvector, whose elements have the same sign and relative magnitudes, is the one that expresses the tendency of the genotypes to maintain, over the years, their relative positions concerning the others (Abeywardena, 1972). For this method,  $r$  is given by Equation 3:

$$r = \frac{\hat{\lambda}_1 - 1}{\eta - 1} \quad (\text{Equation 3})$$

where  $\hat{\lambda}_1$  is the eigenvalue of the covariance ( $\hat{\Gamma}$ ) or correlation matrix ( $\hat{R}$ ).

For the CPCOV method,  $r$  is given by Equation 4:

$$r = \frac{\hat{\lambda}_1 - \hat{\sigma}_Y^2}{\hat{\sigma}_Y^2 - (\eta - 1)} \quad (\text{Equation 4})$$

where  $\hat{\sigma}_Y^2$  is the phenotypic variance (Cruz et al., 2012). Obtaining  $r$  using the AECOR method presents only conceptual differences regarding the methods mentioned above that are based on principal components. Based on the AECOR method,  $r$  is given by Equation 5:

$$r = \frac{\hat{\alpha}' \hat{R} \hat{\alpha} - 1}{\eta - 1} \quad (\text{Equation 5})$$

in which  $\hat{\alpha}$  is the eigenvector associated with the highest  $\hat{R}$  eigenvalue (Mansour et al., 1981).

The minimum number of measurements required to predict the actual value of individuals (n) was determined based on a pre-established (80 and 85%) coefficient of determination ( $R^2$ ) using Equation 6:

$$n = \frac{R^2(1-r)}{(1-R^2)r} \quad (\text{Equation 6})$$

where r is obtained according to either of the different methods mentioned above (Cruz et al., 2012). The genotypic  $R^2$ , which represents the percentage of certainty for predicting the actual progenies value based on  $\eta$ , was obtained by Equation 7:

$$R^2 = \frac{\eta \times r}{1 + r(\eta - r)} \quad (\text{Equation 7})$$

For a better interpretation of the results, the number of measurements associated with different  $R^2$  for fruit yield in soursop obtained from each proposed methodology was represented graphically. The statistical data analyses were performed using Genes (Cruz, 2013) and Microsoft Excel®.

## RESULTS AND DISCUSSION

Table 1 expresses the summary of analysis of variance for the number of fruits and dry mass of seeds evaluated in 75 Brazil nut genotypes during five measurements. The significant effect of genotypes reveals the presence of genetic variability in the population and the possibility of selecting genotypes to increase fruit yield. The F-test also revealed a significant effect for measurements, which indicates that there has been a change in the mean of the genotypes over the years. These results are similar to those obtained for other perennial crops (Lopes et al., 2001; Degenhardt et al., 2002; Danner et al., 2010; Bruna et al., 2012; Negreiros et al., 2014; Lessa et al., 2014; Teodoro et al., 2016) and reinforce the need to study the minimum number of measurements to select superior Brazil nut genotypes for a productive increase.

**Table 1.** Summary of ANOVA for fruit yield in 75 Brazil nut genotypes over five measurements.

Source of variation	Degrees of freedom	Number of fruits	Dry mass of seed
Measurement	4	206,737.81*	2,361.37*
Genotype	74	17,943.39*	160.67*
Residue	296	4,314.76	41.49

\*Significant at 1% probability based on the F-test.

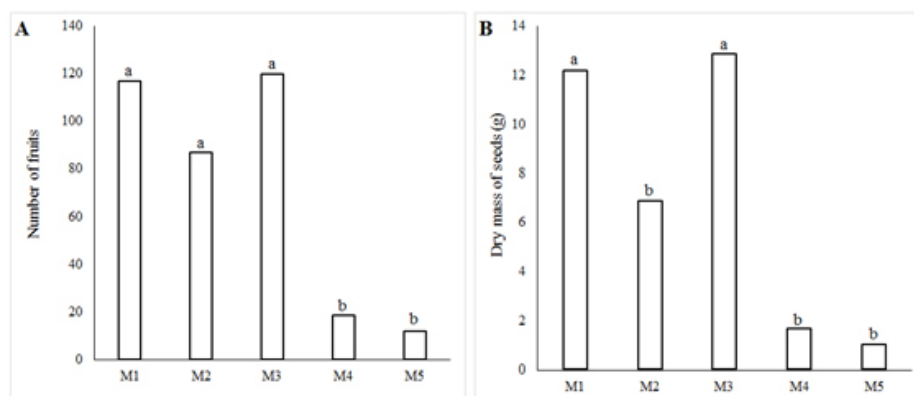
The Scott-Knott test clustered the 75 Brazil nut genotypes in three homogeneous groups for the number of fruits and two for the dry mass of seeds (Table 2). The genotypes 29, 33, and 60 constituted the group with the highest mean for some fruits over the measurements, besides belonging to the group with the highest dry mass of seeds. These results suggest that

these genotypes can be used as matrices for formatting base populations since they are sources of favorable alleles for the productive traits. Thus, these genotypes are promising for future genetic breeding programs, conservation and to be used in the species propagation, either in the production of seedlings or grafting.

**Table 2.** Clustering based on a Scott-Knott test applied to the number of fruits and dry mass of seeds in 75 Brazil nut genotypes over five measurements.

Group	Number of fruits		Dry mass of seeds (g)	
	Genotypes	Mean	Genotypes	Mean
I	29, 33, and 60	224.80	2, 9, 16, 18, 24, 27, 29, 31, 33, 40, 43, 46, 48, 49, 50, 56, 59, 60, 62, 65, 66, 68, 69, 71, 73, 74, and 75	13.57
II	2, 9, 16, 18, 24, 27, 31, 40, 43, 46, 48, 49, 50, 56, 59, 62, 65, 66, 68, 69, 71, 73, 74, and 75	129.70	1, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 19, 20, 21, 22, 23, 25, 26, 28, 30, 32, 34, 35, 36, 37, 38, 39, 41, 42, 44, 45, 47, 51, 52, 53, 54, 55, 57, 58, 61, 63, 64, 67, 70, and 72	3.20
III	1, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 19, 20, 21, 22, 23, 25, 26, 28, 30, 32, 34, 35, 36, 37, 38, 39, 41, 42, 44, 45, 47, 51, 52, 53, 54, 55, 57, 58, 61, 63, 64, 67, 70, and 72	31.70	-	-

It is possible to verify in Figure 2 that the number of fruits was higher in the first three measurements, while the dry mass of seeds was higher in the first and third measurements. These results indicate a cyclical behavior among the Brazil nut genotypes for these traits and suggest caution for selecting superior genotypes. However, when we analyze the repeatability coefficient ( $r$ ) and the coefficient of determination ( $R^2$ ) contained in Table 3, we verified that the estimates are close to or above the reliably established limits -  $r \geq 0.40$ ,  $R^2 \geq 0.75$  - proposed by Bergo et al. (2013) and Oliveira and Moura (2010), respectively. These results indicate that the significance of the factor measurements did not cause serious changes in the ranking of the best genotypes over the years, even though there was a cyclical behavior for both traits.

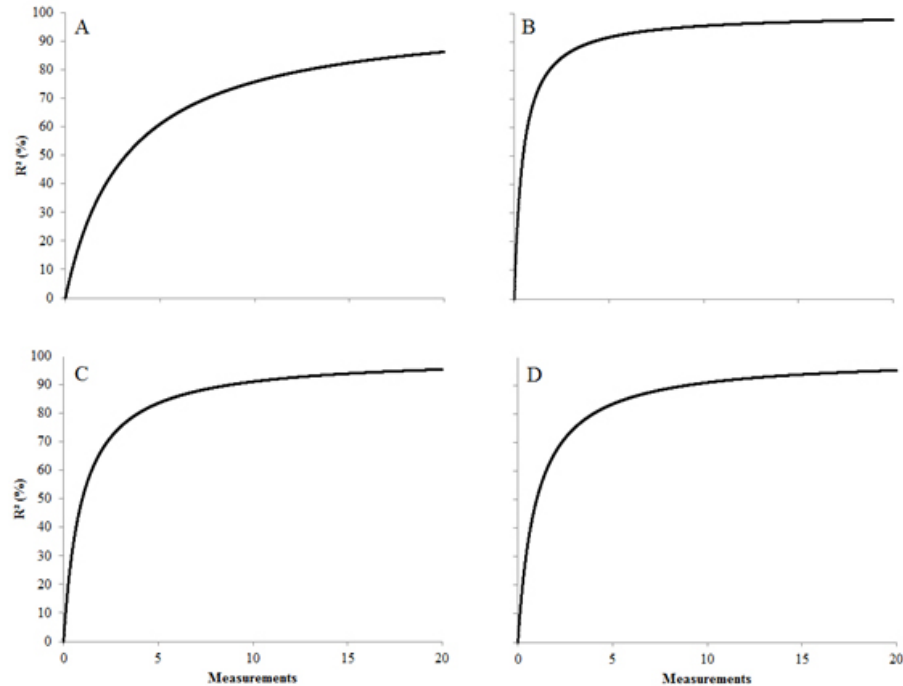


**Figure 2.** Means of fruit yield of 75 Brazil nut genotypes at each measurement (M) point. Bars marked with the same lower-case letter are not significantly different according to the Scott-Knott test at 5% probability.

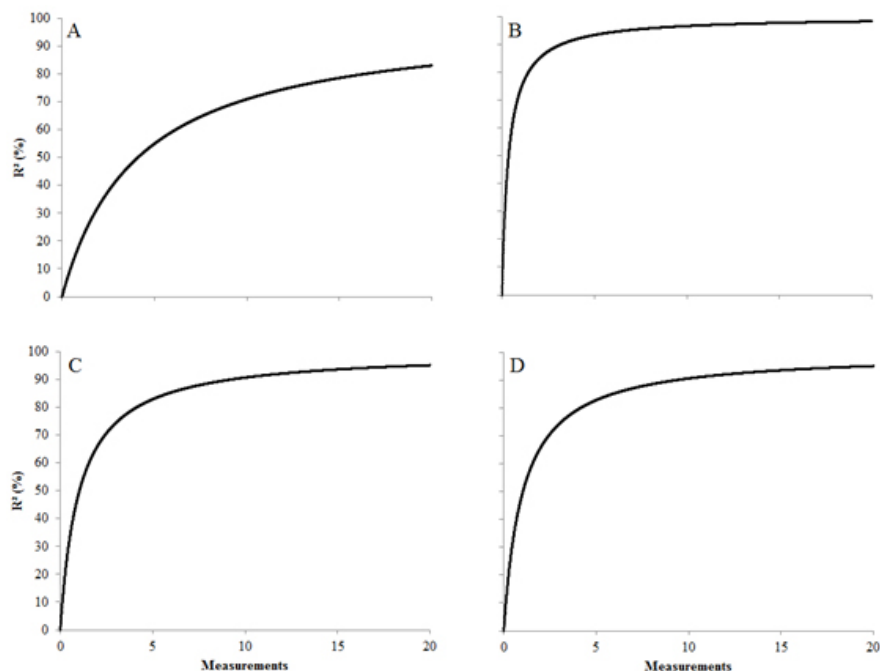
**Table 3.** Estimates of repeatability coefficient ( $r$ ) and coefficient of determination ( $R^2$ ) for the number of fruits and dry mass of seeds in 75 Brazil nut genotypes over five measurements.

Method	$r$	$R^2$	$r$	$R^2$
	Number of fruits		Dry mass of seeds	
Analysis of variance	0.38	0.76	0.35	0.74
Principal components - covariance (CPCOV)	0.69	0.92	0.74	0.93
Principal components - correlation (CPCOR)	0.50	0.83	0.49	0.82
Structural analysis - correlation	0.50	0.83	0.49	0.82

The estimates of  $r$  by ANOVA were lower than those observed with the principal component methodology and close to the structural analysis (Table 3). The principal component methodology based on the covariance matrix provided the highest estimate of  $r$ , which resulted in a lower number of measurements needed to identify superior Brazil nut genotypes for the number of fruits (Figure 3) and dry mass of seeds (Figure 4). Based on this methodology, three measurements are required to predict the true value of the Brazil nut genotypes with a minimum accuracy of 85% (Table S1). The other methods used suggest the use of at least four measurements for predictability of 80%. As the coefficient of determination ( $R^2$ ) increases, a principal component methodology based on the covariance matrix becomes even more efficient, indicating the need for a smaller number of measurements concerning the others.



**Figure 3.** Number of measurements associated with different coefficients of determination ( $R^2$ ) for the number of fruits in the Brazil nut associated with analysis of variance (A), principal components - covariance (B), principal components - correlation (C), and structural analysis - correlation (D).



**Figure 4.** Number of measurements associated with different coefficients of determination ( $R^2$ ) for the dry mass of seeds in the Brazil nut associated with analysis of variance (A), principal components - covariance (B), principal components - correlation (C), and structural analysis - correlation (D).

A significant effect of the measurement factor suggests that these can be affected by regular, irregular, or systematic physiological changes. As this effect can vary in different ways and intensities among the Brazil nut genotypes, ANOVA does not make it possible to isolate this factor, which has its effects mixed with the experimental error. This leads to low estimates of repeatability and high estimates of the number of measurements required. In this case, principal component method, which considers the cyclic trait behavior, is the most recommended to estimate  $r$  with greater accuracy. This is because, with this methodology, the eigenvector is the one that expresses the progeny tendencies to maintain their relative positions in the measurement periods (Abeywardena, 1972; Cruz et al., 2012).

Similar results have been reported on other perennial fruit plants that also expressed cyclic behavior, such as acerola (Lopes et al., 2001), guava (Degenhardt et al., 2002), araçazeiro and pitangeira (Danner et al., 2010), peach (Bruna et al., 2012), sweet orange (Negreiros et al., 2014), and banana (Lessa et al., 2014), in which the methodology based on principal components provided the highest estimates of  $r$  and  $R$ . Pedrozo et al. (2015) estimated repeatability coefficients based on two-trait mixed models of Brazil nut production in Roraima. The number of fruits per plant and the fresh weight of seeds were evaluated in two native forest populations, for 5 and 8 consecutive years, in 51 and 85 adult trees, respectively. Repeatability coefficients were low for both the number of fruits per plant and the fresh weight of seeds, which led to the need for 9 years of evaluation to select the higher plants. In Acre, Assis (2016) also estimated repeatability coefficients for the Brazil nut based on mixed models. The study was conducted in two permanent plots, evaluating the number



of fruits per plant from 140 and 175 Brazil nut trees, over the period of 13 and 6 years, respectively. Repeatability coefficients were considered moderate and high, respectively, for both populations and, considering a coefficient of determination of 90%, would be necessary for the two populations 7 and 6 years, respectively, of evaluation of the number of fruits per plant for selecting superior genotypes.

Thus, due to the results obtained and its consistency with published research, it is recommended to use at least three measurements to identify superior Brazil nut genotypes regarding the number of fruits and dry mass of seeds.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Research supported by grants from Embrapa (Empresa Brasileira de Pesquisa Agropecuária), CNPq (National Counsel of Technological and Scientific Development), and FAPEMAT (Fundação de Amparo à Pesquisa do Estado de Mato Grosso). F.D. Tardin is the recipient of a research fellowship from CNPq.

## REFERENCES

- Abeywardena V (1972). An application of principal component analysis in genetics. *J. Genet.* 61: 27-51. <https://doi.org/10.1007/BF02984099>
- Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, et al. (2013). Köpen's climate classification map for Brazil. *Meteorol. Z.* 22: 711-728. <https://doi.org/10.1127/0941-2948/2013/0507>
- Assis VRA (2016). Melhoramento genético de *Bertholletia excelsa*: seleção, produção de mudas e sistema reprodutivo. Tese de Doutorado. Universidade Federal do Acre, Rio Branco.
- Bergo CI, Negreiros JRS, Migueloni DP and Lunz AMP (2013). Estimativas de repetibilidade de caracteres de produção em pupunheiras para palmito da raça Putumayo. *ver. Bras. Fruticult.* 35: 829-836. <https://doi.org/10.1590/S0100-29452013000300020>
- Bruna ED, Moreto AL and Dalbó MA (2012). Uso do coeficiente de repetibilidade na seleção de clones de pessegueiro para o litoral sul de Santa Catarina. *Rev. Bras. Frutic.* 34: 206-215. <https://doi.org/10.1590/S0100-29452012000100028>
- Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci. Agron.* 35: 271-276. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- Cruz CD, Regazzi AJ and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. UFV, Viçosa.
- Danner MA, Raseira Md CB, Sasso SAZ, Citadin I, et al. (2010). Repetibilidade de caracteres de fruto em araçazeiro e pitangueira. *Cienc. Rural* 40: 2086-2091. <https://doi.org/10.1590/S0103-84782010005000163>
- Degenhardt J, Ducroquet JP, Md REIS, Guerra MP, et al. (2002). Efeito de anos e determinação do coeficiente de repetibilidade de características de frutos de goiabeira-serrana. *Pesqui. Agropecu. Bras.* 37: 1285-1293. <https://doi.org/10.1590/S0100-204X2002000900012>
- Kainer KA, Wadt LHO and Staudhammer CL (2007). Explaining variation in Brazil nut fruit production. *For. Ecol. Manage.* 250: 244-255. <https://doi.org/10.1016/j.foreco.2007.05.024>
- Lessa LS, Ledo CAS, Amorim EP and de Oliveira S (2014). Estimativas de repetibilidade de híbridos diploides (AA) de bananeira. *Pesqui. Agropecu. Bras.* 49: 109-117. <https://doi.org/10.1590/S0100-204X2014000200005>
- Lopes R, Bruckner CH, Cruz CD, Lopes MTG, et al. (2001). Repetibilidade de características do fruto de aceroleira. *Pesqui. Agropecu. Bras.* 36: 507-513. <https://doi.org/10.1590/S0100-204X2001000300015>
- Lorenzi A (2002) Árvores brasileiras: manual de identificação e cultivos de plantas arbóreas do Brasil. 4. ed. Nova Odessa, SP: Instituto Plantarum, v.1.
- Mansour H, Nordheim EV and Rutledge JJ (1981). Estimators of repeatability. *Theor. Appl. Genet.* 60: 151-156. [PubMed https://doi.org/10.1007/BF00264520](https://doi.org/10.1007/BF00264520)

- Mori SA and Prance GT (1990). Taxonomy, ecology, and economic botany of the Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.: Lecythidaceae). *Adv. Econ. Bot.* 8: 130-150.
- Negreiros JRS, Neto Rd CA, Miqueloni DP and Lessa LS (2014). Estimativa de repetibilidade para caracteres de qualidade de frutos de laranjeira-doce. *Pesqui. Agropecu. Bras.* 49: 40-48. <https://doi.org/10.1590/S0100-204X2014000100006>
- Oliveira MSP and Moura EF (2010). Repetibilidade e número mínimo de medições para caracteres de cacho de bacabi (*Oenocarpus mapora*). *Rev. Bras. Frutic.* 32: 1173-1179. <https://doi.org/10.1590/S0100-29452010005000120>
- Pedrozo CA, Tonini H, Resende MDV and Jordão SMS (2015). Repeatability of fruits and seeds production and selection of Brazil nut genotypes in native populations in Roraima. *Rev. Arvore* 39: 863-871. <https://doi.org/10.1590/0100-67622015000500009>
- Teodoro PE, Costa RD, Rocha RB and Laviola BG (2016). Número mínimo de medições para a avaliação acurada de características agrônômicas de pinhão-mansão. *Pesqui. Agropecu. Bras.* 51: 112-119. <https://doi.org/10.1590/S0100-204X2016000200003>
- Tonini H and Pedrozo CA (2014). Variações anuais na produção de frutos e sementes de castanheira-do-brasil (*Bertholletia excelsa* Bonpl., Lecythidaceae) em florestas nativas de Roraima. *Revista Árvore. Viçosa, MG.* 38: 133-144.
- Wadt LHO, Kainer KA, Staudhammer CL and Serrano ROP (2008). Sustainable forest use in Brazilian extractive reserves: Natural regeneration of Brazil nut in exploited populations. *Biol. Conserv.* 141: 322-346. <https://doi.org/10.1016/j.biocon.2007.10.007>

## Supplementary material

**Table S1.** Estimates of the number of measurements associated with different coefficients of determination ( $R^2$ ) for the number of fruits and dry mass of seeds in 75 Brazil nut genotypes over five measurements.