

Estimates of repeatability coefficients and the number of the optimum measure to select superior genotypes in *Annona muricata* L.

C.F.B. Sánchez¹, R.S. Alves², A.d.P. Garcia¹, P.E. Teodoro², L.A. Peixoto², L.A. Silva², L.L. Bhering² and M.D.V. Resende²

¹Universidad Nacional de Colombia, Medellin, Colombia

²Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: L.L. Bhering

E-mail: leonardo.bhering@ufv.br

Genet. Mol. Res. 16 (3): gmr16039753

Received June 13, 2017

Accepted August 8, 2017

Published September 21, 2017

DOI <http://dx.doi.org/10.4238/gmr16039753>

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. The aim of this study was to evaluate repeated measures over the years to estimate repeatability coefficient and the number of the optimum measure to select superior genotypes in *Annona muricata* L. The fruit production was evaluated over 16 years in 71 genotypes without an experimental design. The estimation of variance components and the prediction of the permanent phenotypic value were performed using REML/BLUP proceedings. The coefficient of determination, accuracy, and selective efficiency increased when measures increased. The coefficient of determination of 80% was reached beyond 8 crop seasons with high accuracy and selective efficiency. Thus, the evaluation of 8 crop seasons can be suitable to select superior genotypes in the *A. muricata* L. breeding program. Predicted selection gain had a high magnitude for fruit production indicating that it is possible to take a progressive genetic advance for this trait over cycle breeding.

Key words: Soursop; REML; Mixed models; Longitudinal data; Genetic selection

INTRODUCTION

Annona muricata L. is a perennial crop cultivated in tropical climates and is adapted to several environments conditions. Its fruit has numerous uses, which can be stand out the *in natura* consumption and the cosmetic fabrication (Badrie and Schauss, 2010). Despite its economic importance in Colombia, there are no commercial varieties cultivated in this country. Besides that, it is very challenging to do experiments with perennial species due to the need to evaluate several crop seasons to identify superior genotypes for an important trait. Thus, much time is spent as well as the huge labor investment to lay the experiment over crop seasons. Therefore, the use of statistical methods more suitable to analyze field experiments is important because they can help breeders to make better decisions about the experiment, and consequently maximizing the selection efficiency in the breeding program.

The prediction of genetic parameters such as heritability and repeatability is one of the most important contributions of the quantitative genetics for the plant breeding (Resende, 2015). Currently, the standard analytic proceeding used to quantitative genetic approaches and selection plants is REML/BLUP (restricted maximum likelihood/best linear unbiased prediction). According to Resende and Rosa-Perez (2001), the main advantages of the mixed model methods (REML/BLUP) to predict genotypic values and to estimate genetic parameters are: can be applied to unbalanced data, allows to analyze data without experimental design, allows to use a large number of data that come from several experiments, generating information more accurately, allows to analyze the data taking off fixed effects, and predicts genotypic values unbiased and accurate, consequently maximizing the selection gain.

Experimental analysis considering repeated measures in the same genotypes over crop seasons has some particularities since these measurements are correlated and may be covariance heterogeneity between crop seasons (Resende, 2007; Mariguele et al., 2012). Interestingly, this kind of analysis has been used to predict genotypic values in the specific point among crop seasons or for all crop seasons.

An efficient model to analyze data that have repeated measures is the multivariate model in which each crop season is assigned as a different variable. However, if more than three crop seasons are considered, it is hard to adjust this model (presenting convergence problems) since it is super-parameterized (Resende et al., 2014). When data from individual plants with repeated measures without experimental design are used, the basic repeatability model without experimental design can be performed efficiently (Resende, 2007).

Based on the above considerations, the objective of this study was to evaluate repeated measures over crop seasons for estimating the repeatability coefficient and the optimum number of measurements to be carried out in the selection of superior *A. muricata* L. genotypes.

MATERIAL AND METHODS

The trial was carried out at campus Halcones Farm, in Cerritos - Risaralda, Colombia (4°48'48.00"N, 75°42'58.63"W, at approximately 1345 m in altitude) (WGS 84). Region climate, according to Köppen's classification, is AF, with rainfall scattered throughout the year, with average temperatures of 23°C, annual rainfall of 2100 mm and relative humidity around 75%. Planting was carried out using 8 x 8 m spacing, without an experimental design. The cultural practices were carried out according to the recommended cultural management for *A. muricata* L. Sixteen measurements of the fruit yield of 71 *A. muricata* L. genotypes were carried out from 2000 to 2016.

The basic repeatability model without experimental design is given by the Equation 1:

$$y = Xm + Wp + e \quad (\text{Equation 1})$$

where y is the vector of data; m is the vector of measurements (assigned as fixed) plus the general mean; p is the vector of permanent effects of plants (genotypic effects plus permanent environmental effect) (assigned as random); and e is the vector of errors or residuals (assigned as random). X and W are the incidence matrix for m and p , respectively, (Resende, 2007).

The significance of the difference in the adjustment of different models was tested using the likelihood ratio test (LRT) proposed by Wilks, and defined by the Equation 2:

$$\lambda = 2 \left[\text{Log}_e L_{p+1} - \text{Log}_e L_p \right] \quad (\text{Equation 2})$$

where L_{p+1} and L_p are the peaks of the likelihood function associated with the full model and the reduced model, respectively. Thus, λ should be compared with the probability density function (χ^2 , Table 1) for a determined number of degrees of freedom and error probability (Dobson, 1990). The number of degrees of freedom is defined by the difference between the number of parameters or variance components between the evaluated models.

The efficiency of the use of m measurements in each plant compared with the use of one measurement in each plant regarding genetic gain with selection is defined by the Equation 3:

$$f = \left[\frac{m}{1 + (m-1)\rho} \right]^{1/2} \quad (\text{Equation 3})$$

where f is the selective efficiency; m is the number of measurements, and ρ is the repeatability coefficient. This expression is valid for asexual and sexual strategies for reproducing the selected genotype. The m value to reach an f fraction of the maximum coefficient of determination is defined by Equation 4:

$$m = \frac{f(1-\rho)}{(1-f)\rho} \quad (\text{Equation 4})$$

The mean components (individual BLUPs) based on the permanent phenotypic effect of 71 evaluated soursop genotypes were obtained aiming to rank and select superior genotypes.

The basic repeatability model without experimental design was adjusted using the Selegen REML/BLUP software (Resende, 2016).

RESULTS

Table 1 shows the deviance and the LRT for the parameters estimated based on statistical analysis for fruit production evaluated in 71 *A. muricata* L. genotypes over 16 crop

seasons. The null hypothesis was the full model, and the reduced model did not differ. The permanent effect was significant, i.e., the full model fitted the data better than the reduced model.

Table 1. Deviance and the likelihood rate test (LRT) for parameters of the statistical analysis evaluating 71 *Annona muricata* L. genotypes over 16 crop seasons.

Effect	Deviance	LRT
Permanent ¹	11064.79	324.74**
Residual	-	-
Full model	10740.05	-

¹Genotypic effects plus permanent environment effects; **significant at 1% of probability by the chi-squared test.

Variance components (individual REML) and the coefficient of determination or reliability coefficient (Goddard, 1992) are in Table 2. Heritability was moderate, and the reliability coefficient was high.

Table 2. Permanent phenotypic variance among plants (genotypic variance plus permanent environment variance over crop seasons) (V_{pp}), temporary environment variance (V_{te}), individual phenotypic variance (V_p), individual repeatability ($r = h^2$) and its confidence interval, repeatability of the average of crop seasons or repeated measures (rm), selection accuracy based on the average of m crop seasons or repeated measures (Ac_m), and selection reliability.

V_{pp}	V_{te}	V_p	$r = h^2$	rm	Ac_m	Selection reliability
2368.30	4388.00	6757.30	0.35 ± 0.05	0.90	0.95	0.90

The coefficient of determination, selection accuracy, and selective efficiency between the use of m measures compared with the use of just one measure were calculated aiming to establish the optimum number of measures to make selection for fruit production in *A. muricata* L. To reach $f = 0.80$ or 80% of the maximum coefficient of determination are needed 7.43 8 measurements (Figure 1). On the other hand, four measurements are needed to reach selection accuracy greater than 0.80 (Figure 2). The selective efficiency increased over the eighth measurements and stabilized after that (Figure 3).

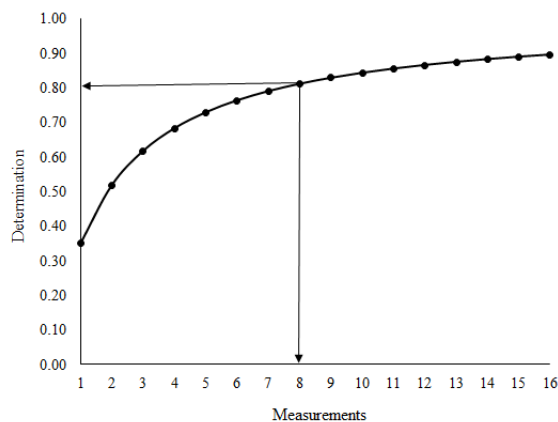


Figure 1. Coefficient of determination in function of the number of measurements for fruit production evaluated in 71 *Annona muricata* L. genotypes.

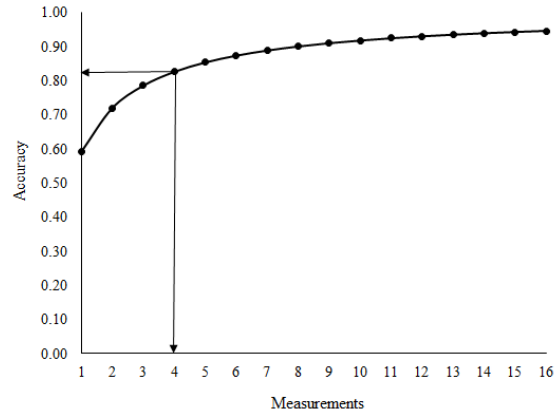


Figure 2. Selection accuracy in function of the number of measurements for fruit production evaluated in 71 *Annona muricata* L. genotypes.

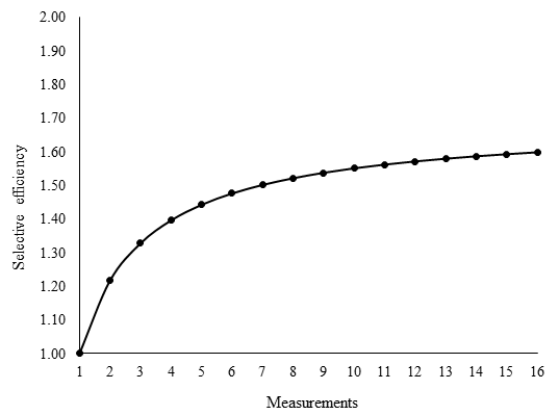


Figure 3. Selective efficiency in function of the number of measurements for fruit production evaluated in 71 *Annona muricata* L. genotypes.

The mean components (individual BLUP) based on the permanent phenotypic effect were obtained considering the selection of 10 genotypes (14.08%), and the predicted selection gain for fruit production was 84.06% (Table 3).

Table 3. Permanent phenotypic effect (P_p), permanent phenotypic value ($u + P_p$), and selection gain (%) based on the selection of 10 superior *Annona muricata* L. genotypes for fruit production.

Rank	Genotype	P_p	$u + P_p$	Selection gain (%)
1	35	129.87	264.72	129.87
2	54	126.44	261.28	128.16
3	65	104.47	239.31	120.26
4	4	81.45	216.29	110.56
5	24	75.60	210.44	103.57
6	15	72.40	207.24	98.37
7	39	69.83	204.67	94.29
8	71	61.54	196.38	90.20
9	1	60.89	195.73	86.94
10	38	58.12	192.96	84.06

DISCUSSION

Techniques for genetic evaluation involve the prediction of genotypic values and the estimation of variance components simultaneously. The optimum proceeding to predict the genotypic values at the individual level is called BLUP (Henderson and Quaas, 1976). The BLUP prediction assumes that the variance components are known. However, in the practice views are needed truly estimates of the variance components (genetic parameters) to obtain the empirical BLUP (Harville and Carriquiry, 1992). Therefore, the optimum proceeding to make the genetic evaluation is REML/BLUP, also defined as mixed model methodology (Resende, 2007).

Besides the mixed model that is suitable for genetic estimation, the likelihood principle also allows comparing several models since they have a hierarchical structure. When the full model performs exactly the phenotypic data, as shown in this study, the difference between two adjusted models can be tested by the LRT (Resende et al., 2014). Thus, the LRT (Table 1) revealed that the genotypic effects plus the permanent environmental effects were significant pointing out that the full model is the best model to fit the data.

The estimative of the individual repeatability ($r = h^2$) is one of the most important findings in this study since there are few types of research in perennial plants that have estimated this parameter. In this study, the estimate individual repeatability for fruit yield is classified as a moderate magnitude and the mean repeatability of crop seasons or repeated measurements are classified as high magnitude (Resende, 2002). Similar findings were reported for other fruit crops such as acerola (Lopes et al., 2001), guava (Degenhardt et al., 2002), araçá and pitanga (Brazilian fruit crops) (Danner et al., 2010), peach (Della Bruna et al., 2012), sweet orange (Negreiros et al., 2014), and banana (Lessa et al., 2014).

The proximity between the predicted genotypic values and the true genotypic value for genotypes can be evaluated based on the parameters called selective accuracy (Vleck et al., 1987), which means the correlation between the predicted genotypic values and the true genotypic values. The selective accuracy depends on the heritability and repeatability, the number and the quality of the information, and the methodology used to predict the genotypic values. The selective accuracy is the principal component for the genetic progress since it is a parameter associated with the selection precision, which breeders can change aiming to maximize the genetic gain (Resende, 2002). In this study, the selective accuracy based on the average of the crop seasons or repeated measurements (Ac_m) was high (0.95) indicating high precision according to the classification proposed by Resende and Duarte (2007), and consequently, high selection reliability (0.90), since it is a coefficient of determination that corresponds to the square of the selective accuracy (Goddard, 1992).

The coefficient of determination, the selective accuracy, and the selective efficiency increased when the number of measurements increased as expected. With the evaluation of 8 crop seasons, it was possible to reach 80% of the maximum coefficient of determination ($m = \infty$) with high accuracy and high selective efficiency. Therefore, the 8 crop seasons evaluated can be suitable to select superior genotypes in the *A. muricata* L. breeding aiming to increase the fruit yield.

Predicted gains with selection were high magnitudes for yield fruit, which the selection of the 10 superior genotypes made a predicted gain of 84.06%. Moreover, the determination of the optimum number of genotypes that should be selected to compose the population to produce seeds and to maintain the population effective size N_E (usually $30 \leq N_E \leq 50$) may consider the endogamy (Resende and Bertolucci, 1995). Thus, we suggest that in future

researches evaluating *A. muricata* L., statistic and genetic designs should be performed to allow the estimation of genetic parameters more accurately. Besides, the superior *Annona muricata* L. identified in this research can be used for diallel crossing aiming to obtain hybrids with high-yield fruit.

CONCLUSIONS

The coefficient of determination, accuracy, and selective efficiency increase when the number of measures increased. Eight crop seasons were enough to obtain 80% of the maximum coefficient of determination with high accuracy and high selective efficiency.

The predicted selection gain was high for fruit production, which indicates that it is possible to obtain genetic advance in this trait over breeding cycles.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais), Funarbe (Fundação Arthur Bernardes), and the Federal University of Viçosa for financial support.

REFERENCES

- Badrie N and Schauss A (2010). Soursop (*Annona muricata* L.): composition, nutritional value, medicinal uses, and toxicology. In: Bioactive Foods in Promoting Health: Fruits and Vegetables (Watson RR and Preedy V, eds.). Academic Press, Oxford, 621-643.
- Danner MA, Raseira MDCB, 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, Reis MD, 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>
- Della Bruna E, 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>
- Dobson AJ (1990). An introduction to generalized linear models. Chapman and Hall, Melbourne.
- Goddard ME (1992). Genetic evaluation in the dairy industry. In: Animal breeding: the modern approach (Hammond K, Grasser H-U and McDonald A, eds.). Post-Graduate Foundation of Veterinary Science, Sydney, 95-101.
- Harville DA and Carriquiry AL (1992). Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* 48: 987-1003. <https://doi.org/10.2307/2532693>
- Henderson CR and Quaas RL (1976). Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43: 1188-1197. <https://doi.org/10.2527/jas1976.4361188x>
- Lessa LS, da Silva Ledo CA, 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>
- Mariguelo KH, Resende MDV, Viana JMS, Fonseca F, et al. (2012). Métodos de análise de dados longitudinais para o melhoramento genético da pinha. *Pesqui. Agropecu. Bras.* 46: 1657-1664. <https://doi.org/10.1590/S0100-204X2011001200011>

- Negreiros JRS, Neto RDCA, Miqueloni DP and Lessa LS (2014). Estimativa de repetibilidade para caracteres de qualidade de frutos de laranja-doce. *Pesqui. Agropecu. Bras.* 49: 40-48. <https://doi.org/10.1590/S0100-204X2014000100006>
- Resende MDV (2002). Genética biométrica e estatística no melhoramento de plantas perenes. Embrapa Informação Tecnológica, Brasília.
- Resende MDV (2007). Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo, Embrapa Florestas.
- Resende MDV (2015). Genética Quantitativa e de Populações. Suprema, Visconde do Rio Branco.
- Resende MDV (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed. Appl. Biotechnol.* 16: 330-339. <https://doi.org/10.1590/1984-70332016v16n4a49>
- Resende MDV and Bertolucci FLG (1995). Maximization of genetic gain with restriction on effective population size and inbreeding in *Eucalyptus grandis*. In: IUFRO Congress on eucalypt plantations 167-170.
- Resende MDV and Rosa-Perez JRH (2001). Genética e Melhoramento de Ovinos. Ed. UFPR, Curitiba.
- Resende MDV and Duarte JB (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui. Agropecu. Trop.* 37: 182-194.
- Resende MDV, Silva FF and Azevedo CF (2014). Estatística Matemática, Biométrica e Computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência. 1. ed. Suprema, Visconde do Rio Branco.
- Vleck LD, Pollak EJ and Oltenacu EAB (1987). Genetics for the animal sciences. WH Freeman, New York.