GMR

Minimum number of measurements for evaluating soursop (*Annona muricata* L.) yield

C.F.B. Sánchez[1], P.E. Teodoro[2], S. Londoño[1], L.A. Silva[2], L.A. Peixoto[1] and L.L. Bhering[2]

[1]Departamento Ciencias Agronómicas,
Universidad Nacional de Colombia, Medellin, Colombia
[2]Laboratório de Biometria, Departamento de Biologia Geral,
Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: L.L. Bhering
E-mail: leonardo.bhering@ufv.br

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 soursop (*Annona muricata* L.) genotypes based on fruit yield. Sixteen measurements of fruit yield from 71 soursop genotypes were carried out between 2000 and 2016. In order to estimate r with the best accuracy, four procedures were used: analysis of variance, principal component analysis based on the correlation matrix, principal component analysis based on the phenotypic variance and covariance matrix, and structural analysis based on the correlation matrix. The minimum number of measurements needed to predict the actual value of individuals was estimated. Principal component analysis using the phenotypic variance and covariance matrix provided the most accurate estimates of both r and the number of measurements required

for accurate evaluation of fruit yield in soursop. Our results indicate that selection of soursop genotypes with high fruit yield can be performed based on the third and fourth measurements in the early years and/or based on the eighth and ninth measurements at more advanced stages.

INTRODUCTION

Soursop (*Annona muricata* L.) is a fruit tree species cultivated in tropical climates that is adapted to different soils. Its trees can reach 10 m, but commercial crops are grown up to 2.5 m height. Its main commercial product is its fruits, which are composed of 80% edible pulp, 16% seeds, and 4% peduncle. They are consumed *in natura* and used for the cosmetics production (Badrie and Schauss, 2010).

In research on perennial species, such as the soursop, several measurements of the trait of interest over the course of the growth cycle are required to identify superior individuals. In these studies, correct discrimination of genotypes depends on the evaluation conditions and intensities that allow both precision (less dispersion) and accuracy (small bias between the estimate and the parameter) of estimates. Appropriate dimensioning of the number of measurements to be performed facilitates characterization of the genotypes more precisely and several years may be necessary to measure trait expression manifested over time (Teodoro et al., 2016). Dimensioning of the number of measurements to be performed using the repeatability coefficient (r) is fundamental for perennial species. It provides an approximation of the maximum value that the broad sense heritability of the trait of interest can have. In addition, this information is also important to determine the shortest possible selection cycle, which has a direct impact on the human and financial resources of the research (Cruz et al., 2012).

Dimensioning the number of measurements has been performed for several perennial crops, such as acerola (Lopes et al., 2001), guava (Degenhardt et al., 2002), araçá and pitanga (Danner et al., 2010), peach (Della Bruna et al., 2012), sweet orange (da Silva Negreiros et al., 2014), banana (Lessa et al., 2014), and jatropha (Teodoro et al., 2016). However, there are no reports in the literature on the number of measurements needed to evaluate fruit yield in soursop. This study aimed to identify the most efficient method to estimate r and to predict the minimum number of measurements needed for accurate evaluation of soursop genotypes based on fruit yield.

MATERIAL AND METHODS

The trials were carried out at campus Halcones Farm, in Cerritos, Risaralda, Colombia (4°48'48.00"N, 75°42'58.63"W, at approximately 1345 m altitude) (WGS 84). The climate in the region, according to Köppen's classification, is AF, with scattered rainfall throughout the year, average and minimum temperatures of 23 and 18°C, respectively, an annual rainfall of 2100 mm, and a relative humidity around 75%. Planting was carried out using 8 x 8 m spacing, using a completely randomized experimental design. The cultivation was conducted without irrigation and the other cultural practices were carried out according to the recommended cultural management for soursop.

Fruit yield was measured 16 times in 71 soursop genotypes between the years 2000 and 2016. In order 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 tested using ANOVA, where the statistical model adopted considered two factors of variation according to Equation 1:

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

where Y_{ij} is the observation for the i-th genotype for the j-th measurement; μ 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 ε_{ij} is the experimental error established by the temporary effects of the environment at the j-th measurement of the i-th progeny. The genotypes were clustered using the Scott and 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 the ANOVA, estimates of the mean squared of the measurements, mean squared of genotypes (MS_G), mean squared error (MS_E), and F-test value for the genotype were obtained. Then, the r was estimated following Equation 2:

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

where η 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 approximate magnitudes, is the one that expresses the tendency of the genotypes to maintain, over the years, their relative positions in relation to the others (Abeywardena, 1972). For this method, r is given by Equation 3:

$$r = \frac{\hat{\lambda}_1 - 1}{\eta - 1}$$ (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:

$$r = \frac{\hat{\lambda}_1 - \hat{\sigma}_Y^2}{\hat{\sigma}_Y^2 - (\eta - 1)}$$ (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 in relation to the above mentioned methods that are based on principal components. Based on the AECOR method, r is given by Equation 5:

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

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:

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

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 η, was obtained by Equation 7:

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

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. An intraclass correlation network was used to graphically express the functional relationship between the 16 measurements of fruit yield, in which the proximity between the nodes (traces) is proportional to the absolute value of the correlation (r_{ij}) between these nodes. The thickness of the edges was controlled by a cutoff value of 0.60, which means that only edges with $|r_{ij}| \geq 0.60$ are highlighted. The statistical data analyses were performed using Genes (Cruz, 2013) and Microsoft Excel®.

RESULTS AND DISCUSSION

In order to verify the genetic variability of fruit yield evaluated in 71 soursop genotypes during 16 years, an ANOVA was performed. The significant effect of genotype indicates that there is genetic variability in the population and, therefore, it is possible to select soursop genotypes with high fruit yield (Table 1). The F-test revealed a significant effect of measurement, indicating that changes in genotype means occurred over time. These results highlight the need to study the minimum number of measurements needed to select superior soursop genotypes for fruit yield.

The Scott and Knott cluster test divided the 71 soursop genotypes into five homogeneous groups with respect to fruit yield (Table 2). Group I consisted of genotypes 59, 124, and 45, characterized by the highest measurement mean. Group II contained genotypes 1, 10, 15, 24, 47, 79, 81, and 179, which also showed relatively high measurement means. The aim of the soursop breeding program is to select genotypes with the highest performance and genetic diversity to participate in crosses to obtain segregating populations. The results found in this study suggest that crosses between individuals from Groups I and II can contribute to generate populations with high variability and mean for fruit yield.

Table 1. Summary of ANOVA for fruit yield in 71 soursop genotypes over 16 measurements.

Source of variation	Degrees of freedom	Mean squared
Measurement	15	93,981.61*
Genotype	70	42,282.05*
Residue	1,050	4,388.99

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

Table 2. Clustering based on a Scott and Knott test applied to fruit yield in 71 soursop genotypes over 16 measurements.

Group	Genotypes	Mean ± standard deviation
I	59, 124, and 45	269.03 ± 12.56
II	1, 10, 15, 24, 47, 79, 81, and 179	209.20 ±10.07
III	18, 46, 51, 54, 77, 89, 125, and 126	173.39 ± 10.44
IV	2, 12, 16, 22, 23, 36, 40, 41, 42, 48, 53, 82, 90, 107, 116, 117, 118, 119, 120, 121, 122, 123, 129, 130, 131, 132, 144, 166, 167, 168, and 169	129.57 ± 13.79
V	3, 11, 19, 20, 21, 35, 38, 44, 49, 50, 52, 55, 57, 58, 78, 83, 115, 133, 135, 136, and 164	80.44 ± 13.39

Figure 1 shows that fruit yield was highest for measurements 3, 4, 13, and 14. These results suggest a cyclical behavior among soursop genotypes for fruit yield and indicates the need for caution when selecting superior genotypes based on this trait. Phenotypic stabilization occurred between the sixth and eleventh measurements. This is important to note because when a repeatability study is performed with genotypes not yet stabilized less accurate estimates can be found. This could lead the researcher to believe that increasing the number of measurements is the solution to the problem (Laviola et al., 2013). In some cases, the absence of measurements in early stages, when there is no complete manifestation of the genetic potential of the material studied, may increase the estimate of r.
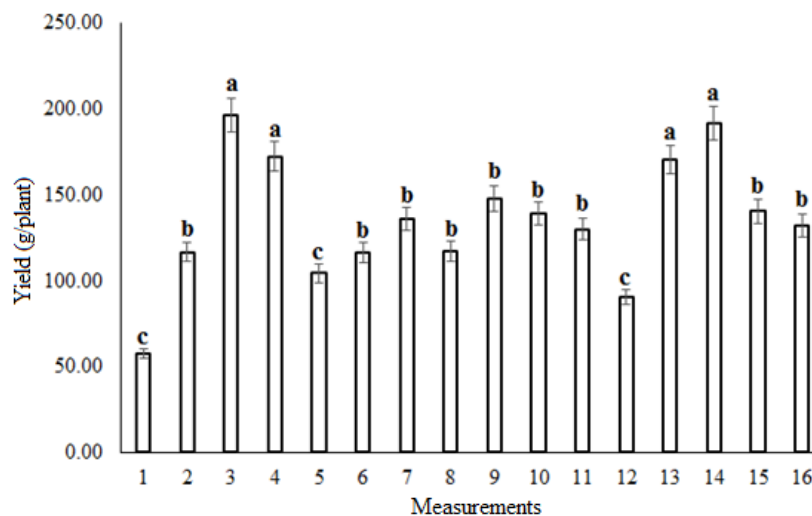

Figure 1. Means of fruit yield of 71 soursop genotypes at each measurement point. Bars marked with the same lower case letter are not significantly different according to a Scott and Knott test at 5% probability.

The estimates of r using the ANOVA were lower than those observed with the principal components (CPCOV) technique but were close to those obtained using AECOR (Table 3). The CPCOV provided the highest estimate of r, which resulted in a lower number of measurements needed to identify superior soursop genotypes (Figure 2). Based on this methodology, six measurements are required to predict the actual value of soursop genotypes with a minimum accuracy of 80% (**Table S1**). The other methods used, suggested that seven measurements were required for a predictability of 80%. As the R^2 increases, CPCOV becomes even more efficient, indicating a need for an even smaller number of measurements compared to the other techniques.

Table 3. Estimates of repeatability coefficient (r) and coefficient of determination (R^2) for fruit yield in 71 soursop genotypes over 16 measurements.

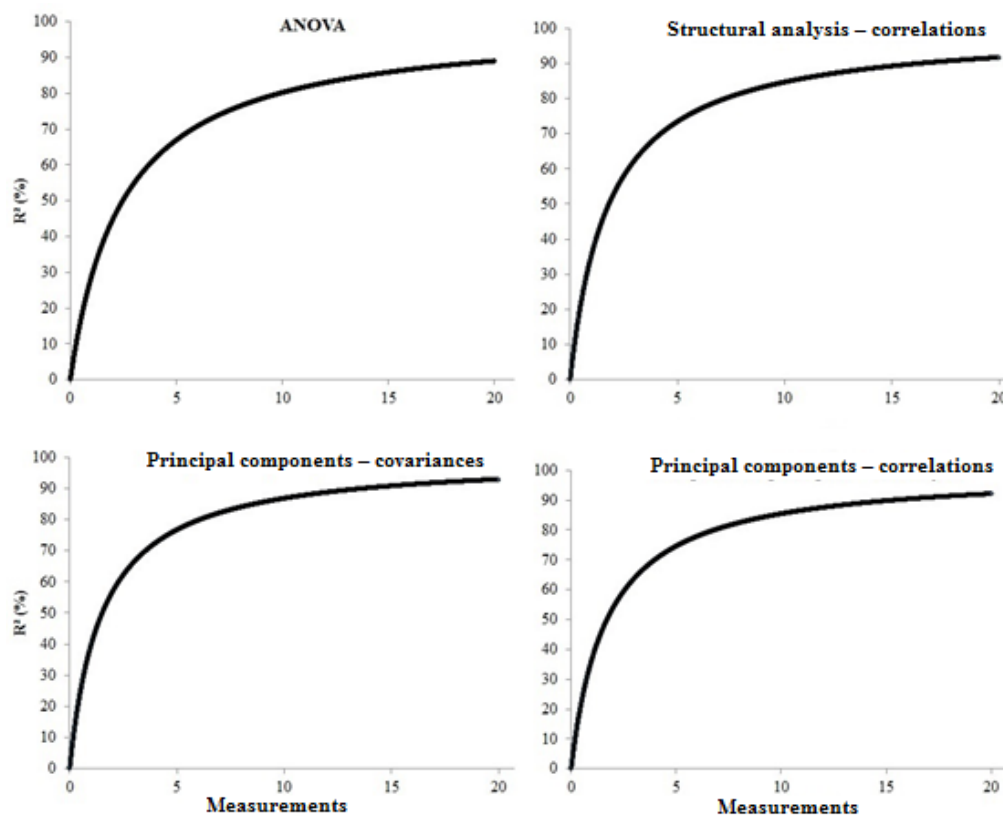| Method	r	R^2
Analysis of variance	0.3505	0.8962
Principal components - covariance (CPCOV)	0.3979	0.9136
Principal components - correlation (CPCOR)	0.3705	0.9034
Structural analysis - correlation	0.3566	0.8986

Figure 2. Number of measurements associated with different coefficients of determination (R^2) for fruit yield in soursop.

That the factor measurement was significant suggests that the measurements can be affected by regular, irregular or systematic physiological changes. This effect can vary in different ways and intensities among soursop genotypes. The ANOVA does not allow for isolating this factor because its effects are mixed with experimental error. This may lead to an underestimation of the repeatability and an overestimation of the number of measurements needed. In this case, the principal component method (CPCOV), which takes into account the cyclical behavior of the trait, is better suited to estimate r with the highest accuracy. With this technique, the eigenvector is the one that expresses the tendency of the progenies to maintain their relative positions over time.

Similar results have been reported on other perennial fruit plants that also express cyclical behavior, such as acerola (Lopes et al., 2001), guava (Degenhardt et al., 2002), araçá and pitanga (Danner et al., 2010), peach (Della Bruna et al., 2012), sweet orange (da Silva Negreiros et al., 2014), and banana (Lessa et al., 2014), in which methodologies based on principal components provided the highest estimates of r and R^2. Thus, based on the results obtained here and their consistency with previously published research, this method (CPCOV) is recommended for future studies of soursop.

Based on the intraclass correlation network between the measurements over the years, we found that the first three measurements (M1, M2, and M3) were similar to each other but more distant from later measurements. This indicates that the most productive genotypes in the first harvests tend to not keep their potential after the first few years (Figure 3). These results are relevant for genetic breeding of soursop and indicate that early selection for increased fruit yield should be avoided.
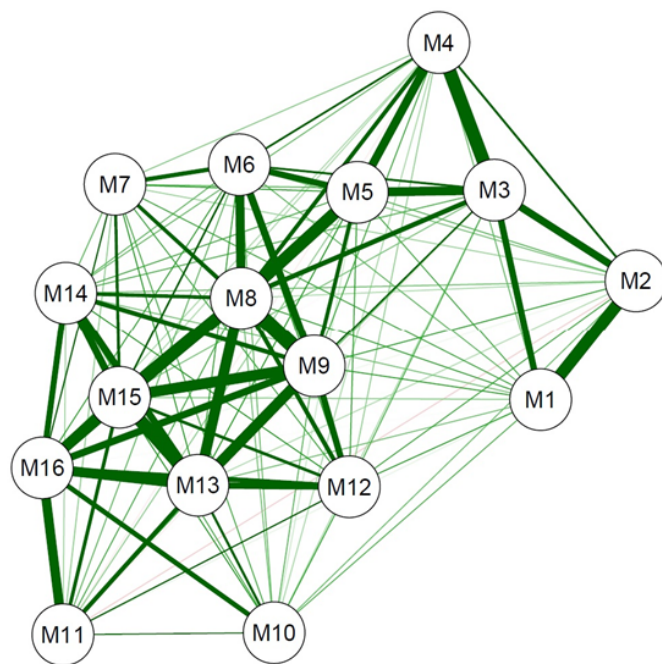

Figure 3. Intraclass correlation network among the 16 measurements (M1-M16) taken for fruit yield in 71 soursop genotypes.

The largest association occurred between the eighth and ninth measurements (M8 and M9) (Figure 3). The selection of soursop genotypes based on these measurements would allow a predictability of 85.40% and an r of 0.7452 (**Table S2**). This situation would be ideal because fruit yield is a quantitative trait determined by a large number of genes that can be expressed to a greater or lesser degree based on the development stage of the genotypes. However, a greater demand for time and labor would be necessary if the selection occurred at this stage. Alternatively, an evaluation of genotypes at measurements 3 and 4 would promote a predictability of 82.76%, associated with an r of 0.6915 (**Table S2**). These estimates are higher than the established reliable limits ($r \geq 0.40$, $R^2 \geq 0.75$) proposed by Bergo et al. (2013) and Oliveira and Moura (2010), respectively. Therefore, it is clear that the increase achieved between the fourth and ninth measurement does not justify the expenditure of time and hand labor.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Abeywardena V (1972). An application of principal component analysis in genetics. *J. Genet.* 61: 27-51. https://doi.org/10.1007/BF02984099

Badrie N and Schauss A (2010). Soursop (*Annona muricata* L.): composition, nutritional value, medicinal uses, and toxicology. Bioactive foods in promoting health: fruits and vegetables (Watson RR and Preedy VR, eds.). Elsevier Inc., Oxford, UK, 621-643.

Bergo CI, Negreiros JRS, Migueloni DP and Lunz AMP (2013). Estimativas de repetibilidade de caracteres de produção em pupunheiras para palmito da raça Putumayo. *Rev. Bras. Frutic.* 35: 829-836. https://doi.org/10.1590/S0100-29452013000300020

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.

da Silva Negreiros JR, 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

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: 1285-1294.

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

Laviola BG, Oliveira AMC, Bhering LL, Alves AA, et al. (2013). Estimates of repeatability coefficients and selection gains in Jatropha indicate that higher cumulative genetic gains can be obtained by relaxing the degree of certainty in predicting the best families. *Ind. Crops Prod.* 51: 70-76. https://doi.org/10.1016/j.indcrop.2013.08.016

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

Mansour H, Nordheim EV and Rutledge JJ (1981). Estimators of repeatability. *Theor. Appl. Genet.* 60: 151-156. https://doi.org/10.1007/BF00264520

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

Teodoro PE, Costa RD, Rocha RB and Laviola BG (2016). Número mínimo de medições para a avaliação acurada de características agronômicas de pinhão-manso. *Pesqui. Agropecu. Bras.* 51: 112-119. https://doi.org/10.1590/S0100-204X2016000200003

Supplementary material

Table S1. Estimates of the number of measurements associated to different coefficients of determination (R^2) for fruit yield in 71 soursop genotypes over 16 measurements.

Table S2. Estimates of repeatability coefficient (r) and coefficient of determination (R^2) associated with different number of measurement (N) and intervals to fruit yield in soursop genotypes.