Crop Breeding and Applied Biotechnology 8: 79-84, 2008
Brazilian Society of Plant Breeding. Printed in Brazil

Determination of the number of years in Arabic coffee progenies selection through repeatability

Júlio César Mistro[1*], Luiz Carlos Fazuoli[1], Oliveiro Guerreiro Filho[1], Maria Bernadete Silvarolla[1], and Masako Toma-Braghini[1]

Received 02 October 2007

Accepted 07 April 2008

ABSTRACT - *The purpose of this study was to estimate the repeatability coefficient of grain yield in Arabic coffee and to determine the minimal number of annual harvests required for a satisfactory level of reliability at the plant level. The experiment was conducted in Campinas (SP) in a randomized complete block design with 21 F_3RC_2 progenies of Icatu germplasm, for nine years, in nine replications. The repeatability coefficients were estimated by the method of analyses of variance, principal components based on correlation and covariance matrices and structural analyses based on correlation matrices. The selection of F_3RC_2 progenies of Icatu germplasm can be carried out with an accuracy of 80% in the three first harvests. The repeatability coefficient estimated by the covariance matrix method was the most adequate methodology to estimate this coefficient, since it minimizes the effect of biennial bearing.*

Key words: analyses of variance, principal components, coefficient of determination, breeding, *Coffea arabica* L.

INTRODUCTION

Arabic coffee (*Coffea arabica* L.), a perennial species, has some peculiar aspects, such as a long reproduction cycle and pronounced variation in annual yield and trait expression over several years (Sera 2001). The breeding process is slow, since several years are needed for the selection of superior genotypes in each generation, totaling an average of 20 to 25 years for the release of a new cultivar. As little time as possible should be lost with the evaluation of genotypes in breeding programs of perennial species, to avoid the waste of ever-scarcer labor and capital resources. On the other hand, the evaluations cannot be concluded in a very short time, since haste can cause errors in the identification of the genotypes of interest.

In this context, the repeatability coefficient is an indispensable information source for breeders of perennial plants, underlying the determination of the number of phenotypic data of annual yield required from each tree for an efficient selection of genotypes with less cost and labor. The selected genotype is expected to maintain the initial performance for the entire lifetime. This coefficient also represents the highest possible value of trait heritability (in the broad sense). It is by far easier to be estimated, since no controlled crosses and study of progenies are needed. The concept of repeatability can be defined as the correlation between the measures of a particular trait in the same tree, evaluated by measurements replicated in time or space. It expresses the proportion of the total variance explained by the variations caused by the genotype and the permanent alterations attributed to the common environment (Falconer 1987, Cruz and Regazzi 1997).

[1] Instituto Agronômico do Estado de São Paulo (IAC/APTA), Centro de Café "Alcides Carvalho", C.P. 28, 13.001-970, Campinas, SP, Brasil. *E-mail: mistrojc@iac.sp.gov.br

A high estimated repeatability value of a trait indicates the possibility of predicting the real value of a tree with a relatively small number of measurements (Cornacchia et al. 1995). A low repeatability on the other hand, means that a high number of replications are necessary for efficient selections. According to Shimoya et al. (2002), estimates of the repeatability coefficient of over 0.5 and of over 80% of the coefficient of determination can be considered satisfactory.

The traditionally most commonly used method to estimate the repeatability coefficient is the analysis of variance. Other methods using Principal components (Abeywardena 1972, Rutledge 1974) and structural analyses (Mansour et al. 1981) have been suggested in view of the greater accuracy (Cruz and Regazzi 1997).

These methods have been applied in several crops, e.g., pine (Cornacchia et al. 1995), alfafa (Ferreira et al. 1999), rubber tree (Gonçalves et al. 1990), elephant grass (Shimoya et al. 2002), and tomato (Cargnelutti Filho et al. 2004). There are very few studies of repeatability in coffee. Fonseca et al. (2004) determined the minimal number of harvests to optimize selection in coffee robusta clones, another commercial species of the genus *Coffea*, by the methods of analysis of variance and Principal components.

This study aimed to estimate the repeatability coefficient of the trait grain yield of Arabic coffee and the minimal number of evaluations required for an accurate prediction of the real value of the trees, determining the repeatability coefficient by different methodologies.

MATERIAL AND METHODS

For nine years, the coffee cherry yield of 21 coffee progenies of cultivar Icatu were evaluated in generation F_3RC_2, in an experiment conducted at the Centro Experimental Central (CEC) of the Instituto Agronômico of the State of São Paulo (IAC/APTA), in Campinas (lat 22°54'S, long 47°05'W, alt 674 m asl), at an average annual temperature of 22 °C and annual rainfall of 1.380 mm, in a random block design, with nine replications with two plants per plot, at a spacing of 4.0 m between rows and 1.5 m between plants.

The coffee yield was obtained by weighing the coffee cherries per plant of the harvests of nine years.

The estimates of the coefficients of repeatability (r) were obtained by the methods of analysis of variance (ANOVA), Principal components based on the Matrices of correlations (CP_{cor}) of covariances (CP_{cov}) and the Structural analysis based on the Matrices of correlations (AE_{cor}). The minimum number of measurements required to predict the real value of the plants, based on the pre-established coefficients of determination (R^2) (0.80, 0.85, 0.90, 0.95 and 0.99), was obtained as proposed by Cruz and Regazzi (1997).

a) Method of the Analysis of Variance (ANOVA)

The repeatability coefficient was estimated using the results of the analysis of variance, according to the model:

$Y_{ij} = u + G_i + A_j + E_{ij}$, where:

Y_{ij} = observation of the i th progeny in the j th environment;

u = general mean;

G_i = random effect of the i th progeny under the influence of the permanent environment (i = 1, 2, ..., p; p = 21).

A_j = fixed effect of the temporary environment on the j th measurement (j = 1, 2, ..., n; n = 9).

E_{ij} = experimental error given by the temporary effects of the environment in the j th measurement of the i th progeny.

The repeatability coefficient was calculated as follows:

$$r = \frac{Cov(Y_{ij}, Y_{ij'})}{\sqrt{\hat{V}(Y_{ij})\hat{V}(Y_{ij'})}} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2}$$, where

$\hat{\sigma}_g^2$ = genotypic variance and $\hat{\sigma}_e^2$ = environmental variance

b) Method of Principal Components

This method, used by Abeywardena (1972), estimates the repeatability coefficient in two ways: firstly by the correlation matrix and secondly by the matrix of phenotypic variances and covariances.

- Based on the Correlation Matrices (CP_{cor})

The method is based on the correlation matrix between genotypes in each pair of years. The normalized eigenvalues (λ) and eigenvectors (α) of R are determined in this matrix. An eigenvector with the same signal and similar magnitudes expresses the tendency of maintaining the relative positions of the genotypes in the different intervals of years.

The repeatability coefficient is represented as follows:

$$r = \frac{\lambda_1 - 1}{\eta - 1}, \text{ where}$$

λ_1 = is the highest eigenvalue associated to the eigenvector whose elements have the same signal and similar magnitudes. It is obtained from the matrix of correlations between data;

η = number of years evaluated.

- Based on Matrices of Covariances (CP_{cov})

The repeatability coefficient is estimated based on the matrix of phenotypic variances and covariances (Γ).

In this case, the estimate of the repeatability coefficient is obtained by the following equation:

$$r = \frac{\lambda_1 - \hat{\sigma}^2_Y}{\hat{\sigma}^2_Y(\eta - 1)}, \text{ where}$$

λ_1= is the eigenvalue (Γ) associated to the eigenvector whose elements have the same signal and similar magnitudes;

$$\hat{\sigma}^2_Y = \hat{\sigma}^2_g + \hat{\sigma}^2_e$$

- Based on the Matrices of Correlations (AE_{cor})

This method, proposed by Mansour et al. (1981), considers R as parametric correlation matrix between genotypes in each pair of evaluations and $\hat{R}$ as estimator. The estimator of the repeatability coefficient based on structural analysis is expressed as follows:

$$r = \frac{\alpha' R \alpha^2 - 1}{\eta - 1}, \text{ where } \alpha' = \left[1/\sqrt{\eta} \cdots 1/\sqrt{\eta} \right]$$

is the eigenvector with parametric elements, associated to the highest eigenvalue of R.

The minimum number of measurements required to predict the real value of the genotypes, based on the predicted coefficient of determination (R^2) (0.80; 0.85; 0.90; 0.95; 0.99), was calculated by the following formula:

$$\eta_0 = \frac{R^2(1-r)}{(1-R^2)r},$$ where η_0= number of necessary measurements to predict the real value; R^2 = coefficient of determination; r = repeatability coefficient determined by one of the methods.

The coefficient of determination (R^2) was computed as:

$$R^2 = \frac{\eta r}{1 + r(\eta - 1)}$$

All statistical analyses were performed using the software Genes (Cruz 2001).

RESULTS AND DISCUSSION

The mean squares of the analyses of variance using the model with two factors of variation, years and progenies are shown in Table 1. Significant differences among the progenies were detected at 1% probability, indicating the existence of genetic variability, prior condition for the selection of superior genotypes in the breeding program. The coefficient of experimental variation was 31.63%, considered acceptable for coffee, especially when environmental effects are accumulated over nine years of harvest.

The estimates of the repeatability coefficients (r) and respective coefficients of determination (R^2) for coffee cherry yield in nine crop years, obtained by four different methods, are given in Table 2. According to the method the repeatability values differed. The value of the repeatability coefficient estimated by ANOVA was lowest (0.26) while the highest was obtained by the principal components, based on the covariance matrix (0.63). The coefficients obtained by the methods of principal components, based on the correlation matrix, and by structural analysis, based on the correlation matrix, were 0.41 and 0.35, respectively. These results agree with Fonseca et al. (2004), who evaluated clones of robust coffee (*Coffea canephora* P.) and stated that the highest repeatability coefficient was estimated by

Table 1. Analysis of variance for the coffee cherry yield of 21 Icatu progenies of Arabic coffee evaluated during nine years at the Instituto Agronômico in the state of São Paulo (APTA/IAC)

Sources of variation	df	MS
Years	8	56516248.9167
Progenies	20	3365726.3238**
Error	160	808114.6375
Mean (g)	2.160	
CV (%)	31.630	

**significant at 1% probability

principal components based on the covariance matrix (0.52), and the lowest by ANOVA (0.32). For pine, Cornacchia et al (1995), and for tomato Cargnelutti Filho et al. (2004) also verified that the estimate of the repeatability coefficient by the principal components, based on the covariance matrix, was higher, compared to the methods of analysis of variance and structural analysis. On the other hand, in elephant grass (Shimoya et al. 2002), cacao (Dias and Kageyama 1998) and sugarcane (Ferreira et al. 2005) no differences were found between the estimated repeatability coefficients obtained by these methods. These results confirm the need of using the different available methods to estimate this coefficient to obtain a precise interval, within which the real value of this parameter is most likely to be found. The definition of the most appropriate method is relevant, since each method considers particularities related to the different phenotypic manifestations, in function of the species and the trait studied.

In many species, as in the case of coffee, yields after the fourth or fifth harvest are alternately high and low, due to the weakening of the plants in the years of higher yields. This alternation between yields is designated biennial bearing. Climatic adversities, such as drought or frost, can anticipate biennial bearing. According to Cruz and Regazzi (1997) biennial bearing may vary in mode and intensity between genotypes and ANOVA, used to estimate the repeatability coefficient, cannot eliminate this additional component from the experimental error and, consequently, the estimator of repeatability would be underestimated, as verified here.

The mean values of the coefficients of determination (R^2), which demonstrate the reliability of a phenotypic value to predict the real value of genotypes, were over 80% (92.74 to 83.00%), except by ANOVA (76%). Fonseca et al. (2004) stated R^2 of 65% by ANOVA, 75% by the principal components based on the correlation matrix and 82%, using the method of principal components based on the covariance matrix. According to Resende (2002) values of over 80% can be considered adequate when a group of plants is selected.

The estimates of the number of measurements required, based on five pre-established coefficients of determination, to predict the real value of the genotypes for the coffee cherry yield evaluated by the four methodologies used to estimate the repeatability coefficients are presented in Table 3. Considering the method of the principal components based on the covariance matrix and a coefficient of determination of 80%, it can be concluded that three harvests would be necessary to infer the superiority of one progeny over another. However, if a greater reliability is desired (R^2 = 85 and 90%), four and six harvests would be necessary, respectively, by the principal components method based on the covariance matrix. Values of over 90% would however make the selection time-consuming and expensive. According to Fonseca et al. (2004), by the method of principal components based on the covariance matrix, four successive harvests would be enough to select coffee robusta genotypes for grain yield, at an accuracy of 80%. Evaluating the coffee cherry yield of 72 progenies of Arabic coffee in the F_5 generation, derived from cultivar Acaiá during eight years, verified that the first three harvests were enough to predict 79% of the yield variation. In this study was possible to pointed out that the minimum period of evaluation to perform selection also depends on the generation of the progenies. These periods of selection can be shorter, in cases where the plant material has a greater genetic variability, or longer, for genotypes with greater homozygosis. This fact was confirmed here, since the Icatu progenies in question were in the F_3 generation, where genetic variability is still considerable.

Thus, in improvement programs of Arabic coffee with Icatu germplasm and in early generations, it can be concluded that three harvests, when estimated by the principal component method based on the

Table 2. Estimates of the coefficients of repeatability (r) and of determination (R^2), considering four methods, of nine years of production of 21 Icatu progenies of Arabic coffee

Method	r	R^2 (%)
ANOVA	0.26	76.00
Principal components based on the covariance matrix	0.63	92.74
Principal components based on the correlation matrix	0.41	86.13
Structural analysis based on the correlation matrix	0.35	83.00

Table 3. Number of harvests (η), considering four methods to estimate the repeatability coefficient associated to the different coefficients of determination (R^2) required for the selection of Icatu progenies of Arabic coffee

R^2	Method	η
80	ANOVA	11
	Principal components based on the covariance matrix	3
	Principal components based on the correlation matrix	6
	Structural analysis based on the correlation matrix	7
85	ANOVA	16
	Principal components based on the covariance matrix	4
	Principal components based on the correlation matrix	8
	Structural analysis based on the correlation matrix	10
90	ANOVA	26
	Principal components based on the covariance matrix	6
	Principal components based on the correlation matrix	13
	Structural analysis based on the correlation matrix	17
95	ANOVA	54
	Principal components based on the covariance matrix	13
	Principal components based on the correlation matrix	27
	Structural analysis based on the correlation matrix	35
99	ANOVA	281
	Principal components based on the covariance matrix	70
	Principal components based on the correlation matrix	143
	Structural analysis based on the correlation matrix	182

covariance matrix, the discrimination of the most productive genotypes is possible with a probability of 80% of being based on the real value of the tree. The application of a lower number of harvests is desirable in terms of saving time and labor, provided that the success of the program of genetic breeding of the crop is not affected.

CONCLUSIONS

The differences in the values of the estimates of the repeatability coefficients indicate the importance of determining the most appropriate method, since each method considers particularities related to the species and the trait studied.

The estimate of the repeatability coefficient by principal components based on the covariance matrix was most appropriate, due to the minimization of the effect of biennial bearing on yield.

The first three harvests were enough to select the best Icatu progenies of Arabic coffee in the 'F_3RC_2' generation, with an accuracy of 80%, to predict the real value of the plant.

ACKNOWLEDGEMENTS

The authors thank the CNPq for an award for Excellence in Research and the Consórcio Brasileiro de Pesquisa e Desenvolvimento de Café for the partial funding of the research project.

Determinação do número de anos na seleção de progênies de café arábica através da repetibilidade

RESUMO - *O presente trabalho teve como objetivos estimar o coeficiente de repetibilidade da característica produção de grãos em café arábica e determinar o número mínimo de colheitas capaz de proporcionar a predição do valor real do indivíduo com nível de certeza aceitável. O experimento, conduzido durante nove anos de colheita em Campinas (SP), em delineamento de blocos ao acaso, com 21 progênies F_3RC_2 do germoplasma Icatu e nove repetições. O coeficiente de repetibilidade foi estimado mediante a utilização dos métodos análise de variância, análise dos componentes principais* *baseada na matriz de covariância e de correlações e na análise estrutural baseada na matriz de correlações. É possível realizar com eficácia de 80% a seleção de progênies F_3RC_2 do germoplasma Icatu utilizando-se os três primeiros anos de produção. Os componentes principais, baseado na matriz de covariâncias, foi a melhor metodologia para a estimativa do coeficiente de repetibilidade, por minimizar o efeito da bienalidade.*

Palavras-chave: análise de variância, componentes principais, coeficiente de determinação, melhoramento, *Coffea arabica*

REFERENCES

Abeywardena V (1972) An application of principal component analysis in genetics. **Journal of Genetics 61**: 27-51.

Cargnelutti Filho A, Radin B, Matzenauer R and Storck L (2004) Número de colheitas e comparação de genótipos de tomateiro cultivados em estufa de plástico. **Pesquisa Agropecuária Brasileira 39**: 953-959.

Cornacchia G, Cruz CD and Pires IE (1995) Estimativa do coeficiente de repetibilidade para características fenotípicas de procedências de *Pinus tecunumanii* (Schw.) Eguiluz & Perry e *Pinus caribaea* var. hondurensis Barret & Golfari. **Revista Árvore 19**: 333-345.

Cruz CD (2001) **Programa Genes: aplicativo computacional em genética e estatística**. UFV, Viçosa. 648p.

Cruz CD and Regazzi AJ (1997) **Modelos biométricos aplicados ao melhoramento genético**. 2nd ed., UFV, Viçosa, 390p.

Dias LAS and Kageyama PT (1998) Repeatability and minimum harvest period of cacao (*Theobroma cacao* L.) in Southern Bahia. **Euphytica 102**: 29-35.

Falconer DS (1987) **Introdução à genética quantitativa.** UFV, Viçosa, 279p.

Ferreira A, Barbosa MHP, Cruz CD, Hoffmann HP, Vieira MAS, Bassinello AI and Silva MF (2005) Repetibilidade e número de colheitas para a seleção de clones de cana-de-acúcar. **Pesquisa Agropecuária Brasileira 40**: 761-767.

Ferreira RP, Botrel MA, Pereira AV and Cruz CD (1999) Avaliação de cultivares de alfafa e estimativas de repetibilidade de caracteres forrageiros. **Pesquisa Agropecuária Brasileira 34**: 995-1002.

Fonseca AFA, Sediyama T, Cruz CD, Sakiyama NS, Ferrão RG, Ferrão MAG and Bragança SM (2004) Repeatability and number of harvests required for selection in robusta coffee. **Crop Breeding and Applied Biotechnology 4**: 325-329.

Gonçalves PS, Cardoso M and Saes LA (1990) Estimativas de repetibilidade na seleção de árvores adultas de seringueira. **Pesquisa Agropecuária Brasileira 25**: 1031-1038.

Mansour H, Nordheim EV and Rutledge JJ (1981) Estimators of repeatability. **Theoretical and Applied Genetics 60**: 151-156.

Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa Informação Tecnológica, Brasília, 975p.

Rutledge JJ (1974) A scaling which removes bias of Abeywardena's estimator of repeatability. **Journal of Genetics 61**: 247-250.

Sera T (2001) Coffee genetic breeding at IAPAR. **Crop Breeding and Applied Biotechnology 1**: 179-200.

Shimoya A, Pereira AV, Ferreira RP, Cruz CD and Carneiro PCS (2002) Repetibilidade de características forrageiras do capim-elefante. **Scientia Agrícola 59**: 227-234.