PRINCIPAL COMPONENT ANALYSIS ON MORPHOLOGICAL TRAITS IN JUVENILE STAGE ARABICA COFFEE

Alexsandro Lara Teixeira[1], Flávia Maria Avelar Gonçalves[2], Juliana Costa de Rezende[3], Rodrigo Barros Rocha[4], Antonio Alves Pereira[5]

(Recebido: 7 de fevereiro de 2012; aceito: 4 de julho de 2012)

ABSTRACT: In plant breeding and selection, it is necessary to evaluate a great number of traits that are naturally associated. Thus, the use of multivariate analysis, such as Principal Component Analysis, can assist data interpretation. The aim of this study was to identify which morphological traits are of greatest importance to discriminate juvenile *Coffea arabica* genotypes. Cultivars, hybrids and some wild genotypes were evaluated from a coffee genebank maintained by the Empresa de Pesquisa Agropecuária de Minas Gerais, in Patrocínio, MG, Brazil. A randomized block experimental design was used with two replicates, spacing of 3.5 x 0.8 meters and ten plants per plot. The morphological characters were measured at 12 months after the plantation, at the plants juvenile phase. Only the first two major components explained 80.04% of total variance. Among the morphological evaluated traits, the length of the first plagiotropic branch, the plant vitality, the stem diameter and the nodes number of the first plagiotropic branch are the most important variables to discriminate the genotypes. These evaluations are important to the early selection of the genotypes of higher potential allowing to concentrate efforts in their evaluating, discarding other minor traits.

Index terms: Early selection, multivariate, coffee breeding.

ANÁLISE DE COMPONENTES PRINCIPAIS EM CARACTERES MORFOLÓGICOS DE CAFÉ ARÁBICA EM ESTÁDIO JUVENIL

RESUMO: *Para a seleção de plantas que reúnam uma série de características favoráveis, faz-se necessário a avaliação um grande número de variáveis, que naturalmente estão associadas. Desse modo, o emprego da análise multivariada, como a análise de Componentes Principais, pode ser útil na interpretação dos dados. Objetivou-se,no presente estudo, identificar quais são os caracteres morfológicos de maior importância na discriminação entre genótipos juvenis de Coffea arabica. Foram avaliados 250 acessos (cultivares, híbridos e alguns genótipos selvagens) oriundos do banco de germoplasma de café, instalado na fazenda experimental da Empresa de Pesquisa Agropecuária de Minas Gerais, em Patrocínio, MG. O delineamento utilizado foi o de blocos casualizados com duas repetições, espaçamento de 3,5 x 0,8 metros, com parcelas de dez plantas. Os caracteres morfológicos foram avaliados em 2006, 12 meses após a implantação da cultura, com as plantas ainda na fase juvenil. Por meio dos resultados constatou-se que os dois primeiros componentes principais explicaram 80,04% da variação total. Dentre as características morfológicas avaliadas, o comprimento do 1º ramo plagiotrópico, vigor, diâmetro do caule e número de nós do 1º ramo plagiotrópico são as variáveis de maior importância na distinção dos acessos. Essa informação permite uma maior concentração de esforços na avaliação das mesmas, descartando outras variáveis de menor importância.*

Termos para indexação: *Seleção precoce, multivariada; melhoramento do cafeeiro.*

1 INTRODUCTION

The search for new cultivars that hold a number of favorable characteristics has been the main focus of coffee breeding programs in Brazil (CARVALHO et al., 2010; MARTINEZ et al., 2007). Besides the direct selection based on productivity, other strategies have been used to maximize the gains from selection. This includes early assessment of morphological characters in order to discriminate and identify the most promising genotypes characteristics of greatest importance in the characterization of genetic materials. These tools provide breeders to better guidance in choosing the best criteria to be used in breeding programs (CRUZ; REGAZZI; CARNEIRO, 2004).

In coffee culture, it is common to evaluate a large number of variables, there is often significant correlations between them. Thus, the use of multivariate analysis such as principal component analysis, is useful for the interpretation of the data. The result of this technique is a graph representing the variation in multiple characteristics of only two Cartesian axes. The closer two points in the graph greater the similarity between them and the greater the proximity of two samples means

[1]Embrapa Rondônia - BR 364 - Km 5,5 - Zona Rural - Cx.P. 127- 76815-800 - Porto Velho-RO alexteixeira@cpafro.embrapa.br
[2]Universidade Federal de Lavras/ UFLA - Departamento de Biologia/DBI - Cx.P. 3037 - 37200-000 Lavras-MG - avelar@dbi.ufla.br
[3]Empresa de Pesquisa Agropecuária de Minas Gerais/EPAMIG - Cx. P. 176 - 37200-000 - Lavras -MG julianacosta@epamig.ufla.br
[4]Embrapa Rondônia - BR 364 - Km 5,5 - Zona Rural - Cx.P. 127 - 76815-800 - Porto Velho-RO rodrigo@cpafro.embrapa.br
[5]Epamig, Unidade Regional da Zona da Mata, Campus da UFV S/N - Cx.P. 216 - 36570-000 -Viçosa-MG pereira@epamig.ufv.br

more similarity between them, variables in relation to the studied variables (RICCI; COSTA; OLIVEIRA, 2011).

The principal components analysis is used as a criterion to judge the importance of their own original variables chosen, ie, the original variables with more weight in the linear combination of the first principal components are the most important from the statistical point of view (MOITA NETO; MOITA, 1998).

Despite being a relatively old, only recently, with the development of faster computer processors, is that multivariate analysis could be used routinely and incorporated into the data analysis (HAIR et al., 1998).

The objective of this work was to identify what are the most important morphological characters to discriminate genotypes juvenile Arabica coffee, with the aid of the technique of principal component analysis.

2 MATERIALS AND METHODS

The experiment was conducted in the municipality of Patrocínio, MG, located at coordinates 18 ° 56'38 "S and 46 º 59'34" W. The climate is classified as tropical of altitude, Cwa (Köppen), with an average annual temperature of 20.2 º C, average rainfall of 1,620 mm, with 65% to 70% of the total concentrated during the period from December to March, and average altitude of 972 m.

We evaluated 250 accessions of Coffea arabica L., represented by cultivars, F1 hybrids (obtained from crosses between elite lines) and some wild genotypes introduced from other countries, coming from the genebank coffee, installed in 2005, at the experimental farm of Empresa de Pesquisa Agropecuária de Minas Gerais, in Patrocínio, MG. The experimental design was a randomized block design with two replications, plots of ten plants and spacing of 3.5 x 0.8 meters. Morphological characters were evaluated in 2006, 12 months after implantation of culture, with plants still in the juvenile phase. The experiments were conducted in accordance with the recommendations of fertilizer for coffee plantations. Were adopted management practices commonly employed in the culture.

The traits were: a) vegetative vigor [VIGOR], assigning grades as arbitrary scale of 10 points, and the note one, given the worst plants with reduced vigor and very pronounced symptom of impoverishment and note 10 to the plants with great force, and markedly more enfolhadas vegetative growth of productive branches, as suggested by Carvalho, Mônaco and Fazuoli (1979); b) plant height [ALT] measured from the plant to the apical stem in centimeters, c) number of pairs plagiotropic [NP (RP)], evaluated by counting all pairs of primary lateral branches who had longer than five centimeters, d) length of the first plagiotrophycal [COMP (1 RP)], evaluated by measuring the first plagiotrophycal above the top of the plant in centimeters, e) number of nodes of the first plagiotrophycal [NN (1 RP)], obtained by counting all the nodes of the branch f) length of the fourth pair of leaves [COMP (4th FP)], in centimeters; g) width of the fourth pair of leaves [wIDTH (4th FP)], in centimeters.

The data relating to morphological characters and yield were subjected to ANOVA assumptions, when it was checked the data for normality by Lilliefors test and homogeneity of variance by Bartlett's test, both at the level of 5% probability. Subsequently, all patients were submitted to analysis of variance with the significance of the effects observed by F test at 5% probability. The selective accuracy () determined by the expression: = (1-1 / f) 1/2, where F is the value of the F test for the effect of Snedecor genotype (RESENDE; DUARTE, 2007). Analyses of variance were performed on the statistical software SAS (Statistical Analysis System). The genotypic correlations and principal components analysis were performed using the software GENES (CRUZ, 2006).

3 RESULTS AND DISCUSSION

All traits showed significant differences between genotypes ($p \leq .05$). Estimates of accuracy were high magnitude ($65 < \hat{r}_{gg} < 90$) for all the variables in question, indicating good experimental precision. The use of accuracy as a measure of experimental precision, suggested by Resende and Duarte (2007), has the advantage of not depending on average, which provides greater safety in the use of phenotypic expression as an indicator of genotypic variation. Accuracy values above 70% indicate a high experimental precision.

Estimates of genotypic correlation (r g) among the eight variables are shown in Table 1. The highest correlation was between COMP (4th FP) and LARG (4th FP) (0,88). The VIGOR, feature commonly used as selection criteria, also showed high correlation with NP (RP), COMP (1 RP) and NN (1 RP) (Table 1).

TABLE 1 - Estimation of genotypic correlation coefficients between eight morphological characters *Coffea arábica* Patrocínio - MG, 2010.

Variables[#]	VIGOR	ALT	DIAM	NP (RP)	COMP (1ºRP)	NN (1ºRP)	COMP (4ºPF)	LARG (4ºPF)
VIGOR	1,00	0,48	0,74	0,76	0,68	0,73	0,52	0,60
ALT		1,00	0,65	0,47	0,73	0,51	0,45	0,39
DIAM			1,00	0,80	0,76	0,80	0,46	0,46
NP(RP)				1,00	0,63	0,80	0,29	0,32
COMP(1ºRP)					1,00	0,70	0,46	0,48
NN(1ºRP)						1,00	0,29	0,38
COMP(4ºPF)							1,00	0,88
LARG(4ºPF)								1,00

All correlations were significant at 1% by the bootstrap with 500 simulations.

[#]VIGOR: vegetative vigor (score 1-10), ALT: plant height (cm); DIAM: crown diameter (cm), NP (RP): number of pairs of reproductive branches (units), COMP (1 RP): length first plagiotrophycal (cm), NN (1 RP): number of nodes of the first plagiotrophycal (unit), COMP (4 FP): length of the fourth pair of leaves (cm), wIDTH (4 FP): width of the fourth pair of leaves (cm).

Relying only on photosynthetic capacity and without direct assessment of grain production, Freitas et al. (2007) found that early selection in length plagiotropic at 12 months of age can be used to identify superior genotypes. Similar results were found by Severino et al. (2002) reporting an association between vegetative vigor and yield in data accumulated during the first three years of production (r = .90). It is also evident that the number of characters we plagiotrophycal the first, number of pairs of reproductive branches and stem diameter are highly correlated.

Reviews morphological characteristics and their correlation estimates are important for breeding programs, since they help the breeder in selecting and discarding early hits. However it is not always possible to evaluate all the desired features, since it is a job that requires costly time and skilled labor. Thus, the characterization of the most important variables from those with smaller contribution to the selection, the best use of resources in the breeding program, allowing to efficiently discriminate genotypes in juvenile stage using as few variables as possible.

Estimates of the eigenvalues with their respective individual and cumulative variances are presented in Table 2 . The first two principal components explained 80.04 % of the total variance .

This result can be considered good , since the second and cross Regazzi (2001) variations in total above 80% obtained with the two or three first principal components analysis allow genotype groups using scatterplots . The dispersion , using only the values of the first two principal components represented in Figure 1. Bosselmann et al. (2009) evaluated morphological, agronomic and some sensory attributes in Arabica coffees grown in agroforestry systems in southern Colombia. Through the results , it was observed that the first two components (CP1 - CP2 and sensory attributes - elevation and grain size) were not sufficient to discriminate genotypes , explaining only 42 % of the variation found in the data. One possible explanation is that within the same component , there was a low correlation between the variables , altitude and percentage of small grains . In general , sensory attributes are influenced by shading and physical attributes are influenced by altitude.

The most important variables in the first vector, which explains 63.80% of the total variability, were, in descending order, COMP (1 RP), VIGOR, NP (RP), and negative values variables COMP (4th PF), DIAM, followed by the others (Table 3).

TABLE 2 - Estimates of the variance of the eigenvalues, percentage of variance and cumulative variance of the principal components, obtained eight morphological characters *Coffea arabica* Patrocínio - MG, 2010.

Canonic Variable	Autovalor	Explanation (%)	Cumulative explanation (%)
1	5.1041	63.8016	63.8016
2	1.2994	16.2421	80.0438
3	0.6916	8.6446	88.6884
4	0.2595	3.2436	91.9321
5	0.2273	2.8416	94.7737
6	0.1790	2.2377	97.0115
7	0.1472	1.8403	98.8517
8	0.0919	1.1483	100.00

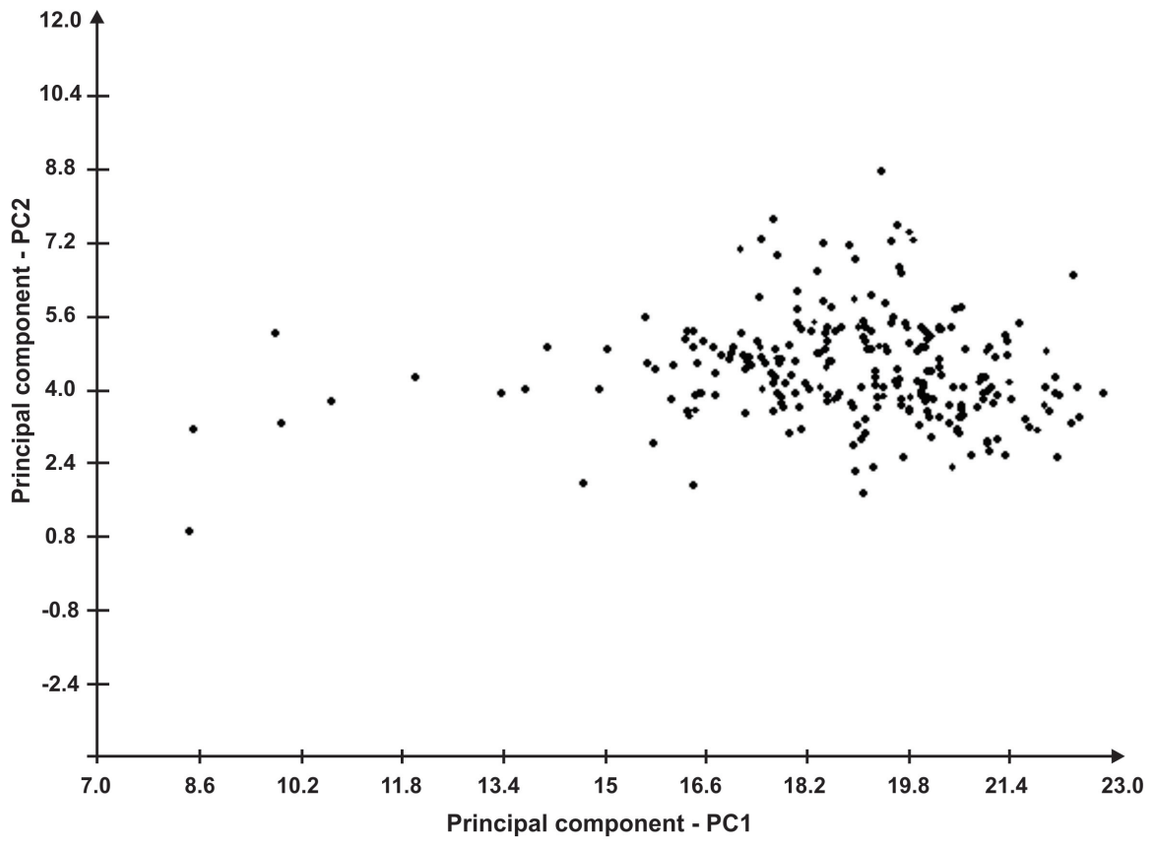

FIGURE 1 -Two-dimensional scatter plot obtained with the values of the first two components using 250 accesses *Coffea arabica* Patrocínio - MG, 2010.

TABLE 3 - Estimates of autovectors (Main components - CP) and the respective wieghts of each variable for its constitution. Patrocínio - MG, 2010.

Variables[#]	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
VIGOR	0.3859	0.3239	0.3996	0.3603	0.3824	0.3697	0.2886	0.3009
ALT	-0.0256	-0.0073	-0.1794	-0.3404	-0.0897	-0.3184	0.627	0.5905
DIAM	-0.3558	0.7636	0.0156	-0.3066	0.3466	-0.1909	-0.0301	-0.1963
NP(RP)	0.3508	-0.2201	-0.4101	-0.3997	0.626	0.132	-0.2779	0.1189
COMP(1ºRP)	0.6055	0.2082	-0.1969	0.288	-0.0472	-0.6531	-0.0252	-0.1975
NN(1ºRP)	-0.2763	-0.4591	0.385	0.1605	0.5396	-0.4513	0.1582	-0.1376
COMP(4ºPF)	-0.3784	0.0509	-0.6388	0.6156	0.1955	0.0712	0.0718	0.1369
LARG(4ºPF)	-0.1205	0.082	0.2117	0.1136	-0.0461	-0.2698	-0.6439	0.6569

All correlations were significant at 1% by the bootstrap with 500 simulations.

[#]VIGOR: vegetative vigor (score 1-10), ALT: plant height (cm); DIAM: crown diameter (cm), NP (RP): number of pairs of reproductive branches (units), COMP (1 RP): length first plagiotrophycal (cm), NN (1 RP): number of nodes of the first plagiotrophycal (unit), COMP (4 FP): length of the fourth pair of leaves (cm), wIDTH (4 FP): width of the fourth pair of leaves (cm).

That is, as among the highest values obtained (positive and / or negative) are the variables COMP (1 RP) and VIGOR, the main component differentiates efficiently, clusters due to characteristics related to the size of plant. The second vector (Principal Component 2 - CP2) explains 16.24% of the variability of the data and the highest values (positive and / or negative) were obtained with the variables DIAM and NN (1 RP) (Table 3).

Through the results it can be seen that the COMP (1 RP), VIGOR, DIAM and NN (1 RP) are the most important variables in discriminating between genotype Arabica coffee still in juvenile stage. With this information one can concentrate his effort in evaluating them, discarding the evaluation of other variables that contribute little to separate the materials.

These results highlight the importance of using multivariate techniques to identify characters, which must be evaluated based on a prior study of their contribution to the variability (CRUZ; REGAZZI; CARNEIRO, 2004; SILVA; SBRISSIA, 2010). Interestingly, these analyzes lies in the possibility of disposal of characters that contribute little to the discrimination of genotypes (Pereira et al., 2010).

This is associated with the possibility of distinguishing genotypes early, still in the juvenile phase, streamlines breeding programs. In the early stages of the selection process, the number of genotypes to be evaluated is very large requiring much time and labor. Therefore, the fewer variables can be assessed and sooner it is possible to discriminate the genotypes, the more efficient it becomes the breeding program, maximizing profits and reducing costs and labor.

4 CONCLUSIONS

The model used for multivariate analysis allowed to identify efficiently among all morphological characteristics studied, the length of the 1st plagiotrophycal, vigor, stem diameter and number of nodes of the 1st plagiotrophycal are the most important variables and considered sufficient to genotype distinction in arabica coffee, still in juvenile stage.

5 THANKS

To EPAMIG and UFLS for making the equipment and infrastructure available and to the promotion agencies Capes, CNPq, Fapemig and Secretaria da Ciência e Tecnologia do Estado de Minas Gerais, for the financial possibility of this project.

6 REFERENCES

BOSSELMANN, A. S. et al. The influence of shade trees on coffee quality in small holder coffee agroforestry systems in Southern Colombia. **Agriculture Ecosystems & Environment**, Amsterdam, v. 129, n. 1/3, p. 253-260, Jan. 2009.

CARVALHO, A.; MÔNACO, L. C.; FAZUOLI, L. C. Melhoramento do café XL: estudos de progênies e híbridos de café Catuaí. **Bragantia**, Campinas, v. 38, n. 22, p. 202-216, set. 1979.

CARVALHO, A. M. D. et al. Correlação entre crescimento e produtividade de cultivares de café em diferentes regiões de Minas Gerais, Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 3, p. 269-275, mar. 2010.

CRUZ, C. D. **Programa Genes:** estatística experimental e matrizes. Viçosa, MG: UFV, 2006. 285 p.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, MG: UFV, 2001. 390 p.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. Divergência genética. In: _____. **Modelos biométricos aplicados ao melhoramento**. Viçosa, MG: UFV, 2004. v. 1, p. 377-413.

FREITAS, Z. M. T. S. et al. Avaliação de caracteres quantitativos relacionados com o crescimento vegetativo entre cultivares de café arábica de porte baixo. **Bragantia**, Campinas, v. 66, n. 2, p. 267-275, abr. 2007.

HAIR, J. F. et al. **Multivariate data analysis**. Englewood Cliffs: Prentice-Hall, 1998. 730 p.

MARTINEZ, H. E. P. et al. Crescimento vegetativo de cultivares de café (*Coffea arabica* L.) e sua correlação com a produção em espaçamentos adensados. **Acta Scientiarum Agronomy**, Maringá, v. 29, n. 4, p. 481-489, out. 2007.

MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, São Paulo, v. 21, n. 6, p. 467-469, nov. 1998.

PEREIRA, M. C. et al. Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica L.*) in the region of upper Paranaiba. **Acta Scientiarum Agronomy**, Maringá, v. 32, n. 4, p. 635-641, out. 2010.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Goiânia, v. 37, n. 3, p. 182-194, set. 2007.

RICCI, M. D. S. F.; COSTA, J. R.; OLIVEIRA, N. G. D. Utitlização de componentes principais para analisar o comportamento do caffeiro a pleno sol e sombreado. **Coffee Science**, Lavras, v. 6, n. 1, p. 44-54, jan. 2011.

SEVERINO, L. S. et al. Associações da produtividade com outras características agronômicas de café (*Coffea arabica* L. "Catimor"). **Acta Scientiarum Agronomy**, Maringá, v. 24, n. 5, p. 1467-1471, dez. 2002.

SILVA, S. C. D.; SBRISSIA, A. F. Análise de componentes principais entre características morfogênicas e estruturais em capim-marandu sob lotação contínua. **Ciência Rural**, Santa Maria, v. 40, n. 3, p. 690-693, mar. 2010.