DISCRIMINATION OF VARIETAL GROUPS AND HYBRIDS OF COFFEA CANEPHORA SPECIES USING MULTIVARIATE ANALYSIS

Gabriela França OLIVEIRA[1]
Taiana Lopes Rangel MIRANDA[1]
Ana Carolina Campana NASCIMENTO[1]
Moysés NASCIMENTO[1]
Eveline Teixeira CAIXETA[2]
Letícia de Faria SILVA[3]
Emilly Ruas ALKIMIM[4]
Felipe Lopes da SILVA[5]

- ABSTRACT: Coffee growing is one of the most important agricultural activities in the world market. Coffea canephora is one of the most commercially relevant species, which can be divided into the varietal groups Conilon and Robusta. These varietal groups have complementary agronomic interests. Therefore, hybrids are obtained through crosses between these groups. Given the difficulty in differentiating the two varietal groups of genotypes in the field, the correct discrimination is essential for the definition of crosses in breeding programs. In this context, the present work aimed to apply a discriminant analysis (DA) to define functions to differentiate between varietal groups and hybrids of C. canephora, and identify the most relevant phenotypic traits in these functions. Data from 165 genotypes from the Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural e do Centro Agronómico Tropical de Investigación y Enseñanza were used to measure different plant traits. It was applied the quadratic DA with the best performance for genotype discrimination, with an average apparent error rate of 0.0364. Cercosporiose incidence, rust incidence, vegetative vigor, plant height and diameter of the canopy projection were the most important traits in the varietal groups' discrimination.

- KEYWORDS: Discriminant analysis; coffee; Conilon; Robusta.

[1] Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: *gabriela.franca@ufv.br; taiana.miranda@ufv.br; ana.campana@ufv.br; moysesnascim@ufv.br*

[2] Empresa Brasileira de Pesquisa Agropecuária - EMRAPA, Embrapa Café, CEP: 70770-901, Brasília-DF, Brasil. E-mail: *eveline.caixeta@embrapa.br*

[3] Universidade Federal de Viçosa - UFV, Departamento de Biologia Geral, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: *leticiafaria785@gmail.com*

[4] Universidade Federal do Triangulo Mineiro - UFTM, Campus de Iturama, Departamento de Agronomia, CEP: 38280-000, Iturama, MG, Brasil, E-mail: emilly.alkimim@uftm.edu.br

[5] Universidade Federal de Viçosa - UFV, Departamento de Agronomia, CEP: 36570-900, Viçosa, MG, Brasil. E-mail*: felipe.silva@ufv.br*

1 Introduction

Coffee growing is one of the most important agricultural activities in the world market (CENCI *et al.*, 2012). Brazil is the largest coffee producer and exporter in the world (BARBOSA *et al.*, 2019; VOLSI *et al.*, 2019).

There are approximately 124 species of the genus Coffea L., but only two are commercially exploited, Coffea arabica L. and Coffea canephora Pierre ex A. Froehner (DAVIS, 2011; RIBEIRO *et al.*, 2014). C. arabica is responsible for the aroma and sweet flavor and provides a better-quality drink. C. canephora has large amounts of caffeine and soluble solids and gives thickness to the drink (BABOVA *et al.*, 2016), which makes this species the main raw material in the production of instant coffee (CORSO *et al.*, 2016). C. canephora is also very consumed in the form of mixtures with the arabica coffee (blends), whose importance and cultivation has increased in recent years.

Due to its large natural geographic distribution and allogamous nature, C. canephora presents wide genetic variability. The study of this available variability has unquestionable importance for the management and genetic breeding of this crop (GOMEZ *et al.*, 2009). Characterization and evaluation of the C. canephora gene pool are also necessary for effective conservation strategies (HUDED *et al.*, 2020).

Berthaud (1986) described for the first time the C. canephora diversity and identified two major distinct genetic groups, Guinean and Congolese. Due to the domestication process, the Congolese group was divided into five subgroups, named SG1, SG2, B, C and Uganda (MUSOLI *et al.*, 2009; LABOUISSE *et al.*, 2020). In Brazil, coffees from two Congolese subgroups were introduced, namely, Conilon (subgroup SG1) and Robusta (subgroup SG2).

The Conilon varietal group comprehends the genotypes with the best adaptation in Brazil, with smaller leaves, less vigor and smaller size, small fruits, tolerance to drought and, generally, susceptibility to rust. The coffee trees of the Robusta varietal group are taller, vigorous plants, with larger leaves and fruits, tolerance to coffee rust, but greater sensitivity to drought (MARRACCINI *et al.*, 2012). Given the complementarity of these groups, they are crossed for the generation of hybrids.

Although the varietal groups exhibit distinct phenotypic traits, the discrimination of these materials in the field is not an easy task. Such limitation is due to the allogamous coffee reproduction, high phenotypic and heterozygous amplitude and the likely occurrence of natural crosses between the two varietal groups (FERRÃO *et al.*, 2013; REN *et al.*, 2013; FERRÃO *et al.*, 2015). Thus, the correct discrimination between these groups is essential for crop management and the use of genotypes in breeding programs (ALKIMIM *et al.*, 2020).

The discrimination of Conilon, Robusta or hybrid genotypes, based on phenotypic traits, can be performed using the multivariate statistical technique, known as discriminant analysis (DA). This technique is useful for the differentiation and/or classification of individuals in a population within two or more previously known groups. For this purpose, rules based on the group traits are elaborated, which will be used for the classification of new individuals in the existing groups (MINGOTI, 2005).

The potential of the discriminating function has been successfully explored in the literature. Nogueira *et al.* (2008) applied discriminant analysis to identify and evaluate new

characteristics for the differentiation of soybean cultivars, and observed an apparent error rate of 16.36%. More recently, Das *et al.* (2018) applied quadratic discriminant analysis to discriminate rice genotypes, and seventeen wavelengths were evaluated, with an apparent error rate of 2%. In this context, the discriminant analysis has been successfully applied to discriminate genotypes from different species. However, this methodology has not been applied to discriminate coffee genotypes.

This study aimed to apply discriminant analysis (DA) to define functions for the differentiation and classification of C. canephora varietal and hybrid groups and identify the most important phenotypic traits for the discrimination of these genotypes.

2 Materials and methods

The data used in this study were obtained from Conilon and Robusta varietal groups and hybrid families originated from crosses between these groups. Conilon genetic material was provided by the Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) and Robusta material was obtained from the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). These genetic materials belong to the breeding program of the Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), in association with the Universidade Federal de Viçosa (UFV) and the Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café), located in the city of Oratórios/MG.

The genotypes were evaluated for seven phenotypic traits. The evaluations were performed at physiological maturity of the coffee fruits. Five categorical traits were evaluated: vegetative vigor (Vig), field evaluation of rust incidence (Rus), cercosporiosis incidence (Cer), fruit maturation time (Mat) and fruit size (FS). The continuous traits evaluated were the following: plant height (PH) and diameter of the canopy projection (DC). The procedures for measuring phenotypic traits were described by Alkimim *et al.* (2020).

The varietal groups Conilon and Robusta were composed of 45 and 37 genotypes, respectively, besides 83 interpopulation hybrids.

For the discrimination of Conilon, Robusta or hybrid genotypes, based on phenotypic traits, it was used the multivariate statistical technique, known as discriminant analysis (DA). DA is a technique used to classify and/or differentiate individuals from a sample or population (BARROSO *et al.*, 2013). It is also used to classify a new observation into one of k different groups, based on functions of the observed variables, which aim to minimize the likelihood of poor classification. Unlike other multivariate approaches, in the discriminant analysis, the groups in which individuals are classified must be previously known. Thus, the rules of discrimination or classification are generated based on the characteristics of the groups (MINGOTI, 2005).

In this work, two types of discriminant functions were evaluated: Fisher's linear (1936) and quadratic.

In order to establish the classification of genotypes in one of the three populations, based on Fisher's linear discriminant function, each population π_i is considered to have a vector of means μ_i and a matrix of variances and homogeneous covariance $\left(\sum_1 = \sum_2 = \sum_3 = \cdots = \sum_p \right)$. According to Ferreira (2018), one should classify *x* in the population π_i

for which the value $D_i(X_i) = \mu_i' \sum_{p}^{-1} x - \frac{1}{2} \mu_i' \sum_{p}^{-1} \mu_i$, be the maximum in relation to all possible values of *i*. Since three groups were considered in this study (Conilon, Robusta and hybrids), we have that $i = \{1,2,3\}$.

On the other hand, to establish the classification of genotypes in one of the three populations, using the functions of quadratic discriminants, it is considered that each population π_i has vector of means μ_i, $\sum_i$ the heterogeneous variance and covariance matrix $(\sum_1 \neq \sum_2 \neq \sum_3)$ and p_i the a priori probability of the genotypes belonging to the population π_i. The quadratic discriminant functions were generated according to Varella (2004): $D_i(x) = ln(p_i) + \frac{1}{2}(x - \mu_i)' \Sigma_i^{-1}(x - \mu_i)$, where $D_i(x)$ is the classification score of the i-th population, x is the vector of variables representing the characters involved in the analysis, $\sum_i$ is the matrix of variances and covariance of the population π_i and, in this study, they were considered a priori identical probabilities $p_i = 1/3$. From these functions, the i-th genotype is classified in the population π_k, k = $\{1,2,3\}$, if $D_i(X_i) = maxD_k(X_k)$.

Thus, according to the discriminate function used, the samples were classified as belonging to the population for which they obtained the highest classification score. Additionally, a multivariate analysis of variance (MANOVA) was performed, with the seven characteristics in three groups. And then, a univariate analysis of variance (ANOVA) was carried out for each characteristic evaluated, in order to verify which variables were most important for the discrimination of the groups (Conilon, Robusta and hybrid), considering a significance level of 5%.

A cross-validation method was used to avoid the underestimation of classification errors. This method allows the evaluation of the generalization capacity of a predictive model in a data set (JAMES *et al.*, 2013). The K-Folds cross-validation method was used, considering K equal to five. In the discriminant analysis, 132 genotypes were used as a training population, while the remaining 33 genotypes were considered the validation population. The discriminant functions were defined through the training population, and the genotypes of the validation population were later classified, assuming that its classification was unknown and using the functions already obtained in the training population. This procedure was repeated 5 times so that all groups were used in the validation population only once.

The evaluation of the efficiency of discriminant functions for the classification of varietal groups and hybrid was based on the apparent error rate (AER), calculated through the quotient between the number of incorrectly rated observations and the number of ratings (correct and incorrect); and also, on the proportions of incorrectly classification observations, which represent the total estimated probability of incorrect ratings.

The statistical analyses were performed in the R software (R CORE TEAM, 2020). The "MASS" package in version 7.3-51.4 and the "lda" and "qda" functions were used to carry out the discriminant analysis.

3 Results and discussions

The discriminatory capacity results of the C. canephora varietal groups and hybrids, considering the linear and quadratic discriminant functions, are shown in Table 1. The correct and incorrect classifications of the genotypes are presented, considering the sum of the results of the cross-validation and the seven traits evaluated (vegetative vigor, rust incidence, cercosporiose incidence, fruit maturation time, fruit size, plant height and canopy diameter).

Fisher's linear discriminating function is indicated for cases of nonnormality and requires the covariance matrix to be homogeneous (FERREIRA, 2008). According to the Box-M test (p-value < 0.01), the hypothesis that covariance matrices are homogeneous was rejected, and in this sense, the use of quadratic discriminant functions is recommended (MINGOTI, 2005). However, the use of the quadratic function requires the assumption of data normality. The Shapiro-Wilk multivariate normality test (p-value < 0.01) revealed that the data do not follow a multivariate normal distribution. Thus, in this study, both functions were considered in the analysis.

In Table 1, the main diagonal exhibits the correct classifications, while for the other cells, the classifications are incorrect. The identification of an incorrect classification is confirmed in the table rows. It was observed that the linear function presented 11 incorrect classifications, four belonging to the Conilon group, six to the Robusta group, and one belonging to the hybrids. When the quadratic function was applied (results in parentheses), there were six incorrect classifications, two of Conilon group, two of Robusta group and two hybrids (Table 1).

Table 1 - Summary of the classification of varietal groups and hybrids of C. canephora species, by linear and quadratic discriminant function (in parentheses)

Varietal Group	Classification according to linear and quadratic discriminant analysis (in parentheses)			Total observations
	Conilon	Robusta	Hybrids	
Conilon	41 (43)	4 (2)	0 (0)	45
Robusta	6 (2)	31 (35)	0 (0)	37
Hybrids	1 (1)	0 (1)	82 (81)	83
Apparent correct rate = 0.9333 (0.9636)				
Apparent error rate= 0.0667 (0.0364)				

The efficiency of the discrimination of functions is associated with the quantity and quality of variables observed in discrimination (CRUZ et al., 2004). The apparent error rate

(APER) is a method used for the analysis of these functions, which was 0.0667 and 0.0364 for linear function and quadratic function, respectively (Table 1). Both of them presented relatively low APER values, which indicates that the discriminant functions were efficient in distinguishing the varietal groups and hybrid. However, the data revealed that the linear function exhibited almost twice the APER value observed by the quadratic function. According to De Carvalho et al. (2018), the break in the assumption of data normality has little influence on the method efficiency when using the quadratic function. On the other hand, the lack of homogeneity of the covariance matrices seems to have influenced the results of classification of the linear function.

Table 2 exhibits the incorrect classification proportions for the two discriminating functions: linear and quadratic. Approximately 8.89% of the Conilon genotypes evaluated were classified incorrectly by the linear discriminant function, and 4.44% for the quadratic function. For Robusta, 16.22% of the genotypes were incorrectly classified when using the linear function, and 5.41%, when using the quadratic function. For hybrids, the proportion of incorrect classification was 1.21% by the linear function and 2.41%, for the quadratic function. The discrimination of Conilon and Robusta were more challenging than hybrid discrimination. This result can be explained by the reproductive system of the species (cross-fertilization), high phenotypic and heterozygous range, and possible natural crosses between the two varietal groups (FERRÃO et al., 2013; REN et al., 2013; FERRÃO et al., 2015). Although it is not an easy task, the evaluation and characterization of genetic variability and discrimination genotypes are essential to manage the crop and develop genetic breeding programs (ANAGBOGU et al., 2019).

Table 2 - Incorrect classification ratio by linear and quadratic discriminant functions

Group	Linear	Quadratic
Conilon	8.89%	4.44%
Robusta	16.22%	5.41%
Hybrids	1.21%	2.41%

Table 3 presents the coefficients of the linear discriminating functions, thus allowing to measure the relative contribution from the traits in the separation of the groups. Based on these functions, the phenotypic traits cercosporiosis incidence (Cer) and rust incidence (Rus) presented higher coefficients in absolute values. Another important variable for the differentiation of these genotypes is the vegetative vigor (Vig). Those traits provided the greatest contribution to the discrimination of the varietal and hybrid groups. Rus and Vig were previously mentioned as important traits for varietal group discrimination (FERRÃO et al., 2015). In general, Robusta coffees are resistant to rust and present higher vigor than Conilon coffees, which are susceptible to this disease. On the other hand, Conilon coffees present competitive advantages in Brazil, due to their tolerance to drought and greater adaption to the conditions of the country (SOUZA et al., 2013).

Table 3 - Coefficients of the linear discriminant functions

Traits	LD1	LD2
Vig	-0.0384	-0.1310
Rus	-0.4900	1.8147
Cer	-2.8449	-0.4221
PHl	-0.0028	-0.0060
DCo	-0.0069	0.0045
Mat	-0.0024	-0.0520
FS	-0.0009	0.0470

LD1: linear discriminant functions 1; LD2: linear discriminant functions 2.

Further analysis using MANOVA for seven traits across three groups (Conilon, Robusta and Hybrids) indicated that the groups are significantly different (p-value < 0.01). Aiming to determinate which traits are responsible for the observed difference, it was performed a univarite analysis of variance (ANOVA). ANOVA revealed significant differences between the groups for "Vig", "Fer", "Cer", "AP1", "DCo", which means that these traits are the most important to discriminate the groups (Table 4).

Table 4 – Univariate F test for characteristics under study

Variables	Vig	Fer	Cer	APl	DCo	Mat	TFr
p-value	0.01*	0.01*	0.01*	0.01*	0.01*	0.29	0.47

* significant at 0.05 probability.

Since the phenotypic variables "Mat" and "TFr" were not significant for the discrimination of populations, the discriminant functions were reestimated without considering the information related to these variables.

Table 5 summarizes the varietal and hybrid groups' classification of this new analysis, in which the main diagonals are the correct classifications, and the incorrect ones are in the other cells. It was observed that, in this case, linear function resulted in 10 misclassifications, 5 of which belonged to the Conilon group; 4, to the Robusta group; and 1, to the hybrids (Table 5). When the quadratic function was used, 11 incorrect classifications occurred; 5 belonged to the Conilon group; 4, to the Robusta group; and 2, to the hybrids (Table 5).

Table 5 - Summary of the classification of varietal groups and hybrids of Coffea Canephora species, by Fisher's linear and quadratic discriminant function (in parentheses)

Varietal Group	Classification according to linear and quadratic discriminant analysis (in parentheses)			Total observations
	Conilon	Robusta	Hybrids	
Conilon	40 (40)	5 (5)	0 (0)	45
Robusta	4 (4)	33 (33)	0 (0)	37
Hybrids	1 (1)	0 (1)	82 (81)	83
Apparent correct rate = 0.9394 (0.9333)				
Apparent error rate= 0.0606 (0.0667)				

As the "Mat" and "TFr" variables were removed, the linear discriminant analysis obtained a performance similar to that of the previous analysis, which demonstrates that the variables "Mat" and "TFr" are not significant for the discrimination of these groups. This performance can be verified through the APER, in which, for linear discriminant function with all variables, the APER was 0.0667 (Table 1). When the variables were excluded, it decreased to 0.0606 (Table 5). The quadratic function in this scenario was less efficient in discriminating varietal groups and hybrids, with an APER of 0.0606 (Table 5), which was greater than that observed in the previous analysis (APER = 0.0364). The removal of these characteristics ("Mat " and " TFr ") for the construction of the discriminant function can be interesting, since it will reduce labor in the field.

In practice, in *C. canephora* breeding process, the correct classification of its botanical varieties is extremely important, since each of its varieties requires specific crop management procedures (MARCOLAN *et al.*, 2009). Inadequate cultural treatments due to possible classification errors can reduce crop productivity (PEREIRA *et al.*, 2000). Another issue is related to the correct recommendation of cultivars for obtaining hybrids. Successful hybridization is directly linked to the correct classification of parents (FERRÃO *et al.*, 2013).

Thus, the discriminant analysis technique for the purpose of discriminating the varietal and hybrid groups of *C. canephora* was successfully applied in this study. In other cultures, this technique has already been applied efficiently. For example, Nogueira *et al.* (2008) used the discriminant analysis to differentiate eleven soybean cultivars. In their study, four experiments were carried out at different times (two in the summer and two in the winter), based on 7 phenotypic characteristics. As a result, they obtained an apparent error rate of (12.73% and 9.09%) and (1.82% and 16.36%), respectively, for summer (December and February) and winter sowing seasons (May and June). Thus, using discriminant analysis, it was possible to distinguish soybean cultivars. Das *et al.* (2020) used the quadratic discriminant analysis to assess the discriminative power of 17 wavelengths of 14 rice genotypes. As a result, the quadratic discriminate analysis was efficient for discriminating

rice genotypes, with a relative error rate of 2%. The Fisher's discriminant analysis was also used to classify forest communities in the pampa biome. In their study, eight structural variables were used in order to classify five types of forests. The discriminant analysis classified all samples correctly in their respective predicted groups, that is, the apparent error rate was 0% (KILCA *et al.*, 2015).

In coffee data, this technique had already been evaluated for classification purposes, but with different objectives, for which the discriminant analysis also obtained satisfactory results. Campos *et al.* (2016) selected groups of seven quality characteristics of coffee seedlings, using six selection criteria. He used Fisher's discriminant analysis to transform the groups of characteristics into a new variable and then compare the results obtained with the univariate analysis of the new variable with the multivariate analysis. As a result, the discriminant analysis proved to be a viable option for treatment discrimination.

Other multivariate techniques have already been successfully used in studies of *C. canephora*. Ivoglo *et al.* (2008) used cluster analysis to quantify the genetic divergence of populations of *C. canephora*. In their study, they demonstrated that, when there is a database with different characteristics of interest, multivariate analysis proves to be a powerful tool for integrating multiple information and assisting the breeding program in the choice of divergent parents that will be more likely to promote superior results. Da Fonseca *et al.* (2006) applied cluster analysis and canonical variables in studies with *C. canephora* clones, in order to separate and classify genotypes of varietal groups from two environmental regions of Espírito Santo, based on 7 phenotypic characteristics, which corroborates the vital role of a correct classification in breeding programs.

4 Conclusions

The discriminant analysis was efficient for distinguishing between varietal groups (Conilon and Robusta) and Coffea canephora hybrids. The variables vegetative vigor, evaluation of the incidence of coffee rust in the field, incidence of cercosporiosis, height of the plant and diameter of the canopy projection were the phenotypic characteristics that proved to be important in the discrimination of these groups. Therefore, with this phenotypic information for new individuals, we can obtain their classification into varietal and hybrid groups efficiently.

▪ *RESUMO. A cafeicultura é uma das atividades agrícolas mais importantes do mercado mundial. Coffea canephora é uma das espécies de maior relevância comercial, podendo ser dividida nos grupos varietais Conilon e Robusta. Esses grupos varietais têm interesses agronômicos complementares. Portanto, os híbridos são obtidos por meio de cruzamentos entre esses grupos.*

Dada a dificuldade em diferenciar os dois grupos varietais de genótipos no campo, a correta discriminação é essencial para a definição de cruzamentos em programas de melhoramento. Nesse contexto, o presente trabalho teve como objetivo aplicar análise discriminante (AD) para definir funções de diferenciação entre grupos varietais e híbridos de C. canephora, e identificar os traços fenotípicos mais relevantes nessas funções. Dados de 165 genótipos do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural e do Centro Agronómico Tropical de Investigación y Enseñanza foram usados para medir diferentes características de plantas. A AD quadrática foi aquela de melhor desempenho para discriminação dos genótipos, com taxa de erro aparente média de 0,0364. Incidência de cercosporiose, incidência de ferrugem, vigor vegetativo, altura da planta e diâmetro da projeção da copa foram as características mais importantes na discriminação dos grupos varietais.

▪ *PALAVRAS-CHAVE: Análise discriminante; café; Conilon; Robusta.*

References

ALKIMIM, E. R. *et al.* Selective efficiency of genome-wide selection in Coffea canephora breeding. *Tree Genetics & Genomes*, v.16, n.3, 2020.

ANAGBOGU, C. F. *et al.* Genetic diversity and re-classification of coffee (Coffea canephora Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. *Genetic Resources and Crop Evolution*, v.66, n.3, p.685-696, 2019.

BABOVA, O.; OCCHIPINTI, A.; MAFFEI, M. E. Chemical partitioning and antioxidant capacity of green coffee (Coffea arabica and Coffea canephora) of different geographical origin. *Phytochemistry*, v.123, p.33-39, 2016.

BARBOSA, I. P; DA COSTA, W. G.; NASCIMENTO, M.; CRUZ, C. D.; DE OLIVEIRA, A. C. B Recommendation of Coffea arabica genotypes by factor analysis. *Euphytica*, v.215, n.10, p.178, 2019.

BARROSO, L. M. A.; NASCIMENTO, M. ; NASCIMENTO, A. C. C.; SILVA, F. F. E. ; FERREIRA, R. P. Uso do método de Eberhart e Russell como informação a priori para aplicação de redes neurais artificiais e análise discriminante visando a classificação de genótipos de alfafa quanto à adaptabilidade e estabilidade fenotípica. *Revista Brasileira de Biometria*, v. 31, p.176-188, 2013.

BERTHAUD, J. *Les ressources génétiques pour l'amélioration des caféiers africains diploïdes: évaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs: Conséquences pour l'application*. Paris: ORSTOM, 1986.

CAMPOS, K. A.; MORAIS, A. R. D.; PAIXÃO, C. A. Alternative for the evaluation of coffee seedlings using Fisher's discriminant analysis. *Revista Ciência Agronômica*, v.47, n.2, p.299-306, 2016.

CENCI, A.; COMBES, M. C.; LASHERMES, P. Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. *Plant molecular biology,* v.78, n.1-2, p.135–145, 2012.

CORSO, M. P.; VIGNOLI, J. A.; BENASSI, M. D. T. Development of an instant coffee enriched with chlorogenic acids. *Journal of Food Science and Technol* v.53, p.1380–1388, 2016.

CRUZ, C. D. *et al. Modelos biométricos aplicados ao melhoramento genético II.* Viçosa: UFV, 2004. 480p.

DA FONSECA, A. F. A. *et al.* Genetic divergence in conilon coffee [Divergência genética em café Conilon]. *Pesquisa Agropecuária Brasileira*, 2006.

DAS, B. *et al.* Discrimination of rice genotypes using field spectroradiometry. *Geocarto International*, v.35, n.1, p.64-77, 2018.

DAVIS, A. P. *et al.* Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. *Botanical Journal of the Linnean Society*, v.167, n.4, p.357-377, 2011.

DE CARVALHO, V. P. *et al.* Discrimination of populations under covariance matrix heterogeneity and non-normal random vectors in genetic diversity studies. *Científica*, v.46, n.4, p.344-352, 2018.

FERRÃO, L. F. V. *et al.* New EST–SSR markers of Coffea arabica: transferability and application to studies of molecular characterization and genetic mapping. *Molecular Breeding*, v.35, n.1, p.31, 2015.

FERRÃO, L. F. V. *et al.* Comparative study of different molecular markers for classifying and establishing genetic relationships in Coffea canephora. *Plant systematics and evolution*, v.299, n.1, p.225–238, 2013.

FERRÃO, L. F. V; CAIXETA, E T; PENA, G; ZAMBOLIM, E. M; CRUZ, C. D; ZAMBOLIM, L; FERRÃO, M. A; SAKIYAMA, N. S. Marcadores microssatélites em estudo de diversidade, mapelamento genético e análises de QTLs em Coffea canephora. p. 134, 2013.

FERREIRA, D. F. *Estatística multivariada*. 2.ed. Lavras: Editora Ufla, 2008.

FERREIRA, D. F. *Estatística multivariada*. 3.ed. Lavras: Editora Ufla, 2018.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v.7, p.179-188, 1936.

GOMEZ, C.; DUSSERT, S.; HAMON, P.; HAMON, S.; DE KOCHKO, A.; PONCET, V. Current genetic differentiation of Coffea canephora Pierre ex A.Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology,* v.9, n.1, p.167, 2009.

HUDED, A. K. C. *et al.* Genetic Diversity and Population Structure Analysis of Coffee (Coffea canephora) Germplasm Collections in Indian Gene Bank Employing SRAP and SCoT Markers. *International Journal of Fruit Science*, v.1, p.1-28, 2020.

IVOGLO, M. G.; FAZUOLI, L. C.; OLIVEIRA, A. C. B.; GALLO, P. B.; MISTRO, J. C.; SILVAROLLA, M. B.; TOMA-BRAGHINI, M. Divergência genética entre progênies de café robusta. *Bragantia*: v.67, n.4, p.823-831, 2008.

JAMES, G; WITTEN, D; HASTIE, T; TIBSHIRANI, R. *An introduction to statistical learning: With applications in R*, New York: Springer2013.

KILCA, R. V.; LONGHI, S. J.; SCHWARTZ, G.; SOUZA, A. M.; WOJCIECHOVSKI, J. C. Application of fisher's discriminant analysis to classify forest communities in the Pampa biome. *Ciência Florestal,* v.25, n.4, p.885-895, 2015.

LABOUISSE, J. P. *et al.* New insights on spatial genetic structure and diversity of Coffea canephora (Rubiaceae) in Upper Guinea based on old herbaria. *Plant Ecology and Evolution*, v.153, n.1, p.82-100, 2020.

MARCOLAN, A. L. *et al.* Cultivo dos cafeeiros Conilon e Robusta para Rondônia. *Embrapa Rondônia*, n.3, p.61, 2009.

MARRACCINI, P. *et al.* Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of Coffea canephora. *Journal of Experimental Botany,* v.63, n.11, p.4191–4212, 2012.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005. 297p.

MUSOLI, P. *et al.* Genetic differentiation of wild and cultivated populations: diversity of Coffea canephora Pierre in Uganda. *Genome,* v.52, p.634–646, 2009.

NOGUEIRA, A. P. O. *et al.* Novas características para diferenciação de cultivares de soja pela análise discriminante. *Ciência Rural*, v.38, n.9, p.2427-2433, 2008.

PEREIRA, R. C. A. *et al.* Avaliação de genótipos dos cafeeiros Arabica e Robusta na Amazônia Ocidental. *Acta Amazonica*, 2000.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/., 2020.

REN, J. *et al.* Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *International Journal of Molecular sciences*, v.14, n.4, p.7061–7088, 2013.

RIBEIRO, B. B. *et al.* Avaliação química e sensorial de blends de Coffea canephora Pierre e Coffea arabica L. *Coffee Science*, v.9, n.2, p.178-186, 2014.

SOUZA, F. D. F. *et al.* Molecular diversity in Coffea canephora germplasm conserved and cultivated in Brazil. *Crop Breeding and Applied Biotechnology*, v.13, n.4, p.221-227, 2013.

VARELLA, C. A. A. *Estimativa da produtividade e do estresse nutricional da cultura do milho usando imagens digitais*. 107f. Tese (Doutorado em Engenharia Agrícola), Universidade Federal de Viçosa, Viçosa, 2004.

VOLSI, B.; TELLES, T. S.; CALDARELLI, C. E.; CAMARA, M. R. G. D. The dynamics of coffee production in Brazil. *PLoS one*, v.14, n.7, p.e0219742, 2019.