Genetics and Molecular Research

Assessment of genetic divergence among coffee genotypes by Ward-MLM procedure in association with mixed models

W.P. Rodrigues[1]*, H.D. Vieira[1]*, P.E. Teodoro[2] F.L. Partelli[3] and D.H.S.G. Barbosa[4]

[1]Universidade Estadual do Norte Fluminense Darcy Ribeiro,
Centro de Ciências e Tecnologias Agropecuárias,
Campos dos Goytacazes, RJ, Brasil
[2]Departamento de Biologia Geral,
Universidade Federal de Viçosa, Viçosa, MG, Brasil
[3]Universidade Federal do Espírito Santo,
Centro Universitário Norte do Espírito Santo, São Mateus, ES, Brasil
[4]Embrapa Mandioca e Fruticultura, Cruz das Almas, BA, Brasil

*These authors contributed equally to this study.
Corresponding author: W.P. Rodrigues
E-mail: wevertonuenf@hotmail.com

Genet. Mol. Res. 15 (2): gmr.15027889
Received October 22, 2015
Accepted December 7, 2015
Published May 6, 2016
DOI http://dx.doi.org/10.4238/gmr.15027889

ABSTRACT. Mixed linear models have been used for the analysis of the genetic diversity and provided further accurate results in crops such as eucalyptus, castor, and sugarcane. However, to date, research that combined this analysis with Ward-MLM procedure has not been reported. Therefore, the aim of the present study was to identify divergent coffee genotypes by Ward-MLM procedure, in association with the mixed-decision models. The experiment was initiated in February 2007, in the northwestern Rio de Janeiro State. The 25 evaluated genotypes were grown with a spacing of 2.5 x 0.8 m, in a randomized block design, with 5 replications, containing 8 plants each. The

following agronomic traits were evaluated: plant height, stem diameter, plagiotropic branch number, and productivity. Four measurements were performed for each character from 2009 to 2012, between May and July. Data were analyzed using REML/BLUP analysis and Ward-MLM procedure. The Ward-MLM procedure in association with mixed linear models demonstrated the genetic variability among the studied coffee genotypes. We identified two groups of most divergent coffee genotypes, which can be combined by crossings and selections in order to obtain genotypes with high productivity and variability.

Key words: BLUP/REML; *Coffea arabica*; Genetic value

INTRODUCTION

In breeding programs of *arabica* coffee plant (*Coffea arabica* L.) the choice of parents to be crossed is necessary, to obtain segregating populations. Artificial hybridization in autogamous plants usually involves cross between two parents. Major limitations of this type of hybridization are limited genetic variation and low recombination, due to the repeated process of self-fertilization. In this regard, an approach for achieving superior progenies is to gather information on the agronomic superiority and genetic diversity, allowing combination of parents and identification of the broader gene pool, and assessing the viability of crosses (Cruz et al., 2004).

The use of multivariate techniques has enabled studies of genetic divergence among coffee genotypes (Carvalho et al., 2012, Guedes et al., 2013). Multivariate analyses are based on algorithms, or distance measurements, which simultaneously consider many features and allow unification of multiple information obtained from a set of characters. Among the available techniques, the agglomerative methods are the most used. They involve grouping of genotypes, such that there is homogeneity within a group and heterogeneity among the different groups (Mohammadi and Prasana, 2003).

Among the multivariate methods, the Ward procedure - modified location model (MLM), proposed by Franco et al. (1998) consists of an excellent strategy for quantifying the variability, using quantitative and/or qualitative variables. The Ward-MLM procedure is performed in two steps. First, the groups are defined by the Ward clustering method (Ward, 1963), using the Gower dissimilarity matrix (Gower, 1971); subsequently, the vector average of quantitative variables for each sub-population is estimated by the MLM procedure.

The MLM procedure, combined with analysis of variance (ANOVA), has been proven to be effective in differentiating maize genotypes (Gutiérrez et al., 2003; Franco et al., 2005; Ortiz et al., 2008), fodder radish (Padilla et al., 2005), tomato (Gonçalves et al., 2009), common bean (Barbé et al., 2010; Cabral et al., 2010), pepper (Sudré et al., 2010), and banana (Pestana et al., 2011). However, high environmental influence on quantitative traits, mostly employed in the selection of genotypes in soybean-breeding programs, makes the results less accurate compared with other techniques.

Considering the above-mentioned factor, use of the mixed linear model in genetic diversity analysis has the advantage of using genotypic instead of phenotypic values, providing more accurate results compared to the conventional statistical methods (Resende, 2004). This

has been proven in the studies on eucalyptus (Rocha et al., 2007), castor (Oliveira et al., 2013), and sugarcane (Lopes et al., 2014). However, despite the importance of using mixed models in the genetic divergence analysis, to date no research combining this analysis with Ward-MLM procedure has been reported. Thus, the aim of the present study was to identify divergent coffee genotypes by Ward-MLM procedure using mixed models.

MATERIAL AND METHODS

The experiment was initiated in February 2007, at Panorama 1 Farm, located in the municipality of Varre e Sai, in the northwestern Rio de Janeiro State. The soil was identified as red yellow latosol. The climate was typical of high tropics, with hot summers and colder winters, with an average annual temperature 19.0°C and 1600 mm average annual precipitation (Martorano et al., 2003).

The 25 genotypes (Table 1), evaluated in five replicates, were grown with a 2.5 x 0.8-m spacing, in a randomized block design, containing 8 plants each. The following agronomic traits were evaluated: plant height (PH), stem diameter (SD), plagiotropic branch number (PBN), and productivity (PROD). Four measurements were made for each character between May and July, from 2009 to 2012. PH and SD were measured with the aid of a tape measure. For PROD estimation, the collected volume was converted to bags processed per ha (bags/ha) using the scale 480 L coffee cherries per 60-kg bag processed.

Table 1. Number of genotypes, origin, and provenience of the coffee genotypes evaluated.

No.	Genotype	Origin	Provenience
G1	Catucaí Vermelho 785/15	Icatu Vermelho x Catuaí Amarelo	PROCAFÉ
G2	Catucaí Amarelo 2 SL	Icatu Vermelho x Catuaí Amarelo	PROCAFÉ
G3	IPR/Iapar	Villa Sarchi x Híbrido do Timor	IAPAR
G4	Catiguá MG 02	Catuaí Amarelo 86 x Híbrido do Timor	EPAMIG
G5	IPR 99/Iapar	Villa Sarchi x Híbrido do Timor	IAPAR
G6	Acauã	Mundo Novo 388-17 x Sarchimor 1668	PROCAFÉ
G7	Araponga MG 01	Catuaí Amarelo 86 x Híbrido do Timor	EPAMIG
G8	Palma II	Catuaí x Catimor	IBC
G9	Sabiá 398	Acaiá x Catimor	PROCAFÉ
G10	IPR 103/Iapar	Icatu x Catuaí	IAPAR
G11	IPR 100/Iapar	Villa Sarchi x Híbrido do Timor	IAPAR
G12	H 4193-3-3-716-4-1	Catuaí Amarelo x Híbrido do Timor	EPAMIG
G13	H 419-10-6-2-12-1	Catuaí Amarelo x Híbrido do Timor	EPAMIG
G14	Catucaí Amarelo 24/137	Icatu Vermelho x Catuaí Vermelho	PROCAFÉ
G15	Iapar 59	Villa Sarchi x Híbrido do Timor	IAPAR
G16	Oeiras	Seleção de Catimor	EPAMIG
G17	Catuaí Vermelho144	Caturra Amarelo 7476 x Mundo Novo 374	IAC
G18	Catucaí Amarelo 20/15	Icatu Vermelho x Catuaí Vermelho	PROCAFÉ
G19	Catiguá MG 01	Catuaí Amarelo 86 x Híbrido do Timor	UFV
G20	H 419-10-6-2-5-10-1	Catuaí Amarelo x Híbrido do Timor	EPAMIG
G21	IPR 104/Iapar	Villa Sarchi x Híbrido do Timor	IAPAR
G22	Sacramento MG 01	Catuaí Vermelho 81 x Híbrido do Timor	EPAMIG
G23	Bourbon Amarelo LCJ 10	Bourbon Vermelho x Amarelo de Botucatu	IAC
G24	Pau Brasil	Catuaí Vermelho 141 x Híbrido do Timor	EPAMIG
G25	H 419-10-6-2-5-1	Catuaí Amarelo x Híbrido do Timor	EPAMIG

Data were analyzed by REML/BLUP analysis, using the statistical model 55 of the Selegen software (Resende, 2007), given by Equation 1:

$$Y = Xm + Zg + Wp + Ti + e \quad \text{(Equation 1)}$$

where, *Y* is the vector data, *m* is the vector of effects of measurement-repeating combinations (assumed to be fixed) added to the overall average, *g* is the vector of the genotypic effects (assumed to be random), *p* is the vector of the permanent environment effect (here, random plots), *i* is the vector of the effects of the genotype x measurement interaction, and *e* is the vector of errors or waste (random). The letters represented the incidence matrices for these purposes.

Prediction of the genotypic values, for each trait, capitalizing the mean interaction (gem) in different environments is given by Equation 2:

$$gem = \hat{\mu}_j + \hat{g}_i + \hat{g}e_m \quad \text{(Equation 2)}$$

where, $\hat{g}e_m$ is calculated by Equation 3:

$$\hat{\mu} + \frac{\left(\frac{\hat{\sigma}_g^2 + \hat{\sigma}_c^2}{n}\right)}{\hat{\sigma}_g^2}\hat{g}_i \quad \text{(Equation 3)}$$

where, $\hat{\mu}$ is the overall mean of all environments; *n* is the number of environments, and $\hat{g}_i$ is the genotypic effect of the genotype i.

The data were analyzed simultaneously using the Ward-MLM method to create the groups through clustering. For the Ward clustering method, the distance matrix was constructed by Gower's algorithm (Gower, 1971). The definition of the ideal number of groups was according to the pseudo-F and pseudo-t^2 criteria. Differences among the groups, the correlation between the variables, and the canonical (CAN) variable were examined graphically. The distance for the distribution of the traits proposed by Franco et al. (1998) was used to determine the dissimilarity among the groups. All analyses were carried using the statistical SAS software (SAS Institute, 2009).

RESULTS AND DISCUSSION

The ideal number of groups obtained from the procedure of the likelihood function (pseudo-F and pseudo-t^2), was equal to three. Gonçalves et al. (2009) and Barbé et al. (2010) reported that the analysis of the likelihood function can bring forth further precise criteria for the formation of groups, resulting in the determination of less subjective groups. Barbé et al. (2010), Cabral et al. (2010), and Pestana et al. (2011) observed results of similar magnitude.

Group I consisted of 13 coffee genotypes (Table 2) and was characterized by the lowest average for genotypic characters with regard to PH, SD, and PBN. Group II consisted of 11 genotypes that presented the highest average for genotypic traits PBN and PROD. Only G23 composed the group III, which had the highest genetic means for the traits PH and SD.

Table 2. Number of genotypes per group and predicted genotypic values for plant height (PH, cm), stem diameter (SD, mm), plagiotropic branch number (PBN), and productivity (PROD, bag/ha) evaluated in four measurements in 25 coffee genotypes.

Group	Genotypes	PH	SD	PBN	PROD
I	G1, G3, G5, G6, G8, G9, G12, G13, G15, G16, G20, G21, and G25	152.80	43.10	61.66	40.67
II	G2, G4, G7, G10, G11, G14, G17, G18, G19, G22, and G24	165.72	45.37	65.25	42.50
III	G23	191.41	50.86	63.79	28.19

From Table 3 it can be observed that the genotypes from group II displayed significant similarity to those from the groups I and III. Thus, to prevent restriction of genetic variability and offset of the gains achieved by selection, crosses between the genotypes of these groups are not recommended for breeding by hybridization. Since genetically related parents tend to have many genes or alleles in common, when two such parents are crossed there is little advantage, given the low level of allelic heterozygosis upon crossing (Cruz et al., 2014).

Table 3. Distance between the groups formed by the Ward-MLM procedure.

Group	I	II	III
I	0	15.26	102.90
II		0	47.43
III			0

The longest distance (102.90) was found between groups I and III, i.e., genotypes of this groups are the most divergent. This bullish divergence, in principle, allows the crossing between these pairs, to maximize heterosis in the progeny and increases the possibility of segregating the individuals in the subsequent generations because of different numbers of loci in which the dominance effects are evident (Cruz et al., 2014). Moreover, considering the productivity, which is extremely important for the selection of superior genotypes in breeding programs, individuals of these groups had the highest average for genotypic traits. Thus, it may be possible to generate genotypes with high heterosis owing to different numbers of loci in which the dominance effects are evident.

In the analysis based on the CAN variables, it was observed that the first two variables accounted for 86.58% of the variance (Figure 1). Therefore, the two-dimensional representation was most suitable for representing data set because the accumulated variance in CAN 1 and CAN 2 was larger than 80%, the minimum limit recommended by Cruz et al. (2014) for use in this technique. The distance between the genotypes of the groups I and III reinforces the prospect of obtaining highly productive genotypes with the crossing of individuals from these groups.

In conclusion, Ward-MLM procedure in association with mixed linear models demonstrated the genetic variability among the studied coffee genotypes.

We identified two groups of the most divergent genotypes, which could be combined by crossings and selections to obtain further productive and diverse genotypes.

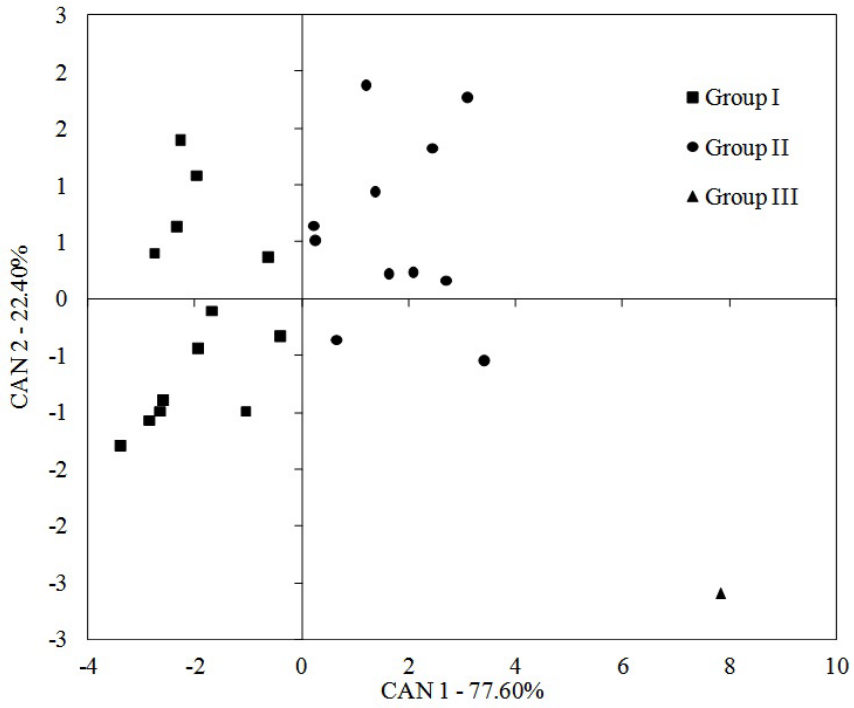

Figure 1. Scores of the first two canonical variables (CAN1 and CAN2) for the three groups of coffee genotypes formed by Ward-MLM procedure.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) and the National Council for Scientific and Technological Development (CNPq).

REFERENCES

Barbé TC, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, et al. (2010). Association between advanced generations and genealogy in inbred lines of snap bean by the Ward-Modified Location Model. *Euphytica* 173: 337-343. http://dx.doi.org/10.1007/s10681-009-0089-z

Cabral PDS, Soares TCB, Gonçalves LSA, Amaral Júnior AT, et al. (2010). Quantification of the diversity among common bean accessions using Ward-MLM strategy. *Pesq. Agropec. Bras.* 45: 1124-1132. http://dx.doi.org/10.1590/S0100-204X2010001000011

Carvalho AM, Mendes ANG, Botelho CE, Oliveira ACB, et al. (2012). Desempenho agronômico de cultivares de café resistentes à ferrugem no Estado de Minas Gerais, Brasil. *Bragantia* 71: 481-487. http://dx.doi.org/10.1590/S0006-87052013005000007

Cruz CD, Regazzi AJ and Carneiro PCS (2004). Modelos biométricos aplicados ao melhoramento genético. Vol. 1. 3rd edn. Editora UFV, Viçosa.

Cruz CD, Carneiro PCS, Regazzi AJ. (2014) Modelos Biométricos Aplicados ao Melhoramento Genético. Vol. 2. 3rd edn. Editora UFV, Viçosa.

Franco J, Crossa J, Villaseñor J, Taba S, et al. (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696. http://dx.doi.org/10.2135/cropsci1998.0011183X003800060045x

Franco J, Crossa J, Taba S and Shands H (2005). A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45: 1035-1044. http://dx.doi.org/10.2135/cropsci2004.0292

Gonçalves LS, Rodrigues R, do Amaral Júnior AT, Karasawa M, et al. (2009). Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet. Mol. Res.* 8: 364-374. PubMed http://dx.doi.org/10.4238/vol8-1gmr549

Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-874. http://dx.doi.org/10.2307/2528823

Guedes J, Vilela DJM, Rezende JC and Silva FL (2013). Divergência genética entre cafeeiros do germoplasma Maragogipe. *Bragantia* 72: 127-132. http://dx.doi.org/10.1590/S0006-87052013000200003

Gutiérrez L, Franco J, Crossa J and Abadie T (2003). Comparing a preliminary racial classification with a numerical classification of the maize landraces of Uruguay. *Crop Sci.* 43: 718-727. http://dx.doi.org/10.2135/cropsci2003.0718

Lopes VR, Filho JCB, Daros E, Oliveira RA, et al. (2014). Divergência genética entre clones de cana-de-açúcar usando análise multivariada associada a modelos mistos. *Semin. Cienc. Agrar.* 35: 125-134. http://dx.doi.org/10.5433/1679-0359.2014v35n1p125

Martorano LG, Rossiello ROP, Meneguelli NA, Lumbreras J F, et al. (2003). Aspectos climáticos do noroeste fluminense, RJ. Documento, 43. Embrapa Solos, Rio de Janeiro.

Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248. http://dx.doi.org/10.2135/cropsci2003.1235

Oliveira RS, Silva SA, Brasileiro BP, Medeiros EP, et al. (2013). Genetic divergence on castor bean using the ward-mlm strategy. *Rev. Cienc. Agron.* 44: 564-570. http://dx.doi.org/10.1590/S1806-66902013000300019

Ortiz R, Crossa J, Franco J, Sevilla R, et al. (2008). Classification of Peruvian highland maize races using plant traits. *Genet. Resour. Crop Evol.* 55: 151-162. http://dx.doi.org/10.1007/s10722-007-9224-7

Padilla G, Cartea ME, Rodríguez VM and Ordás A (2005). Genetic diversity in a germplasm collection of *Brassica rapa* subsp. *rapa* L. from northwestern Spain. *Euphytica* 145: 171-180. http://dx.doi.org/10.1007/s10681-005-0895-x

Pestana RKN, Amorim EP, Ferreira CF, Amorim VBO, et al. (2011). Agronomic and molecular characterization of gamma ray induced banana (*Musa* sp.) mutants using a multivariate statistical algorithm. *Euphytica* 178: 151-158. http://dx.doi.org/10.1007/s10681-010-0329-2

Resende MDV (2004). Métodos estatísticos ótimos na análise de experimentos de campo. Embrapa Floresta, Colombo, 57.

Resende MDV (2007). Matemática e estatística na análise de experimentos e no melhoramento genético. Embrapa Floresta, Colombo, 561.

Rocha MGB, Pires IE, Rocha RB, Xavier A, et al. (2007). Seleção de genitores de *Eucalyptus grandis* e de *Eucalyptus urophylla* para produção de híbridos interespecíficos utilizando REML/ BLUP e informação de divergência genética. *Rev. Arvore* 31: 977-987. http://dx.doi.org/10.1590/S0100-67622007000600001

SAS Institute (2009). SAS language and procedures: usage. Version SAS 9.2.

Sudré CP, Gonçalves LS, Rodrigues R, do Amaral Júnior AT, et al. (2010). Genetic variability in domesticated *Capsicum* spp as assessed by morphological and agronomic data in mixed statistical analysis. *Genet. Mol. Res.* 9: 283-294. PubMed http://dx.doi.org/10.4238/vol9-1gmr698

Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244. http://dx.doi.org/10.1080/01621459.1963.10500845