

# Factor analysis for plant and production variables in *Coffea canephora* in the Western Amazon

Gabi Nunes Silva<sup>1</sup> , Laís Mayara Azevedo Barroso<sup>1</sup> , Cosme Damião Cruz<sup>2</sup> ,  
Rodrigo Barros Rocha<sup>3</sup> , Fábio Medeiros Ferreira<sup>4</sup> 

<sup>1</sup>Fundação Universidade Federal de Rondônia/UNIR, Campus Ji-Paraná, Departamento Acadêmico de Matemática e Estatística, Ji-Paraná, RO, Brasil

<sup>2</sup>Universidade Federal de Viçosa/UFV, Departamento de Biologia Geral, Viçosa, MG, Brasil

<sup>3</sup>Empresa Brasileira de Pesquisa Agropecuária/Embrapa, Centro de Pesquisa Agroflorestal de Rondônia, PortoVelho, RO, Brasil

<sup>4</sup>Instituto de Ciências Exatas e Tecnologia da Universidade Federal do Amazonas/UFAM, Campus Itacoatiara, Itacoatiara, AM, Brasil

Contact authors: gabi.silva@unir.br; lais.barroso@unir.br; cdcruz@ufv.br; rodrigo.rocha@embrapa.br; fmferreira77@gmail.com

Received in November 25, 2021 and approved in May 9, 2022

## ABSTRACT

The evaluation of morphological characters related to the hulled coffee yield subsidizes the selection of *Coffea canephora* plants that combine a set of favorable traits. However, the greater the number of traits considered, the more difficult the selection process becomes. In this context, multivariate analyzes can be useful to overcome this problem. The aim of this study was to identify, in a set of agronomic traits of *Coffea canephora*, the determining factors of biological phenomena and use these factors to recognize patterns of diversity and similarity from biological complexes of interest to the breeder. To this, eleven morphological descriptors were evaluated of 130 clones of the botanical varieties Conilon and Robusta and intervarietal hybrids over two crop years in the experimental field of Embrapa, in the municipality of Ouro Preto do Oeste, state of Rondônia (RO). To group the traits, the multivariate technique of Factor Analysis was used. The effect of genotype x year interaction was significant for the eleven traits analyzed. Based on the scree plot, three factors were established. Factors were interpreted as architecture, vigor and grains with a satisfactory percentage of explained variability. The interpretation of the factors highlighted the importance of the Conilon variety to improve the architecture of the Robusta botanical variety. These results show that it is possible to use factor scores to identify varieties and traits that favor higher production of hulled coffee.

**Key words:** Coffee; Conilon; Robusta; Intervarietal hybrids; Multivariate analyzes.

## 1 INTRODUCTION

Coffee is one of the most consumed beverages in the world. According to the United States Department of Agriculture (USDA, 2021), Brazil, a major producer and exporter of this bean, is responsible for approximately 30% of world exports. Of the many species of the genus *Coffea* L., the two most widely cultivated are *Coffea arabica* L. and *Coffea canephora* Pierre ex A. Froehner (Ribeiro et al., 2014). About 45% of world coffee production is *Coffea canephora*, and this species is cultivated in Central and West African countries as well as in Southeast Asia and South America (Marcolan; Espindula, 2015; USDA, 2021).

The two botanical varieties of *Coffea canephora* are Conilon, which is more drought tolerant but has greater susceptibility to diseases, and Robusta, which are taller and more vigorous plants, with larger leaves and fruits, and has greater resistance to pests and diseases, but is less drought tolerant (Rocha et al., 2021). In Rondônia, in the Western Amazon, both botanical varieties are cultivated. However, cultivation of Robusta is more common, because in this region, the rains are abundant and well distributed throughout the year, which benefits the cultivation of more vigorous varieties and plants that demand more water (Marcolan; Espindula, 2015).

Studies reveal that the distinct genetic backgrounds of the varieties provide different characteristics associated with

plant development, responses to biotic and abiotic stresses, production, and beverage quality (Marcolan; Espindula, 2015).

The strategies to obtain genetic progress of *Coffea canephora* are through intrapopulation improvement in Conilon and Robusta varieties or by exploring heterosis (hybrid vigor) and complementary characteristics to form intervarietal hybrids (Resende, 2015). Regardless of the breeding strategy adopted, the promising genotypes must have different agronomic and commercial attributes.

One of the greatest challenges for the breeder is to simultaneously combine desirable phenotypes of the desired agronomic traits, which can be highly correlated. These associations do not always follow the respective directions of selection, do not allow always satisfactory indirect genetic gains, and demand greater intellectual effort, costs, and time. Therefore, the greater the number of variables to be considered in the selection process, the greater the difficulty in selecting the outstanding genotypes, because these variables, when correlated, can represent different biological and physiological aspects of the selected genotypes.

Factor Analysis (FA) is a multivariate method that to evaluate the original variability of a vector through a smaller number of variables, called a factor, which will represent supercharacters or biological complexes to group the original correlated characteristics (Mingoti, 2005). In

breeding, supercharacters can be used to predict indirect genetic gains to select and complement simultaneous selection techniques based on selection indices (Ferreira et al., 2010).

To improve plant species, Reis et al. (2017) used FA to develop indices that allow selection of maize genotypes, and Barbosa et al. (2019) used FA to select *Coffea arabica* genotypes with high potential for traits with commercial interest.

Hence, this work aimed to identify, in the agronomic characters of 130 *Coffea canephora* clones, the determining factors of biological phenomena for genotypic evaluation and use these factors to recognize patterns of diversity and similarity of biological complexes of interest to the breeder. Furthermore, we intended to confirm whether the biological complexes are affected by different agronomic traits, bearing in mind the different genetic backgrounds of each group of varieties evaluated.

## 2 MATERIAL AND METHODS

### 2.1 Experimental data

The data are from an experiment conducted in the municipality of Ouro Preto do Oeste, RO (10°37'03"S and 62°51'50"W) in an Embrapa experimental field. The statistical design adopted was a randomized block design with four replications, four plants per plot, and a spacing of 3 × 2 m. Characteristics were evaluated of 130 clones of the botanical varieties Conilon and Robusta, and hybrids between these two varieties. These included 76 clones of the Conilon variety, 36 clones of the Robusta, and 18 intervarietal hybrids. Cultural treatments were performed according to Marcolan et al. (2009).

The 11 traits evaluated over two years (2014 and 2015), namely plant height in meters (H); leaf length (LL) and leaf width (LW), in centimeters, were estimated from the evaluation of 10 leaves collected from the middle third of the plant and measured using a digital caliper; length of plagiotropic branches in meters (LPB); distance between rosettes of the intermediate part of the plagiotropic branch in centimeters (DR); number of fruits per rosette of the intermediate part of the plagiotropic branch (NFR); maturation time was determined using the criterion that the plant has 70% of fruits in the cherry stage, with the record of the number of days between flowering and harvest (ND); number of productive branch (NPB); number of rosettes per plagiotropic branch (NRPB); grain size was evaluated individually using a set of 12 different sieves from samples of 250g of processed coffee through the trait average sieve (AS); production were estimated based on the weight of processed grains in bags increased per hectare (PROD).

### 2.2 Statistical analysis

The statistical model used for the analysis of variance (ANOVA) was:

$$Y_{ijk} = \mu + B_k + G_i + A_j + GA_{ij} + \varepsilon_{ijk} \quad (1)$$

In which,  $Y_{ijk}$  is the observation of the  $i$ -th genotype in the  $j$ -th year and the  $k$ -th block;  $\mu$  overall mean of the experiment;  $B_k$  effect of block  $k$ ;  $G_i$  effect of genotype  $i$ ;  $A_j$  effect of year  $j$ ;  $GA_{ij}$  effect of the interaction between genotype  $i$  and year  $j$ ;  $\varepsilon_{ijk}$  random error associated with observation  $ijk$ . To estimate genetic parameters, environments and genotypes were considered as fixed effects in the model.

To group the most correlated variables and thus obtain a factor that represents them, reducing the number of variables, the multivariate technique of Factor Analysis was used. For this purpose, the joint FA model used was:

$$X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \dots + l_{im}F_m + \varepsilon_i \quad (2)$$

In which  $i = 1, 2, \dots, 11$ , with 11 being the number of original variables observed;  $l_{ij}$  represents the factor loading associated with the  $i$ -th variable  $X_i$  and the factor  $F_j$ , with  $j = 1, 2, \dots, m$ , where  $m$  is the number of factors;  $F_j$  is a random vector ( $m \times 1$ ) of common unobservable latent factors, and  $\varepsilon_i$  is the random error vector that is associated only with the  $i$ -th phenotypic variable.

The number of factors used in the model ( $m$ ) was determined by the scree-plot graph (Cattell, 1966). The scree plot uses the eigenvalues in descending order and looking for a "big gap" in this graph (Zhu; Ghodsi, 2006).

The initial factor loadings were estimated by the spectral decomposition of the correlation matrix and the final factor loadings were established using the varimax rotation method. To maximize the variability of factor loadings and help interpret the distribution of variables in the respective factors, the Varimax rotation was employed (Barbosa et al., 2019). According to the literature, the characteristics that present factor loadings greater than 0.5 must be considered in each factor (Peterson, 2000; Merenda 1997).

Statistical analyzes were performed using the software Genes and its integration with the software R (Cruz, 2016; R Core Team, 2021).

## 3 RESULTS

The analysis of variance indicated that the effect of the years wasn't significant for LL, ND and AS. The source of variation Between CixYears wasn't significant for ND and Production (Table 1). For the others sources of variation were significant according to the F test at 5% probability (Table 1).

**Table 1:** Summary of the joint analysis of variance under the simple factorial scheme with specification of genotype classes (G) where each class Ci is defined by C1: Conilon; C2: Robusta and C3: Hybrid for the characteristics: plant height in meters (H), leaf length in centimeters (LL), length of plagiotropic branches in meters (LPB), distance between rosettes in centimeters (DR), number of fruits per rosette (NFR), leaf width in centimeters (LW), number of days to maturate (ND), number of productive plagiotropic branches (NPB), number of rosettes per plagiotropic branches (NRPB), average sieve (AS), production in bags increased per hectare (PROD).

SV	DF	Mean squares					
		H	LL	LPB	DR	NFR	LW
Block	3	0.01	0.31	0.01	0.29	26.33	0.27
Clones	129	0.26**	26.79**	0.07**	2.71**	69.74**	8.75**
C1	75	0.19**	15.35**	0.06**	1.69**	31.30**	3.51**
C2	35	0.30**	16.39**	0.08**	3.82**	132.03**	5.96**
C3	17	0.41**	6.04**	0.07**	1.62**	104.94**	1.98**
Between Ci	2	0.63**	813.85**	0.40**	31.05**	122.35**	311.87**
Years	1	65.16**	3.71 <sup>ns</sup>	0.22**	24.60**	1217.81**	3.04*
Clones x Years	129	0.05**	1.35**	0.01**	1.00**	46.25**	0.50**
C1xA	75	0.04**	1.19**	0.01**	1.02**	34.95**	0.47**
C2xA	35	0.06**	0.91**	0.01**	0.76**	23.59**	0.30**
C3xA	17	0.05**	2.85**	0.01**	1.29**	51.49**	0.85**
Between CixYears	2	0.13**	2.60**	0.05**	1.80**	822.22**	2.15**
Residual	777	0.01	0.43	0.01	0.13	5.69	0.12

SV	DF	Mean squares				
		ND	NPB	NRPB	AS	PROD
Block	3	145.18	833.78	4.54	1.11	304.42
Clones	129	1926.53**	6514.42**	21.79**	9.10**	2352.60**
C1	75	1759.01**	5213.42**	17.47**	6.89**	1761.57**
C2	35	1184.55**	5459.09**	34.02**	8.38**	1782.75**
C3	17	2400.91**	3413.33**	15.07**	7.87**	1981.31**
Between Ci	2	17160.79**	100129.47**	26.67**	115.20**	37644.83**
Years	1	597.06 <sup>ns</sup>	186238.90**	829.31**	1.01 <sup>ns</sup>	95987.19**
Clones x Years	129	284.09**	1098.93**	12.12**	1.99**	646.89**
C1 x A	75	278.88**	1029.05**	11.35**	1.87**	539.66**
C2 x A	35	323.11**	1318.60**	15.33**	2.08**	572.37**
C3 x A	17	243.24**	884.18*	7.36**	1.86**	1336.00**
Between CixYears	2	144.16 <sup>ns</sup>	1700.49*	25.22*	6.12**	114.68 <sup>ns</sup>
Residual	777	96.34	479.40	3.89	0.67	165.69

\*and \*\* = significant at 1% and 5% probability by the F test, respectively.

Table 2 shows the mean values and coefficient of variation corresponding to clones and groups Conilon, Robusta and Hybrids for the eleven characteristics evaluated.

The average production was higher in the plant population with hybrids (52.67 sc/ha), followed by Conilon (47.47 sc/ha). The productivity of the Robusta population was lower than the others (29.85 sc/ha), although the plants have presented close numbers of fruits per rosette (NFR) and of rosettes per plagiotropic branches (NRPB) (Table 2).

The low estimates of the coefficient of variation indicated good experimental precision (Table 2).

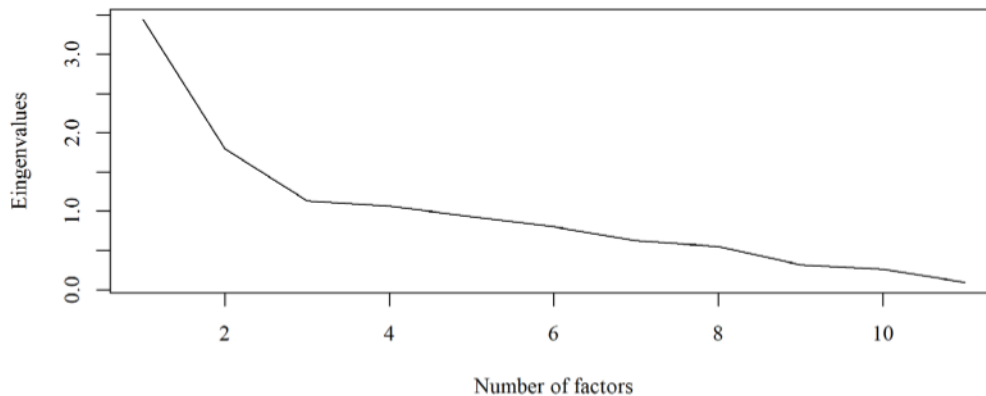
Factor analysis considered the average performance of 130 *C. canephora* clones. To determine the ideal number of factors, as suggested by Catell (1966), the scree-plot was used (Figure 1).

The final factor loadings and commonalities after rotation via the Varimax method are presented in Table 3.



**Table 2:** Mean values and coefficient of variation (CV) corresponding to clones and groups, Conilon (C1), Robusta (C2), and Hybrids (C3) for the characteristics: plant height in meters (H), leaf length in centimeters (LL), length of plagiotropic branches in meters (LPB), distance between rosettes in centimeters (DR), number of fruits per rosette (NFR), leaf width in centimeters (LW), number of days to mature (ND), number of productive plagiotropic branches (NPB), number of rosettes per plagiotropic branches (NRPB), average sieve (AS), production (PROD). N= number of observations.

	N		H	LL	LPB	DR	NFR	LW
Total Clones	130	Mean	1.55	12.77	0.78	4.91	17.00	5.42
		CV	6.65	5.15	9.44	7.26	14.03	6.38
C1 - Conilon	76	Mean	1.52	11.99	0.77	4.78	16.74	4.94
		CV	6.75	5.48	9.64	7.45	14.25	7.00
C2 - Robusta	36	Mean	1.60	14.79	0.83	5.30	17.78	6.67
		CV	6.42	4.44	8.92	6.72	13.42	5.18
C3 - Hybrid	18	Mean	1.52	12.00	0.76	4.66	16.56	4.95
		CV	6.74	5.48	9.70	7.64	14.41	6.99
	N		ND	NPB	NRPB	AS	PROD	
Total Clones	130	Mean	301.94	81.33	10.33	15.09	43.31	
		CV	3.25	26.92	19.09	5.43	29.72	
C1 - Conilon	76	Mean	299.06	89.91	10.40	14.71	47.47	
		CV	3.28	24.35	18.97	5.57	27.12	
C2 - Robusta	36	Mean	311.06	58.91	10.00	15.78	29.85	
		CV	3.15	37.17	19.72	5.19	43.12	
C3 - Hybrid	18	Mean	295.87	89.97	10.70	15.32	52.67	
		CV	3.32	24.34	18.43	5.35	24.44	



**Figure 1:** Scree-plot.

Considering that each characteristic must be inserted into only one factor, the factors were built with the characteristics that presented an absolute factor loading greater than 0.5 (Peterson, 2000; Merenda 1997).

The first factor F1 was more related to the characteristics that describe plant architecture, because the highest factor loadings were for the characteristics: plant height (0.75), leaf length (0.71), length of plagiotropic branches (0.80), distance between rosettes (0.73), and leaf width (0.73) (Table 3). This factor was biologically interpreted as representative of the biological complex that defines architecture of the plant.

The second factor F2 was interpreted as Production, because it presented highest factor loadings for the characteristics: number of grains per rosette (-0.57), number of days to maturation (-0.56), average sieve (-0.53), and production (0.56) (Table 3).

The third factor F3 obtained highest factor loadings for the characteristics: number of productive plagiotropic branches (0.71) and number of rosettes per plagiotropic branches (0.64); therefore, F3 was interpreted as Plant vigor.

Figure 2 are graphs with the scores obtained from the Factor Analysis, discriminating the botanical variety of the evaluated clones.

**Table 3:** Final factor loadings (after Varimax rotation) and commonality values obtained from the Factor Analysis of the 11 characteristics evaluated.

Characteristics	Factor loadings after rotation			Commonality
	F1	F2	F3	
H	<b>0.75</b>	0.15	0.38	0.73
LL	<b>0.71</b>	-0.49	-0.06	0.74
LPB	<b>0.80</b>	0.06	0.04	0.65
DR	<b>0.73</b>	-0.10	-0.14	0.56
NFR	0.06	<b>-0.57</b>	0.55	0.64
LW	<b>0.73</b>	-0.51	-0.16	0.82
ND	0.03	<b>-0.56</b>	-0.10	0.32
NPB	-0.38	0.27	<b>0.71</b>	0.72
NRPB	0.10	0.15	<b>0.64</b>	0.45
AS	-0.07	<b>-0.53</b>	-0.32	0.39
PROD	-0.15	<b>0.56</b>	0.06	0.34
Mean	--		--	0.58

H: plant height, LL: leaf length, LPB: length of plagiotropic branches, DR: distance between rosettes, NFR: number of fruits per rosette, LW leaf width of, ND: number of days to maturate, NPB number of productive plagiotropic branches, NRPB: number of rosettes per plagiotropic branches, AS: average sieve, PROD: production.

The scores associated with the architecture separate the genotypes of the Conilon and Robusta populations, with an overlap between the Conilon and hybrid genotypes.

For the productive characteristics, greater similarities were observed between the hybrid individuals and those

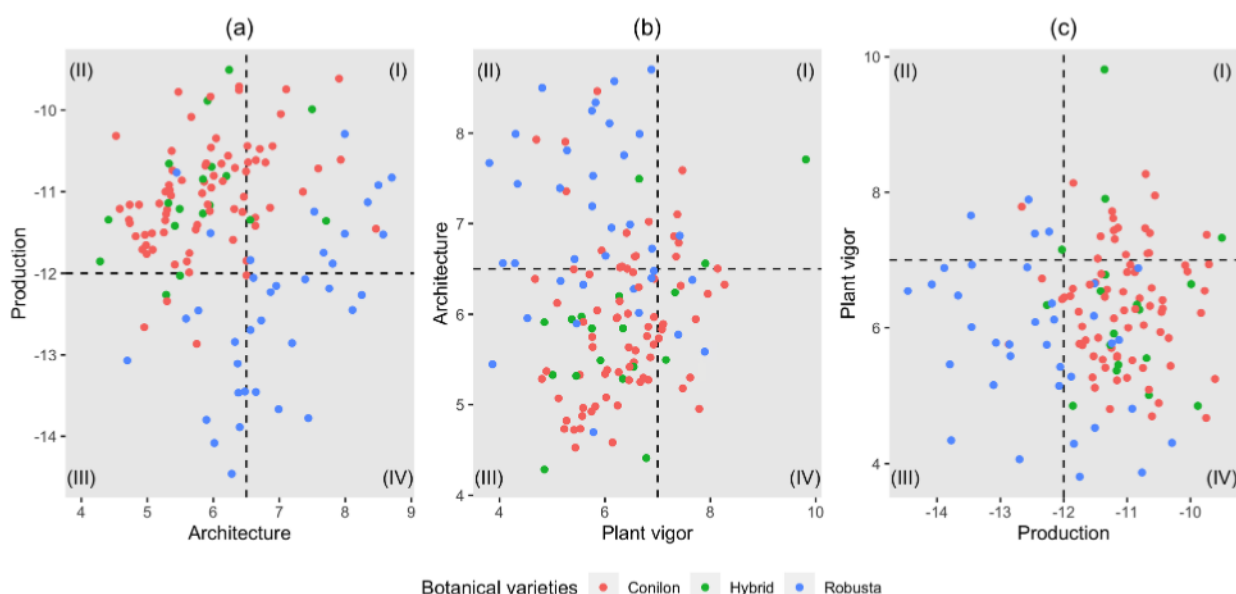
of Conilon. Individuals of the Conilon botanical variety were distinguished not only by the shorter size and smaller structures, but also by the better performance of their productive characteristics.

## 4 DISCUSSION

Table 1 shows the summary of the joint analysis of variance under the simple factorial scheme. The sources of variation that were significant indicate wide variability within each genotypic class analyzed, between clones of the Conilon, Robusta, and hybrid varieties, once in the field, individuals of the Robusta botanical variety characteristically have greater height, larger leaf and branch size, and more distance between the rosettes, which do not necessarily reflect in greater productivity (Souza, 2015).

Hybrid vigor, understood as the superiority of individuals from divergent crosses, was reported by Charrier and Berthaud (1985) in the 1980s and used in the following years for plant selection (Ivoglo et al., 2008). More recently, the breeding program of Embrapa developed a cultivar with hybrid characteristics suitable for planting in regions of the Western Amazon (Teixeira et al., 2017). However, few studies have focus on understanding the contribution of agronomic characteristics in the expression of this superiority, considering the diversity among accessions and the beans to be obtained with the selection.

Bean yield is a characteristic of quantitative expression, and the expression of traits associated with bean yield can result in higher yields in various ways (Table 2).



**Figure 2:** Graph of the scores obtained from the Factor Analysis. (a) Architecture Factor × Production Factor; (b) Plant Vigor Factor × Architecture Factor; (c) Production Factor × Plant Vigor Factor. The colors represent the botanical varieties Conilon (Red), Hybrid (Green), and Robusta (Blue). The values of I, II, III, and IV indicate the quadrants.

In scree-plot graph, the eigenvalues are in descending order, and the ideal number of factors is estimated considering the index of the last component before the plot flattens, which statistically represents a decrease in importance in relation to the total variance (Catell, 1966). This greatest decrease occurred in the third factor ( $m=3$ ).

When using the Factor analysis methodology, commonality is a useful measure to determine the proportion of the variance of each characteristic explained by established common factors (Teixeira et al., 2016).

The estimated average commonality of 0.58 indicates that the identified factors can be interpreted in a biological context, based on the rotated factor loadings (Table 3). When evaluating different scenarios, Figueiredo Filho and Silva Júnior (2010) suggest that only estimates close to or greater than 0.50 are interpreted.

Analyzing the factors obtained, it can be observed that although the superior performance of *C. canephora* coffee trees with hybrid characteristics between the botanical varieties Conilon and Robusta is known, few studies have sought to better understand the expression of the characteristics associated with the production of this coffee tree. When interpreting estimates of direct and indirect effects of secondary traits in relation to the main trait – bean yield – Spinelli et al. (2018) observed that the number of plagiotropic branches and the number of rosettes per productive branch were the characteristics with the greatest direct effect on the processed coffee yield. Positive associations between the number and length of productive branches were also observed by Ferrão et al. (2007), who considered estimates of simple and partial, unstructured correlations.

The grouping of characteristics according to their nature is the main differential of factor analysis. The information contained in the results allows us to discriminate the relationship between groups of characteristics that are expressed together and influence the final productivity of accessions.

The factor loading can be defined as the correlation of the variable with each of the factors. Positive estimates indicate that the variable is positively correlated with the factor, while negative estimates indicate that this association is negative or inversely proportional. Thus, factor analysis is performed to reduce the dimension of the correlation matrix between the studied variables, by grouping the characteristics into the observed factors that were associated with architecture, production, and plant vigor.

The morphological and architectural characteristics of plants were positively associated in the first factor. Hence, the typical characteristics of the Robusta botanical variety, with the larger plant structures, tend to express themselves together, with the heavier plants associated with the longer plagiotropic branches, the greater distance between rosettes,

and the larger leaves. Associations between plant architectural features can help or even hinder plant breeding. The greater distance between rosettes decreases the productive potential of the plants, and genotypes with longer plagiotropic branches typical of the Robusta botanical variety are desired, with less distance between the rosettes which is a characteristic that is typical of the Conilon botanical variety.

Studies of genetic diversity in *C. canephora* demonstrated, through principal component analysis and discriminant analysis, that representative characteristics of architecture, plant vigor, and production satisfactorily helped discriminate accessions of the botanical varieties Conilon and Robusta. Intervarietal hybrids were more similar to the Conilon group (Oliveira et al., 2018; Ferrão et al., 2021; Souza et al., 2021).

The second factor was associated with productive characteristics, exhibiting negative associations between yield and number of fruits per rosette (NFR), number of days to maturity (ND), and bean size (AS). Fewer fruits per rosette is a typical characteristic of the Robusta botanical variety, while the less days to maturation and the smaller bean size characterize the Conilon botanical variety. Therefore, the complementarity between the contrasting characteristics is as important as the hybrid vigor that manifests itself in divergent hybridizations (Rocha et al., 2015).

The third factor associated with plant vigor indicates the importance of selecting plants with complementary characteristics between the two botanical varieties. This factor indicates an association between the number of plagiotropic branches and the number of rosettes, which is mainly expressed in the hybrid plant population, since their parents in the populations of Conilon and Robusta have these different characteristics.

According to Cruz et al. (2014), the identification of genetic diversity between parents helps in the conservation and strategic use of genetic resources and aims to identify the hybrid combinations with greater heterotic effect and greater heterozygosity.

The interpretation of the factor analysis scores can group the more similar genotypes separately considering the factors associated with architecture, production, and plant vigor (Figure 2).

The clones that stand out in quadrants II and IV of Figure 2 are characterized as good, but divergent for the factors under study, because they only have one outstanding factor. The Robusta variety, for example, has good scores for the Architecture factor but lower scores for the Production factor (Figure 2a). This variety maintains the same pattern for the Architecture and Plant Vigor factors (Figure 2b). On the other hand, the Conilon variety exhibited greater concentration in quadrant IV relating to the Production and Plant Vigor factors (Figure 2c), and in quadrant III, for the Plant Vigor and



Architecture factors (Figure 2b), showing that this botanical variety has good scores for these factors.

Although the factor scores associated with plant vigor also show a greater divergence between the populations of Conilon and Robusta, the hybrid genotypes overlapped between these two populations. The greater heterozygosity of this population from divergent hybridizations is also expressed in the field, with individuals distributed over all quadrants of the dispersion (Ferrão et al., 2021).

The separation of Conilon and Robusta varieties in relation to scores associated with the architecture (Figures 2a and 2b) is related with the characteristics of these botanical varieties, since Robusta genotypes are taller and more vigorous plants, with larger leaves and fruits than Conilon genotypes (Rocha et al., 2021), which justifies the distribution of Robusta in quadrants I and IV (Figure 2a) and in quadrant II (Figure 2b).

An overlap between the Conilon and hybrid genotypes seen in Figure 2 was an expected relationship between these two varieties, since the change in architecture observed in the intervarietal hybrid plants is one of the main characteristics obtained from crossing Robusta and Conilon, while maintaining vigor, the hybrids have a slower growth rate than plants of the Robusta botanical variety (Moraes et al., 2021).

The arrangement of the clones evaluated in the quadrants shown in Figure 2 also evidenced the productive potentials of the Conilon botanical variety. The good productive potential of the Conilon botanical variety has been observed in different regions of the country (Ferrão et al., 2020).

The interpretation of the factors allows us to observe the importance of the Conilon variety for improving the architecture of the Robusta botanical variety, obtaining shorter plants, with less space between rosettes, with more plagiotropic branches, while also maintaining the longer branches and the larger beans, characteristics of the Robusta botanical variety (Marcolan; Espindula, 2015).

## 5 CONCLUSIONS

The discrimination of botanical varieties of the *canephora* trees directly assists coffee breeding programs, as it allows the crossing between these varieties and supports the systematic production of hybrid progenies, to obtain and select improved genotypes.

Factor analysis managed to reduce the 11 characteristics initially evaluated to only 3 factors related to architecture, vigor, and production, which had a satisfactory percentage of explained variability.

## 6 ACKNOWLEDGEMENTS

The authors are grateful to National Council for Scientific and Technological Development (CNPq - Conselho

Nacional de Desenvolvimento Científico e Tecnológico), to Minas Gerais State Research Foundation (FAPEMIG - Fundação de Pesquisa do Estado de Minas Gerais), to Amazonas State Research Foundation (FAPEAM - Fundação de Amparo à Pesquisa do Estado do Amazonas) and to Coordination for the Improvement of Higher Education Personnel (CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES – Finance Code 001) for financial support.

## 7 AUTHORS' CONTRIBUTIONS

GNS, LMAB and CDC wrote the manuscript and conducted all statistical analyses. RBR Performed the experiment and co-work the manuscript, review and approved the final version of the work. FMF co-work the manuscript, review and approved the final version of the work.

## 8 REFERENCES

- BARBOSA, I. D. P. et al. Recommendation of *Coffea arabica* genotypes by factor analysis. **Euphytica**, 215(178), 2019.
- CATTELL, R. B. The scree test for the number of factors. **Multivariate behavioral research**, 1(2):245-276, 1966.
- CHARRIER, A.; BERTHAUD J. Botanical classification of coffee. In: CLIFFORF, M. N. E.; WILLSON, K. C. **Coffee: Botany, biochemistry, and production of beans and beverage**. New York: Croom Helm, p.13-47,1985.
- CRUZ, C. D. et al. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, Minas Gerais: Editora UFV, 2014. 560p.
- CRUZ, C. D. Genes software-extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum Agronomy**, 38(4):547-552. 2016.
- FERRÃO, M. A. G. et al. Characterization and genetic diversity of *Coffea canephora* accessions in a germplasm bank in Espírito Santo, Brazil. **Crop Breeding and Applied Biotechnology**, 21(2):e36132123, 2021.
- FERRÃO, R. G. et al. (Ed.). **Café Conilon**. Vitória: Incaper, 2007. 702p.
- FERRÃO, R. G. et al. Cultivares de cafés Conilon e Robusta. **Informe Agropecuário. Cafés Conilon e Robusta: potencialidades e desafios**, 41(309):17-25, 2020.
- FERREIRA, F. M. et al. Formação de supercaracteres para seleção de famílias de maracujazeiro amarelo. **Acta Scientiarum Agronomy**. Maringá, 32(2):247-254, 2010.

- FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Visão além do alcance: uma introdução à análise fatorial. **Opinião Pública**, 16(1):160-185, 2010.
- IVOGLO, M. G. et al. Divergência genética entre progênies de café robusta. **Bragantia**, 67(4):823-831, 2008.
- MARCOLAN, A. L.; ESPINDULA, M.; **Café na Amazônia**. 1 ed. Brasília, Distrito Federal: Embrapa, 2015. 474p.
- MARCOLAN, A. L. et al. **Cultivo dos cafeeiros Conilon e Robusta para Rondônia**. 3. ed. rev. atual. Porto Velho: Embrapa Rondônia: Emater-RO, 2009. 67p. (Embrapa Rondônia. Sistema de produção, 33).
- MERENDA, P. F. A Guide to the Proper Use of Factor Analysis in the Conduct and Reporting of Research: Pitfalls to Avoid. **Measurement and Evaluation in Counseling and Development** 30. 156-164, 1997.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte, Minas Gerais: Editora UFMG, 2005. 297p.
- MORAES, M. S. et al. Environmental stratification and performance of *Coffea canephora* clones grown in the Western Amazon. **Coffee Science**, 16:e161911, 2021.
- OLIVEIRA, L. N. L. et al. Selection of *Coffea canephora* parents from the botanical varieties Conilon and Robusta for the production of intervarietal hybrids. **Ciência Rural**, 48(4):1-7, 2018.
- PETERSON, R. A. A Meta-Analysis of Variance Accounted for and Factor Loadings in Exploratory Factor Analysis. **Marketing Letters**, 11:261-275, 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/>. 2021.
- REIS, G. G. et al. Tropical maize selection indexes genotypes for efficiency in use of nutrients: phosphorus. **Revista Ceres**, 64:266-273, 2017.
- RESENDE, M. D. V. **Genética Quantitativa e de Populações**. 1ed. Visconde do Rio Branco, Minas Gerais: Suprema, 2015. v.1, 422p.
- RIBEIRO, B. B. et al. Avaliação química e sensorial de blends de *Coffea canephora* Pierre e *Coffea arabica* L. **Coffee Science**, 9(2):178-186, 2014.
- ROCHA, R. B. et al. Melhoramento de *Coffea canephora* – considerações e metodologias. In: MARCOLAN, A. L.; ESPINDULA, M. C. (Ed.). **Café na Amazônia**. Brasília: Embrapa, 2015. p.100-126.
- ROCHA, R. B. et al. *Coffea canephora* breeding: estimated and achieved gains from selection in the Western Amazon, Brazil. **Ciência Rural**, 51(5):e20200713, 2021.
- SOUZA, F. de F. et al. de Aspectos gerais da biologia e da diversidade genética de *Coffea canephora*. In: MARCOLAN, A. L.; ESPINDULA, M. C. (Ed.). **Café na Amazônia**. Brasília: Embrapa, p.84-98, 2015.
- SOUZA, M. S. et al. Classification of *Coffea canephora* clones in botanical varieties by discriminant analysis of the k-nearest neighbors. **Revista Ceres**, 68(5):371-374, 2021.
- SPINELLI, V. M. et al. Contribution of agronomic traits to the yield of *Coffea canephora* Pierre ex A. Froehner Hulled coffee. **Coffee Science**, 13(3):333-340, 2018.
- TEIXEIRA A. L. et al. Performance of intraspecific hybrids (Kouillou x Robusta) of *Coffea canephora* Pierre. **African Journal of Agricultural Research**, 12:2675-2680, 2017.
- TEIXEIRA, F. R. F. et al. Factor analysis applied to genome prediction for high-dimensional phenotypes in pigs. **Genetics and Molecular Research**, 15:1-10, 2016.
- BARROS, S. **Coffee annual report**. United States Department of Agriculture: Foreign Agricultural service. São Paulo: ATO, 2021.
- ZHU, M.; GHODSI, A. Automatic dimensionality selection from the scree plot via the use of profile likelihood. **Computational Statistics & Data Analysis**, 51:918-930, 2006.