Ensaios
e Ciência
ISSN 1415-6938 e ISSN 2178-695X

DOI: https://doi.org/10.17921/1415-6938.2025v29n4p952-967

Employment of the REML/BLUP and Multivariate in Dwarf Cashew Clones in Pacajus, Ceará

Uso de REML/BLUP e Multivariada em Clones de Cajueiro-Anão em Pacajus, Ceará

Received: June 2, 2025
Accepted: November 6, 2025

Gilberto Ken Iti Yokomizo: Embrapa Amapá, Ap, Brazil. E-mail: gilberto.yokomizo@embrapa.br

Dheyne Silva Melo: Embrapa Agroindústria Tropical, CE, Brazil.

Kuang Hongyu: Universidade Federal do Mato Grosso. MT, Brazil.

Luiz Augusto Lopes Serrano: Embrapa Agroindústria Tropical. CE, Brasil.

Francisco das Chagas Vidal Neto: Embrapa Agroindústria Tropical. CE, Brasil.

Abstract

There is a progressive worldwide interest in cashew nuts, demanding new cultivars that are more productive and resistant to pests and diseases, but a complicating factor is the number of characteristics evaluated. One solution is the use of multivariate statistics, which makes it possible to interpret these together. Therefore, employing the REML/BLUP (best linear unbiased prediction/restricted maximum likelihood) procedure along with multivariate analysis, this work aims to identify clones with the best performance. A total of 18 dwarf cashew clones were evaluated in an experimental design of randomized blocks, three replications, four plants per plot, spacing of 8 x 6 m, with the following characteristics being evaluated: NAN: visual note for the presence of anthracnose ; NMP: visual note for the presence of black mold; NOC: visual note for the presence of powdery mildew on the chestnut; NOI: visual note for the presence of powdery mildew on the inflorescence; PRO: nut productivity; PMC: average nut weight; IPP: Production precocity index. The main conclusions are: clones 1, 3, 4, 5, 7, 8, 9, 10, 11, and 18 stood out in four characteristics and 5, 8 and 10 in more than four; by Biplot, except for PMC, the clones show high GxA interactions; in the Box Plot the effect of years directly and positively influences the characteristics, mainly PRO, IPP, NMP, NOI, therefore greatly influenced by stochastic factors while PMC is the one with the least environmental influence, this confirms what was observed for results of heritability, accuracy and CVr ratio with higher values for PMC.

Keywords: *Anarcardium occidentale* L. Main Components. Plant Breeding. Genetic Variation. Genetic Gain.

Resumo

Existe um progressivo interesse mundial pela castanha de caju, demandando novos cultivares mais produtivos e resistente às pragas e doenças, porém um complicador é a quantidade de características avaliadas, uma solução é o uso da estatística multivariada, que possibilita interpretar estas conjuntamente. Desta forma empregando o procedimento REML/BLUP (melhor predição linear não tendenciosa/máxima verossimilhança restrita) juntamente com uma análise multivariada, este trabalho tem como objetivo identificar clones de melhor desempenho. Um total de 18 clones de cajueiro-anão foram avaliados em um delineamento experimental de blocos ao acaso, três repetições, quatro plantas por parcela, espaçamento de 8 x 6 m, sendo avaliadas as características: NAN: nota visual para a presença de antracnose; NMP: nota visual para a presença de mofo preto; NOC: nota visual para a presença de oídio na castanha; NOI: nota visual para a presença de oídio na inflorescência; PRO: produtividade de castanha; PMC: peso médio da castanha e; IPP: Índice de precocidade de produção. As principais conclusões são: destacaram-se os clones 1, 3, 4, 5, 7, 8, 9, 10, 11, e 18 em quatro características e 5, 8 e 10 em mais de quatro; pelo Biplot, exceto para PMC, os clones apresentam interação GxA altos; no Box Plot o efeito de anos influenciam direta e positivamente as características, principalmente PRO, IPP, NMP, NOI, portanto muito influenciados pelos fatores estocásticos enquanto que PMC é a de menor influência ambiental, isso ratifica o observado para resultados de herdabilidade, acurácia e relação CVr com valores maiores para PMC.

Palavras-chave: *Anarcardium occidentale* L. Componentes Principais. Melhoramento vegetal. Variação Genética. Ganho Genético.

---

1 Introduction

The cashew tree (*Anacardium occidentale* L.), considered to be of Brazilian origin, is a fruit-bearing species of great socioeconomic importance in the Northeastern region of Brazil (Brainer, 2022). Furthermore, it is highly adapted to environments characterized by low soil fertility, high average temperatures, and water stress, being able to survive under low annual rainfall conditions—around 500 mm—and even in areas with high precipitation exceeding 3,700 mm (Anjos Júnior *et al.*, 2016; Kahlmann; Kohn, 2018).

This species has multiple uses and is cultivated on a large scale across four continents: South America, Australia, Asia, and Africa. Globally, the leading producer of cashew nuts is Côte d'Ivoire, with approximately 850 thousand tons in 2020, followed by India (about 780 thousand tons), Vietnam (350 thousand tons), and then Burundi, the Philippines, and Cambodia, among others. Brazil ranks only 11th in world production (Brainer, 2022).

In Brazil, the largest cashew-producing areas are concentrated in the states of Ceará, Piauí, and Rio Grande do Norte. According to IBGE (2022), cashew nut production in 2022 totaled 121,680 tons, representing a 9.9% increase compared to 2021. The average yield reached 287 kg/ha, nearly 11% higher than the previous year, generating total sales of BRL 476.952 million. Ceará accounted for 56.61% of this production, followed by Piauí (20.66%)

and Rio Grande do Norte (13.68%). The states of Pernambuco, Paraíba, Bahia, Alagoas, and some regions of the Midwest together represented 9% of national production (Brainer, 2022).

The part commonly referred to as the “fruit” is, in fact, the hypertrophied peduncle, which constitutes the edible pulp known as the “pseudofruit.” The so-called “nut” is the true fruit of the cashew tree, from which both the kernel and the cashew nut shell liquid (CNSL) are obtained (Moraes *et al*., 2013), both of which have numerous applications.

Processing of cashew fruit parts is carried out in a few large-scale factories and in more than a hundred small-scale units known as mini-factories (MF). Together, they form an industrial complex with an installed capacity to add value to 270 thousand tons per year, directly generating nearly 40 thousand rural jobs and 15 thousand industrial jobs (Instituto Caju Brasil – ICB, 2020). Indirectly, the sector is estimated to benefit approximately 1.5 million people (Faria *et al*., 2018), highlighting the substantial socioeconomic importance of the cashew value chain.

Of particular note is the growing global demand for cashew nuts, which has driven increased production needs. The main consumers are the United States, India, and China. Cashew nuts accounted for 16% of the global tree nut and dried fruit market during the 2021/22 harvest season, ranking fourth in consumer preference after almonds, walnuts, and pistachios (INC, 2022).

This growing international demand necessitates the development of new cultivars with higher productivity and greater resistance to pests and diseases. However, research efforts in selection are often hindered by the need to evaluate a large number of traits, as genotypes that excel in one trait may perform poorly in another. A valuable tool to interpret multiple traits simultaneously is multivariate statistical analysis, which enables the joint assessment of correlated variables, providing a more comprehensive interpretation than univariate analysis, which considers each trait in isolation (Cruz; Carneiro; Regazzi, 2014).

Two prominent multivariate techniques used for such analyses are Principal Component Analysis (PCA) and Canonical Discriminant Analysis (CDA). PCA transforms a set of correlated variables into a new set of uncorrelated variables, called principal components, which linearly reduce data dispersion while retaining the maximum amount of total variance information and arranging the components in decreasing order of importance (Johnson;

Wichern, 2019;). CDA, similar to PCA, reduces data dimensionality by standardizing variables and identifying those that best discriminate among evaluated groups (Hair *et al.*, 2014; Johnson; Wichern, 2019).

It should also be noted that, as the cashew tree is a perennial species, quantitative genetic studies aimed at selection should employ standard analytical procedures involving the estimation of variance components using Restricted Maximum Likelihood (REML) and the prediction of genetic values through Best Linear Unbiased Prediction (BLUP), also referred to as the mixed-model methodology (Sturion; Resende, 2010).

Based on these considerations, the present study aimed to identify superior cashew clones and determine which evaluated traits most contribute to observed differences through multivariate analysis, emphasizing these traits in the selection process.

2 Material and Methods

A total of 18 dwarf cashew tree clones (Table 1), exhibiting good potential for nut and pseudofruit production and previously selected for the edaphoclimatic conditions of the coastal region of Ceará, were evaluated from 2016 to 2019, when they were between three and six years old, i.e., at fruit-bearing age. The planting was carried out on March 16, 2013, at the Embrapa Experimental Field in Pacajus, located at the geographical coordinates 4°11'26.62" S; 38°29'50.78" W, at an altitude of 60 m above sea level. The area has flat topography and a Red-Yellow Argisol soil type, with an average annual rainfall of 939.1 mm and mean temperatures ranging from 26 °C to 28 °C (FUNCEME, 2023).

Table 1 – Identification of the 18 dwarf cashew tree clones. Pacajus, CE

Clones	Treatment	Clones	Treatment
A + C 276/1	1	HI 58-92-2	10
END II 6-9	2	PRO 740/4	11
PRO 555/2	3	HB 58	12
PRO 553/2	4	HB 124/4	13
H 84/92/2	5	PRO 106/2	14
A + A 134/1	6	CP 76	15
HB 116/4	7	BRS 226	16
HB 33	8	PRO 111/3	17
PRO 611/1	9	HB 135/1	18

Source: research data.

Table 2 presents the total and monthly rainfall values for the evaluation period, as precipitation represents the main stochastic component of variation in the experimental area.

Table 2 – Monthly and annual total rainfall in Pacajus, C

Months	**2016**	**2017**	**2018**	**2019**
January	181.0	38.0	108.3	183.8
February	54.0	251.0	186.9	265.2
March	56.0	445.0	110.9	343.0
April	177.0	207.0	239.1	287.3
May	71.0	124.0	112.1	161.9
June	26.0	59.0	0.0	28.2
July	4.0	112.0	0.0	21.4
August	2.0	0.0	0.0	3.5
September	0.0	0.0	0.5	5.9
October	5.5	14.0	0.5	0.0
November	9.0	0.0	0.0	0.0
December	31.0	0.0	30.2	25.0
Total	**616.5**	**1250.0**	**788.5**	**1325.2**

Source: research data.

The experimental design consisted of a randomized block design with 18 treatments (clones) of dwarf cashew trees, three replications, and four plants per plot, spaced 8 × 6 m apart. The evaluated traits included five phytosanitary and three productivity-related characteristics:

Phytosanitary characteristics:

- **NAN:** Visual score for the presence of anthracnose, with a scale ranging from 0 (absence of symptoms) to 4 (more than 25% of the leaf area affected).
- **NMP:** Visual score for the presence of black mold, with the same scale from 0 to 4.
- **NOC:** Visual score for the presence of powdery mildew on nuts, from 0 to 4.
- **NOI:** Visual score for the presence of powdery mildew on inflorescences, from 0 to 4.

Productivity characteristics:

- **PRO:** Nut yield throughout the harvest season, in kg·ha^{-1}.
- **PMC:** Average nut weight, based on the mean of 100 nuts per plot/clone, after harvest, in grams.
- **IPP:** Precocity index of production, estimated according to Cilas et al. (2011) using the equation:

$$IPP = \frac{ny_1 + (n-1)y_2 + 2y_{n-1} + y_n}{nY},$$

where y_i is the yield per plant in the i-th harvest within a series of size n; Y is the sum of yields per plant over the evaluated period; and n is the number of harvests.

All crop management practices were conducted in accordance with the recommendations for commercial cashew cultivation described in *Sistema de Produção do Caju* (Serrano *et al.*, 2016).

Estimation of variance components and prediction of genetic values were performed using the REML/BLUP procedure (Restricted Maximum Likelihood/Best Linear Unbiased Prediction). All analyses were conducted with the *Selegen-REML/BLUP* software, based on Model 29 (Resende, 2007, 2016).

The statistical model used for evaluating clones in a randomized block design with one observation per plot and multiple years was: $y = Xm + Zg + Wp + e$,where y is the data vector; m is the vector of fixed effects for the measurement–replication combinations plus the overall mean; g is the vector of genotypic effects (random); p is the vector of permanent environmental effects (plots, random); and e is the vector of residual errors (random). The capital letters represent the incidence matrices for the respective effects. The vector m accounts for all measurements across replications and adjusts simultaneously for replication, measurement, and replication × measurement interaction effects.

To observe and understand differences in clone performance generated by the evaluated traits simultaneously, a multivariate analysis was performed. Two methodologies were used: Principal Component Analysis (PCA) and Canonical Discriminant Analysis (CDA), as described below.

PCA was performed using the Pearson correlation matrix, which standardizes each variable (mean equal to zero and variance equal to one) to account for differing units and scales. The dependent relationship among variables was estimated by the correlation matrix (R). For the joint analysis, all data were standardized according to the equation: $z = \frac{x-\bar{x}}{s}$, where z is the standardized value of x, $\bar{x}$ is the mean, and s is the standard deviation.

The absolute value of an eigenvector associated with the eigenvalue of each component (i.e., evaluated trait) determines the importance level of that variable within a given principal component (Fraga *et al.*, 2016 ; Rencher; Christensen, 2012).

The first principal component (PC1) explains the greatest proportion of total additive genetic variance, followed by PC2 and subsequent components. In a dataset with p variables, the eigenvector is estimated as $x' = [x_1, x_2, \ldots, x_p]$ from the correlation matrix (R), with eigenvalue–eigenvector pairs (λ_i, e_i) for $i = 1,2,\ldots,p$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$(Rencher; Christensen, 2012; Johnson; Wichern, 2019). The i-th principal component is calculated as: $PC_i = e_i^t x = e_{i1}x_1 + e_{i2}x_2 + \cdots + e_{ip}x_p$,where e_{ip} represents the i-th eigenvector and x_p the p-th variable from the original data.

Canonical Discriminant Analysis (CDA), also referred to as Canonical Variable Analysis, was applied to the PCA results to identify the most relevant traits. CDA begins with verification of assumptions, followed by a Multivariate Analysis of Variance (MANOVA) to test for significant differences among treatments using multivariate tests such as Wilks' Lambda, Pillai's Trace, Hotelling–Lawley, and Roy's Largest Root (Johnson; Wichern, 2019).

CDA aims to estimate functions of the variables $X = (X_1, X_2, \ldots, X_p)$ that best discriminate *g* groups by using *Z*, a linear combination of the original variables, such that the mean of *Z* differs across groups, thus enabling group differentiation. The method constructs linear combinations of variables $X_i$with coefficients $a = (a_1, a_2, \ldots, a_p)$that maximize the F-ratio in a one-way ANOVA (Hair *et al.*, 2009 ; Johnson; Wichern, 2007): $Z = a_1X_1 + a_2X_2 + a_3X_3 + \ldots + a_pX_p$.

Based on the obtained values, a biplot was constructed for the first two canonical variables (CAN1 and CAN2), allowing joint differentiation of treatments regarding the most important traits. Additionally, based on the canonical values (discriminant function coefficients) generated by CDA for each year and the canonical correlations of the studied traits, a box plot was produced to analyze the distribution, frequency, variability, and central tendency of values, thus enabling comparisons among groups (Naes; Brockhoff; Tomic, 2010).

All analyses were conducted using *R* software version 3.4.1 (R Core Team, 2020), adopting a 5% significance level.

3 Results and Discussion

Table 3 presents the results of the Variance Components Analysis (Individual REML) in dwarf cashew trees, highlighting that only for the traits PRO and PMC the genetic factors surpassed the environmental ones. This is evidenced by the higher values of heritability ± standard deviation of heritability ($H^2g\% \pm s$), heritability adjusted for clone means ($H^2mc\%$), and clone selection accuracy (ACC). It is also apparent that the selection process based on these traits is easier to perform, since the relative coefficient of variation (CVr) exceeds one. This finding is particularly important, as these traits are related to the quantitative yield of cashew nuts, indicating that through selection it is possible to obtain and transmit improved performance to subsequent generations.The remaining traits, however, are strongly influenced by environmental or stochastic factors (Maia *et al.*, 2019), making effective selection unfeasible.

Table 3 – Estimates of variance coefficients, heritability, accuracy, and coefficients of variation for eight traits in dwarf cashew trees resulting from the Variance Components Analysis (Individual REML)

Trait	**Vg**	**Ve**	**Vf**	**$H^2g\%$ ± s**	**$H^2mc\%$**	**ACC**	**CVg%**	**CVe%**	**CVr**	**Mean**
PRO	28,891.23	20,742.36	81,985.62	35.24 ± 19.13	52.77	0.7265	30.1445	25.5420	1.1802	563.8646
PMC	2.2798	0.1824	2.7078	84.20 ± 29.58	91.92	0.9587	14.6353	4.1392	3.5358	10.3169
IPP	0.0002	0.0036	0.0135	1.43 ± 3.85	2.56	0.1599	2.7439	11.8208	0.2321	0.5058
%CF	0.2403	26.3455	44.0668	0.55 ± 2.38	1.44	0.1198	8.6642	90.7239	0.0955	5.6576
ANT	0.0328	0.3346	0.4517	7.26 ± 8.69	22.32	0.4724	16.9944	54.2771	0.3131	1.0657
NMP	0.0123	0.1784	0.1927	6.39 ± 8.15	28.35	0.5324	14.1687	53.9125	0.2628	0.7834
NOC	0.0008	0.1096	0.1575	0.52 ± 2.33	1.59	0.1262	1.9753	22.7971	0.0866	1.4522
NOI	0.0039	0.1268	0.2821	1.39 ± 3.80	3.01	0.1736	9.9136	56.4319	0.1757	0.6310

Legend: Vg – Genetic variance; Ve – Environmental variance; Vf – Phenotypic variance; $H^2g\%$ ± s – Broad-sense heritability ± standard deviation; $H^2mc\%$ – Heritability adjusted for clone means; ACC – Accuracy of clone selection; CVg% – Genetic coefficient of variation; CVe% – Environmental coefficient of variation; CVr – Relative coefficient of variation; Mean – Trait mean.
Source: research data.

It is noteworthy that the genotypic values (m + g) are those that breeding programs aim to estimate, regardless of environmental influences, as they represent the likely true performance of the genotypes.

The new mean estimates correspond to the predictions generated by the BLUP method under field conditions, taking into account environmental effects; thus, the REML/BLUP methodology provides these predicted values (Borges *et al.*, 2010; Pordeus *et al.*, 2013).

As shown in Table 4, the genotypic values (m + g), in general—except for the first clone, which exhibited superior performance for the evaluated trait—were lower than the new mean, indicating that environmental factors undoubtedly contribute to the final performance of the clones.

It is also observed that most traits exhibited genotypic values close to their new mean, and vice versa, consistent with the findings of Pordeus *et al.* (2013), with the exception of PRO and PMC, which showed wider differences.

This suggests that traits related to disease evaluations were relatively stable, while for these two production-related traits there was substantial variation, allowing effective selection and identification of clones with superior performance.

Table 4 – Mean Components (Individual BLUP) of dwarf cashew tree clones (Cl) and their respective genetic values (g), genotypic values (m + g), gain (G), and new means (NM) for eight traits

		PRO						**PMC**		
Cl	**g**	**m+g**	**G**	**NM**		**Cl**	**g**	**u+g**	**G**	**NM**
7	370.22	934.08	370.22	934.08		11	1.88	12.20	1.88	12.20
3	111.22	675.09	240.72	804.58		3	1.78	12.10	1.83	12.15
1	72.72	636.58	184.72	748.58		9	1.73	12.04	1.80	12.11
9	48.08	611.94	150.56	714.42		8	1.26	11.58	1.66	11.98
5	26.67	590.54	125.78	689.65		18	0.64	10.96	1.46	11.77
13	24.34	588.21	108.88	672.74		7	0.61	10.93	1.32	11.63
17	9.08	572.95	94.62	658.48		5	0.60	10.91	1.21	11.53
18	2.02	565.88	83.04	646.91		6	0.55	10.87	1.13	11.45
16	-0.24	563.62	73.79	637.65		14	0.54	10.86	1.07	11.38
10	-4.06	559.81	66.01	629.87		12	0.36	10.68	1.00	11.31
11	-16.49	547.37	58.51	622.37		10	0.30	10.62	0.93	11.25
14	-20.17	543.70	51.95	615.81		4	-0.18	10.14	0.84	11.16
15	-32.03	531.83	45.49	609.35		1	-0.54	9.78	0.73	11.05
6	-76.92	486.94	36.75	600.61		16	-0.85	9.46	0.62	10.94
4	-88.83	475.03	28.37	592.24		17	-1.61	8.71	0.47	10.79
12	-98.30	465.56	20.46	584.32		13	-1.85	8.46	0.33	10.64
8	-111.65	452.21	12.69	576.55		15	-1.85	8.46	0.20	10.51
2	-215.65	348.21	0.00	563.86		2	-3.36	6.96	0.00	10.32
		IPP						**%CF**		
Cl	g	u+g	G	NM		Cl	g	u+g	G	NM
10	0.0045	0.5103	0.0045	0.5103		15	0.106	5.763	0.106	5.763
17	0.0038	0.5095	0.0041	0.5099		7	0.075	5.732	0.090	5.748
2	0.0033	0.5091	0.0039	0.5096		2	0.061	5.718	0.080	5.738
11	0.0012	0.5069	0.0032	0.5090		17	0.055	5.713	0.074	5.732
13	0.0008	0.5065	0.0027	0.5085		13	0.028	5.685	0.065	5.722
4	0.0005	0.5063	0.0023	0.5081		1	0.013	5.671	0.056	5.714
9	0.0001	0.5059	0.0020	0.5078		6	0.012	5.670	0.050	5.708
15	0.0001	0.5059	0.0018	0.5076		9	0.003	5.660	0.044	5.702
6	-0.0003	0.5055	0.0015	0.5073		16	0.001	5.659	0.039	5.697
3	-0.0004	0.5054	0.0014	0.5071		18	-0.002	5.656	0.035	5.693
14	-0.0005	0.5053	0.0012	0.5070		11	-0.008	5.650	0.031	5.689
16	-0.0009	0.5049	0.0010	0.5068		10	-0.018	5.640	0.027	5.685
1	-0.0012	0.5046	0.0008	0.5066		4	-0.030	5.628	0.023	5.680
12	-0.0013	0.5045	0.0007	0.5065		12	-0.046	5.612	0.018	5.675
18	-0.0016	0.5042	0.0005	0.5063		14	-0.049	5.609	0.013	5.671
8	-0.0023	0.5035	0.0004	0.5061		8	-0.062	5.595	0.009	5.666
5	-0.0026	0.5031	0.0002	0.5060		5	-0.066	5.592	0.004	5.662
7	-0.0032	0.5026	0.0000	0.5058		3	-0.072	5.585	0.000	5.658
		ANT						**NMP**		
Cl	g	u+g	G	NM		Cl	g	u+g	G	NM
5	0.151	1.217	0.151	1.217		6	0.098	0.881	0.098	0.881
14	0.104	1.170	0.127	1.193		7	0.074	0.857	0.086	0.869
7	0.088	1.154	0.114	1.180		13	0.065	0.848	0.079	0.862
17	0.080	1.146	0.106	1.172		15	0.065	0.848	0.075	0.859
12	0.076	1.142	0.100	1.166		18	0.037	0.820	0.068	0.851
1	0.025	1.091	0.087	1.153		2	0.026	0.810	0.061	0.844
11	0.020	1.086	0.078	1.144		3	0.018	0.802	0.055	0.838
16	0.020	1.086	0.071	1.136		16	0.009	0.792	0.049	0.832
9	0.017	1.082	0.065	1.130		12	-0.005	0.779	0.043	0.826
10	-0.016	1.050	0.057	1.122		5	-0.013	0.770	0.037	0.821
8	-0.023	1.043	0.049	1.115		10	-0.033	0.751	0.032	0.814

4	-0.037	1.028	0.042	1.108		14	-0.033	0.751	0.026	0.809
13	-0.051	1.014	0.035	1.101		4	-0.033	0.750	0.021	0.804
2	-0.072	0.994	0.027	1.093		9	-0.047	0.737	0.016	0.800
15	-0.075	0.991	0.021	1.086		11	-0.047	0.737	0.012	0.795
18	-0.075	0.991	0.015	1.080		17	-0.047	0.737	0.008	0.792
3	-0.100	0.966	0.008	1.074		1	-0.065	0.718	0.004	0.787
6	-0.132	0.934	0.000	1.066		8	-0.069	0.714	0.000	0.783
		NOC						**NOI**		
Cl	g	u+g	G	NM		Cl	g	u+g	G	NM
7	0.0063	1.459	0.0063	1.459		6	0.0257	0.657	0.026	0.656
4	0.0040	1.456	0.0051	1.457		2	0.0111	0.642	0.018	0.649
5	0.0035	1.456	0.0046	1.457		4	0.0095	0.641	0.015	0.646
9	0.0016	1.454	0.0038	1.456		3	0.0030	0.634	0.012	0.643
17	0.0016	1.454	0.0034	1.456		9	0.0000	0.631	0.010	0.641
3	0.0012	1.453	0.0030	1.455		10	0.0000	0.631	0.008	0.639
6	0.0009	1.453	0.0027	1.455		11	0.0000	0.631	0.007	0.638
12	0.0006	1.453	0.0024	1.455		12	0.0000	0.631	0.006	0.637
18	0.0006	1.453	0.0022	1.454		13	0.0000	0.631	0.006	0.636
13	0.0003	1.452	0.0020	1.454		14	0.0000	0.631	0.005	0.636
11	-0.0004	1.452	0.0018	1.454		15	0.0000	0.631	0.005	0.635
14	-0.0007	1.452	0.0016	1.454		16	0.0000	0.631	0.004	0.635
15	-0.0007	1.452	0.0014	1.454		17	0.0000	0.631	0.004	0.635
16	-0.0014	1.451	0.0012	1.453		18	0.0000	0.631	0.004	0.635
10	-0.0017	1.451	0.0010	1.453		8	-0.0011	0.630	0.003	0.634
8	-0.0046	1.448	0.0007	1.453		5	-0.0088	0.622	0.004	0.633
2	-0.0054	1.447	0.0003	1.453		1	-0.0148	0.616	0.001	0.632
1	-0.0055	1.447	0.0000	1.452		7	-0.0245	0.606	0.000	0.631

Source: research data.

The Individual BLUP mean components for eight traits evaluated in dwarf cashew tree clones from the Pacajus, CE experimental site are presented in Table 4. Data analysis indicated that clones 7, 3, 1, 9, 5, 13, 17, and 18 showed the best performance for the PRO trait. For PMC, the superior clones were 11, 3, 9, 8, 18, 7, 5, 6, 14, 12, and 10. Regarding IPP, the best-performing clones were 10, 17, 2, 11, 13, 4, 9, and 15. For %CF, the top clones were 3, 5, 8, 14, 12, 4, 10, 11, and 18. For ANT, clones 6, 3, 18, 15, 2, 13, 4, 8, and 10 stood out. Regarding NMP, the superior clones were 8, 1, 17, 11, 9, 4, 14, 10, 5, and 12. For NOC, the best clones were 1, 2, 8, 10, 16, 15, 14, and 11, and for NOI, 7, 1, 5, and 8 performed best. Clones 1, 3, 4, 5, 7, 8, 9, 10, 11, and 18 were superior in at least four traits, while clones 5, 8, and 10 excelled in more than four traits. Conversely, Pordeus *et al.* (2013) did not observe genotype coincidences among the traits evaluated.

Regarding the Multivariate Analysis of Variance (MANOVA), according to Song (2013), four main statistical methods are commonly used — Wilks' Lambda, Pillai's Trace, Hotelling-

Lawley Trace, and Roy’s Largest Root — to test the hypothesis that all samples originate from the same genetic material with identical mean vectors.

The multivariate tests (Table 5) detected significant differences using the F-test among the different evaluation years, indicating the influence of distinct environmental or stochastic factors (Maia *et al*., 2019) on the observed variations. Specifically, the years 2016 and 2018 exhibited rainfall patterns that differed from those recorded in 2017 and 2019.

Table 5 – Summary of the multivariate analysis of variance, using four statistical tests, indicating numerator degrees of freedom (NGL), denominator degrees of freedom (DGL), and F-test probability for seven response variables evaluated simultaneously: PRO, PMC, IPP, NAN, NMP, NOC, and NOI

Effect	**Statistical Test**	**Value**	**F Value**	**NGL**	**DGL**	**Prob F**
Years	Wilks’ Lambda	0.027	61.827	24	595.16	**
	Pillai’s Trace	2.027	53.892	24	621	**
	Hotelling–Lawley Trace	8.104	68.770	24	611	**
	Roy’s Largest Root	5.135	132.87	9	207	**

PRO: productivity (kg·ha^{-1}); PMC: average nut weight (g); IPP: precocity index; ANT: score for anthracnose; NMP: score for black mold; NOC: score for brown powdery mildew; and NOI: score for powdery mildew on the inflorescence.
Source: research data.

In Table 6, it can be observed that, regarding the canonical variables, there were statistically significant differences at the 1% probability level, as indicated by the F-test probability values. Additionally, it is verified that only three canonical variables were sufficient to explain 100% of the total data variability, with the first two accounting for more than 80% of the cumulative variance, which is considered adequate to ensure reliability and a proper interpretation of the results.

This outcome allows for an efficient understanding of the joint behavior of the traits evaluated in the experiment (Cruz *et al*., 2012).

The cumulative percentage in the first two canonical variables is similar to that observed by Teixeira *et al*. (2013) in coffee plants and by Morais *et al*. (2019) in *Dalbergia ecastaphyllum*, confirming that the use of two canonical variables is sufficient and reliable to interpret the behavior of dwarf cashew clones for the analyzed traits.

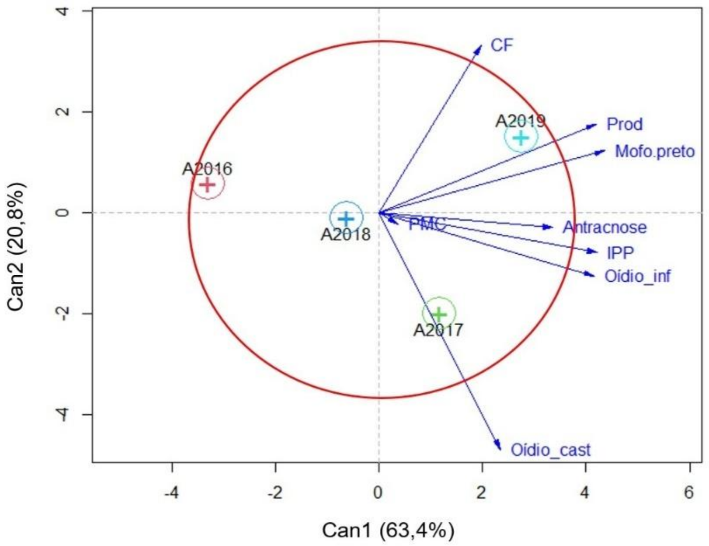
Table 6 – Canonical variables (CV), canonical discriminant function (Can), F-test approximation (Fap), eigenvalues (λi), percentage of variance explained by canonical variables (VVC), and cumulative percentage (aVVC) for eight traits of dwarf cashew clones. Pacajus, CE

CV	**Can**	**Fap**	**λ_i**	**VVC (%)**	**aVVC (%)**
CV1	0.8370	61.827**	5.1351	63.366	63.366
CV2	0.6271	43.452**	1.6820	20.755	84.121
CV3	0.5627	44.395**	1.2868	15.879	100.000

Significant at 1% probability level by the F-test. **0**
Source: research data.

A canonical biplot involving eight traits of dwarf cashew clones associated with different evaluation years is presented in Figure 1. The purpose of the canonical biplot is to illustrate the main differentiations based on the discrimination among treatment groups (dwarf cashew clones and years), according to González-Martin *et al*. (2016). In canonical discriminant analysis, a direct approach is adopted, visually assessing the projection of the two canonical discriminant functions (Sorbolini *et al*., 2016).

Figure 1 – Canonical discriminant biplot showing the correlation between years and evaluated traits in dwarf cashew clones



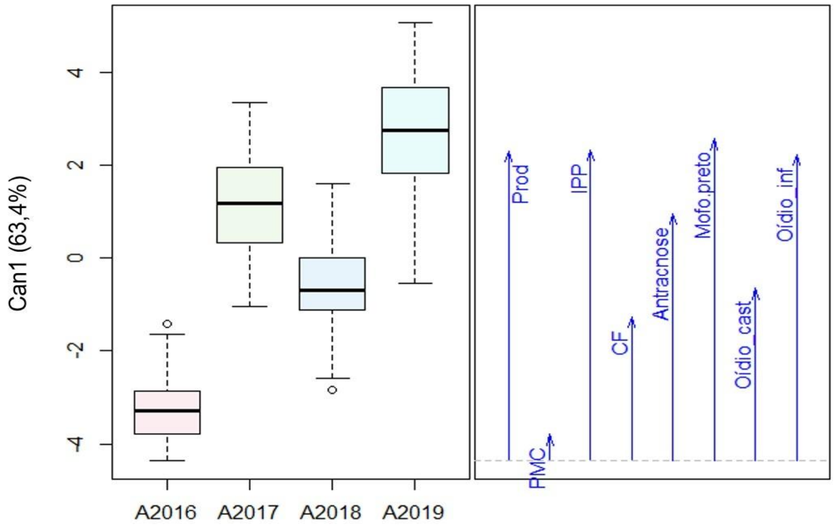
Source: reserch data.

The axis of the first canonical discriminant function encompasses an important and significant portion of the variance, accounting for 63.4% of the total variation and allowing the ordering of the effects of years and evaluated characteristics. It is noted that the variations inherent to the years 2017 and 2019 contributed the most in Figure 1 to distinguish the variables in different areas of the plot, as they were positioned farther from the axes, while 2018 showed the smallest contribution, located at the opposite extreme. The year 2017 directly influenced the NOI characteristic, whereas 2019 was more strongly associated with PRO and NMP. Furthermore, except for PMC — which undoubtedly exhibited the greatest genetic effect compared to non-genetic ones — the clones interacted with environmental effects, showing high G×A (genotype-by-environment) interactions that contributed to the observed differences in the other traits.

In summary, based on the Biplot, the variations in 2017 and 2019 contributed more strongly to overall variation compared to the clones, while 2018 had the least influence. The year 2017 directly affected NOI, whereas 2019 mainly influenced PRO and NMP.

Using the group scores obtained through the Biplot, a Box Plot (Figure 2) was constructed to represent the distribution of the data set and to assess its symmetry and dispersion. This type of graph is particularly suitable for comparing two or more data sets — in this case, the different years — allowing visualization of how the canonical variable distinguished the groups (years) and which characteristics were most relevant for this differentiation.

Figure 2 – Box plot showing the distribution of canonical analysis scores by year for the evaluated characteristics in dwarf cashew clones



Source : research data.

Based on the Box Plot structured with the first canonical function, it is observed that all years directly and positively influenced all evaluated characteristics (Figure 2), differing only in intensity, with 2019 showing the greatest discriminant capacity.

Regarding the characteristics, all years contributed positively to differentiation; however, the greatest contributions were observed for PRO, IPP, NMP, and NOI, which were therefore strongly influenced by stochastic factors. Intermediate behavior was noted for CF, ANT, and NOC, while PMC showed the lowest variation, indicating minimal environmental influence. This behavior contrasts with that reported by Nascimento *et al.* (2021), in which some traits contributed positively and others negatively.

It is important to emphasize that the canonical discriminant analysis of the full set of characteristics was effective in identifying the relative importance of each variable in discriminating among the four year groups.

When associating the behavior of clones across the evaluation years with rainfall data (Table 2), it is observed that 2016 was indeed the most restrictive year, while 2017 and 2019 had greater water availability, favoring higher contributions of the evaluated characteristics, including those related to disease incidence. This indicates that water availability is an important factor influencing both clone productivity and phytosanitary aspects. In contrast, PMC was minimally affected, confirming its strong association with genetic (heritability) factors in the clones.

4 Conclusion

The clones 1, 3, 4, 5, 7, 8, 9, 10, 11, and 18 stood out as superior in at least four traits, while clones 5, 8, and 10 showed superiority in more than four traits. The traits PRO (productivity) and PMC (average nut weight) demonstrated greater ease of selection and heritability across generations.

The years had a direct and positive influence, with stronger effects observed for PRO, IPP, NMP, and NOI, which were more affected by stochastic factors, whereas PMC was the least influenced by environmental variation.

References

ANJOS JÚNIOR, O.R. et al. O rendimento da castanha de caju: uma análise espacial para o estado da Paraíba. *Reflexões Econôm.*, v.2, n.2, p.37-54, 2016.

BORGES, V. Seleção de clones de batata-doce pelo procedimento REML/BLUP. *Acta Scientiarum. Agronomy Maringá*, v. 32, n. 4, p. 643-649, 2010.

BRAINER, M.S.C.B. Cajucultura. *Cad. Setorial Etene*, v.7, n.230, 2022.

CILAS, C.; MONTAGNON, C.; BAR-hEN, A. Yield stability in clones of Coffea canephora in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Gen. Genom.*, v.7, p.421-429, 2011.

CRUZ, C.D.; CARNEIRO, P.C.S.; REGAZZI, A.J. *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: Editora UFV, 2014.

CRUZ, M.C.M. et al. Fluorescência da clorofila em folhas de tangerineira 'Ponkan' e limeira ácida 'Tahiti' submetidas ao estresse hídrico. *Rev. Bras. Fruticul.*, v.31, p.896-901, 2009. doi: https://doi.org/10.1590/S0100-29452009000300037.

FARIA, A.P. *et al.* CAJU: o sabor do Nordeste. *Rev Gastron.*, p. 1-15, 2018.

FRAGA, A.B. *et al.* Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. *Trop. Anim. Health Prod.*, v.48, n.3, p.533-538, 2016.

FUNCEME - Fundação Cearense de Meteorologia e Recursos Hídricos. *Portal Hidrológico do Ceará – Calendário das Chuvas no Estado do Ceará*. 2023. Disponível em: http://www5.funceme.br/app/calendario/produto/municipios/maxima/anual.

GONZÁLEZ-MARTIN, M.I. *et al.* The role of the canonical biplot method in the study of volatile compounds in cheeses of variable composition. *Grasas Aceites*, v.67, n.1, p.1-8, 2016. doi: https://doi.org/10.3989/gya.0250151

HAIR JUNIOR, J.F. et al. *Análise multivariada de dados*. Porto Alegre: Bookman, 2009.

HAIR, J.F. et al. *Multivariate Data Analysis*. Harlow: Pearson Education, 2014.

IBGE 2022 *Levantamento Sistemático da produção agrícola. 2022.* Disponível: https://sidra.ibge.gov.br/home/lspa/brasil

ICB - Instituto Caju Brasil. *O agronegócio caju em números*. 2020. Disponível em: https://cajubrasil.org/wp-content/uploads/2020/05/Boletim-ICB_7.pdf

INC - International Nut and Dried Fruit Council Foundation. *Nuts & Dried Fruits Statistical Yearbook.* 80p., 2022. https://www.nutfruit.org/files/tech/1651579968_Statistical_Yearbook_2021-2022.pdf

JOHNSON, A.R.; WICHERN, D.W. *Applied Multivariate Statistical Analysis*. Prentic Hall 2007.

OHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. New Jersey: Pearson. 2019.

KAHLMANN, K.; KOHN, M. *USDA/FAS Food for Progress LIFFT-Cashew*. SeGaBi Cashew Value Chain Study, n.2, 2018.

MAIA, M.C.C. et al. Principal component and biplot analysis in the agro-industrial characteristics of Anacardium spp. *Euro. Sci. J.*, v.15, p.21-31, 2019.

MORAES, I.V.M. et al. *Aproveitamento industrial do pedúnculo de caju. In: Agronegócio caju: práticas e inovações*. Fortaleza: Embrapa, 2013.

MORAIS, D.V. et al. Leaf geometric morphometrics among populations of Dalbergia ecastaphyllum (L.) Taub. *Biosc. J.*, v.35, n.6, p.1789-1798, 2019. doi: http://dx.doi.org/10.14393/BJ-v35n6a2019-3981.

NAES, T.; BROCKHOFF, P.B.; TOMIC, O. *Statistics for sensory and consumer science*. United Kingdom: John Wiley and Sons, 2010.

NASCIMENTO, A.S. et al. Qualidade Físico-Química e Compostos Bioativos do Mel de Melipona scutellaris Produzido em uma Região Urbano-Industrial. *Rev. Virt. Quím.*, v.13, n.6, p.1268-1277, 2021.

PORDEUS, R.V. et al. Estimativa de parâmetros genéticos do cajueiro anão precoce em um solo arenoso pelo procedimento reml/blup. *Rev. Verde Agroecol. Desenvol. Sust.*, v.8, n.3, p.41-51, 2013.

RENCHER, A.C.; CHRISTENSEN, W.F. *Methods of multivariate analysis*. New York: Wiley & Sons, 2012.

RESENDE, M.D.V. *Software SELEGEN-REML/BLUP : sistema estatístico e seleção genética computadorizada via modelos lineares mistos*. Colombo: Embrapa Florestas, 2007. 359 p

RESENDE, M.D.V. Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, 16: 330-339, 2016.

SERRANO, L.A.L. *Sistema de produção do caju*. Fortaleza: Embrapa Agroindústria Tropical, 2016.

SONG, F.N. *Técnicas de análise multivariada com aplicações a dados de natureza Biológica*. Rio Claro: Universidade Estadual Paulista, 2013.

SORBOLINI, S. et al. Use of canonical discriminant analysis to study signatures of selection in cattle. *Genet. Selec.Evol.*, v.48, n.1, 2016. doi: https://doi.org/10.1186/s12711-016-0236-7

STURION, J.A.; RESENDE, M.D.V. Avaliação genética e análise de deviance em um teste desbalanceado de procedência e progênie de Ilex paraguariensis. *Pesq. Florestal Bras.*, v.30, p.157-160, 2010

TEIXEIRA, A.L. et al. Análise de componentes principais em caracteres morfológicos de café arábica em estádio juvenil. *Coffee Sci.,* v.8, n.2, p.205-210, 2013.