



# Genetic diversity for agronomic and bromatological traits in forage cactus

Mariana Santos de Jesus<sup>1\*</sup>, Adriana Rodrigues Passos<sup>1</sup> and Rafael Parreira Diniz<sup>2</sup>

<sup>1</sup>Instituto Federal de Sergipe, Av. Padre Airton Gonçalves Lima, 1140, 49500-000, São Cristóvão, Campus Itabaina, Itabaiana, Sergipe, Brazil. <sup>2</sup>Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil. \*Author for correspondence. E-mail: [maryanamell@gmail.com](mailto:maryanamell@gmail.com)

**ABSTRACT.** This work aimed to estimate the genetic diversity in accessions of *Opuntia ficus-indica* collected in 13 regions of the semiarid region of Bahia. A total of 65 accessions were evaluated in a randomized complete block design, with three replications, at the Rio Seco experimental station belonging to the State University of Feira de Santana, Amélia Rodrigues-BA. Characterization of the accessions was made through the evaluation of 17 descriptors, namely 11 agronomic and six bromatological. The average Euclidean distance was used to estimate the genetic diversity among accessions. The shortest distances were obtained for accessions from the same collection site while the largest were observed in accessions 54 and 62 (10.32 DE) and 63 and 3 (10.22 DE). The analysis of canonical variables indicated cladodes total number (CTN), plant width (PW), CL, plant height (PH), ether extract (EE), and dry weight (DW) for discard as they presented the lowest contribution of the data set variation. Principal component analysis and *K-means* method were used to establish the clusters, and the formation of four groups was indicated. The first two principal components captured 52.5% of the total variation present in accessions. The descriptors with the greatest contribution to the variation observed in *O. ficus-indica* were total cladode photosynthetic area (TCPA), cladode area (CA), and cladodes width (CW). There is divergence between cactus forage accessions collected in the semiarid region of Bahia. This information will allow the use of these materials for the formation of segregating populations in the genetic improvement program of the State University of Feira de Santana. The accessions of groups III and IV should be explored by the forage cactus breeding program, as they presented greater productive potential.

**Keywords:** cactus forage; genetic improvement; multivariate analysis.

Received on March 3, 2022.

Accepted on October 6, 2022.

## Introduction

The semiarid climate is present throughout the Brazilian Northeast region and North of Minas Gerais, representing an approximate area of 982,563.3 km<sup>2</sup>. In Bahia, the semiarid region covers 360,000 km<sup>2</sup>, which correspond to 64% of its total area, encompassing 258 municipalities. Across the state, 762,620 agricultural establishments are presented where approximately 2,078,469 people are employed, with meat farming and the production of milk and its derivatives representing the main source of income and employment (IBGE, 2017).

Semiarid climate is characterized by low rainfall rates, where rains show high seasonal variations and are concentrated between 3-5 months throughout the year (Leite, Silva, Andrade, Pereira, & Ramos, 2014). Climatic factors, as well as low forage support capacity typical of the Caatinga, impose serious restrictions on the extensive production system practiced in most of the state of Bahia, since periods of drought cause serious reductions in the availability and quality of pasture available to herds (Galvão Júnior, Silva, Morais, & Lima, 2014). In this context, the use of adapted forage species, such as the cactus forage (*Opuntia ficus-indica* Mill.) enables productivity to be maintained during periods when natural pastures are reduced, constituting an important strategy for food security for herds (Ramos et al., 2015). Pereira et al. (2021), noted that the use of spineless cactus in the form of silage in sheep diets, resulted in higher intakes of dry matter, organic matter, neutral detergent fiber, ether extract, non-fibrous carbohydrates, and total digestible nutrients; and higher digestibility coefficients of dry matter, organic matter, and total digestible nutrients.

Cactus forage production in Brazil corresponds to about 3.59 billion tons per year, where the Northeast region provides 62% of this productivity, while Bahia appears as the largest national producer with 42% of the

production (IBGE, 2017). Under semiarid conditions, the forage cactus has an average phytomass production of 470,000 kg ha<sup>-1</sup>, approximately (Silva, 2014; Silva et al., 2016).

The cactus forage is a perennial plant, vegetatively propagated, easy to grow, with high water content, and palatable, having around 75% digestibility coefficient (Silva et al., 2016). Regardless of its genus, this species has total carbohydrate contents ranging from 73.8 to 85.2%, non-fibrous carbohydrates from 42.3 to 65.0%, and minerals from 8.1 to 17.7% (Cavalcante, Santos, Silva, Fagundes, & Silva, 2014). On the other hand, it shows low dry matter contents (6.1 to 17.1%), crude protein contents (2.9 to 6.0%), neutral detergent fiber (20.1 to 32.8%), and acid detergent fiber (9.5 to 22.5%) (Wanderley et al., 2012; Pessoa et al., 2013). Therefore, there is genetic variability for several important traits and genetic improvement of this species can favor the selection of superior genotypes in terms of agronomic performance and bromatological composition considering the possibility of improving these attributes.

Genetic improvement programs for this species are being developed in the States of Pernambuco, Paraíba, and Bahia, in order to increase its productivity and nutritional traits, allowing expanding its genetic base and generating information for the use and conservation of its genetic resources. At the State University of Feira de Santana, in Bahia State, Brazil, the genetic improvement program has been developed by the department of biology sciences, program is developing activities focused on the cactus forage, which involve characterization, genetic evaluation, and selection. The next steps involve hybridization of the most divergent and promising genotypes, as well as the conduction of the program through selection by hybridization, thus expanding its genetic base and allowing advances in the improvement of this species, in order to develop superior genotypes that allow to select superior clones to be released to producers.

One of the basic premises for the development of improvement research is the presence of genetic variability in the population under selection. Several studies and techniques to quantify the genetic diversity in plant germplasm are used, such as the estimation of genetic divergence among genotypes (Cruz & Ragazzi, 2012). On the other hand, as the characterization of accessions is carried out considering several traits, those methodologies that capture a large part of the genetic divergence, such as multivariate analysis, can increase the efficiency of this step and generate more precise results on the genetic structure of the population.

Multivariate techniques have been an important tool in the evaluation of genetic diversity and represent a set of methods that allow a global study of the evaluated traits, pointing out correlations, similarities, or differences between them, and preventing the least possible loss of information (Freitas et al., 2013). Among the multivariate techniques, the principal component analysis (PCA) and the canonical variable analysis (CVA) stand out. PCA represents an important methodology in the discrimination of accessions and in reducing the variables to a limited number of new variables (Maciel et al., 2018). Moreover, CVA is a technique that allows the reduction of the dimensionality of the data, similar to PCA, but being used especially in discriminant analyzes carried out from samples with repeated observations (Cruz & Ragazzi, 2012). According to Silva, Santos, Pinto, and Ramos (2020), multivariate techniques are important tools for genetic improvement, since they allow combining the multiple information contained in the experimental unit, so that it is possible to make a selection based on a complex of variables and also to discriminate more promising materials in different contexts.

Thus, this research aimed to quantify the genetic diversity in cactus forage (*O. ficus-indica*) accessions collected in the State of Bahia, by using agronomic and bromatological descriptors, in order to identify potential parents for the formation of segregating populations that can be used in the genetic improvement program of this species.

## Material and methods

### Access collection

Expeditions to collect accessions of cactus forage belonging to the species *O. ficus-indica* var. gigante were carried out in 13 populations, distributed through different regions of the State of Bahia, covered by semiarid climate: Caetité (CA), Vitória da Conquista (VC), Ruy Barbosa (RB), Feira de Santana (FS), Jeremoabo (JE), Ipirá (IP) Morro do Chapéu (MC), Riachão do Jacuípe (RJ), Valente (V), Juazeiro (JU), Senhor do Bonfim (SB), Jaguarari (JA), and Irecê (IR). In each population, three cladodes from five different plants were collected, obeying a distance of about 20 meters between the plants and considering different phenotypic aspects. The most vigorous plants that showed good phytosanitary conditions were selected. After collection, cladodes were left in the sun for 15 days to heal the cutoff point and prevent attacks from fungi present in the soil.

### Location and agronomic experimentation

The experiment was carried out at the Rio Seco Experimental Station, belonging to the State University of Feira de Santana-BA, which is located in the municipality of Amélia Rodrigues, at 12°23'30" S, 38°45'24" W, and 217 m. During the validity period of the experiment, a rainfall level of 1,270 mm<sup>3</sup> and an average temperature of 24.6°C were registered (INMET, 2017). The climatic condition of Amélia Rodrigues is classified as tropical rainforest (Af) according to the Köppen index. Before planting, a soil analysis of the experimental area was carried out, nonetheless, corrections of its chemical properties were not required. Soil preparation was made by plowing and sieving.

The experiment was installed under a randomized complete block design, with three replications and an experimental plot of five plants, with a spacing of 1.0 m between rows and 0.5 m between plants. At planting, one cladode per hole was used, arranged in a vertical position, having its cut part facing the ground, deep enough to prevent cladodes from falling.

Agronomic and bromatological descriptors were evaluated 12 months after planting. Regarding the morphological characterization, the following descriptors were evaluated: plant height in cm (PH); plant width in cm (PW); cladodes total number (CTN); cladodes length in cm (LC); cladodes width in cm (CW); cladodes diameter in mm (CD); green mass in g (GM); dry weight in g (DW); percentage of dry matter (DM) obtained by the formula  $DM = [(dry\ mass\ (g)/wet\ mass\ (g)) \times 100]$ ; each cladode area (CA) was determined by the expression:  $[CA = length \times width \times 0.632]$ . The total cladode photosynthetic area (TCPA) was obtained by the following formula:  $[TCPA = CA \times CTN]$ . Measurements of plants and cladodes were obtained using a measuring tape and caliper, while for green and dry weight a high precision digital scale was used.

The following bromatological descriptors were evaluated: crude protein content (CPC), obtained using the Kjeldahl method (AOAC, 2019); Furthermore, ether extract (EE), neutral detergent fiber (NDF), and acid detergent fiber (ADF) were calculated by the method of Van Soest (AOAC). Besides, the content of total digestible nutrients (TDN) was estimated by the expression  $\%TDN = 87.84\% - (0.70\% + \%ADF)$  and non-fibrous carbohydrates (NFC) by the expression  $\%NFC = 100 - (CPC + EE + NDF + Ashes)$ . All data were subjected to analysis of variance, taking into account the assumptions of the analysis, such as the normality of the residuals. From this information, adjusted phenotypic means of all traits were obtained for each accession.

In order to evaluate the significance of the effect of the accessions, a deviation analysis by using LRT (likelihood ratio test) was performed. For this, the following equation was used (Sturion & Resende, 2010):  $D = -2\text{Log}L$ , where D is the deviance value, while L is the maximum point of the residual likelihood function. The analysis was performed by using the Selegen-REML/BLUP software (Resende, 2016).

Before quantifying the genetic divergence among the 65 cactus forage accessions, the adjusted phenotypic means of the 17 evaluated traits were submitted to the analysis of canonical variables in order to discard redundant descriptors that have low explanation of the genetic variance. The recommendation for discarding redundant quantitative descriptors was based on the method by direct selection (Jolliffe, 2014), which suggests discarding any descriptor that presents the highest weighting coefficient in absolute value in the principal component with the lowest eigenvalue, starting from the last component to the one whose eigenvalue does not exceed 0.70.

Subsequently, the Mantel (1967) test was performed to verify the significance of the geographic distance on the distance between accessions. After discarding the descriptors that contributed the least to the explanation of genetic variation, the Euclidean distance was estimated. Data standardization, using the Z index, was performed in order to eliminate the differences between the scales used in the assessment of descriptors, according to the following formula:  $Z_i = (X_i - X) / S_i$ , where  $Z_i$  is the standardized value for a given descriptor 'i' with magnitude  $X_i$ , general mean  $X$ , and standard deviation  $S_i$ , with  $\sim N(0,1)$ . Graphical representation through the heat map was obtained through the Euclidean distances between the accessions. The principal components analysis (PCA) was processed from the covariance matrix of the original variables, obtaining from it the eigenvalues that built the eigenvectors (Hair, William, Babin, Anderson, & Tatham, 2009). The definition of the number of clusters to group the 65 accessions was made through the *k-means* method. All analyzes were performed using the *easynova*, *factoextra*, *tidyverse*, and *cluster* packages, installed in the R software version 3.5.1 (R Development Core Team, 2018).

## Results

### Genetic variability in cactus forage

The deviance analysis revealed significance for the effect of accessions for all evaluated descriptors (Table 1), which evidence the existence of genetic variability.

**Table 1.** Summary of deviance analysis for the effect of accessions and the complete model, and maximum likelihood ratio test for agronomic and bromatological descriptors evaluated in 65 cactus forage accessions.

Descriptors	Deviance		LRT <sup>‡</sup> (Chi-Squared)
	Genotype	Complete model	
Plant height (cm)	1480.92	1306.96	173.96**
Plant width (cm)	1547.88	1432.88	115.00**
Cladodes total number	673.84	656.76	17.08**
Cladodes length (cm)	1043.67	1022.80	20.87**
Cladodes width (cm)	822.67	813.56	9.11**
Cladodes diameter (mm)	1147.32	1029.12	118.20**
Green mass (g)	2424.03	2338.00	86.03**
Dry weight (g)	1482.45	1377.96	104.49**
Dry mass (%)	446.52	310.48	136.04**
Cladode area (cm <sup>2</sup> )	2078.48	2063.20	15.28**
Total cladode photosynthetic area (cm <sup>2</sup> )	2907.65	2889.00	18.65**
Crude protein content (%)	1528.02	399.72	1128.30**
Acid detergent fiber (%)	840.25	430.88	409.37**
Neuter detergent fiber (%)	905.79	225.14	680.65**
Ether extract (%)	1205.09	470.48	734.61**
Total digestible nutrients (%)	675.50	180.74	494.76**
Non-fibrous carbohydrate	1018.72	612.92	405.80**

<sup>‡</sup>LRT – Maximum likelihood ratio test for tabulated chi-squared.

In general, the maximum values for Euclidean distance (ED) are associated with accessions collected in different regions. Accessions 54 and 56 showed the greatest distance (10.32 ED), being the most diverse from each other, coming from the regions of Senhor do Bonfim and Irecê, respectively. Accessions collected in Irecê were the most divergent in relation to other collection sources. The smallest distances were observed between accessions from the same collection origin, with the exception of accessions 6, 22, and 40, which presented smaller distances when compared with accessions collected from other populations (Table 2).

**Table 2.** Maximum and minimum Euclidean distances values between 65 accessions of *O. ficus-indica* collected in 13 localities in the state of Bahia and growing at the Rio Seco Experimental Station, Amélia Rodrigues, Bahia State, Brazil.

*Loc	Gen	Gen.Max	Dist.	Gen.Min	Dist.	Loc	Gen	Gen.Max	Dist.	Gen.Min	Dist.
Caetité	1	61	8.09	5	2.02	Riachão do Jacuípe	34	19	7.64	35	2.57
	2	63	8.77	1	2.27		35	19	7.63	34	2.57
	3	63	8.77	1	2.27		36	62	6.74	37	2.87
	4	61	9.71	2	2.56		37	19	7.67	36	2.87
	5	63	7.98	1	2.02		38	62	8.45	39	1.93
Vitória da Conquista	6	60	7.56	24	2.83	Valente	39	62	9.19	38	1.93
	7	62	8.25	9	2.17		40	63	7.84	22	2.62
	8	60	7.19	10	2.66		41	61	7.19	45	2.36
	9	60	7.79	7	2.17		42	17	8.52	43	2.26
	10	60	7.68	9	2.34		43	62	8.65	44	2.24
Rui Barbosa	11	63	7.81	13	2.56	Juazeiro	44	62	9.76	43	2.24
	12	61	7.66	13	1.77		45	3	6.88	41	2.36
	13	61	8.26	12	1.77		46	19	8.07	48	2.50
	14	63	8.66	13	1.99		47	44	6.96	50	2.31
	15	44	8.24	11	3.26		48	44	6.91	49	2.46
Feira de Santana	16	62	6.97	20	2.45	Senhor do Bonfim	49	24	7.74	48	2.46
	17	3	9.54	18	3.78		50	40	6.86	47	2.31
	18	3	8.50	19	2.85		51	61	7.57	55	2.83
	19	62	8.88	18	2.85		52	62	7.62	55	2.24
	20	62	7.87	16	2.45		53	61	9.56	55	2.46
Jeremoabo	21	60	7.73	23	2.68	54	62	10.32	53	2.84	
	22	60	7.60	40	2.62	55	61	8.12	52	2.24	

	23	60	7.77	21	2.68		56	32	7.84	57	2.73
	24	60	8.34	25	2.21		57	3	8.94	60	2.63
	25	60	8.47	24	2.21	Jaguarari	58	44	9.68	60	2.81
	26	44	7.07	27	2.38		59	3	8.69	60	2.37
	27	44	8.24	26	2.38		60	3	9.19	59	2.37
Ipirá	28	54	8.32	30	2.43		61	54	10.26	62	2.28
	29	44	6.74	22	2.88		62	54	10.32	61	2.28
	30	44	8.04	28	2.43	Irecê	63	3	10.22	65	2.26
Morro	31	58	8.54	33	2.14		64	54	8.77	65	2.18
Chapéu	32	17	8.63	33	3.28		65	3	8.76	64	2.18
	33	60	8.15	31	2.14						

\*Loc: Genotype collections site; Gen: genotype; Gen.Max: Genotype with maximum distance; Gen.Min: Genotype with minimum distance; Dist: Euclidian distance.

The geographic and dissimilarity matrices evidenced significant correlation when analyzed by T and Mantel tests (1 and 5%), however of low magnitude (Table 3).

**Table 3.** Correlation between geographic data and the dissimilarity matrix observed in 65 accessions of *O. ficus-indica* collected in different populations of the State of Bahia.

Matrix correlation	
Correlation	0.341
Number of data	2080
T-test	0.00**
Mantel test	0.01359++

\*\* : Significant at 1% probability by t-test; ++ : Significant at 1% probability by Mantel test based on 10,000 simulations.

### Discarding features with low variation

Descriptors were discarded based on weighting coefficients associated with canonical variables of eigenvectors under 0.70. Six descriptors were indicated for discarding, namely five agronomic and one bromatological. Descriptors of lesser importance, in order of discarding, were CTN, PW, CL, PH, EE, and DW, whose highest eigenvalues of the modulus occurred in high canonical variables. Therefore, those descriptors are of lesser relative importance in the analysis of the data set and for determination of the genetic diversity among the accessions. The remaining descriptors were selected for their relevance in the discrimination of *O. ficus-indica* accessions. The first four canonical variables accumulated 99.99% of the variation observed in the data set (Table 4).

**Table 4.** Estimation of the weighting coefficients associated with the canonical variables of eigenvectors less than 0.70 with identification of the discarded descriptor and estimation of the eigenvalues associated with the canonical variables and their total and accumulated variances, obtained from the 17 descriptors evaluated.

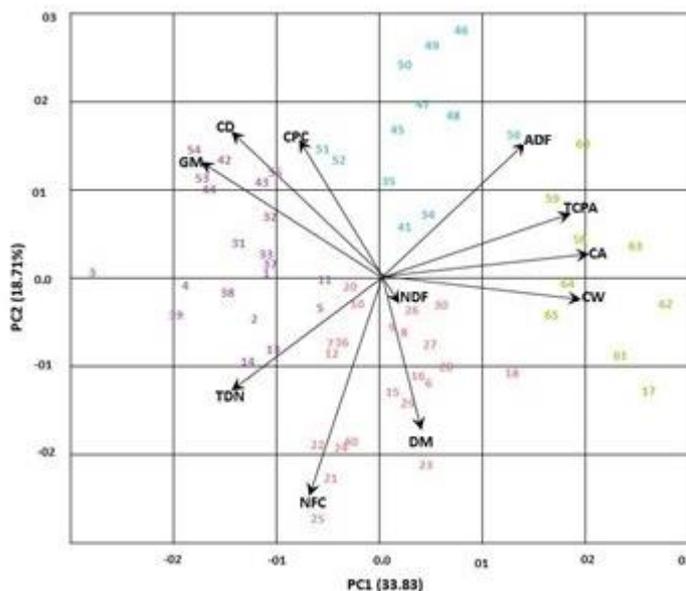
Discard	Descriptor	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9
4	PH	-0.01	0.04	0.14	-0.09	-0.06	0.08	-0.18	-0.24	0.28
2	PW	-0.21	0.09	0.30	0.18	-0.05	0.28	-0.31	0.27	-0.04
1	CTN	-0.02	-0.06	0.06	0.17	0.07	0.21	-0.01	-0.12	0.18
3	CL	-0.18	0.06	0.28	0.01	-0.17	-0.12	0.22	-0.27	0.18
	CW	1.08	-0.01	-0.03	0.02	0.01	-0.17	-0.90	0.78	-0.16
	CD	0.15	-0.60	0.36	0.34	-0.07	-0.02	0.11	0.27	0.78
	GM	-0.06	0.41	0.06	0.05	0.46	0.91	0.29	0.23	0.22
6	DW	-0.17	-0.05	0.53	-0.06	0.03	-0.05	-1.32	0.99	-0.11
	DM	-0.017	0.81	0.07	0.12	-0.12	-0.042	-1.32	1.38	0.07
	CA	0.15	0.02	0.29	0.30	0.47	-0.24	1.01	-0.13	-0.14
	TCPA	-0.09	0.19	-0.56	1.18	-0.34	0.02	-0.62	0.69	-0.24
	CPC	-0.09	0.05	0.55	0.76	-0.32	0.01	-1.51	1.53	-0.57
	ADF	-0.18	-0.12	-0.42	-0.38	0.04	0.15	0.18	-0.08	0.06
	NDF	-0.09	0.04	1.29	0.15	-0.15	-0.10	4.00	-4.25	1.42
5	EE	-0.02	0.10	-0.58	-1.14	-1.02	-0.04	1.83	-1.87	0.75
	TDN	-0.02	-0.01	1.24	1.69	2.01	-0.05	1.08	-1.18	0.50
	NFC	-0.01	0.08	-1.58	-0.10	-0.35	0.02	0.76	-0.91	0.25
	Eigenvalues	170803.39	43380.04	14720.15	1078.13	5.97	4.93	3.51	1.86	170803.39
	Variance (%)	74.26	18.26	6.40	0.47	0.00	0.00	0.00	0.00	74.26
	Accumulated variance (%)	74.26	96.12	99.52	99.99	99.99	99.99	100	100	74.26

Discard	Descriptor	CV10	CV11	CV12	CV13	CV14	CV15	CV16	CV17	CV18
4	PH	0.54	-0.42	.	.	-17.93	-0.60	6.98	11.46	0.54
2	PW	-0.11	-0.31	.	.	.	.	-8.62	10.57	-0.11
1	CTN	-0.38	0.08	.	.	.	.	.	13.90	-0.38
3	CL	0.43	-0.35	.	.	.	1.47	-0.31	0.02	0.43
	CW	-0.02	-0.21	0.01	0.25	0.05	-0.01	0.19	0.13	-0.02
	CD	-0.24	-0.28	0.01	0.04	0.02	0.03	-0.03	0.18	-0.24
	GM	-0.57	-0.16	0.00	-0.17	0.01	0.01	-0.12	0.06	-0.57
6	DW	0.28	-0.61	0.02	0.19	0.20	0.05	0.27	0.62	0.28
	DM	0.23	-0.44	0.00	0.30	0.01	0.01	0.21	0.01	0.23
	CA	-0.95	0.09	0.01	0.21	0.06	0.02	0.20	0.20	-0.95
	TCPA	-0.33	0.17	0.00	-0.18	0.01	0.01	-0.10	0.02	-0.33
	CPC	-1.02	0.50	0.00	-0.19	-0.02	-0.01	-0.15	0.09	-1.02
	ADF	0.39	1.20	0.00	0.65	0.00	0.00	0.46	0.02	0.39
	NDF	0.84	-1.56	0.00	0.41	0.01	0.00	0.30	0.04	0.84
5	EE	1.30	0.93	.	-0.99	0.01	0.01	-0.70	0.07	1.30
	TDN	-4.04	1.94	0.00	-0.40	0.00	0.00	-0.29	0.00	-4.04
	NFC	-1.19	2.22	0.00	0.68	-0.01	0.00	0.48	0.07	-1.19
Eigenvalues		1.86	0.84	0.69	0.52	0.42	0.25	0.24	0.012	1.86
Variance (%)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Accumulated variance (%)		100	100	100	100	100	100	100	100	100

PH – Plant height (cm); PW – Plant width (cm); CTN – Cladodes total number; CL – Cladodes length (cm); CW – Cladodes width (cm); CD – Cladodes diameter (mm); GM – Green mass (g); DW – Dry weight (g); DM – Dry mass (%); CA – Cladode area (cm<sup>2</sup>); TCPA – Total cladode photosynthetic area (cm<sup>2</sup>); CPC – Crude protein content (%); ADF – Acid detergent fiber (%); NDF – Neuter detergent fiber (%); EE – Ether extract (%); TDN – Total digestible nutrients (%); NFC – Non-fibrous carbohydrates (%).

### Clustering and Principal Components analysis

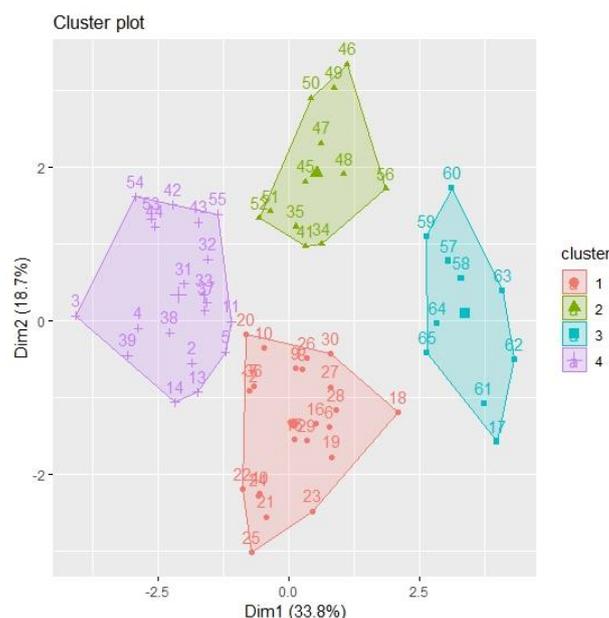
The two first principal components captured 52.54% of the data variation (Figure 1). It can be considered that the two-dimensional percentage map presented is adequate for the assessment of the relationships between the descriptors, since it explains a large part of the data variability with a percentage greater than 50% in the first two components.



**Figure 1.** Biplot of the first two main components (PC1 and PC2) of the principal component analysis of 65 *O. ficus-indica* accessions, based on agronomic and bromatological characteristics, Amelia Rodrigues, Bahia State, Brazil. PH – Plant height (cm); PW – Plant width (cm); CTN – Cladodes total number; CL – Cladodes length (cm); CW – Cladodes width (cm); CD – Cladodes diameter (mm); GM – Green mass (g); DW – Dry weight (g); DM – Dry mass (%); CA – Cladode area (cm<sup>2</sup>); TCPA – Total cladode photosynthetic area (cm<sup>2</sup>); CPC – Crude protein content (%); ADF – Acid detergent fiber (%); NDF – Neuter detergent fiber (%); EE – Ether extract (%); TDN – Total digestible nutrients (%); NFC – Non-fibrous carbohydrates (%).

The descriptors ADF x TDN and CPC x DM presented a negative association according to the angle of approximately 180°. Regarding the seven selected descriptors, TDN had the lowest eigenvector which indicated a smaller contribution to data variation, being also amenable to being discarded for the study of genetic variability in accessions of *O. ficus-indica*, when subjected to growing conditions similar to those

observed at the Station Experimental Rio Seco. Grouping was performed using the hierarchical *k-means* method, resulting in the formation of four clusters (Figure 2).



**Figure 2.** Grouping of 65 *O. ficus-indica* accessions, based on the *k-means* method from adjusted phenotypic means of five agronomic and two bromatological characteristics, Amélia Rodrigues, Bahia State, Brazil.

In general, clusters grouped accessions from different collection sources. Accessions collected in the same region showed a greater clustering tendency, along with those from the closest geographic regions. The number of accessions per group ranged between 10 (cluster 3) to 23 (cluster 1). Cluster 3 consisted of accessions collected in Irecê and Jaguarari (Figure 1). Group I gathered accessions with average values for DM and NFC. The NFC descriptor showed the greatest eigenvector regarding the relative contribution for discrimination in this group. Group II contains the accessions that showed the highest mean values for CD and the lowest mean values for DM and NFC. Group III had the highest mean values for CW, CA, TCPA, NDF, and ADF. Group 4 presented high values for CPC, GM, and TDN, and low values for TCPA and DM, to the detriment of the negative correlations established between these descriptors (Figures 1 and 2) (Table 5).

**Table 5.** Characteristics of the four groups formed by accessions of *O. ficus-indica* estimated through the phenotypic evaluation of agronomic and bromatological descriptors, Amélia Rodrigues, Bahia State, Brazil.

Descriptor.	Average $\pm$ Standard deviation				General average
	Group I	Group II	Group III	Group IV	
CW	14.52 $\pm$ 0.06	14.93 $\pm$ 0.35	15.98 $\pm$ 1.09	13.4 $\pm$ 0.74	14.43
CD	20.79 $\pm$ 1.02	24.81 $\pm$ 1.82	19.03 $\pm$ 2.27	23.96 $\pm$ 1.22	22.23
GM	531.46 $\pm$ 33.77	624.39 $\pm$ 31.93	470.27 $\pm$ 77.05	661.60 $\pm$ 58.21	579.23
DM	9.15 $\pm$ 0.60	7.64 $\pm$ 0.46	8.20 $\pm$ 0.07	7.75 $\pm$ 0.38	8.29
CA	229.58 $\pm$ 9.23	243.65 $\pm$ 7.19	285.52 $\pm$ 36.79	205.80 $\pm$ 19.54	233.48
TCPA	1066.37 $\pm$ 54.79	1280.23 $\pm$ 92.42	1540.61 $\pm$ 280.54	952.79 $\pm$ 135.10	1143.86
CPC	8.50 $\pm$ 0.62	10.28 $\pm$ 0.63	8.93 $\pm$ 0.32	10.95 $\pm$ 0.49	9.38
ADF	23.69 $\pm$ 0.56	25.69 $\pm$ 0.86	26.41 $\pm$ 1.36	23.71 $\pm$ 0.55	24.48
NDF	29.37 $\pm$ 0.22	29.77 $\pm$ 0.05	30.31 $\pm$ 0.44	29.69 $\pm$ 0.00	29.68
TDN	69.84 $\pm$ 0.12	68.92 $\pm$ 0.53	68.51 $\pm$ 0.82	70.51 $\pm$ 0.59	69.67
NFC	45.59 $\pm$ 0.73	42.39 $\pm$ 1.52	43.95 $\pm$ 0.42	44.95 $\pm$ 0.28	44.55

CW – Cladodes width (cm); CD – Cladodes diameter (mm); GM – Green mass (g); DM – Dry mass (%); CA – Cladode area (cm<sup>2</sup>); TCPA – Total cladode photosynthetic area (cm<sup>2</sup>); CPC – Crude protein content (%); ADF – Acid detergent fiber (%); NDF – Neuter detergent fiber (%); TDN – Total digestible nutrients (%); NFC – Non-fibrous carbohydrates (%).

When considering the most important attributes for forage cactus, taking into account its greater aptitude as a forage, it is observed that the accessions that are allocated in group III, followed by group IV, express greater productive potential, in view of presenting greater value for these descriptors. In this sense, the characters cladodes width, cladode area, total cladode photosynthetic area, crude protein content, acid detergent fiber, neuter detergent fiber and total digestible nutrients.

## Discussion

The genetic distances between accessions of *O. ficus-indica* evidenced the existence of variability within and between accessions collected in different locations of the state of Bahia. The indication of four groups by the hierarchical *K-means* method showed the tendency to group accessions with the same origin of collection, showing the greatest intrapopulation genetic similarity. Clusters also revealed that genetic diversity may not be structured according to collection sites, since accessions from different origins were gathered under the same group, demonstrating the genetic proximity between them. *O. ficus-indica* has its center of origin and diversity reported for the central region of Mexico (Gois, Silva, & Ribeiro, 2013). This species was introduced in Brazil during the late nineteenth century and its propagation occurs mainly asexually, by the planting of its cladodes. In this mode of reproduction, genetic variability tends to narrow, which explains the low genetic diversity among accessions from the same region and the genetic similarity between accessions from different places. The observed genetic diversity can be attributed to the introduction of new materials, allogamy, seedling production through seeds, and the occurrence of mutations. A similar result has already been reported by Ferreira et al. (2003), who also used multivariate techniques to estimate the genetic divergence between cactus forage accessions.

The weighting coefficients related to the canonical variables (CV) identified those descriptors that presented the greatest relative contribution to the assessment of genetic diversity among accessions of *O. ficus-indica*. The two first CV accumulated 96.12% of the variation present in the data. The eigenvalue criterion value under 0.70 suggested discarding the descriptors CTN, PW, CL, PH, EE, and DW as they showed a higher weighting coefficient value in the last canonical variables. Following Johnson and Wichern (2013), the exclusion of redundant descriptors allows for a reduction in the execution time of activities inherent to the experiment, labour, time, and cost spent on the analysis and interpretation of experimental data.

Descriptors discarding can also be carried out through principal component analysis (PCA), based only on the individual information of each accession and obeying the exclusion criterion of variables that show eigenvalue under 0.70. However, PCA is not suitable for the case where data came from an experiment with more than one information per accession. Therefore, canonical variables were used to carry out the discard of descriptors, being the most appropriate method for experiments that have data with repetition, and with the possibility of estimating a matrix of residual variances and covariances required in the analysis (Cruz & Ragazzi, 2012).

PCA reduced the data set allowing a better analysis of the interrelationships between the variables, as well as explaining them regarding their inherent dimensions. The first three principal components captured 67% of the total variation observed in the data, so those variables with higher eigenvalue in these principal components should be prioritized in the assessment of the genetic diversity in *O. ficus-indica* following the methodology described by Khattree and Naik (2000) and Jolliffe (2014). Thus, it is possible to reduce the dimension of 11 original variables to only three latent variables that represented the largest fraction of the variation. According to Hair et al. (2009) the percentage representation of the data variance must be represented by, at most, three principal components.

The Biplot graphic (Figure 1) represents the first two principal components that explain 52.5% of the variation observed in the data set. The two principal components summarized an important portion of the total sample variance, so they were used to evaluate the data set. The number of components selected depends on the researcher. Nevertheless, Jolliffe (2014) and Khattree and Naik (2000) suggested that components that explained a minimum of 60% of the variation in the data should be chosen. Tobar-Tosse, Santos, Ferraudo, Charlo, and Braz (2015), carrying out the characterization of soybean genotypes by principal component analysis, reported the capture of 59% of the data variation in the first two principal components. Jolliffe (2014) pointed out that, despite being an ancient technique, the principal component analysis, along with other multivariate analysis techniques, is not frequently used for the interpretation of research data with forage plants. Furthermore, Silva and Sbrissia (2010) highlighted the potential of the principal component analysis in the interpretation of data from forage plants.

Dimensionality of the data revealed the existence of positive correlations, due to the angle less than 90°, observed between the characters CPC x CDxGM, TCPA x CA x CW, CTN x NFC x NDF x DM. Negative correlations with angles close to 180°, and therefore with high magnitude, were observed between CPC x NDF, CPC x DM, CPC x NFC, ADF x TDN, ADF x NFC, TCPA x TDN, and CA x TDN (Figure 1).

Variables CA, TCPA, and CW showed similar contributions to PC1 being the most important for this component, as they present the eigenvector with the greatest length and proximity to the CP1 axis in the first

principal component (Figure 1). This component is related to plant's light-interception capacity, represented by variables that are related to the increase in the light-gathering surface. Variables CA, TCPA, and CW are highly correlated with each other as a function of the acute angle observed between their eigenvectors and also showed positive correlations, although of less magnitude, with ADF and NDF.

CP2 shows the productivity of green matter and protein in function that variables GM, CD, and CPC have presented the largest eigenvectors in this component and high correlation between these characters. This component also demonstrates the contrast of these variables in relation to DM, which demonstrated that the increase in GM, CD, and CPC negatively influence DM production. This reduction in DM content, as a consequence of the production of green mass, may be related to the saving of nutrients and water that the plant carries out when it is under adverse climatic conditions, taking into account that up to 90% of its green mass is attributed to the accumulation of water in the cladodes.

Principal component analysis is a multivariate technique that is based on modeling the covariance structure in order to analyze the correlation structures established between the variables. Therefore, PCA promotes groups based on correlations through the linear combination of all original variables, which are independent of each other, as well as the estimation of groups with the purpose of retaining the maximum amount of information in terms of the total variation contained in the data (Ricci, Costa, & Oliveira, 2013). On the other hand, the *K-means* method carried out the groupings based on the sum of the distances to the centroids of each cluster, based on the Euclidean distance from the centroids for each accession. The two methods were in agreement regarding the indication of the accessions for the formation of the groups.

It was observed that individuals in group I stood out as they registered the highest mean values for DM and NFC, with these two descriptors being highly correlated. Besides, group II gathered those accessions with the highest mean values for CD and values above the mean for the other descriptors, but DM, TDN, and NFC. Furthermore, in Group III, accessions that showed the highest values for TCPA, CD, CW, ADF, and NDF were clustered, however, these accessions had a lower potential for green matter production. Finally, group IV included individuals with the highest percentages of CPC and TDN as well as the highest mean value for GM and lowest DM content (Table 5).

Aiming at the genetic progress of the crop and the probability of extracting superior lines through segregating crossings and expanding the genetic base of *O. ficus-indica*. The selection of parents should occur among the different groups, which present the highest average values for a genetic characteristic. Therefore, to increase traits of high commercial value such as CPC and DM, priority should be given to crossing individuals from groups II x IV and I x III, respectively (Table 5). It is noteworthy that CPC and DM showed significant negative correlations and that the increase in one characteristic will imply the reduction of the other.

The genetic diversity registered between the accessions, the categorization into distinct groups, and the identification of the performance of these accessions regarding the evaluated variables will allow the selection of the most divergent accessions among each group as parent candidates for the formation of segregating populations, with greater probability of promoting satisfactory results for the *O. ficus-indica* genetic improvement program.

Furthermore, research focused on the genetic divergence present in *O. ficus-indica* accessions enable the regeneration, characterization, conservation, and exploitation of the variability available in the work collection belonging to the State University of Feira de Santana, guiding future actions of this and other genetic improvement programs for this species.

Larger cladodes exhibit the potential to carry out higher photosynthetic rates, in addition to providing more product for animal feed, normally used in the form of green mass. The selection for these traits should be taken into account when you want to increase green mass, resulting in increments in dry mass (Table 5).

Protein, an important macronutrient for food, whether animal or human, would not be the biggest highlight of the crop. However, small increments are decisive in the choice of accessions that can be used by the improvement program, considering the use of hybridization, bearing in mind the need to expand the genetic base and possible search of this variability (Table 5).

Acid and neutral detergent fiber is an important indicator of forage cactus digestibility. In this sense, it is observed that the accessions of group III were more expressive, demonstrating selection potential for this descriptor. Furthermore, the highest total nutrient digestibility was observed in the accessions of Group IV, confirming that the accessions of these two groups can be exploited within the breeding program, allowing the evaluation of the hybridization potential of these two groups (Table 5).

## Conclusion

Quantification of genetic diversity among the 65 accessions of *O. ficus-indica* revealed the existence of variability among the collection sites, resulting in the formation of four groups. It is inferred that, for breeding purposes, crosses between accessions from different groups, with the highest mean values for a given characteristic, can generate segregating populations with greater genotypic variability and superior to their parents. In this sense, the accessions that presented the highest Euclidean distance values and the highest average values for a given characteristic are the most suitable for hybridization. The accessions of groups III and IV should be explored by the forage cactus breeding program, as they presented greater productive potential.

## References

- Association of Official Agricultural Chemists [AOAC]. (2019). *Official Methods of analysis* (21st ed.). Washington, DC: AOAC.
- Cavalcante, L. A. D., Santos, G. R. A., Silva, L. M., Fagundes, J. L., & Silva, M. A. (2014). Respostas de genótipos de palma forrageira a diferentes densidades de cultivo. *Pesquisa Agropecuária nos Trópicos*, 44(4), 424-433. DOI: <https://doi.org/10.1590/S1983-40632014000400010>
- Cruz, C. D., & Regazzi, A. J. (2012). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, MG: UFV.
- Ferreira, C. A., Santos, D. C., Santos, M. V. F., Silva, J. A. A., Lira, M. A., & Molica, S. G. (2003). Utilização de técnicas multivariadas na avaliação da divergência genética entre clones de palma forrageira (*Opuntia ficus-indica* Mill.). *Revista Brasileira de Zootecnia*, 32(6), 1560-1568. DOI: <https://doi.org/10.1590/S1516-35982003000700004>
- Freitas, I. L. J., Amaral Junior, A. T., Viana, A. P., Pena, G. F., Cabral, P. S., Vittorazzi, C., & Silva, T. R. C. (2013). Ganho genético avaliado com índices de seleção e com REML/Blup em milho-pipoca. *Pesquisa Agropecuária Brasileira*, 48(11), 1464-1471. DOI: <https://doi.org/10.1590/S0100-204X2013001100007>
- Galvão Júnior, J. G. B., Silva, J. B. A., Morais, J. H. G., & Lima, R. M. (2014). Palma forrageira na alimentação de ruminantes: cultivo e utilização. *Acta Veterinaria Brasilica*, 8(2), 78-85. DOI: <https://doi.org/10.21708/avb.2014.8.2.3490>
- Gois, G. C., Silva, F. C. S., & Ribeiro, W. S. (2013). *Descrição morfológica, origem, domesticação, dispersão da palma forrageira e sua introdução no Brasil*. In W. S. Ribeiro, F. C. S. Silva, E. R. Costa, S. M. Silva, & G. C. Eculica (Eds), *A Palma e sua importância no nordeste brasileiro* (p. 9-20). Brasília, DF: Editora Kiron.
- Hair, J. F., William, C. B., Babin, J. B., Anderson, E. R, & Tatham, R. L. (2009). *Análise multivariada de dados* (5. ed.). Porto Alegre, RS: Bookman.
- Instituto Brasileiro de Geografia e Estatística [IBGE]. (2017). *Censo agropecuário*. Retrieved on July 20, 2018 from <https://censoagro2017.ibge.gov.br>
- Instituto Nacional de Meteorologia [INMET]. (2017). Retrieved on Dec. 15, 2018 from [www.inmet.gov.br](http://www.inmet.gov.br)
- Johnson, R. A., & Wichern, D. W. (2013). *Applied multivariate statistical analysis* (6th ed.). New Jersey, NY: Pearson.
- Jolliffe, I. T (2014). *Principal component analysis*. New York, NY: Springer-Verlag.
- Khattree, R., & Naik, D. N. (2000). *Multivariate data reduction and discrimination with SAS software*. New York, NY: BBU and John Wiley Sons.
- Leite, M. L. M. V., Silva, D. S., Andrade, A. P., Pereira, W. E., & Ramos, J. P. F. (2014) Caracterização da produção de palma forrageira no Cariri paraibano. *Revista Caatinga*, 27(2), 192-200.
- Maciel, G. M., Peixoto, J. V.M., Rocha, J. P. R., Almeida, R. S., Oliveira, C. S., & Mendonça, T. F. N. (2018). Multivariate techniques in the determination of genetic diversity in pest-resistant mini tomato genotypes. Scientific communication. *Horticultura Brasileira*, 36(4), 504-509. DOI: <https://doi.org/10.1590/S0102-053620180413>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209-220.
- Pereira, G. A., Santos, E. M., Oliveira, J. S., Araújo, G. G. L., Paulino, R. S., Perazzo, A. F., ... Leite, G. M. (2021). Intake, nutrient digestibility, nitrogen balance, and microbial protein synthesis in sheep fed

- spineless-cactus silage and fresh spineless cactus. *Small Ruminant Research*, 194, 1-8.  
DOI: <https://doi.org/10.1016/j.smallrumres.2020.106293>
- Pessoa, R. A. S., Ferreira, M. A., Silva, F. M., Bispo, S. B., Wanderley, W. L., & Vasconcelos, P. C. (2013). Diferentes suplementos associados à palma forrageira em dietas para ovinos: consumo, digestibilidade aparente e parâmetros ruminais. *Revista Brasileira de Saúde e Produção Animal*, 14(3), 508-517.
- Ramos, J. P. F., Santos, E. M., Pinho, R. M. A., Bezerra, H. F. C., Pereira, G. A., Beltrão, G. R., & Oliveira, J. S. (2015). Vegetative growth of cactus forage depending on organic fertilization. *REDVET - Revista Electrónica de Veterinaria*, 16(12), 1-12.
- R Development Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.
- Resende, M. D. V. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, 16(4), 330-339. DOI: <https://doi.org/10.1590/1984-70332016v16n4a49>
- Ricci, M. D. S. F., Costa, J. R., & Oliveira, N. G. D. (2011). Utilização de componentes principais para analisar o comportamento do cafeeiro a pleno sol e sombreado. *Coffee Science*, 6(1), 44-54.  
DOI: <https://doi.org/10.1038/bjc.2011.292>
- Silva, A. S., Santos, E. M., Pinto, I. M. P., & Ramos, J. P. F. (2020). Análise multivariada da palma forrageira: características morfoprodutivas sob correlações canônicas. *Agrarian*, 13(47), 100-106.  
DOI: <https://doi.org/10.30612/agrarian.v13i47.10621>
- Silva, J. A. Donato, S. L. R., Donato, P. E. R., Souza, E. S., Padilha Júnior, M. C., & Silva Junior, A. A. (2016). Produtividade e crescimento vegetativo da palma forrageira sob espaçamentos e adubações química. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 20(6), 564-569. DOI: <https://doi.org/10.1590/1807-1929/agriambi.v20n6p564-569>
- Silva, M. A. (2014). Respostas de genótipos de palma forrageira a diferentes densidades de cultivo. *Pesquisa Agropecuária nos Trópicos*, 44(4), 424-433. DOI: <https://doi.org/10.1590/S1983-40632014000400010>
- Silva, S. C., & Sbrissia, A. F. (2010). Análise de componentes principais entre características morfogênicas e estruturais em capim-marandu sob lotação contínua. *Ciência Rural*, 40(3), 690-693.  
DOI: <https://doi.org/10.1590/S0103-84782010000300034>
- Sturion, J. A., & Resende, M. D. V. (2010). *Melhoramento genético da erva-mate*. Colombo, PR: Embrapa Florestas.
- Tobar-Tosse, D. E., Santos, R. C. W., Ferraudo, C. A. S., Charlo, H. C.O., & Braz, L. T. (2015). Caracterização de genótipos de soja-hortaliça por análise de componentes principais. *Ciência Rural*, 45(7), 1214-1219.  
DOI: <https://doi.org/10.1590/0103-8478cr20140597>
- Wanderley, W. L., Ferreira, M. A., Batista, A. M. V., Vêras, A. S. C., Bispo, S. V., Silva, F. M., & Santos, V. L. F. (2012). Consumo, digestibilidade e parâmetros ruminais em ovinos recebendo silagens e fenos em associação à palma forrageira. *Revista Brasileira de Saúde e Produção Animal*, 13(2), 444-456.  
DOI: <https://doi.org/10.1590/S1519-99402012000200013>