Universidade Federal Rural do Semi-Árido
Pró-Reitoria de Pesquisa e Pós-Graduação
https://periodicos.ufersa.edu.br/index.php/caatinga
ISSN 1983-2125 (online)

http://dx.doi.org/10.1590/1983-21252024v3712466rc

Scientific Article

Conventional morphological descriptors and artificial neural networks for characterizing biofortified lettuce germplasm

Redes neurais artificiais para descritores morfológicos em germoplasma de alface biofortificada

Ana C. P. Jacinto[1]*, Renata Castoldi[2], Isadora G. da Silva[3], Diesiele C. S. Mota[2], Leticia G. Moreira[2], Gabriel M. Maciel[2], Camila S. de Oliveira[2], Hamilton C. de O. Charlo[4]

[1]Institute of Agricultural Sciences, Universidade Federal de Uberlândia, Uberlândia, MG, Brazil. [2]Institute of Agricultural Sciences, Universidade Federal de Uberlândia, Monte Carmelo, MG, Brazil. [3]Department of Agriculture, Universidade Federal de Lavras, Lavras, MG, Brazil. [4]Department of Agronomy, Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro, Uberaba, MG, Brazil.

ABSTRACT - The classification based on morphological descriptors in lettuce is considered a complex activity and proves to be efficient for studying phenotypic characteristics. Therefore, the objective of this study was to analyze the biofortified lettuce germplasm bank at the Universidade Federal de Uberlândia using both conventional morphological descriptors and artificial neural networks. The experiment was conducted in the field. The experimental design employed was a randomized complete block design, consisting of 14 treatments (11 genotypes of mini lettuce, and the cultivars Purpurita, UDI 10.000, and Pira 72) with four replications. Nine morphological descriptors were evaluated. Following the data acquisition, dissimilarity matrix analyses, principal component analysis, dendrogram construction, and artificial neural network (ANN) analyses were performed. The genotypes exhibited phenotypic variability when compared to the parental strains UDI 10.000 and Pira 72. The purple color of the leaves and anthocyanin presence across the entire leaf surface were predominant among the genotypes. Descriptors such as leaf intensity and color, as well as anthocyanin intensity, coloration, and distribution, were the most influential in assessing genetic variability. The Self-Organizing Map (SOM) demonstrated greater sensitivity in discriminating between genotypes compared to the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). While the UPGMA clustering method grouped genotypes into three clusters, the SOM method grouped into five clusters. The use of genetic distance analyses and SOM dendrogram proved to be effective in selecting individuals UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#10, and UFU 215#13, which are clustered with the cultivar UFU Mini Biofort.

Keywords: Artificial intelligence. Plant breeding. Vegetables.

RESUMO - A classificação baseada em descritores morfológicos em alface é considerada uma atividade complexa e mostra-se eficiente para estudo de características fenotípicas. Portanto, o objetivo deste estudo foi analisar o banco de germoplasma de alface biofortificada da Universidade Federal de Uberlândia utilizando descritores morfológicos convencionais e redes neurais artificiais. O experimento foi conduzido em campo. O delineamento experimental empregado foi de blocos casualizados, composto por 14 tratamentos (11 genótipos de minialface e as cultivares Purpurita, UDI 10.000 e Pira 72) com quatro repetições. Foram avaliados nove descritores morfológicos. Após a aquisição dos dados, foram realizadas análises de matrizes de dissimilaridade, análise de componentes principais, construção de dendrogramas e análises de redes neurais artificiais (RNA). Os genótipos apresentaram variabilidade fenotípica quando comparados às linhagens parentais UDI 10.000 e Pira 72. A cor roxa das folhas e a presença de antocianinas em toda a superfície foliar foram predominantes entre os genótipos. Descritores como intensidade e cor das folhas, intensidade, coloração e distribuição de antocianinas foram os mais influentes na avaliação da variabilidade genética. O Mapa Auto-Organizável (SOM) demonstrou maior sensibilidade na discriminação entre genótipos em comparação ao Método de Grupos de Pares Não Ponderados com Média Aritmética (UPGMA). Enquanto o método de agrupamento UPGMA agrupou os genótipos em três agrupamentos, o método SOM agrupou em cinco agrupamentos. A utilização de análises de distância genética e dendograma SOM mostrou-se eficaz na seleção dos indivíduos UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#10 e UFU 215#13, que estão agrupados com a cultivar UFU Mini Biofort.

Palavras-chave: Inteligência artificial. Melhoramento de Plantas. Vegetais.

Conflict of interest: The authors declare no conflict of interest related to the publication of this manuscript.

Received for publication in: January 22, 2024.
Accepted in: May 27, 2024.

***Corresponding author:**
<carol.agro.ufu@gmail.com>

INTRODUCTION

Lettuce is a leafy vegetable of great economic importance in Brazil. In recent years, this vegetable has been highly consumed and produced in the country. Despite Brazil's ranking as the third-largest global producer of lettuce, it stands out as a notable contributor, yielding approximately 671,509 tons across an area of 86,799 thousand hectares (CARVALHO et al., 2019).

According to Souza et al. (2019), lettuce cultivars exhibit considerable morphological and genetic variations, yet they are categorically classified into only five major groups that diverge phenotypically.

In Brazil, a growing and promising market emerges for new types of lettuce (SALA; COSTA, 2012). Therefore, lettuce must be characterized according to the morphological traits, as well as the resistance or susceptibility to pests and diseases. Traditionally, basic morphological descriptors suitable for

plant groups with distinct morphological parameters are employed (KORIR et al., 2012).The assessment and description of phenotypic characteristics are tasks of utmost importance for germplasm bank curators and breeders, especially when initiating a breeding program aiming to select genotypes that are more productive and possess qualitative traits meeting the demands of the consumer market (SILVA et al., 2021). Moreover, the use of descriptors have proved to be efficient for studying morphological variability among species (SOUZA et al., 2019; SALA; COSTA, 2012).

The adoption of morphological characteristics in diversity studies is highly encouraged, as traditional approaches to cultivar identification rely on phenotypic data. These data are also included in the procedure for registering and protecting novel cultivars (FARAHANI et al., 2019). Some phenotypic descriptors based on morphological characteristics have been previously utilized for sweet potato (EVANGELISTA et al., 2022) and chili pepper (ESPÍRITO SANTO; MENEZES; CARMO, 2022). This observation underscores that morpho-agronomic characterization enhances our understanding of germplasm banks.

Various statistical methods explore the variability and assist in the selection and combination of generations, with emphasis on analysis through agglomerative methods. This method utilizes dissimilarity measures derived from quantitative and qualitative variables, allocating genotypes into groups in such a way that there is homogeneity within these groups and heterogeneity among them (ALI-SHTAYEH et al., 2017).

The use of artificial neural networks (ANNs) based on morphological descriptors in lettuce remains unexplored. The ANNs mimic the functioning of biological neurons, acquiring knowledge and learning from errors (CARDOSO et al., 2019). ANNs can be used in classification and grouping, prediction of traits of interest, estimation of genetic diversity, fitting of models, study of adaptability and stability, and in genome-wide selection, among others (CARDOSO et al., 2019). One type of ANNs is the Kohonen's self-organizing maps (SOM), which simulate the functioning of the cerebral cortex, recognize patterns and create clusters (with neurons of the network), establishing strong connections with the closest neurons; similarly, close groups are more similar (CRUZ; NASCIMENTO, 2018). This occurs as the method of analysis has high accuracy; hence, even if experimental errors occur that do not meet the premises, as in experiments using univariate statistics, groups of more representative similar and dissimilar materials form in the ANNs, thus reducing subjectivity in the selection of genotypes. This study aimed to analyze the biofortified lettuce germplasm bank at UFU using conventional morphological descriptors and artificial neural networks.

MATERIALS AND METHODS

The experiment was conducted at the Vegetable Experimental Station of the Universidade Federal de Uberlândia (UFU), Campus Monte Carmelo, with geographical coordinates of 18°43'36.03" South latitude, 47°31'28.59" West longitude, and an altitude of 903 m.

The climate of the region, according to the Köppen classification, is humid temperate, characterized by hot summers and dry winters (BECK et al., 2018).

The experiment was conducted from May 4 to July 30, 2019. The experimental design employed was a randomized complete block design, consisting of 14 treatments with four replications. The treatments comprised 11 genotypes from generation F5:6, resulting from the cross between the Uberlândia 10.000 lineage (UDI 10.000) and the Pira 72 cultivar, developed using the genealogical method. These genotypes were UFU 66#3, UFU 66#4, UFU 66#7, UFU 66#8, UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#7, UFU 215#10, UFU 215#13, and UFU MC MINIBIOFORT2, along with the Purpurita mini lettuce cultivar and the parental cultivars Pira 72 and UDI 10.000. Each experimental plot consisted of four planting rows, each 1.05 m long, spaced at 0.15 m between the plants and 0.15 m between rows. For evaluation, ten plants per plot were considered, selected from the two central rows.

The genotypes belong to the Vegetable Germplasm Bank of UFU, Monte Carmelo Campus, and have already been pre-selected for elevated levels of carotenoids in the leaves.

Seeds were sown in expanded polystyrene trays with 200 cells, filled with Maxfertil® coconut fiber-based substrate. The seedlings were kept in a greenhouse (7 m x 4 m), covered with 150-micron UV-resistant transparent plastic, until they reached the transplant stage. Immediately after the seedlings had three to five definitive leaves, they were transplanted into beds. Prior to this, the soil of the experimental area underwent liming to raise the base saturation to 70%. Approximately 40 days later, planting fertilization was conducted, consisting of 30 kg ha^{-1} of N, 300 kg ha^{-1} of P_2O_5, and 18 kg ha^{-1} of K_2O, based on soil analysis results and recommendations by Fontes (1999). Subsequent top dressing fertilizer applications comprised 30 kg ha^{-1} of N and 18 kg ha^{-1} of K_2O at 15 d after planting fertilization and 45 kg ha^{-1} of N and 27 kg ha^{-1} of K_2O at 30 and 45 d, respectively, after transplanting (FONTES, 1999).

Morphological characterization of the plants in the useful plot was conducted when the plants were ready for harvest, i.e., 38 d after transplantation. For this purpose, nine morphological descriptors (Table 1) were used, based on the guidelines for conducting tests on the distinctiveness, uniformity, and stability of lettuce cultivars (MAPA, 2001).

Table 1. Morphological descriptors for lettuce, based on the guidelines for conducting tests on the distinctiveness, uniformity, and stability.

Characteristics[1]	Abbreviation
Leaf coloration	LC
Anthocyanin coloration	AC
Leaf coloration intensity	LCI
Anthocyanin distribution	AD
Anthocyanin coloration intensity	ACI
Leaf crinkling degree	LCD
Blister size	BS
Leaf edge type	LET
Growth habit	GH

[1]MAPA - Ministry of Agriculture, Livestock, and Supply (2001). Guidelines for the execution of tests for distinctness, uniformity, and stability of lettuce cultivars. (*Lactuca sativa* L.).

These descriptors described below were used:

a) Leaf coloration: can be classified as: 1 = green, 2 = purple, or 3 = variegated;

b) Anthocyanin coloration: can be: 1 = absent or 2 = present;

c) Leaf color intensity: using the ordinal descriptive scale, where: 1 = very light, 2 = light, 3 = medium, 4 = dark, 5 = very dark.

d) Anthocyanin distribution: can be: 1 = absent, 2 = localized, or 3 = throughout the surface.

e) Anthocyanin coloration intensity: using the ordinal descriptive scale, where: 1 = very weak, 2 = weak, 3 = medium, 4 = strong, 5 = very strong.

f) Leaf crinkling degree: using the ordinal descriptive scale, where: 1 = absent or very weak, 3 = weak, 5 = medium, 7 = strong, and 9 = very strong.

g) Blister size: using the ordinal descriptive scale, where: 1 = small, 2 = medium, and 3 = large.

h) Leaf edge type: using the ordinal descriptive scale, where: 1 = absent or very weak, 3 = weak, 5 = medium; 7 = strong, and 9 = very strong.

i) Growth habit: can be classified as: 1 = upright, 2 = semi-upright, or 3 = nearly horizontal or prostrate.

After obtaining the scores for qualitative characteristics of the ten genotypes and four lettuce cultivars, genetic distance was determined based on the dissimilarity coefficient of discordance. In addition to the dissimilarity matrix, a principal component analysis and dendrogram construction were carried out. The R Core Team Studio software (2022) was used for these analyses.

To carry out the principal component analysis and dendrogram, the “Multivariate Analysis” package was used. To create the dendrogram, the distance was first calculated, using the discordance method, recommended for multicategory or binary qualitative data. After calculating the distance, the dissimilarity matrix was obtained; with this matrix, the dendrogram was obtained using the UPGMA method and the cut was performed using the Mojena criterion.

The comparison of similarity among treatments was obtained using the artificial neural network technique, employing the SOM. Different combinations were tested before selecting the final parameters, resulting in various simulations based on different combinations. The unsupervised data classification technique was performed using an architecture corresponding to 3000 epochs, with three neurons in the first dimension and two in the second dimension (3x2) and a neighbor radius pattern equal to 1. The topology used was hexagonal, based on Euclidean distance for cluster formation. No specific definition is available regarding the topology to be used, as this choice is a random process, is subjective, and depends on each situation to be analyzed (SANTOS et al., 2019).

RESULTS AND DISCUSSION

In general, the genotypes exhibited variation in scores across all evaluated descriptors, particularly when compared to the parental strains UDI 10.000 and Pira 72, indicating phenotypic variability (Table 2).

Regarding leaf color, the genotypes were classified into three types, with 50% having purple leaves, 35.71% having green leaves, and 14.29% having variegated leaves. These percentages are directly related to the descriptor of anthocyanin coloration, where the materials were mostly categorized with the presence of this pigment.

The higher percentage of genotypes with purple leaves is preferable for the consumer market, given the observed trend of increased consumption in recent years for this type of lettuce. This is because purple lettuce has been used in salad mixes, both for its attractive color and its high anthocyanin content (KIM et al., 2018).

The consumption of foods with high concentrations of anthocyanins, carotenoids, and chlorophyll contributes to cell protection against oxidative stress, preventing some degenerative diseases, cardiovascular issues, and infections (BENDOKAS et al., 2019).

Additionally, other descriptors are associated with leaf color, such as leaf color intensity, anthocyanin coloration intensity, and anthocyanin distribution, with the last two being employed for lettuce cultivars with purple leaves.

Regarding leaf color intensity, a variation from light to dark was observed, with no observations of very light or dark intensities. Only the cultivar UDI 10000 exhibited a light color, while the others showed color intensities ranging from medium (Pira 72, UFU 66#7, UFU 66#8, UFU 215#6, UFU

215#7, UFU 215#13, and Purpurita) to dark (UFU 66#3, UFU 66#4, UFU 215#1, UFU 215#2, UFU 215#10, and UFU Mini Biofort).

For the morphological descriptor "anthocyanin distribution," all levels were observed: absent in cultivars with green leaves (UDI 10.000, UFU 66#3, UFU 66#4, UFU 66#7, and UFU 66#8), localized (Pira 72 and UFU 215#7), and present throughout the leaf surface in the other genotypes (215#1, UFU 215#2, UFU 215#6, UFU 215#10, UFU 215#12, UFU Mini Biofort, and Purpurita).

Table 2. Scores for the qualitative characteristics of the ten genotypes and four lettuce cultivars.

Genotypes	LC[1]	AC	LCI	AD	ACI	LCD	BS	LET	GH
UFU 66#3	1	1	4	1	1	9	3	3	3
UFU 66#4	1	1	4	1	1	7	2	3	3
UFU 66#7	1	1	3	1	1	7	2	3	3
UFU 66#8	1	1	3	1	1	7	2	3	3
UFU 215#1	2	2	4	3	5	1	1	1	2
UFU 215#2	2	2	4	3	5	1	1	1	2
UFU 215#6	2	2	3	3	4	3	1	1	2
UFU 215#7	3	2	3	2	3	7	2	3	2
UFU 215#10	2	2	4	3	5	3	1	1	2
UFU 215#13	2	2	3	3	4	3	1	1	2
UFU Mini Biofort	2	2	4	3	5	3	1	1	2
Purpurita	2	2	3	3	4	1	1	9	1
Pira 72	3	2	3	2	3	9	3	9	1
UDI 10.000	1	1	2	1	1	3	1	1	3

[1]LC: 1 = green, 2 = purple, or 3 = variegated; AC: 1 = absent or 2 = present; LCI: 1 = very light, 2 = light, 3 = medium, 4 = dark, 5 = very dark; AD: 1 = absent, 2 = localized, or 3 = throughout the surface; ACI: 1 = very weak, 2 = weak, 3 = medium, 4 = strong, 5 = very strong; CD: 1 = absent or very weak, 3 = weak, 5 = medium, 7 = strong, 9 = very strong; BS: 1 = small, 2 = medium, and 3 = large; LET: 1 = absent or very weak, 3 = weak; 5 = medium; 7 = strong, 9 = very strong; GH: 1 = upright, 2 = semi-upright, or 3 = nearly horizontal or prostrate.

Conversely, for the descriptor "anthocyanin coloration intensity," no genotypes were present in the germplasm bank with intensity 2, i.e., very weak. Nevertheless, they were classified from absent (35.71%) to very strong (28.57%). According to Santos et al. (2021), research has been conducted to select cultivars with more intense green coloration, indicating a higher chlorophyll content and shiny leaves.

Cultivars characterized by very dark leaves are not preferred by consumers as they are often associated with overripe products. Covre et al. (2020) found a higher preference for lettuce cultivars with a light green color, indicating a preference for lettuce with greater brightness.

The higher percentage of genotypes with medium to dark-colored leaves may have resulted from the prior selection of the germplasm bank for biofortification with high levels of carotenoids in the leaves, a precursor of vitamin A (SOUSA et al., 2019).

According to Klooster et al. (2012), the higher the chlorophyll content, the greater the intensity of leaf coloration, as chlorophyll has a high correlation with total carotenoids.

The domestication of lettuce is not limited to changes in color; there have also been changes in size, shape, texture, leaf flavor, and growth habits (MOU, 2011).

In this context, descriptors associated with shape, size, texture, and growth habit were evaluated. The leaf crinkling degree varied from leaves without crinkling (UFU 215#1, UFU 215#2, Purpurita) to a high degree of crinkling (UFU 66#3, Pira 72).

The cultivars UDI 10.000 and Pira 72, used as controls for being parental strains, had blister sizes classified as small (1) and large (3), respectively. Small blister size was observed in 57.14% of the treatments, followed by 28.57% and 14.28% with medium to large blister sizes, respectively.

The leaf edge type is an important descriptor for characterizing leafy greens as either smooth or curly based on the wave. Therefore, 85.71% of the genotypes had a smooth leaf edge, as they exhibited a very weak to weak wave at the edge. Contrarily, the cultivars Pira 72 and Purpurita had a very strong wave, characterized as a curly leaf edge.

Regarding the growth habit, 50% of the genotypes (UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#7, UFU 215#10, UFU 215#13, and UFU Mini Biofort) were observed to exhibit a semi-upright growth habit. Meanwhile, 35.71% of the genotypes (UFU 66#3, UFU 66#4, UFU 66#7, UFU 66#8, and UDI 10.000) showed an almost horizontal growth habit, and the cultivars Purpurita and Pira 72 exhibited an upright growth habit.

Three growth habits, that is almost horizontal, erect, and semi-erect were observed among the genotypes, with semi-erect being predominant. This behavior is desirable as cultivars with a semi-erect growth habit are less prone to diseases, including tip burn (*Rhizoctonia solani*), a widespread disease in various lettuce-producing areas (DIAS et al., 2021).

Furthermore, these results are consistent with the findings of Bohórquez et al. (2017), who, while working with

lettuce cultivars in the Amazon region, observed the presence of all three types of growth habits.

Based on these qualitative traits, the dissimilarity frequency index was used to create the dissimilarity matrix and dendrogram. This index is recommended for a group of multicategory or binary traits, where the similarity index can be established according to concordance and discordance. It allows for the simultaneous analysis of qualitative traits (CRUZ; CARNEIRO, 2014).

The use of the dissimilarity matrix provides the breeder with information about how similar or divergent two genotypes can be. The closer to 0, the more these genotypes are genetically similar, exhibiting similar characteristics.

The dissimilarity coefficients ranged from 0.10 to 1.00 (Figure 1). The smallest dissimilarity coefficient value (0.10) was observed between the genotypes UFU 66#7 and UFU 66#8; UFU 215#1 and UFU 215#2; UFU 215#6 and UFU 215#13; and UFU 215#10 and UFU Mini Biofort. Conversely, the highest value (1.00) was observed for 17 combinations, including genotypes UFU 66#3 and UFU 66#4 with cv. Purpurita; UFU 215#7 and cv. Pira 72 with cv. UDI 10.000, among others (Figure 1).

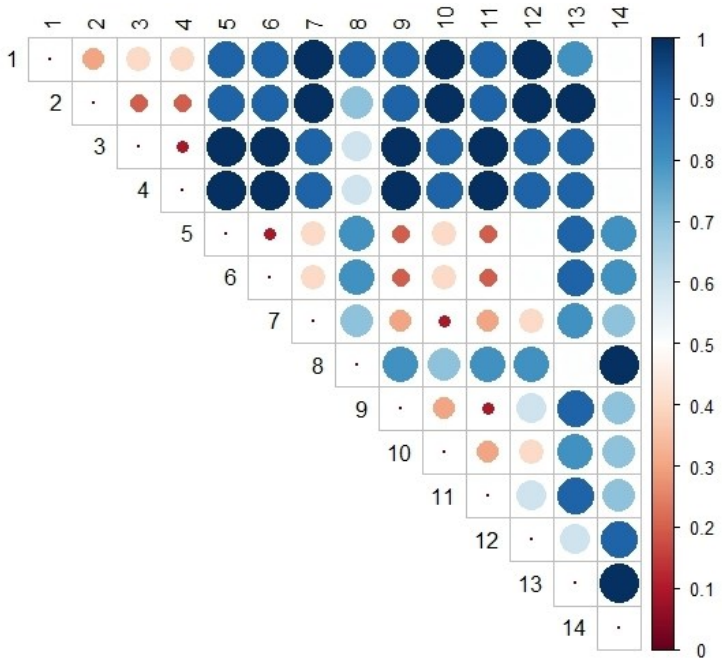

Figure 1. Dissimilarity matrix among ten genotypes and four lettuce cultivars, calculated based on the dissimilarity coefficient, using nine morphological descriptors. 1 – UFU 66#3, 2 – UFU 66#4, 3 – UFU 66#7, 4 – UFU 66#8, 5 – UFU 215#1, 6 – UFU 215#2, 7 – UFU 215#6, 8 – UFU 215#7, 9 – UFU 215#10, 10 – UFU 215#13, 11 – UFU Mini Biofort, 12 – Purpurita, 13 – Pira 72, and 14 – UDI 10.000.

No zero indices were observed, indicating the absence of duplicates. According to Burle and Oliveira (2010), genotypes with zero indices can be eliminated, thereby reducing the maintenance costs of collections and germplasm banks.

Moreover, several combinations of genotypes were observed to have the highest dissimilarity coefficient values, indicating that based on morphological traits, there is a high genetic variability.

After obtaining the dissimilarity matrix, the dendrogram with three separate groups was generated using the Mojena criterion (1977).

The dendrogram based on the dissimilarity coefficient (Figure 2) showed a high cophenetic correlation (94.74%), a value that indicates the high reliability between the original distance matrix and the matrix generated by the UPGMA clustering (CRUZ; CARNEIRO, 2014).

Three groups were formed and the separation of the groups was performed using the Mojena criterion (1977), resulting in a cutoff line of 0.697.

The first group consisted of five genotypes (UFU 66#3, UFU 66#4, UFU 66#7, UFU 66#8, and the cv. UDI 10.000), the second group comprised seven genotypes (UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#10, UFU 215#13) and the cultivars UFU Mini Biofort and Purpurita, whereas the third group consisted of the genotype UFU 215#7 and the cultivar Pira 72.

Controversial results were reported by Sousa et al. (2019) when evaluating 49 biofortified lettuce genotypes resistant to root-knot nematodes (*Meloidogyne* spp.). They identified four distinct groups using the UPGMA method (with a 30% cutoff), with the cultivars Pira 72 and UDI 10.000 belonging to the same group formed by 21 genotypes.

The first two principal components (PCs) were observed to account for 80% of the total variance (Table 3).

PC1 explained the highest variance (56.81%) with the highest eigenvalue (5.11), which was dominated by ACI (0.9792), AD (0.9685), AC (0.9604), LC (0.8138), and LCI (0.2472). As for PC2, the variance of 23.23% can be explained by the traits GH, LCI, ACI, and AD (Tables 3 and 4).

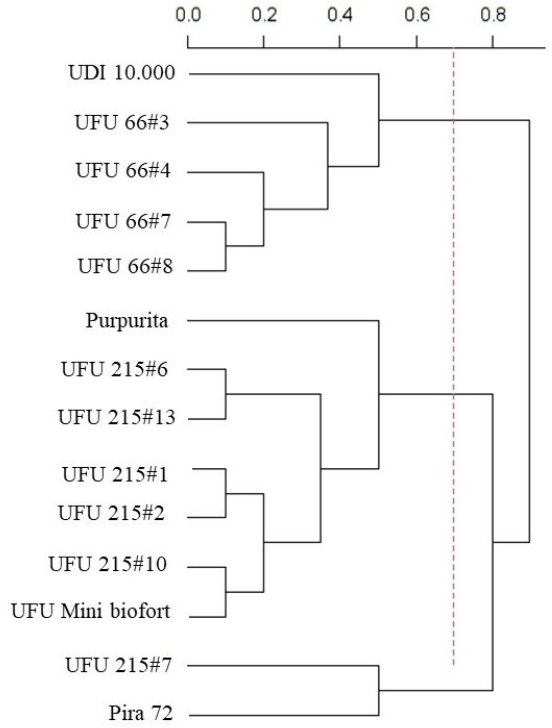

Figure 2. UPGMA dendrogram (based on the dissimilarity index) of ten genotypes and four lettuce cultivars, using nine morphological traits.

Table 3. Cumulative variances and eigenvalues for the first four principal components (PCs).

	[1]PC1	PC2	PC3	PC4
Eigenvalues	5.11	2.09	1.15	0.43
Variance (%)	56.81	23.23	12.78	4.88
Cumulative Variance	56.81	80.04	92.82	97.70

[1]PC = Principal Component.

Table 4. Contribution of variables to the total variability (PC1 to PC4) of lettuce genotypes.

[1]Variables	[2]PC1	PC2	PC3	PC4
LCD	-0.7246	-0.5426	0.2842	-0.2909
BS	-0.6344	-0.6685	0.3380	-0.1187
LET	-0.0428	-0.8831	- 0.2420	0.3957
GH	-0.8008	0.5665	0.1515	-0.1000
AC	0.9604	-0.2006	0.0275	-0.1417
LC	0.8138	-0.4274	0.1170	-0.2413
ACI	0.9792	0.0874	0.1576	-0.0027
AD	0.9685	0.0214	0.0308	-0.0951
LCI	0.2472	0.1333	0.9129	0.2939

[1]For the abbreviations, refer to Table 1. [2]PC = Principal Component.

The genotypes UDI 10.000, UFU 66#3, UFU 66#4, UFU 66#7, UFU 66#8 were grouped in class 1. Classes 3, 4, and 6 grouped the genotypes Pira 72, UFU 215#7, and Purpurita separately. In class 5, 42.86% of the evaluated genotypes were grouped. No genotype was allocated to class 2 (Figure 3).

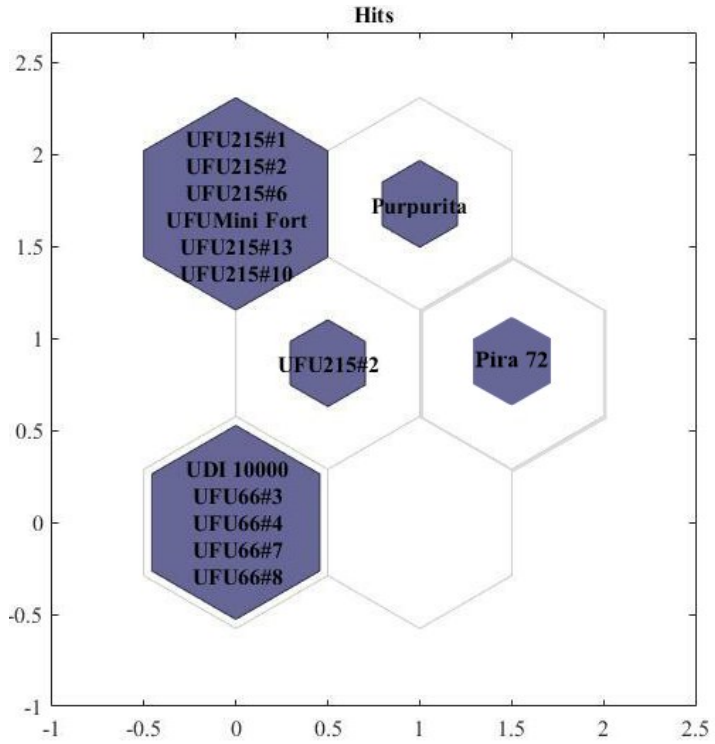

Figure 3. Self-organizing map (SOM) for six classification classes, formed by artificial neural networks. Class 1: row 1 column 1, Class 2: row 2 column 2, Class 3: row 2 column 2, Class 4: row 2 column 2, Class 5: row 3 column 1, and Class 6: row 3 column 2.

The effect of each variable on each group formed in SOM was demonstrated in the representation of the neural topology of the network generated through weights and the association of each input variable with the output neuron (Figure 4).

The characteristics AC, ACI, AD, and LCI were the features that contributed the most to allocate the genotypes UFU 215#1, UFU 215#2, UFU 215#6, UFU Mini biofort, UFU 215#13, and UFU 215#1 in the same grouping by the SOM. , the features LCD and GH were crucial to distinguish the group formed by the genotypes UDI 10,000, UFU 66#3, UFU 66#4, UFU 66#7, and UFU 66#8.

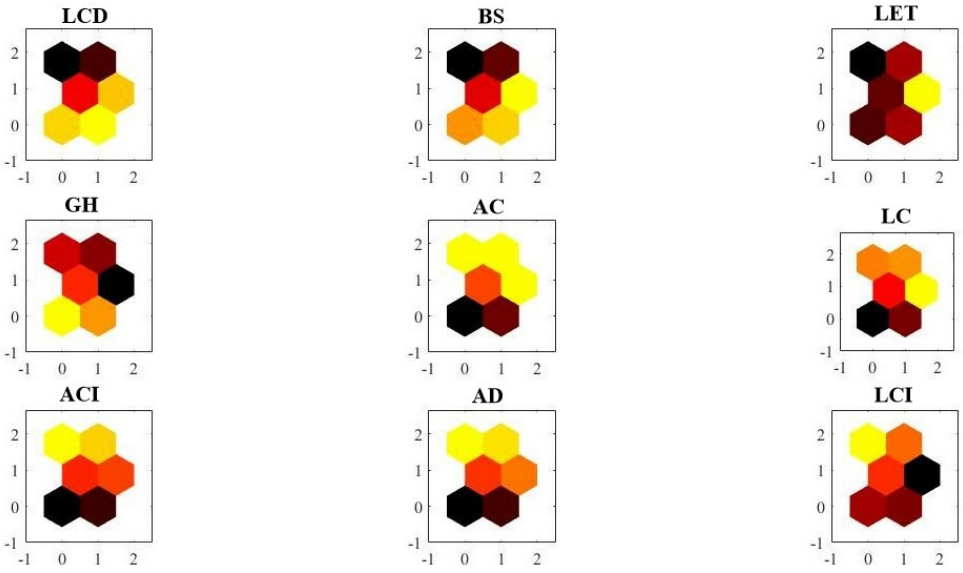

Figure 4. Characteristics and weights in the activation of each neuron in the self-organizing map. Lighter colors represent a greater effect of a variable on the group determined by the neuron.

The features LCD, BS, LET, AC, and LC were the most significant in distinguishing the Pira 72 cultivar from the other genotypes. AC, ACH, and AD were important in distinguishing the group formed by the Purpurita cultivar.

The similarity in color patterns observed between the features LCD and BS; and ACI and AD indicates that these characteristics are closely correlated.

SOM is a promising methodology in the study of genetic dissimilarity obtained through artificial neural networks. These maps enable data classification based on the distance between them, allowing for the visualization of similar patterns (SANTOS et al., 2019).

The SOM demonstrated greater sensitivity in discriminating between genotypes compared to the UPGMA method. While the UPGMA clustering method grouped genotypes into three clusters, the SOM grouped the same genotypes into five clusters. The superior discriminatory ability of SOMs has been reported in cotton, dwarf tomato, and rice germplasms (OLIVEIRA et al., 2021; CARDOSO et al., 2021; SANTOS et al., 2019).

The group formed by the genotypes UDI 10.000, UFU 66#3, UFU 66#4, UFU 66#7, and UFU 66#8 showed consistency between the two methods used in the study of genetic dissimilarity, indicating that these genotypes exhibit high similarity. This same scenario was observed in the clustering of genotypes UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#10, UFU 215#13, and UFU Mini Biofort. The consistency in the clustering of these genotypes underscores the reliability of the results obtained.

Allocation of the cultivars Pira 72 and Purpurita to distinct groups, both from each other and from the other genotypes, highlights that the genotypes from this breeding program are different from those available in the market.

The color patterns observed in the SOM are related to the distance between neurons and the importance of a characteristic in distinguishing a cluster (CARDOSO et al., 2021). The similarity in color patterns between the LCD and BS characteristics emphasizes the correlation between these two traits related to leaf curling. The same observation was made for the ACI and AD characteristics, which are correlated with anthocyanin presence in the leaves.

The results showed the possibility of forming different groups of bloodlines lettuce biofortified using the cluster method. However, the morphological descriptors suggested by the Ministry of Agriculture, Livestock and Supply are merely descriptive, which often makes it difficult to characterize the genotypes and consequently, their replication. Therefore, the use of artificial neural networks can facilitate the process of characterizing germplasm banks.

The proposed methodology was used for mini lettuce lines obtained from the UFU germplasm bank. In this sense, this study supports novel research with more strains and can be used for other lettuce segments.

CONCLUSION

Morphological descriptors were effective in assessing the variability of genotypes, separating them into three distinct groups. The use of genetic distance analyses and SOM dendrogram proved efficient in selecting individuals UFU 215#1, UFU 215#2, UFU 215#6, UFU 215#10, and UFU 215#13, which were grouped with the cultivar UFU Mini Biofort, considering a cultivar with a commercial standard for mini lettuce. SOM demonstrated higher efficiency in representing the genetic dissimilarity among the studied genotypes. The descriptors ACI, AD, AC, LC, and LCI were the most influential in evaluating genetic variability; therefore, they are recommended for studies in germplasm banks.

ACKNOWLEDGMENTS

The authors thank the Agronomy Graduate Program at the Universidade Federal de Uberlândia (UFU) for supporting this research and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for funding (Process number APQ-01952-18).

REFERENCES

ALI-SHTAYEH, M. S. et al. Morphological characterization of snake melon (*Cucumis melo var. flexuosus*) populations from Palestine. **Genetic Resources and Crop Evolution**, 64: 7-22, 2017.

BECK, H. E. et al. Present and future Köppen–Geiger climate classification maps at 1-km resolution. **Scientific Data**, 1: 1-12, 2018.

BENDOKAS, V. et al. Anthocyanins: From plant pigments to health benefits at mitochondrial level. **Critical Reviews in Food Science and Nutrition**, 60: 1-14, 2019.

BOHÓRQUEZ, A. C. C. et al. Seleção de cultivares de alface para produção na mesorregião do Alto Solimões - AM. **Cultura Agronômica**, 26: 115-122, 2017.

BURLE, M. L.; OLIVEIRA, M. S. P. **Manual de curadores de germoplasma - vegetal: caracterização morfológica**. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 2010. 15 p.

CARDOSO, D. B. O. et al. Colored fiber cotton in the Uberlândia region using artificial neural networks for yield assessment. **Genetics and Molecular Research**, 18: 1-13, 2019.

CARDOSO, D. B. O. et al. Use of computational intelligence in the genetic divergence of colored cotton plants. **Bioscience Journal**, 37: 1-9, 2021.

CARVALHO et al. **Anuário Brasileiro de Horti & Fruti 2020**. 2019. Disponível em: <http://www.editoragazeta.com.br/sitewp/wp-content/uploads/2020/05/HORTIFRUTI_2020.pdf>. Acesso em: 10 fev. 2023.

COVRE, E. A. et al. Physical-chemical and sensory characteristics of Brunela lettuce. **Revista Agrarian**, 13: 265-272, 2020.

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa, MG: UFV, 2014. 585 p.

CRUZ, C. D.; NASCIMENTO, M. **Inteligência computacional aplicada ao melhoramento genético**. Viçosa, MG: UFV, 2018.414 p.

DIAS, P. P. et al. Atividade fungitóxica in vitro do óleo essencial de capim limão (*Cymbopogon citratus* (D.C.) Stapf) e controle de *Rhizoctonia solani* em plantas de alface (*Lactuca sativa* L.), cultivar maravilha quatro estações. **Summa Phytopathologica**, 47: 204-208, 2021.

ESPÍRITO SANTO, L. M.; MENEZES, B. R.; CARMO, M. G. F. Genetic variability in *Capsicum spp.* accessions through multicategorical traits. **Revista Ceres**, 69: 195-202, 2022.

EVANGELISTA, R. L. C. A. et al. Caracterização morfológica de 21 genótipos de batata-doce. **Revista AgroFIB**, 2: 55-66, 2022.

FARAHANI, M. et al. Phenotypic diversity among *Morus alba var. nigra* genotypes as revealed by multivariate analysis. **Scientia Horticulturae**, 248: 41-49, 2019.

FONTES, P. C. R. Alface. In: RIBEIRO, A. C.; GUIMARÃES, P. T. G.; ALVAREZ, V. H. A. (Eds.). **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais: 5° Aproximação**. Viçosa, MG: Editora UFV, 1999. v. 1, cap. 18, p. 177.

KIM, D. E. et al. Metabolite profiling of green, green/red, and red lettuce cultivars: Variation in health beneficial compounds and antioxidant potential. **Food Research International**, 105: 361-370, 2018.

KLOOSTER, W. S. et al. Growth and physiology of deciduous shade trees in response to controlle drelease fertilizer. **Scientia Horticulturae**, 135: 71-79, 2012.

KORIR, N. K. et al. Plant variety and cultivar identification: advances and prospects. **Critical Reviews in Biotechnology**, 33: 1-15. 2012.

MAPA - Ministério da Agricultura Pecuária e Abastecimento. **Instruções para execução dos ensaios de Distinguibilidade, Homogeneidade e Estabilidade de cultivares de alface (*Lactuca sativa* L.)**. 2001. Disponível em: https:// www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios /insumos-agricolas/protecao-de-cultivar/olericolas. Acesso em: 15 fev. 2023.

MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **The Computer Journal**, 20: 359-363, 1977.

MOU, B. Mutations in lettuce improvement. **International Journal of Plants Genomics**, 2011: 1-7, 2011.

OLIVEIRA, C. S. et al. Artificial neural networks and genetic dissimilarity among saladette type dwarf tomato plant populations. **Food Chemistry: Molecular Sciences**, 3: 1-9, 2021.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, AT: R Foundation for Statistical Computing, 2022. Disponível em: <https://www.r-project.org/>. Acesso em: 15 fev. 2023.

SALA, F. C.; COSTA, C. P. Retrospectiva e tendência da alfacicultura brasileira. **Horticultura Brasileira**, 30: 187-194, 2012.

SANTOS, C. P. et al. Morphological evaluation of lettuce genotypes grown under hydroponic system. **Horticultura Brasileira**, 39: 312-318, 2021.

SANTOS, I. G. et al. Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. **Acta Scientiarum Agronomy**, 41: 1-9, 2019.

SILVA, E. H. C. et al. Morphoagronomic characterization and genetic diversity of a Brazilian okra [*Abelmoschus esculentus* (L.) Moench] panel. **Genetic Resources and Crop Evolution**, 68: 371-380, 2021.

SOUSA, L. A. et al. Agronomic potential of biofortified crisphead lettuce (*Lactuca sativa*) and its reaction to rootknot nematode. **Australian Journal of Crop Science**, 13: 773-779, 2019.

SOUZA, J. L. et al. Loose-Leaf Lettuce genotypes for the Agreste region in Sergipe, Brazil. **Revista Verde**, 14: 629-634, 2019.