Ciência e Agrotecnologia
Science and Agrotechnology

Genetic diversity in half-sibling progenies of kale: A multivariate approach

Diversidade genética em progênies de meios-irmãos de couve: Uma abordagem multivariada

Orlando Gonçalves Brito[1], Natália Oliveira Silva[2], Valter Carvalho de Andrade Júnior[1]*, Amanda Gonçalves Guimarães[3],
Márcia Regina da Costa[2], Alcinei Mistico Azevedo[4], Marcelo Augusto Vieira Santos[2], Eduardo Alves da Silva[1]

ABSTRACT

Kale is one of the most widely cultivated leafy vegetables in Brazil, but its breeding genetic is still limited. Therefore, it is crucial to develop research that explores the genetic diversity of the available germplasms, aiming to establish promising crosses in the studied populations. The present study aimed to evaluate the genetic divergence between half-sibling progenies of kale, obtained through recurrent selection, using multivariate methods to select the most promising progenies for forming a recombinant population. The plants obtained from the seeds of 25 pre-selected half-sibling progenies and three commercial cultivars were evaluated as treatment groups. These seeds were produced in 2018 by subjecting the clones of the progenies to recombination, followed by the selection of appropriate seeds from plants with superior agronomic performance. In the present study, plants in each plot were evaluated for twelve quantitative and fifteen qualitative traits. Genetic divergence was studied based on canonical variables to obtain quantitative data. The UPGMA method was applied to the quantitative, qualitative, and combined data, and Mahalanobis genetic distance was adopted. The first three canonical variables explained 73.75% of the total cumulative variation compared to the commercial cultivars and 55.29% when considering the half-sibling progenies exclusively. The half-sibling progenies P1, P2, P3, P4, P5, P9, P11, P15, P16, P18, P20, P22, P23, P24, and P25 were the most divergent and should, therefore, be prioritized when composing the recombinant population for the third cycle of recurrent selection because, in addition to genetic variation, these progenies also present characteristics of interest to the kale genetic breeding program. The commercial cultivars exhibited the most divergent genotypes and should, therefore, be included in the next recombination cycle.

Index terms: *Brassica oleracea*; recombination; screening; genetic distance.

RESUMO

A couve-de-folhas é uma das hortaliças folhosas mais cultivadas no Brasil, porém seu melhoramento ainda é limitado. Portanto, é crucial desenvolver pesquisas que explorem a diversidade genética dos germoplasmas disponíveis, visando estabelecer cruzamentos promissores nas populações estudadas. O objetivo deste estudo foi avaliar a divergência genética entre progênies de meios-irmãos de couve-de-folhas obtidas por seleção recorrente e selecionar as progênies mais promissoras para compor uma população recombinante utilizando métodos multivariados. Como tratamentos, foram avaliadas plantas obtidas a partir de sementes de 25 progênies de meio-irmãos pré-selecionadas e três cultivares comerciais. Para produzir essas sementes, clones das progênies foram submetidos a recombinação em 2018, com subsequente seleção de sementes de plantas com desempenho agronômico superior. Foram avaliadas doze características quantitativas e quinze qualitativas em cada parcela, totalizando vinte e sete características avaliadas. O estudo da divergência genética foi realizado através de variáveis canônicas para os dados quantitativos. O método UPGMA foi aplicado a dados quantitativos, qualitativos e combinados, utilizando-se a distância genética de Mahalanobis. As três primeiras variáveis canônicas explicaram 73,75% da variação total acumuladas quando comparadas às cultivares comerciais, e 55,29% quando considerado exclusivamente as progênies meios-irmãos. As progênies de meios-irmãos P1, P2, P3, P4, P5, P9, P11, P15, P16, P18, P20, P22, P23, P24 e P25 foram as mais divergentes e devem ser priorizadas na composição da população recombinante para o terceiro ciclo de seleção recorrente, pois além de variação genética, também apresentam características de interesse ao programa de melhoramento genético da couve. As cultivares comerciais apresentaram-se como os genótipos mais divergentes e deverão ser incluídas no próximo ciclo de recombinação.

Termos para indexação: *Brassica oleracea*; recombinação; triagem; distância genética.

Introduction

Kale (*Brassica oleracea* var. *acephala*) is an important leafy vegetable with a high content of compounds beneficial to human health, such as proteins, carbohydrates, fibers, calcium, iron, vitamins, niacin, carotenoids, antioxidants, and glucosinolates (Biegańska-Marecik, Radziejewska-Kubzdela & Marecik, 2017; Šamec, Urlić & Salopek-Sondi 2018). The high nutritional content of kale has led to a considerable increase in its consumption, and consequently, an increased demand for kale in the consumer market, warranting an expansion of the kale cropping area. In addition, research aimed at the genetic breeding of kale has become important.

Among the breeding methods adopted for kale, recurrent selection is a viable approach that aims to increase the frequency of favorable alleles and thereby increase the mean of the main characteristics (Hallauer, Carena, & Miranda Filho, 2010). The knowledge of the genetic diversity of populations subjected to recurrent selection is essential to determine the strategies that would enhance the selection gains (Cruz, Regazzi, & Carneiro, 2012; Morais Júnior et al., 2017). In addition, the study of genetic divergence in breeding programs allows for estimating the genetic relationships within a population, organizing the available germplasm, identifying superior genotypes for possible crosses, and composing the recombinant population for the next cycle, providing relevant information applicable to the development of better cultivars (Li et al., 2019; Topwal, Singh & Shanker, 2019; Zhang et al., 2023).

Multivariate analysis is an efficient tool for quantifying the degree of diversity among genotypes and integrating the information of different traits by value or in graphs, thereby facilitating the comparison and classification of genotypes based on the set of these evaluated traits (Sultana & Ghafoor, 2009; Uyeda, Caetano & Pennell, 2015; Barth et al., 2020). In this context, the use of agronomic and morphological descriptors is an efficient and inexpensive approach to evaluating the genetic diversity of a population (Zhang et al., 2023). Several techniques may be adopted for such studies, including the analysis of canonical variables (CV) and clustering methods. The study of CV requires advanced knowledge of multivariate statistics, although the results generated are easy to interpret and based on the creation of new variables that represent the original variables measured (Cruz, Carneiro & Regazzi, 2014). Clustering methods, on the other hand, aim to classify the genotypes such that the genotypes within each group are homogeneous and those in different groups are heterogeneous (Hair Jr et al., 2005; Zhang et al., 2023).

However, despite the suitability of the above techniques, a limited number of studies have been conducted to date exploring the genetic divergence in kale using different multivariate techniques. Therefore, conducting such studies is essential to provide important information for crop breeding. In this context, the present study aimed to evaluate the genetic divergence between half-sibling progenies of kale, obtained through recurrent selection, using multivariate methods to select the most promising progenies for forming a recombinant population.

Material and Methods

Obtaining the half-sibling progeny

The first step of obtaining the half-sibling progeny was conducted from January to December 2018 and comprised performing the recombination of 25 pre-selected half-sibling progenies of kale obtained from the C1 cycle of the UFVJM breeding program (Brito et al., 2020). The C0 cycle comprised evaluating the germplasm bank accessions (Azevedo et al., 2012) and the subsequent selection of the superior half-sibling progenies (Azevedo et al., 2021).

Each selected plant gave rise to a new progeny (family), which was evaluated later in the second stage of the study. In the recombination process, four shoots were collected from each of the selected plants and placed in polystyrene trays with 72 cells, each filled with the commercial substrate Carolina Soil®. The trays were then placed in greenhouses for 40 days to allow for the seedlings to root. Afterward, three seedlings were transplanted in a randomized block design, with a spacing of 0.75 m between the plants and 1.00 m between the rows, to allow for recombination to occur. The seeds of each clone were harvested when the silicas were completely dry, followed by their grouping based on the mother plant and storage in paper bags in cold chambers with a temperature of 10 °C and 50% relative humidity until sowing in February 2019.

The second stage of the study was comprised of conducting the agronomic test of the half-sibling progenies obtained in the first stage of the study.

Second stage: Treatments and design

The second stage began in April 2019 and lasted until September 2019. The seeds of the 25 half-sibling progenies (P1-P25) formed the treatment group, while those from three commercial cultivars, namely, Manteiga (C1), Manteiga da Geórgia (C2), and Manteiga Legítima Pé-alto (C3), were used as control group seeds. The experiments were conducted in a randomized block design with four replicates and six plants per plot.

Environmental conditions and implementation of the experiment

The experiment was conducted in two stages, each of which was performed in the Olericulture Sector, Department of Agronomy, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), JK *campus*, Diamantina, state of Minas Gerais, Brazil. The JK campus is located at 18° 9' S latitude and 43° 21' W longitude, at an altitude of 1400 m. The typical ortic quartz-arenic neosol is the predominant soil in this region (Santos et al., 2018). In the study period, which was between January 2018 and September 2019, the average temperature in the study region was 18.87 °C and the average relative humidity was 73.41% (Instituto Nacional de Meteorologia - INMET, 2020).

The soil was prepared with a plow and two harrows. Subsequently, raised beds with a width of 1.00 m and a height of 0.20 m were created. Seeds were then sown in polystyrene trays with 128 cells (one seed per cell), each cell filled with the commercial substrate Carolina Soil®. The trays were then placed in a greenhouse and irrigated daily until transplanting.

The transplanting occurred 50 days after the date of sowing (April 9, 2019), when seedlings with three to four fully expanded leaves appeared from the seeds. These seedlings were then arranged in the field in a double-row bed, with 0.5 m spacing between the plants within the same bed and 1.00 m spacing between the rows of different beds. The management of weeds, pests, diseases, and nutrition in the first and second stages was performed according to the recommendations for the crop (Trani et al., 2015). Fertilization was performed at the time of planting. Top-dressing was performed after each harvest.

Evaluations

The evaluations comprised 12 quantitative and 15 qualitative characteristics at the individual (plant) level. The mean values obtained for each plot were then subjected to a statistical analysis. When evaluating the kale genotypes in this experiment, it was not considered appropriate to include the effect of assessments in the statistical model. This was because these assessments often result in a zero value in the event of no leaves or shoots available for collection, which leads to issues related to ANOVA assumptions, rendering it challenging to apply the traditional parametric approaches. Therefore, the analysis was conducted at the whole plant level.

Harvesting was performed between 30 and 150 days after transplanting (DAT), at biweekly intervals, totaling eight harvests. All evaluations were performed at the individual (per plant) level. The following characteristics were evaluated: i) number of leaves characterized by a length equal to or greater than 15 centimeters, with no damage caused by pests and/or pathogen attack, and without signs of senescence (Azevedo et al., 2014); ii) leaf productivity (t ha^{-1}), calculated based on the total number of commercial leaves produced from each plant multiplied by the population of 20 thousand plants ha^{-1} used in the experiment; iii) average mass per leaf (g), calculated as the ratio between the total leaf mass and the number of leaves; iv) number of shoots.

At the time point of the last harvest, i.e., at 150 DAT, leaf biometry was performed. Three commercial and representative leaves of each plant were evaluated, and the mean values of the following were obtained: v) petiole length (cm); vi) leaf length (cm); vii) leaf blade length (cm); viii) leaf width (cm); ix) diameter of the base of the petiole (mm); x) diameter of the middle of the petiole (mm). At 160 DAT, the following values were measured: xi) plant height (cm); xii) stem diameter (mm).

In addition, the following qualitative characteristics were evaluated at 145 DAT using the scale of scores proposed by the (International Board for Plant Genetic Resources - IBPGR 1990): i) leaf incision; ii) apex shape; iii) leaf blade thickness; iv) leaf angle; v) leaf blade shape; vi) leaf margin; vii) bubbles in the leaf blade; viii) leaf tip; ix) leaf color; x) widening of the petiole; xi) petiole section; xii) stem height; xiii) color of the petiole; xiv) rib color; xv) color of the stem.

Statistical analysis

The statistical analyses were conducted using the R software (R Core Team, 2019). The analysis of variance was conducted at the individual plant level using the "manova" function in the "stats" package. The model considered was $Y_{ijk} = m + t_i + b_j + d_{ij} + eijk$, where Y_{ijk} represents the multivariate observations associated with the i-th treatment in the j-th replication and k-th individual, m is the vector of constants, t_i is the vector of effects of the i-th level of the "treatment" factor, b_j is the vector of effects of the j-th level of the "blocks" factor, d_{ij} is the interaction effect between the i-th treatment and j-th block, and e_{ijk}: is the vector of experimental error effects corresponding to the observations Y_{ijk}.

Next, the canonical variables (CV) were obtained using the "candisc" package. The quantitative data were utilized to obtain the dissimilarity matrix using the Mahalanobis method. The expression considered was $D_{ii'}^2 = V'R^{-1}V$, where $D_{ii'}^2$ represents the Mahalanobis distance between the genotypes i and i', V is the vector resulting from the difference between the response variables of the genotypes i and i', V' is the transpose of vector V, and R^{-1} is the inverse of the residual variance-covariance matrix.

Qualitative data, concerning the rating scales, were compiled into a matrix revealing the percentage occurrence of each rating for each evaluated characteristic, followed by the estimation of the Euclidean distances (Cruz, Carneiro & Regazzi, 2014) using the "Distancia" function in the "MultivariateAnalysis" package.

A joint analysis of the quantitative data and qualitative data was performed by performing the weighted averaging of these matrices, considering the number of evaluated descriptors. Subsequently, these matrices were normalized to values ranging between 0 and 1 using the "*MultivariateAnalysis*" package. After obtaining these matrices, dendrograms were constructed using the *Unweighted Pair Group Method with Arithmetic Mean* (UPGMA) approach, for which the "*hclust*" function in the "stats" package and "*as.phylo*" function in the "*ape*" package were used. The threshold point for the constructed dendrograms was established following Mojena (1977). The correlation between the three dissimilarity matrices that were obtained was estimated and evaluated for significance by conducting the Mantel test with 1000 simulations using the "*MultivariateAnalysis*" package.

Results and Discussion

The results of the multivariate analysis of variance (MANOVA) indicated a significant effect ($p < 0.05$) from all sources of variation studied for the quantitative characteristics (Table 1). The effects of "Genotype" (progenies) and "Between" (genotype x block interaction) variables were consistently significant regardless of the statistical test adopted and the presence or absence of commercial cultivars in the analysis composition. This finding indicated that it is feasible to explore the variation at both family (among the progenies) and individual (within the progenies) levels.

The first three canonical variables explained 73.75% of the total accumulated variation when considering the variation relative to the commercial cultivars (Figure 1A-1C). The first canonical variable (CV1) represented 37.64%, the second (CV2) represented 26.75%, and the third (CV3) represented 9.36% (Figure 1). However, when evaluating the half-sibling progenies exclusively (Figure 1D-1F), the explanation was lower (55.29%).

The above results indicated that, for the evaluated quantitative traits in this population, less average variation existed among the progenies compared to the commercial cultivars. Furthermore, under these conditions, the discrimination of the genotypes for the identification of ideotypes to be used as future parent genotypes might be limited when using this methodology. In order to satisfactorily interpret the genetic divergence using canonical variables, the first canonical variable should explain at least 70% of the total variation in the evaluated dataset. In this context, the first three variables were the most efficient in representing the variability manifested between the genotypes, particularly when considering the control genotypes (commercial cultivars). The inclusion of controls is important as it allows for identifying, in addition to revealing the dissimilarity between progenies, the genotypes most similar to those commonly sold in the market.

In the graphical interpretation of the dispersion of scores for CV1 (Figure 1A and 1B), the dispersion of the commercial cultivars C1 (Manteiga), C2 (Manteiga da Geórgia), and C3 (Manteiga Legítima Pé-alto) was observed. The score results for these commercial cultivars revealed greater divergence compared to the half-sibling progenies. The number of leaves, the mean mass per leaf, and the diameters of the base and middle of the petiole were the characteristics with the highest

correlations to CV1, and only the number of leaves demonstrated a negative correlation (Table 2). The cultivars C1 and C2 (which exhibited negative scores in CV1) differed strongly from C3 as the former had a higher number of leaves while having a lower average mass per leaf and smaller diameters of the base and the middle of the petiole.

In the exclusive evaluation of the half-sibling progenies in VC1 (Figure 1D and E), the progenies P2, P3, P4, P11, P16, and P25 exhibited the greatest divergence from the other progenies. The strongest correlations observed between VC1 and the studied traits were negative, with particularly notable correlations recorded for the mid-petiole diameter, plant height, average mass per leaf, leaf width, leaf productivity, petiole length, and leaf length (Table 2). Accordingly, the progenies P3, P16, and P25 were associated with higher values of these traits, which was different from what was observed for the progenies P2, P4, and P11. This behavior contributed to the greater dissimilarity among these genotypes when the half-sib progenies were evaluated exclusively.

Considering the second canonical variable (CV2) in the compositions with the presence of commercial cultivars (Figure 1A and 1C), the highest correlations were noted for the mean mass per leaf, number of shoots, leaf width, and plant height, and each of these correlations was negative (Table 2). In addition, the three commercial cultivars, C1, C2, and C3 (with positive scores in CV2) presented a lower number of shoots, narrower leaves, and lower plant height, different from what was observed for the P3, P16, and P25 progenies. These characteristics are common in commercial kale cultivars as the selection of compact and productive genotypes is sought in crop breeding programs (Azevedo et al., 2017).

Table 1: Results of the multivariate analysis of variance for the quantitative traits evaluated in the half-sibling progenies and the commercial cultivars (controls) of kale.

| Effect	Statistical test	With control cultivars				Without control cultivars			
		Df[/1]	Valor test	F approx.	P value	Df	Valor test	F approx	P value
Block	Pillai's Trace	3	0.3968	6.00	2.200×10^{-16}***	3	0.44762	6.2	2.200×10^{-16}***
	Lambda Wilks		0.6489	6.10	2.200×10^{-16}***		0.61188	6.3	2.200×10^{-16}***
	Hotelling-Lawley		0.4730	6.20	2.200×10^{-16}***		0.54000	6.3	2.200×10^{-16}***
	Roy's Greatest Root		0.2560	10.20	2.200×10^{-16}***		0.28000	10.0	2.200×10^{-16}***
Genotype	Pillai's Trace	27	2.1657	4.00	2.200×10^{-16}***	24	1.62279	2.8	2.200×10^{-16}***
	Lambda Wilks		0.0656	4.60	2.200×10^{-16}***		0.15989	3.0	2.200×10^{-16}***
	Hotelling-Lawley		3.6100	5.30	2.200×10^{-16}***		2.09000	3.1	2.200×10^{-16}***
	Roy's Greatest Root		1.3160	23.60	2.200×10^{-16}***		0.52000	9.3	2.200×10^{-16}***
Between	Pillai's Trace	81	2.0925	1.30	3.969×10^{-7}***	72	2.08518	1.3	1.252×10^{-6}***
	Lambda Wilks		0.0957	1.30	1.453×10^{-7}***		0.09625	1.3	5.216×10^{-7}***
	Hotelling-Lawley		2.6540	1.30	4.704×10^{-8}***		2.65000	1.3	1.951×10^{-7}***
	Roy's Greatest Root		0.5040	3.00	7.910×10^{-14}***		0.52000	3.1	3.523×10^{-13}***

[/1]Degrees of freedom.
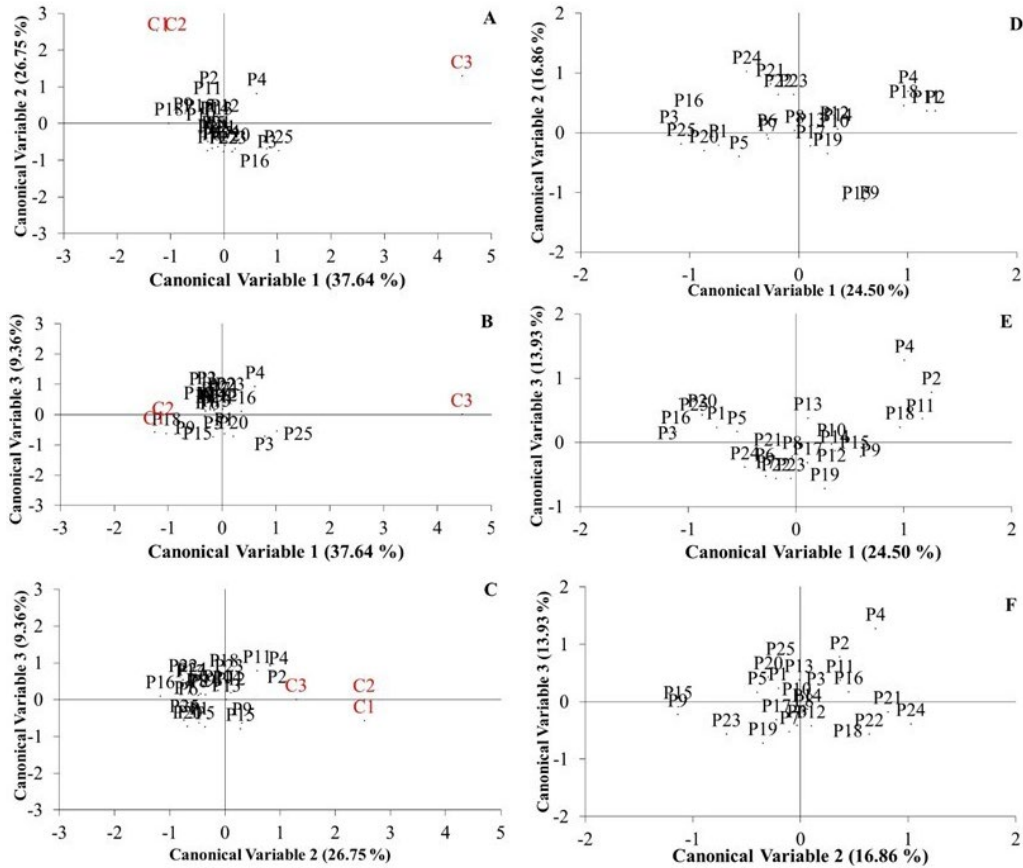
Figure 1: The standardized canonical scores for the first three canonical variables, presented in a bi-dimensional manner, for the half-sibling progenies of kale in the presence (A, B, and C) and absence (D, E, and F) of commercial cultivars in the dissimilarity analysis composition.

Table 2: Correlations of the first three canonical variables (VC1, VC2, and VC3) to the quantitative traits evaluated for the half-sibling progenies and the commercial cultivars of kale.

Characteristics	Correlation					
	With control cultivars			Without control cultivars		
	[1/]CV1	CV2	CV3	CV1	CV2	CV3
Number of leaves	-0.74	0.49	-0.24	0.24	-0.41	-0.57
Leaf productivity	-0.35	-0.28	-0.61	-0.68	-0.22	-0.17
Average mass per leaf	0.58	-0.63	-0.24	-0.78	0.13	0.36
Number of shoots	-0.13	-0.91	0.14	-0.47	0.26	-0.80
Petiole length	-0.30	-0.17	-0.28	-0.65	0.39	-0.02
Leaf length	-0.42	-0.30	-0.26	-0.65	0.17	-0.10
Leaf blade length	-0.41	-0.36	-0.14	-0.38	-0.11	-0.14
Leaf width	0.20	-0.59	-0.18	-0.71	0.26	0.13
Petiole base diameter	0.75	-0.09	-0.25	-0.63	0.13	0.35
Mid-petiole diameter	0.96	0.05	-0.20	-0.80	0.04	0.39
Plant height	-0.04	-0.74	-0.67	-0.78	-0.56	-0.30
Stem diameter	0.44	-0.18	0.05	-0.38	0.46	-0.42

[1/]CV: Canonical variable.

In the exclusive evaluation of the dispersion of half-sib progenies in VC2 (Figure 1D and 1F), P15, P19, P21, and P24 exhibited the greatest divergence from the others. Plant height demonstrated the highest correlation to VC2 (Table 2), thereby contributing significantly to explaining the divergence in this canonical variable. P21 and P24, which exhibited a correlation to this variable, diverged mainly due to their lower plant height values, unlike P15 and P19, which had taller plants. Notably, the recombination of these parentals may be of interest for kale crop breeding as it might facilitate the development of plants with medium height. This would ensure greater ergonomics for the producer, which would facilitate harvesting without compromising plant architecture.

In the evaluation of the dispersion of the half-sibling progenies and commercial cultivars based on the scores of the third canonical variable (CV3) (Figure 1B and 1C), the progenies P1, P2, P3, P4, P5, P9, P11, P15, P16, P18, P20, and P25 and the commercial cultivar C1 were the most distant from the other genotypes. In CV3, the main correlations were noted for leaf yield and plant height, each of which was negative. Therefore, the genotypes C1, P1, P3, P5, P9, P15, P20, and P25 exhibited higher leaf productivity and plant height compared to P2, P4, P11, and P18. Therefore, recombination between these genotypes is of interest as it would increase the possibility of obtaining more productive and shorter genotypes, which is considered beneficial in crop breeding.

In the evaluation of the dispersion of the progenies in VC3 in the absence of commercial cultivars in the analysis (Figure 2E and 2F), the progenies P2, P4, P18, P19, P22, and P23 exhibited greater dissimilarity compared to the others. This was attributed mainly to the high negative correlation between VC3 and the number of shoots (Table 2). Therefore, the progenies P2 and P4 exhibited lower shoot production per plant compared to the other progenies, particularly in comparison to P18, P19, P22, and P23, which were associated with higher shoot production. Plants with lower shoot production are of interest in the development of commercial cultivars as this characteristic facilitates culture practices and ensures the commercialization of seeds (Brito et al., 2020).

The genotypes with characteristics that are divergent from those of the other cultivars are of interest when aiming for new recombinations. However, it should be noted that the selection of commercial cultivars for this purpose could limit the crosses with half-sibling progenies as commercial kale cultivars have reduced flowering capacities (Brito et al., 2020). A solution to this concern would be to perform recombination of the genotypes from cultivars that will flower, including commercial cultivars, and subsequently evaluate the obtained seeds agronomically in a desired location.
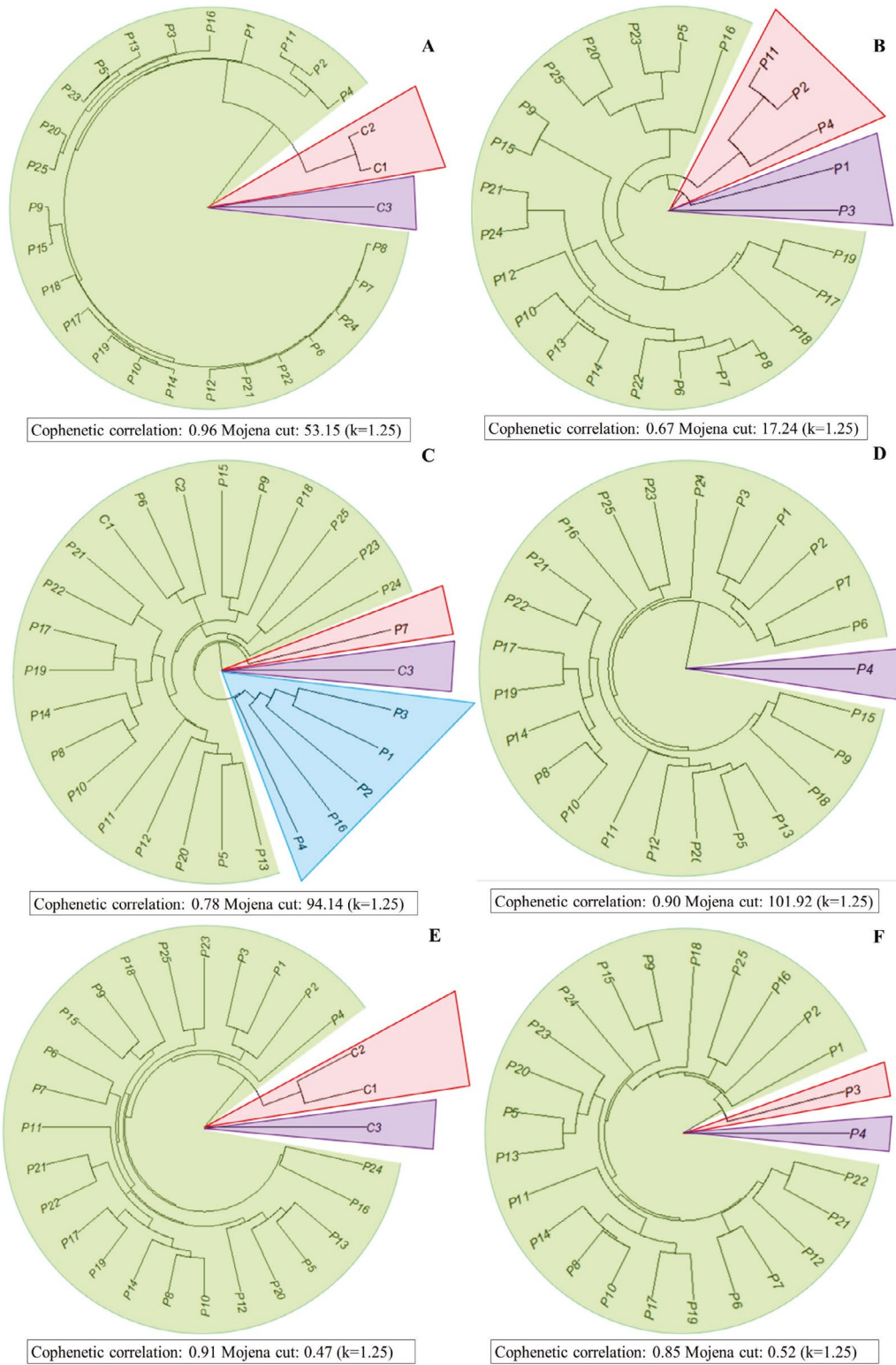
The dendrogram obtained for the quantitative characteristics using the UPGMA method indicated the formation of three groups when considering the dissimilarity relative to the commercial cultivars (Figure 2A). The first group was formed by cultivar C3, the second group by cultivars C1 and C2, and the third group was formed by all the half-sibling progenies evaluated. The clustering of the half-sibling progenies in the same group was indicative of small genetic variability in the quantitative traits of these genotypes.

In the exclusive evaluation of the quantitative data of half-sibling progenies (Figure 2B), the formation of three distinct groups was noted again. The progenies P1, P3, P4, P16, and P18 exhibited the greatest divergence and should, therefore, be prioritized for the next stage of recurrent selection, which involves recombination. These progenies were also revealed to be divergent in the study of canonical variables, particularly the P4, P16, and P18 progenies, highlighting the importance of these genotypes in composing the recombinant population for the next selection cycle. Genetic variation in recombinant populations is essential as it facilitates the identification of the most divergent genotypes, which are expected to have a greater heterotic potential for the next recombination cycle and could result in greater genetic gains (Li et al., 2019).

The dendrogram of the qualitative characteristics of the half-sibling progenies and commercial cultivars revealed greater genetic divergence among these genotypes, with the formation of four distinct groups (Figure 2C). The commercial cultivar C3 and the half-sibling progenies P4, P7, P16, and P24 exhibited the greatest divergence among all evaluated genotypes. Nineteen half-sibling progenies formed a single group with cultivars C1 and C2, indicating that these 19 progenies have qualitative characteristics similar to those of the two commercial cultivars. This finding could be of interest to breeding programs as these progenies might enhance the commercial acceptance of segregating genotypes. In the evaluation of the dispersion of half-sibling progenies based on qualitative data (Figure 2D), only two distinct groups were observed, indicating low genetic diversity in the population for qualitative traits. This was primarily attributed to the successive selection of families with a specific commercial pattern. Therefore, it was inferred that progeny P4, which was the most divergent from the others, has the potential to be used as a parent progeny in future recombinations, contributing to the increased variability in qualitative traits in future generations.

In the evaluation of quantitative and qualitative data, three distinct groups were noted in each dataset, with the presence (Figure 2E) and absence (Figure 2F) of commercial cultivars. Cultivar C3 was the most divergent, followed by cultivars C1 and C2, which were grouped into a single cluster (Figure 2E). When considering the half-sibling progenies exclusively, P3 and P4 emerged as the most dissimilar genotypes (Figure 2F). Brito et al. (2021) studied the first cycle of recurrent selection and reported greater genetic divergence in the joint analysis of quantitative and qualitative traits in the half-sibling progenies evaluated in Diamantina, MG, with the formation of six distinct groups. This reduction in genetic diversity among the genotypes evaluated in successive selection and recombination cycles was expected (Morais Júnior et al., 2017), considering the selection of genotypes with more similar traits.

Cophenetic correlation: 0.96 Mojena cut: 53.15 (k=1.25)

Cophenetic correlation: 0.67 Mojena cut: 17.24 (k=1.25)

Cophenetic correlation: 0.78 Mojena cut: 94.14 (k=1.25)

Cophenetic correlation: 0.90 Mojena cut: 101.92 (k=1.25)

Cophenetic correlation: 0.91 Mojena cut: 0.47 (k=1.25)

Cophenetic correlation: 0.85 Mojena cut: 0.52 (k=1.25)

*Analysis conducted without commercial cultivars (controls). P - Progenies/C - commercial cultivars.

Figure 2: Dendrograms obtained using the UPGMA method from the dissimilarity matrices based on quantitative traits (A; B*), qualitative traits (B; C*), and joint analysis (quantitative + qualitative) (C; D*) of the kale genotypes.

A combined multivariate analysis conducted considering both quantitative and qualitative characteristics provides important information for breeding programs by providing greater knowledge of the genetic divergence among the evaluated genotypes (Machado, Jesus & Ledo, 2015). However, the selection of parents in kale breeding programs prioritizes quantitative traits as these are the main objectives of such programs (Brito et al., 2021). However, when searching for genotypes with different leaf shapes, edges, or colors, qualitative studies are of great relevance. Currently, the consumer market prefers light green leaves, smooth edges, few bubbles, and non-prominent veins.

The cophenetic correlation coefficients determined in the present study were in the range of 0.96–0.67 for quantitative characteristics, 0.78–0.90 for qualitative characteristics, and 0.91–0.85 for the joint analysis (Figure 2A-2F). These coefficients indicated a good correlation between the dissimilarity matrix and the dendrograms generated from it. It must be noted that values above 0.80 indicate high reliability in interpreting the dendrograms (Sokal & Rohlf, 1962; Saraçli, Doğan & Doğan, 2013).

When the clusters generated for the quantitative characteristics using the CV (Figure 1A-1D) and UPGMA (Figure 2A-2F) methods were compared, similarities were observed between the dispersion of the genotypes and the groups formed in the agglomerative analysis method (dendrogram). When considering the commercial cultivars, greater genetic divergence was noted between C3 and the other evaluated genotypes and in the clustering formed by cultivars C1 and C2. When comparing the quantitative data of the half-sibling progenies exclusively, the most divergent genotypes were consistently identified in both methods, each highlighting P1, P2, P3, P4, and P11. However, according to the literature, it is common to obtain different results when using the two methods (Azevedo et al., 2013). The differences between the clusters are justified by the fact that CV considers the residual covariance matrix and the phenotypic covariance between the evaluated characteristics, while the agglomerative methods, such as dendrograms, consider the measures of dissimilarity between genotypes (Cruz, Regazzi & Carneiro, 2012). Therefore, the comparison of the results obtained using different multivariate analysis methods ensures a further accurate interpretation of the obtained results (Azevedo et al., 2015).

The mean Mahalanobis distances determined in the present study ranged from 16.41 to 148.11 for quantitative traits, from 70.91 to 108.82 for qualitative traits, and 0.32 to 0.74 for the joint analysis (Table 3), when the commercial cultivars were considered in the analysis. The Manteiga

Legítima Pé-alto cultivar (C3) exhibited the greatest divergence relative to the other genotypes evaluated in the evaluation of both quantitative characteristics and the joint analysis. when considering the qualitative characteristics, in addition to C3, the progenies P2, P18, and P24 exhibited large distances, i.e., greater divergence. Such information is important as it facilitates the identification of genotypes that should be subjected to recombination in future cycles.

When evaluating the genetic distances among the half-sibling progenies exclusively, a smaller variation was observed in the quantitative data (8.91–24.18) compared to when the commercial cultivars were included. In the qualitative data, the variation was relatively wider (72.26–150.41). However, when considering the two types of data together, the variation was similar and ranged between 0.32 and 0.72. In the quantitative analyses, the progenies P3 and P4 presented the greatest divergent, differing significantly from the other evaluated half-sibling progenies by 28% and 72%, respectively. When considering the qualitative data and the combined dataset, progeny P4 appeared as the most distant from the other genotypes by 92% and 80%, respectively.

In the presence of commercial cultivars, the dissimilarity matrix for the dendrogram of the joint analysis revealed high and significant correlation values with the dissimilarity matrices obtained for the quantitative (0.90) and qualitative (0.84) variables (Table 4). The correlation between the dissimilarity matrices of the quantitative and qualitative characteristics had moderate and significant intensity (0.51), which suggested a lower association between these variables in the study of genetic divergence among the evaluated genotypes. Similar findings were obtained when evaluating the half-sibling progenies exclusively.

In a study exploring genetic divergence in okra genotypes based on the evaluation of quantitative, qualitative, and molecular characteristics, a low association was noted between the different types of descriptors (Massucato et al., 2020). Therefore, despite the formation of different clusters for qualitative and quantitative traits when using the UPGMA analysis, the use of joint analysis in the study of genetic divergence could provide important complementary information for the genetic breeding of crops (Andrade et al., 2017). The understanding and effective utilization of the genetic diversity in kale would enable breeders to create cultivars with greater productivity, resistance to pests and diseases, and adaptability to different climatic conditions. This would, besides contributing to the conservation of kale's genetic resources, also increase the availability of high-yielding varieties with quality leaves.

Table 3: Genetic distances between the half-sibling progenies and kale cultivars.

Progenies	With control cultivars									Without control cultivars								
	Quantitative			Qualitative			Joint			Quantitative			Qualitative			Joint		
	DM[1]	MP[2]	MD[3]	DM	MP	MD	DM	MP	MD	DM	MP	MD	DM	MP	MD	DM	MP	MD
P1	29.21	P5	C3	86.16	P3	P18	0.40	P3	C3	20.93	P8	P4	88.74	P3	P4	0.49	P13	P4
P2	24.09	P11	C3	93.85	P1	P24	0.42	P1	C3	17.44	P11	P3	96.78	P1	P4	0.48	P11	P3
P3	28.38	P25	C3	91.54	P1	P18	0.42	P1	C3	23.51	P25	P18	92.3	P1	P4	0.53	P5	P18
P4	31.11	P2	C3	91.51	P21	P24	0.43	P21	C3	24.18	P2	P9	150.41	P7	P24	0.72	P11	P9
P5	19.17	P23	C3	75.28	P13	P2	0.34	P13	C3	11.63	P13	P4	75.89	P13	P4	0.36	P13	P4
P6	23.81	P7	C3	77.41	C1	C3	0.36	P14	C3	13.77	P7	P4	80.74	P14	P4	0.4	P14	P4
P7	21.83	P8	C3	92.44	P6	P18	0.41	P6	C3	11.47	P8	P4	91.79	P6	P18	0.41	P6	P4
P8	18.33	P7	C3	77.63	P10	C3	0.35	P10	C3	9.64	P7	P3	77.58	P10	P4	0.34	P10	P4
P9	24.39	P15	C3	82.32	P13	C3	0.38	P15	C3	17.15	P15	P4	84.05	P5	P4	0.44	P15	P4
P10	19.28	P14	C3	70.91	P8	C3	0.32	P8	C3	10.24	P14	P3	72.26	P8	P4	0.33	P8	P4
P11	20.67	P2	C3	81.04	P10	C3	0.37	P10	C3	14.42	P2	P3	82.79	P10	P4	0.41	P10	P3
P12	21.30	P14	C3	81.86	P13	C3	0.37	P13	C3	13.7	P14	P4	83.02	P13	P4	0.4	P13	P4
P13	16.41	P5	C3	73.78	P5	C3	0.33	P5	C3	9.25	P14	P3	73.78	P5	P4	0.33	P5	P4
P14	16.82	P8	C3	71.93	P10	C3	0.32	P10	C3	8.91	P13	P1	74.11	P10	P4	0.32	P10	P4
P15	22.03	P9	C3	86.17	P9	C3	0.39	P9	C3	15.83	P9	P4	88.24	P9	P4	0.44	P9	P4
P16	28.08	P17	C3	86.3	P13	C3	0.4	P13	C3	18.05	P25	P9	86.81	P13	P4	0.46	P13	P4
P17	24.38	P22	C3	74.66	P19	C3	0.35	P19	C3	16.18	P10	P3	75.39	P19	P4	0.4	P19	P4
P18	30.31	P10	C3	91.83	P14	C3	0.43	P14	C3	19.92	P10	P3	92.96	P14	P4	0.5	P14	P4
P19	23.40	P10	C3	79.45	P17	C3	0.36	P17	C3	14.41	P10	P4	80.02	P17	P4	0.4	P17	P4
P20	23.73	P25	C3	78.86	P5	C3	0.36	P5	C3	15.15	P25	P18	79.07	P5	P4	0.4	P5	P4
P21	22.36	P8	C3	72.04	P22	C3	0.33	P22	C3	15.96	P24	P4	74.09	P22	P4	0.4	P22	P4
P22	22.46	P8	C3	78.36	P21	C3	0.36	P21	C3	12.69	P24	P4	79.01	P21	P4	0.38	P21	P4
P23	19.99	P5	C3	86.67	P21	C3	0.39	P21	C3	13.67	P13	P18	87.67	P21	P4	0.42	P13	P4
P24	21.17	P7	C3	94.62	P10	C3	0.42	P10	C3	13.69	P21	P4	95.12	P10	P4	0.44	P22	P4
P25	27.05	P20	C3	92.25	P17	C3	0.42	P17	C3	18.48	P20	P18	92.62	P17	P4	0.48	P23	P18
C1	61.46	C2	C3	82.43	P6	P24	0.46	P27	C3	-	-	-	-	-	-	-	-	-
C2	58.43	C1	C3	87.85	P6	P24	0.47	P26	C3	-	-	-	-	-	-	-	-	-
C3	148.11	P4	P18	108.82	P1	P18	0.74	P4	P18	-	-	-	-	-	-	-	-	-

[1]AD: Average distance; [2]MP: Nearest; [3]MD: Most distant.

Table 4: Correlation between the dissimilarity matrices established for different kale genotypes.

Dendrogram	Correlation between dissimilarity matrices					
	With control cultivars			Without control cultivars		
	Quantitative	Qualitative	Joint	Quantitative	Qualitative	Joint
Quantitative	-	0.51**	0.90**	-	0.49**	0.87**
Qualitative		-	0.84**		-	0.86**
Joint			-			-

**Significance based on the Mantel test conducted with 1000 simulations.

Conclusions

Genetic divergence existed between the genotypes evaluated in the second cycle of recurrent selection. The half-sibling progenies P1, P2, P3, P4, P5, P9, P11, P15, P16, P18, P20, P22, P23, P24, and P25 should compose the recombinant population. The evaluated commercial cultivars exhibited greater genetic divergence, and it is, therefore, recommended to incorporate these genotypes into future recombinations. The multivariate techniques adopted in the present study were effective in analyzing the genetic diversity of the population.

Author Contribution

Conceptual idea: Andrade Júnior, V.C; Brito, O.G.; Silva, N.O.; Azevedo, A.M., Methodology design: Andrade Júnior, V.C; Brito, O.G.; Silva, N.O.; Azevedo, A.M.; Guimarães, A.M.; Costa, M.R., Data collection: Silva, N.O.; Santos, M.A.V.; Guimarães, A.M., Data analysis and interpretation: Brito, O.G.; Silva, N.O.; Azevedo, A.M.; Guimarães, A.M.; Silva, E.A. and Writing and editing: Brito, O.G.; Silva, N.O.; Azevedo, A.M.; Andrade Júnior, V.C; Guimarães, A.M.; Costa, M.R.

Acknowledgments

References

Andrade, E. K. V. D. et al. (2017). Genetic dissimilarity among sweet potato genotypes using morphological and molecular descriptors. *Acta Scientiarum. Agronomy*, *39*(4):447-455.

Azevedo, A. M. et al. (2012). Desempenho agronômico e variabilidade genética em genótipos de couve. *Pesquisa Agropecuária Brasileira*, *47*(12):1751-1758.

Azevedo, A. M. et al. (2013). Seleção de genótipos de alface para cultivo protegido: Divergência genética e importância de caracteres. *Horticultura Brasileira*, *31*(2):260-265.

Azevedo, A. M. et al. (2014). Divergência genética e importância de caracteres em genótipos de couve. *Horticultura Brasileira*, *32*(1):51-57.

Azevedo, A. M. et al. (2015). Divergência genética e importância de caracteres em genótipos de batata-doce visando a produção de silagem. *Revista Brasileira de Ciências Agrárias*, *10*(3):479-484.

Azevedo, A. M. et al. (2017). Population parameters and selection of kale genotypes using Bayesian inference in a multi-trait linear model. *Acta Scientiarum. Agronomy*, *39*(1):25-31.

Azevedo, A. M. et al. (2021). Determination of the optimal number of evaluations in half-sib progenies of kale by Bayesian approach. *Horticultura Brasileira, 39*(1):20-25.

Barth, E. et al. (2020). Selection of experimental hybrids of strawberry using multivariate analysis. *Agronomy*, *10*(4):598.

Biegańska-Marecik, R., Radziejewska-Kubzdela, E., & Marecik, R. (2017). Characterization of phenolics, glucosinolates and antioxidant activity of beverages based on apple juice with addition of frozen and freeze-dried curly kale leaves (*Brassica oleracea* L. var. *acephala* L.). *Food Chemistry*, 230:271-280.

Brito, O. G. et al. (2020). Genetic parameters selection gains and genotypic correlations in kale half-siblings progenies. *Emirates Journal of Food and Agriculture*, *32*(8):591-599.

Brito, O. G. et al. (2021). Genetic divergence between halfsibling progenies of kale using different multivariate approaches. *Horticultura Brasileira*, *39*(2):178-185.

Cruz, C. D., Carneiro, P. C. S., & Regazzi, A. J. (2014). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, Minas Gerais, Brasil: Editora UFV, 668p.

Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, Minas Gerais, Brasil: Editora UFV, 514p.

Hair Jr, J. F. et al. (2005). *Análise multivariada de dados*. Porto Alegre, Rio Grande do Sul, Brasil: Bookman, 578p.

Hallauer, A. R., Carena, M. J., & Miranda Filho, J. B. (2010). *Quantitative genetics in maize breeding*. New York, New York, United States: Springer Science/Business Media, 664p.

Instituto Nacional de Meteorologia -INMET. (2020). *Boletins agroclimatológicos*. Available in: <http://www.inmet.gov.br/>.

International Board for Plant Genetic Resources - IBPGR. 1990. *Descriptors for Brassica and raphanus*. Rome, Italy: International Board for Plant Genetic Resources, 51p.

Li, Z. et al. (2019). The evolution of genetic diversity of broccoli cultivars in China since 1980. *Scientia Horticulturae*, *250*:69-80.

Machado, C. D. F., Jesus, F. N., & Ledo, C. A. S. (2015). Genetic divergence of the acessions of passion fruit using quantitative and qualitative descriptors. *Revista Brasileira de Fruticultura*, *37*(2):442-449.

Massucato, L. R. et al. (2020). Genetic diversity among Brazilian okra landraces detected by morphoagronomic and molecular descriptors. *Acta Scientiarum. Agronomy*, *42*:e43426.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal*, *20*(4):359-363.

Morais Júnior, O. P. et al. (2017). Genetic variability during four cycles of recurrent selection in rice. *Pesquisa Agropecuária Brasileira*, 52(11):1033-1041.

R Core Team. (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available in: https://www.R-project.org/

Šamec, D., Urlić, B., & Salopek-Sondi, B. (2018). Kale (*Brassica oleracea* var. *acephala*) as a superfood: Review of the scientific evidence behind the statement. *Critical Reviews in Food Science and Nutrition*, *59*(15):2411-2422.

Santos, H. G. et al. (2018). *Sistema brasileiro de classificação de solos*. Brasília, Distrito Federal, Brasil: Embrapa, 356p.

Saraçli, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, *203*:2-8.

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, *11*(2):33-40.

Sultana, T., & Ghafoor, A. (2009). Botanical and molecular evidences of landraces from the germplasm exclusively collected from Baluchistan, a centre of diversity for lens culinaris. *African Journal of Biotechnology*, *8*(20):5310-5315.

Topwal, M., Singh, D. K., & Shanker, K. (2019). A review on cauliflower (*Brassica oleracea* (L.) var. *botrytis*) genotypes for genetic variability, heritability, genetic advance and correlation coefficients studies. *International Journal of Chemical Studies*, *7*(3):48-54.

Trani, P. E. et al. (2015). *Couve de folha: Do plantio à pós-colheita*. Campinas, São Paulo, Brasil: Instituto Agronômico de Campinas, 36p.

Uyeda, J., Caetano, D., & Pennell, M. (2015). Comparative analysis of principal componentes can be misleading. *Systematic Biology*, *64*(4):677-689.

Zhang, Y. et al. (2023). Phenotypic characteristics and genetic diversity of new broccoli (*Brassica oleracea* var. *italica* Plenck) varieties in China. *Scientia Agricola*, *80*:e20220223.