Revista Ciência Agronômica, v. 56, e202391693, 2025
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Genetic divergence of common bean lines for agronomic traits by hierarchical methods considering multicollinearity[1]

Nerinéia Dalfollo Ribeiro[2]*, Sandra Maria Maziero[3], Henrique da Silva Argenta[2], Greice Godoy dos Santos[2], Fabricio Fuzzer de Andrade[2], Éverton Luis da Silva Celmer[2]

ABSTRACT - Genetic divergence for agronomic traits in common bean lines can be analyzed with and without multicollinearity and by different hierarchical methods, which can lead to errors in the interpretation of results obtained from dendrograms. The objectives of this study were to evaluate the variability and genetic divergence of common bean lines for agronomic traits and to analyze the clustering pattern of hierarchical methods obtained with and without multicollinearity. The study involved data from three experiments in which 12 Mesoamerican bean genotypes were evaluated for 17 agronomic traits. Genetic divergence was analyzed through four hierarchical methods, considering two degrees of multicollinearity: severe (with multicollinearity) and weak (without multicollinearity). Ten of the 17 agronomic traits showed a significant genotype × environment interaction effect, indicating response variability of the genotypes in different growing environments. When performed with and without multicollinearity, the hierarchical methods resulted in varied clustering patterns for the genotypes regarding agronomic traits. When analyzed with and without multicollinearity, the furthest neighbor and Ward methods revealed low cophenetic correlation coefficients (≤ 0.69), explaining why the generated dendrograms do not adequately represent the variation observed between common bean genotypes. The unweighted pair group method with arithmetic mean (UPGMA) and nearest neighbor methods implemented without multicollinearity provide a high cophenetic correlation coefficient (0.76). However, the nearest neighbor method is more discriminative in allocating the common bean genotypes into different groups, allowing for a better recognition of differences regarding agronomic traits.

Key words: *Phaseolus vulgaris* L. Genotype × environment interaction. Multicollinearity diagnostics. Dendrogram analysis.

DOI: 10.5935/1806-6690.20250043
Editor-in-Chief: Profa. Charline Zaratin Alves - charline.alves@ufms.br
*Author for correspondence
Received for publication 08/07/2023; approved on 03/09/2024

[1] This research work was carried out with financial support from National Council for Scientific and Technological and Development (CNPq)

[2] Departamento de Fitotecnia, Federal University of Santa Maria (UFSM), Santa Maria-RS, Brazil, nerineia@hotmail.com (ORCID ID 0000-0002-5539-0160), henriqueargent@gmail.com (ORCID ID 0000-0001-5464-6382), godoy.greice@gmail.com (ORCID ID 0000-0002-9104-0379), fabriciofuzzer@yahoo.com (ORCID ID 0000-0003-3383-0038), evertoncelmer@gmail.com (ORCID ID 0000-0002-7153-7580)
[3] Agronomy Course, Federal University of Fronteira Sul (UFFS), Erechim-RS, Brazil, maziero.sandra@gmail.com (ORCID ID 0000-0003-4811-3445)

INTRODUCTION

Common bean (*Phaseolus vulgaris* L., 2n = 2x = 22) is grown on all continents except Antarctica. It contributes most of the daily intake of calories, protein, and minerals to the diet of thousands of people in many countries (Suárez-Martínez *et al*., 2016). For this reason, breeding programs have put great effort into developing common bean cultivars with high grain yield potential and agronomic traits that meet the demand of producers, namely, an early to normal cycle (65 to 95 days), upright plant architecture, and disease resistance.

To achieve these goals, breeding programs must identify promising parents to be used in crosses to generate populations with genetic variability. Parents with contrasting agronomic traits have been chosen and recommended for use in controlled crosses, through different methods of genetic divergence analysis (Carvalho *et al*., 2016; Ribeiro; Santos; Argenta, 2022; Santos *et al*., 2022). Among the available methods, hierarchical approaches are widely used due to the ease of visually interpreting similarities and differences between genotypes in the generated dendrograms. The most commonly used hierarchical methods for evaluating the genetic divergence of common bean genotypes regarding agronomic and/or morphological traits are the UPGMA (Carvalho *et al*., 2016; Catarcione *et al*., 2023; Hegay *et al*., 2014; Long *et al*., 2020; Nkhata *et al*., 2020; Oliveira *et al*., 2021; Savic *et al*., 2019), nearest neighbor (Barelli *et al*., 2009; Correa; Gonçalves, 2012), furthest neighbor (Cargnelutti Filho; Storck; Ribeiro, 2009; Santos *et al*., 2022), and Ward (Boros; Wawer; Borucka, 2014; Delfini *et al*., 2017; Gupta *et al*., 2021; Kefelegn; Mekibib; Dessalegn, 2020; Nogueira *et al*., 2021).

However, most of the previous studies evaluating the genetic divergence of common bean lines for multiple traits using hierarchical methods failed to mention whether these analyses were processed with or without multicollinearity. When the furthest neighbor method was carried out using different degrees of multicollinearity, there was a change in the clustering pattern of the common bean cultivars for agronomic traits (Cargnelutti Filho; Storck; Ribeiro, 2009). The clustering pattern of common bean cultivars for seed physiological quality traits varied under salt stress conditions, resulting in different compositions of the groups formed by Tocher analysis performed with and without multicollinearity (Guilhen *et al*., 2016). Therefore, in the presence of multicollinearity, it is necessary to identify which traits are causing the greatest multicollinearity problems and establish criteria for excluding these traits before conducting genetic divergence analyses (Cruz; Carneiro, 2003). This strategy can reduce errors in the interpretation of results obtained from the dendrograms.

When genetic divergence analysis is performed with and without multicollinearity, it is hypothesized that the clustering pattern of different hierarchical methods varies for the same set of evaluated common bean genotypes regarding agronomic traits. Applying different hierarchical methods, with and without multicollinearity, on a database of Value for Cultivation and Use experiments from the Southern Brazilian Common Bean Network (VCU-SB), will improve our understanding of the genetic divergence between lines developed by different research institutions, as well as increase the efficiency of planning new controlled crosses. Therefore, the objectives of this study were to evaluate the variability and genetic divergence of common bean lines for agronomic traits and to analyze the clustering pattern of hierarchical methods obtained with and without multicollinearity.

MATERIAL AND METHODS

Description of experiments

Three VCU-SB experiments for the 2020 and 2021 biennium were conducted in Santa Maria, Rio Grande do Sul, Brazil, in the area of the Federal University of Santa Maria (29º42' S latitude, 53º49' W longitude, and 95 m altitude). The region has a humid subtropical climate (Alvares *et al*., 2013), which allows common bean to be sown in two seasons: rainy (September or October) and dry (January or February). The experiments were established on 10/09/2020, 02/23/2021, and 10/08/2021, corresponding to the 2020 rainy, 2021 dry, and 2021 rainy season crops, respectively.

All experiments were laid out in a randomized block design with three replications. Each experimental plot consisted of four 4-m rows spaced 0.5 m apart, with the two central rows constituting the usable area (4 m^2). A total of 12 common bean genotypes were analyzed, comprising six elite lines (LEP 01-16, LEC 03-16, CHC 04-233-2, CHP 12 355-02, LP 08-186, and LP 09-180) and six cultivars (Pérola, BRS Intrépido, FEPAGRO Triunfo, IPR Sabiá, IPR Urutau, and BRS Esteio). These genotypes were obtained by the following breeding programs: Brazilian Agricultural Research Corporation (units: Temperate Climate and Rice and Beans), Rural Development Institute of Paraná, State University of Maringá, Agricultural Research and Rural Extension Corporation of Santa Catarina, and Department of Diagnosis and Agricultural Research of the Secretariat of Agriculture, Livestock, and Rural Development of the State of Rio Grande do Sul. All common bean genotypes evaluated had small or medium-sized grains (mass of 100 grains ≤ 40 g), characteristic of the Mesoamerican gene pool, as well as the grain types most widely produced in Brazil: carioca (beige seed coat with brown streaks) and black.

The three experiments were implemented in a conventional cultivation system. Fertilization involved 200 kg ha^{-1} of the 05-20-20 formulation (urea: 45% nitrogen; superphosphate: 18% P_2O_5; and potassium chloride: 60% K_2O) and 40 kg ha^{-1} of urea (45% nitrogen). The amount of fertilizer used as base fertilization (05-20-20 formulation) and top dressing (urea) was based on the soil analysis report.

Disease and insect control was performed preventively by treating seeds with the fungicide Maxim® (fludioxonil and metalaxil-M) and the insecticide Cruiser® 350 FS (thiamethoxam), both at a dose of 200 mL 100 kg^{-1} of seeds. The use of the insecticide Engeo™ Pleno (thiamethoxam and lambda-cyhalothrin), at a dose of 125 mL ha^{-1}, was also necessary during the development of the crop to control *Diabrotica speciosa*. Competition with weeds was avoided by applying the pre-emergence herbicide Dual Gold® (S-metolachlor) at a dose of 1.25 L ha^{-1} and the post-emergence herbicide Basagran® (bentazone) at a dose of 1.50 L ha^{-1}, along with mechanical control (weeding). Irrigation was only applied when there was a risk of losing the experiments due to the low amount of precipitation recorded during the period affected by the La Niña climatic phenomenon.

Evaluation of agronomic traits

Seventeen agronomic traits were determined in all experiments (Table 1). The traits cycle, common bacterial blight (*Xanthomonas phaseoli*), lodging, and general adaptation score were analyzed in the usable area of the plots at the maturity stage (R9), before the experiments were harvested. The other agronomic traits were evaluated post-harvest.

Statistical analyses

Individual analysis of variance was performed on the data obtained in each experiment, and the significance of the genotype (G) effect was evaluated by the F test at 5% probability. The homogeneity of residual variances of the traits determined in the three experiments was analyzed by Hartley's maximum F test.

Combined analysis of variance was implemented for the 11 genotypes common to all experiments, considering the G effect and mean as fixed. The effects of environment (E), genotype × environment interaction (G × E), and error were analyzed as random. The main effects (G and E) and the G × E interaction were interpreted as significant whenever the F test showed a p value < 0.05 for the different traits.

Table 1 - Description of the 17 agronomic traits determined in 12 common bean genotypes in the experiments conducted in the years 2020 and 2021

Trait	Method description	Unit
1. Cycle	Number of days between plant emergence and maturity	days
2. Common bacterial blight	Grade scale: 1 (absence of lesions) to 9 (< 20% of pods showing lesions)	-
3. Lodging	Grade scale: 1 (all plants upright) to 9 (all plants prostrate, touching the soil)	-
4. General adaptation score	Grade scale: 1 (excellent) to 9 (very poor), based on analysis of plant architecture, disease incidence, and pod production	-
5. Insertion of the first pod	Length between the cotyledonary node and the insertion of the first pod*	cm
6. Insertion of the last pod	Length between the cotyledonary node and the insertion of the last pod*	cm
7. Plant height	Length between the cotyledonary node and the plant apex*	cm
8. Stem diameter	Measured 1 cm above ground level*	mm
9. Number of nodes	Count of the number of nodes on the main branch*	-
10. Number of branches	Count of the number of branches on the main branch*	-
11. Number of branches with pods	Count of the number of branches with pods on the main branch*	-
12. Pod width	Measured at the center of the pod, in five representative pods*	mm
13. Pod length	Measured in five representative pods*	cm
14. Number of pods per plant	Count of the number of pods per plant*	-
15. Number of grains per pod	Ratio between number of grains per plant and number of pods per plant*	-
16. Mass of 100 grains	Weight of 100 grains from three random samples, corrected to 13% moisture	g
17. Grain yield	Weight of grains obtained in the usable area, extrapolated to kg ha-1, with grain moisture standardized to 13%	kg ha^{-1}

* in 10 plants randomly sampled in the usable area

Multicollinearity diagnostics was carried out using data from the phenotypic correlation matrix of the combined analysis of variance. The degree of multicollinearity was analyzed based on the condition number (CN), which measures the ratio between the highest and lowest eigenvalues in the correlation matrix. The CN value was related to three classes of multicollinearity: severe (CN ≥ 1000), moderate to strong (100 < CN < 1000), and weak (CN ≤ 100), as established by Montgomery, Peck and Vining (2012).

The genetic dissimilarity matrix was obtained from the residual variance and covariance matrices generated in combined analysis of variance. The dissimilarity measure used was Mahalanobis' generalized distance with standardized means. To determine which traits made the greatest contribution to genetic divergence, the Singh (1981) method was applied in two scenarios: (1) with multicollinearity (severe multicollinearity; considering all evaluated traits); and (2) without multicollinearity (weak multicollinearity; after deleting the traits that caused multicollinearity effects). To obtain weak multicollinearity, highly correlated traits, traits with a variance inflation factor > 10, and traits with a greater weight in the last eigenvectors were excluded, as proposed by Cruz and Carneiro (2003).

Cluster analyses were undertaken using four hierarchical methods: unweighted pair group method with arithmetic mean (UPGMA), single linkage (nearest neighbor), complete linkage (furthest neighbor), and Ward, considering the two scenarios: with and without multicollinearity. The clustering pattern of the eight generated dendrograms was evaluated by the cophenetic correlation coefficient (CCC). The CCC was obtained from Pearson's linear correlation, from the elements of the dissimilarity matrix and the elements of the cophenetic matrix (Cruz; Carneiro, 2003). The significance of the CCC was checked by Student's t test at 1% probability. Analyses of variance, multicollinearity diagnostics, and other statistical analyses were performed using Genes software (Cruz, 2016).

RESULTS AND DISCUSSION

Individual and combined analyses of variance

Individual analyses of variances revealed that the ratio between the highest and lowest residual mean square was less than seven for all the traits evaluated in the three experiments. Therefore, homogeneous residual variances were obtained, which allowed combined analysis of variance to be performed without the need to adjust the degrees of freedom of the error or the G × E interaction.

Thirteen traits exhibited a significant genotype effect (Table 2). The common bean lines and cultivars differed for most of the agronomic traits, implying the possibility of analyzing their genetic divergence and performing selection. However, 10 of the 17 traits evaluated displayed a significant genotype × environment interaction. This result suggests that when selected common bean lines are grown in different environments, there will be variations in phenological, plant architecture, and production traits. A significant genotype × environment interaction has been reported for several agronomic traits analyzed in common-bean breeding programs (Boros; Wawer; Borucka, 2014; Cabral *et al*., 2011; Delfini *et al*., 2017; Kläsener; Ribeiro; Santos, 2022; Ribeiro; Santos; Argenta, 2022; Santos *et al*., 2022). In this case, the response variability of common bean lines in different environments (G × E interaction) must be considered in the selection and analysis of genetic divergence for multiple traits.

When significant genotype × environment interactions were observed for multiple agronomic traits, distinct groups of common bean genotypes were formed for each evaluated experiment by Tocher's and UPGMA cluster analyses (Ribeiro; Maziero, 2022). As a consequence, the evaluation of genetic divergence based on data from individual experiments will be specific to each environment and will be of little applicability for plant breeding programs. In view of this, these authors recommended the use of data from three and two experiments in Tocher's and UPGMA cluster analyses, respectively, to more assertively analyze genetic divergence for agronomic traits in common bean.

In the present study, cluster analyses were performed using a database of three field experiments. This methodology made it possible to accurately interpret the genetic diversity of common bean cultivars in relation to agronomic traits. Additionally, 12 to 26 common bean genotypes have been evaluated in each biennium in the VCU-SB experiments. The smaller number of common bean genotypes included in the 2020 to 2021 biennium (12) is due to the difficulty encountered by many research institutions in multiplying seeds of the new lines available and in conducting the VCU-SB experiments because of the coronavirus pandemic.

The selective accuracy obtained for common bacterial blight, lodging, and insertion of the first pod showed moderate experimental precision, according to the classes proposed by Resende and Alves (2020). For the other agronomic traits, experimental precision was very high (selective accuracy ≥ 0.89). High selective accuracy has been described for agronomic traits evaluated in common bean lines (Cargnelutti Filho; Ribeiro; Burin, 2010; Pereira *et al*., 2019; Ribeiro; Santos; Argenta, 2022). The magnitude of the selective accuracy observed in the present study indicates low experimental error in the determination of agronomic traits, which contributes to an efficient selection of common bean lines with traits that provide greater agronomic performance.

Table 2 - Combined analysis of variance containing the degrees of freedom (DF), mean squares, mean, coefficient of experimental variation (CEV, %), and selective accuracy (SA) for the following traits: cycle (days), common bacterial blight (Blight), lodging (LDG), general adaptation score (GAS), insertion of the first pod (IFP, cm), insertion of the last pod (ILP, cm), plant height (PH, cm), stem diameter (SD, mm), number of nodes (NN), number of branches (NB), number of branches with pods on the main branch (NBPMB), pod thickness (PT, mm), pod length (PL, cm), number of pods per plant (NPP), number of grains per pod (NGP), mass of 100 grains (M100G, g), and grain yield (Yield, kg ha^{-1}) obtained from 11 common bean genotypes evaluated in the three experiments carried out between 2020 and 2021

	DF	Mean square					
		Cycle	Blight	LDG	GAS	IFP	ILP
Block/ Environment	6	18.75	1.48	1.10	1.32	4.29	21.70
Genotype (G)	10	107.67*	1.98^{ns}	4.24^{ns}	7.73*	5.21^{ns}	96.63*
Environment (E)	2	8119.28*	306.91*	97.04*	112.85*	381.54*	2844.81*
G x E	20	66.34*	1.69^{ns}	3.76^{ns}	2.46*	9.57*	21.94^{ns}
Error	60	8.64	1.24	2.22	1.28	2.93	15.66
Mean		88.47	5.18	4.89	4.70	17.09	46.38
CEV		3.32	21.49	30.50	24.08	10.02	8.53
SA		0.96	0.61	0.69	0.91	0.66	0.92
	DF	PH	SD	NN	NB	NBPMB	PT
Block/Environment	6	13.46	1.02	0.75	0.45	0.26	0.26
Genotype (G)	10	294.31*	1.63*	4.15*	1.63*	1.62*	0.15^{ns}
Environment (E)	2	2785.85*	17.32*	12.04*	21.63*	16.84*	7.47*
G x E	20	35.65^{ns}	0.87*	0.75^{ns}	1.07*	1.06*	0.14^{ns}
Error	60	39.82	0.16	0.44	0.28	0.27	0.11
Mean		64.22	7.28	13.85	2.31	2.14	6.52
CEV		9.82	5.46	4.80	23.08	24.28	5.14
SA		0.93	0.95	0.95	0.91	0.98	0.90
	DF	PL	NPP	NGP	M100G	Yield	
Block/Environment	6	0.19	11.91	0.18	1.38	123244.38	
Genotype (G)	10	0.68*	29.84*	1.24*	27.19*	1240908.86*	
Environment (E)	2	4.69*	419.93*	0.19^{ns}	181.54*	13280216.06*	
G x E	20	0.12*	44.03*	0.12^{ns}	8.78*	168693.00*	
Error	60	0.05	6.03	0.09	1.80	81617.26	
Mean		9.49	15.74	4.59	23.05	1852.77	
CEV		2.67	15.60	6.71	5.83	15.42	
SA		0.96	0.89	0.96	0.97	0.97	

* Significant by the F test at 0.05 probability. ^{ns}: not significant

Nonetheless, the genotype and genotype × environment interaction effects were not significant for common bacterial blight, lodging, or pod width, so there was no genetic variability for these traits. For this reason, these three traits were excluded from the other statistical analyses. Similarly, in the study by Oliveira *et al*. (2021), after removing the non-polymorphic morphological descriptors, the authors conducted genetic divergence analysis recombinant inbred lines of common bean. Therefore, only the agronomic traits that exhibited genetic variability will be useful to recognize the differences between common bean genotypes in the groups formed in the dendrograms.

Multicollinearity diagnostics and contribution of agronomic traits to genetic divergence

When multicollinearity diagnostics were performed with data from the 14 traits that exhibited significant genotype and/or genotype × environment interaction effects (Table 2), a CN = 2,298,149.4 was obtained, denoting severe multicollinearity as per the classes established by Montgomery, Peck and Vining (2012). The condition of severe multicollinearity constituted scenario 1 (with multicollinearity) in the cluster analyses (Table 3).

With multicollinearity, the following traits showed the greatest contribution to genetic divergence: number of grains per pod (14.07%), pod length (13.98%), mass of 100 grains (12.09%), and number of nodes (11.14%). Together, these four traits accounted for 51.28% of the observed variation in the common bean genotypes. However, number of grains per pod was not the most important agronomic trait in differentiating the common bean genotypes when multicollinearity diagnostics was not performed before processing Mahalanobis' generalized distance (Cabral *et al*., 2011; Carvalho *et al*., 2016). Additionally, the use of a large number of agronomic traits in genetic divergence analysis in common bean resulted in a low contribution percentage for each of the different traits analyzed in the present study, confirming previously published results (Cabral *et al*., 2011; Carvalho *et al*., 2016). This can be attributed to the fact that the traits that caused the greatest multicollinearity problems were not identified and excluded before processing the genetic divergence analyses. These results reveal difficulties in identifying efficient agronomic descriptors for differentiating common bean genotypes when dissimilarity measures are carried out with multicollinearity.

To obtain scenario 2 (without multicollinearity) it was necessary to identify which traits were correlated and, consequently, leading to multicollinearity. The traits number of branches and number of branches with pods were highly correlated (r = 0.9893), with the former exhibiting a greater weight in the last eigenvectors and a greater variance inflation factor. Thus, number of branches was found to be the trait that caused the greatest multicollinearity problems and thus had to be removed. In the current study, we did not maintain all analyzed variables and incorporate a constant (K value) in the diagonal of the matrix due to the difficulty in determining an appropriate value for this constant. Furthermore, the inclusion of a K value resulted in less reliable correlation coefficient estimates compared with those obtained after excluding highly correlated agronomic traits evaluated in soybean genotypes (Del Conte *et al*., 2020).

Table 3 - Relative contribution (S.j) of the traits cycle, general adaptation score (GAS), insertion of the first pod (IFP), insertion of the last pod (ILP), plant height (PH), stem diameter (SD), number of nodes (NN), number of branches (NB), number of branches with pods on the main branch (NBPMB), pod length (PL), number of pods per plant (NPP), number of grains per pod (NGP), mass of 100 grains (M100G), and grain yield (Yield) obtained from Mahalanobis' generalized distance of 11 common bean genotypes evaluated in the three experiments carried out between 2020 and 2021

Trait	With multicollinearity		Without multicollinearity	
	Sj	%	Sj	%
Cycle	158.69	9.50	*	*
GAS	78.29	4.68	*	*
IFP	25.06	1.50	24.52	2.74
ILP	65.36	3.91	82.30	9.20
PH	85.70	5.13	*	*
SD	89.07	5.33	*	*
NN	186.09	11.14	*	*
NB	22.46	1.34	*	*
NBPMB	31.81	1.90	69.86	7.81
PL	233.62	13.98	199.28	22.28
NPP	78.78	4.71	*	*
NGP	235.12	14.07	168.43	18.83
M100G	202.07	12.09	184.64	20.64
Yield	178.77	10.70	165.35	18.49

* Traits after removed multicollinearity diagnostics was carried out

Multicollinearity diagnostics was redone with the remaining traits. This new analysis demonstrated that number of pods per plant should be excluded, applying the same described criteria that resulted in the elimination of number of branches. The procedure was repeated and indicated that the traits general adaptation score, number of nodes, cycle, stem diameter, and plant height should also be discarded in the other analyses to circumvent the effects of multicollinearity. This analysis methodology resulted in the exclusion of seven correlated traits until a CN = 45.74 was obtained, indicating weak multicollinearity according to the criteria of Montgomery, Peck and Vining (2012).

Weak multicollinearity corresponded to scenario 2 (without multicollinearity) in the cluster analyses (Table 3). Without multicollinearity, the following traits showed the greatest contribution to genetic divergence: pod length (22.28%), mass of 100 grains (20.64%), number of grains per pod (18.83%), and grain yield (18.49%). Mass of 100 grains and grain yield were considered useful agronomic descriptors to differentiate the common bean genotypes, after removing the traits that caused moderate to severe multicollinearity in data obtained from three experiments (Ribeiro; Maziero, 2022; Santos *et al*., 2022). These results denote ease in allocating common bean genotypes into different groups when Mahalanobis' generalized distance is obtained without multicollinearity.

Without multicollinearity, four traits (pod length, mass of 100 grains, number of grains per pod, and grain yield) explained, together, 80.24% of the variation observed in the genotypes, exceeding the 51.28% displayed for the four traits recognized in dissimilarity analysis with multicollinearity. Therefore, dissimilarity measures estimated without multicollinearity identified agronomic descriptors with a greater contribution to genetic divergence, i.e. more efficient descriptors to analyze the genetic variability of common bean genotypes. This enables a better differentiation between common bean genotypes clustered into different groups based on the evaluation of agronomic traits important for selection.

Clustering pattern with and without multicollinearity

The CCC values ranged from 0.50 (Ward with multicollinearity) to 0.85 (UPGMA with multicollinearity) (Figures 1 and 2), with significance at 1% probability by the t-test. These results show that the use of different hierarchical methods with and without multicollinearity resulted in variable clustering patterns. A greater range of variation in CCC values was observed when dissimilarity measures were applied to various clustering methods for agronomic traits evaluated in common bean cultivars, in which the effect of multicollinearity was not detected (Cargnelutti Filho; Ribeiro; Burin, 2010). However, after circumventing multicollinearity effects for agronomic traits analyzed in one or more common-bean VCU-SB experiments, the UPGMA method indicated a similar range of variation for the CCC values (Ribeiro; Maziero, 2022). Therefore, the dissimilarity measure employed, the standardization of means, the use of multicollinearity diagnostics (Cargnelutti Filho; Ribeiro; Burin, 2010; Cargnelutti Filho; Storck; Ribeiro, 2009), and the number of experiments from which the data were obtained (Ribeiro; Maziero, 2022) may result in changes in the clustering pattern generated by hierarchical methods for the agronomic traits analyzed in common bean lines.

Higher CCC values are associated with a greater fit between the cophenetic matrix and the genetic dissimilarity matrix (Cabral *et al*., 2011). Cophenetic correlation coefficients close to unity indicate greater consistency of the clustering pattern (Cruz; Carneiro, 2003), and can be used as a parameter for defining the clustering method to be used in studies of genetic divergence of common bean lines for agronomic traits (Cargnelutti Filho; Ribeiro; Burin, 2010). In this case, the magnitude of the CCC values must be evaluated to choose the dendrogram that most assertively represents the genetic diversity of the common bean genotypes regarding the agronomic traits important for the selection.

When genetic divergence analyses were performed using the UPGMA and nearest neighbor methods, the highest CCC values were obtained in the analyses implemented with multicollinearity (Figure 1A and 1C). However, it is important to consider that when cluster analysis is carried out with all agronomic traits evaluated, i.e., in the presence of moderate to severe multicollinearity, multicollinear traits will receive a greater weight in this analysis (Cruz; Carneiro, 2003). The use of a large number of agronomic and/or morphological traits, without previous multicollinearity diagnostics, is frequent in studies of genetic divergence in common bean using the UPGMA (Carvalho *et al*., 2016; Catarcione *et al*., 2023; Hegay *et al*., 2014; Long *et al*., 2020; Nkhata *et al*., 2020; Savic *et al*., 2019) and nearest neighbor (Barelli *et al*., 2009; Correa; Gonçalves, 2012) methods. The multicollinearity effect prevents the expression of the true degree of dissimilarity between the groups generated in the dendrograms based on agronomic traits determined in common bean cultivars (Cargnelutti Filho; Storck; Ribeiro, 2009). In this case, dissimilarity results can be misinterpreted, leading to errors in the identification of promising parents to be used in controlled crosses.

For the furthest neighbor and Ward methods, the analyses without multicollinearity showed the highest CCC values (Figure 2B and 2D). However, CCC values ≤ 0.69 were obtained, indicating a lower fit between the cophenetic matrix and the genetic dissimilarity matrix used to generate the dendrograms. Nonetheless, after excluding the traits that were highly correlated, the use of the furthest neighbor method applied to several agronomic traits evaluated in three experiments generated a CCC = 0.77, which made it possible to easily recognize the differences between the common bean landraces and cultivars and to select promising parents for use in breeding programs (Santos *et al*., 2022). In the present study, the dendrograms generated by the furthest neighbor and Ward methods without multicollinearity did not faithfully represent the real variation observed between the common bean genotypes for the different agronomic traits, due to the low magnitude found for the CCC values (CCC ≤ 0.69). For this reason, the Ward and furthest neighbor methods were not considered suitable for assessing genetic divergence from the database used in the present genetic divergence analysis.

By the UPGMA and nearest neighbor methods analyzed without multicollinearity, CCC was 0.76 (Figure 1B and 1D). The greater magnitude of the CCC value provides greater reliability in the graphical representation of the groups that can be visualized in the dendrogram. When the analyses were implemented without multicollinearity, the common bean genotypes were allocated into two and three groups by the UPGMA and nearest neighbor methods, respectively, considering 70% similarity as a criterion for defining the groups.

Figure 1 - Dendrograms and cophenetic correlation coefficient (CCC) of the unweighted pair group method with arithmetic mean (UPGMA) with (A) and without (B) multicollinearity and of the single-linkage (nearest neighbor) method with (C) and without (D) multicollinearity, using Mahalanobis' generalized distance, based on agronomic traits determined in 11 common bean genotypes evaluated in the three experiments carried out between 2020 and 2021

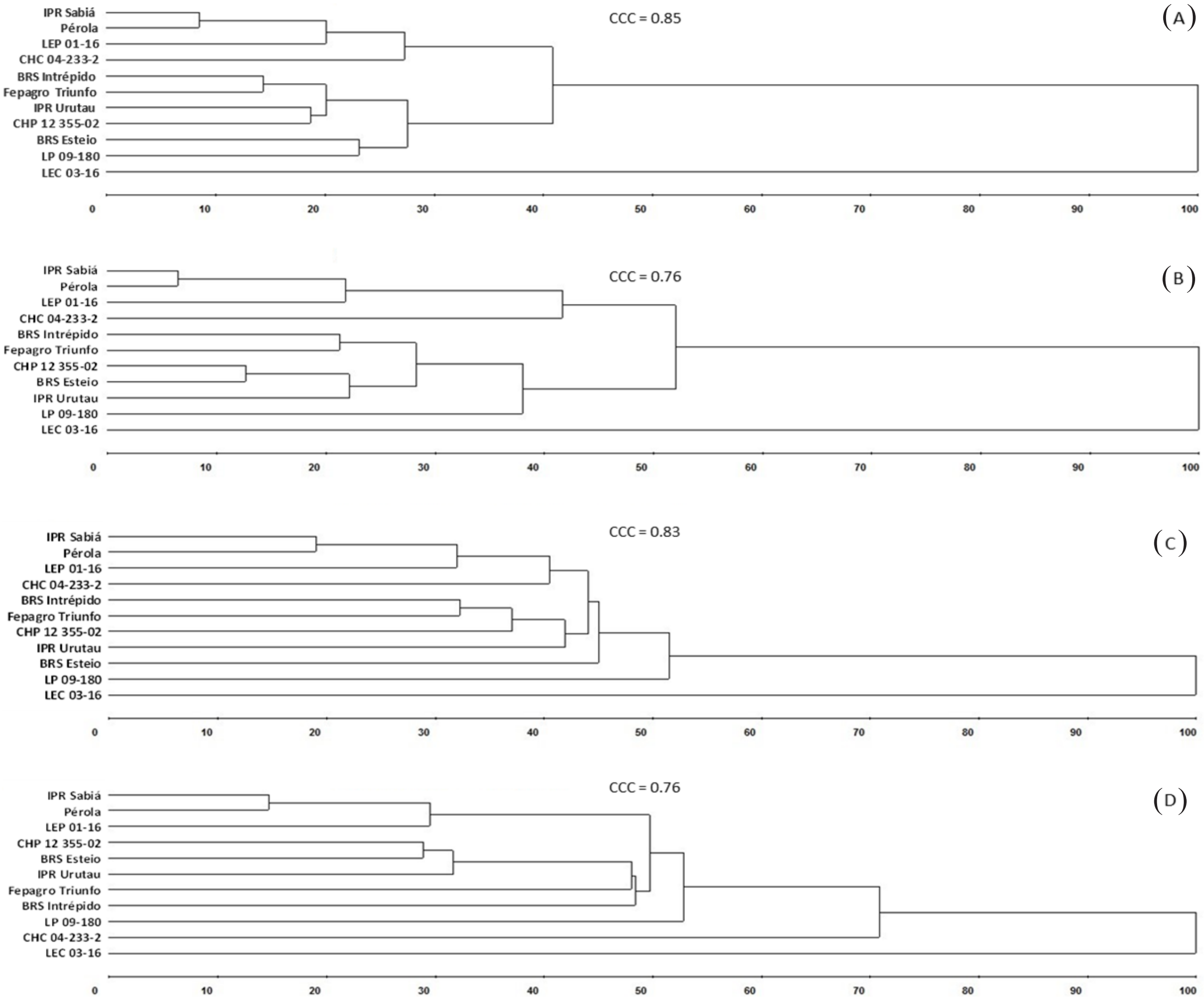
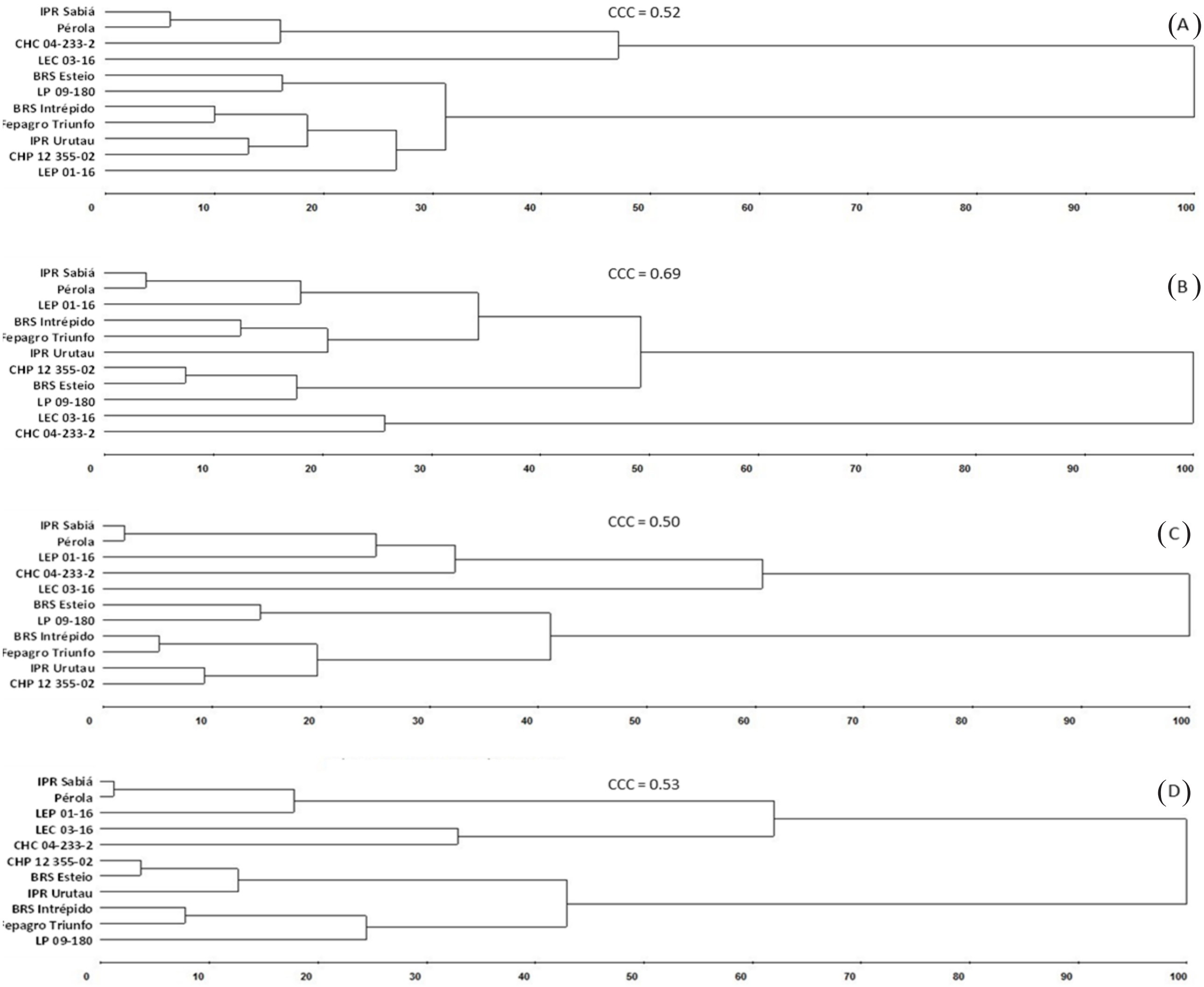


Figure 2 - Dendrograms and cophenetic correlation coefficient (CCC) of the complete-linkage (furthest neighbor) method with (A) and without (B) multicollinearity and Ward method with (C) and without (D) multicollinearity, using Mahalanobis' generalized distance, based on agronomic traits determined in 11 common bean genotypes evaluated in the three experiments carried out between 2020 and 2021



Without multicollinearity, it was possible to allocate the common bean genotypes into different groups based on agronomic traits more efficiently using the UPGMA method (Cargnelutti Filho; Ribeiro; Burin, 2010; Ribeiro; Maziero, 2022). In the present study, when analyzed without multicollinearity, the UPGMA and nearest neighbor methods generated more informative dendrograms regarding the agronomic traits that differentiate common bean genotypes. However, the nearest neighbor method without multicollinearity clustered the common bean genotypes into a larger number of groups (Figure 1 D). Group 1 was formed by line LEC 03-16, which displayed the lowest values for the following traits: number of grains per pod, mass of 100 grains, and grain yield. Group 2 consisted of line CHC 04-233-2, which differed from the other common bean genotypes, mainly due to its greater pod length. Group 3 contained the other evaluated lines and cultivars that exhibited the highest mass of 100 grains and grain yield values.

In the present study, the nearest neighbor method without multicollinearity was the most discriminative in allocating the common bean genotypes into different groups, which allowed for a better differentiation between common bean genotypes regarding agronomic traits. The nearest neighbor method without multicollinearity generated a dendrogram that more assertively represents the variability and genetic diversity of common bean genotypes in terms of agronomic traits important for the breeding program.

CONCLUSIONS

1. When genetic divergence analyses are performed with and without multicollinearity, the clustering pattern of common bean genotypes varies between hierarchical methods;
2. The unweighted pair group method with arithmetic mean and nearest neighbor method implemented without multicollinearity result in a high cophenetic correlation coefficient (0.76);
3. The nearest neighbor method is more discriminative in allocating common bean genotypes into different groups, allowing for a better recognition of differences related to agronomic traits.

ACKNOWLEDGEMENTS

To the National Council for Scientific and Technological and Development (CNPq) and to the Coordination for the Improvement of Higher Education Personnel (CAPES) for financial support and scholarships. To the Rio Grande do Sul Research Support Foundation (FAPERGS) for the grants awarded. The authors would also like to thank Ms. Luise Berleze Penna for her assistance in conducting experiments.

REFERENCES

ALVARES, C. A. *et al*. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

BARELLI, M. A. A. *et al*. Genetic divergence in common bean landrace cultivars from Mato Grosso do Sul State. **Semina Ciências Agrárias**, v. 30, n. 4, p. 1061-1072, 2009. Supplement 1.

BOROS, L.; WAWER, A.; BORUCKA, K. Morphological, phenological and agronomical characterisation of variability among common bean (*Phaseolus vulgaris* L.) local populations from The National Center for Plant Genetic Resources: Polish Genebank. **Journal of Horticultural Research**, v. 22, n. 2, p. 123-130, 2014.

CABRAL, P. D. S. *et al*. Diversidade genética de acessos de feijão comum por caracteres agronômicos. **Revista Ciência Agronômica**, v. 42, n. 4, p. 898-905, 2011.

CARGNELUTTI FILHO, A.; RIBEIRO, N. D.; BURIN, C. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de dissimilaridade e métodos de agrupamento. **Pesquisa Agropecuária Brasileira**, v. 45, n. 3, p. 236-243, 2010.

CARGNELUTTI FILHO, A.; STORCK, L.; RIBEIRO, N. D. Agrupamento de cultivares de feijão em presença e em ausência de multicolinearidade. **Ciência Rural**, v. 39, n. 9, p. 2409-2418, 2009.

CARVALHO, I. R. *et al*. Genetic variability among common black bean (*Phaseolus vulgaris* L.) accessions in southern Brazil. **Australian Journal of Crop Science**, v. 10, n. 10, p. 1474-1483, 2016.

CATARCIONE, G. de *et al*. Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces in the Lazio Region of Italy. **Plants**, v. 12, n. 4, p. 744-767, 2023.

CORREA, A. M.; GONÇALVES, M. C. Divergência genética em genótipos de feijão comum cultivados em Mato Grosso do Sul. **Revista Ceres**, v. 59, n. 2, p. 206-212, 2012.

CRUZ, C. D. Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum Agronomy**, v. 38, n. 4, p. 547-552, 2016.

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa, MG: Editora UFV, 2003. 585 p.

DEL CONTE, M. V. *et al*. Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. **Plos One**, v. 15, p. 1-15, 2020.

DELFINI, J. *et al*. Distinctness of Brazilian common bean cultivars with carioca and black grain by means of morphoagronomic and molecular descriptors. **Plos One**, v. 12, n. 11, p. 1-22, 2017.

GUILHEN, J. H. S. *et al*. Physiological characteristics in seeds of the common bean under multicollinearity and conditions of salinity. **Revista Ciência Agronômica**, v. 47, n. 1, p. 127-134, 2016.

GUPTA, C. *et al*. Genetic diversity analysis of common bean (*Phaseolus vulgaris* L.) collected from north west Himalaya for agro-morphological traits. **Plant Archives**, v. 21, n. 1, p. 1475-1481, 2021.

HEGAY, S. *et al*. Genetic diversity analysis in *Phaseolus vulgaris* L. using morphological traits. **Genetic Resources and Crop Evolution**, v. 61, n. 3, p. 555-566, 2014.

KEFELEGN, N.; MEKIBIB, F.; DESSALEGN, Y. Genetic advancement and variability of released common bean (*Phaseolus vulgaris* L.) varieties from 1974–2009 GC in Ethiopia. **Advances in Agriculture**, v. 2020, p. 1-7, 2020.

KLÄSENER, G. R.; RIBEIRO, N. D.; SANTOS, G. G. Genetic divergence and selection of common bean cultivars based on plant architecture and grain yield. **Revista Caatinga**, v. 35, n. 4, p. 818-828, 2022.

LONG, J. *et al*. Genetic diversity of common bean (*Phaseolus vulgaris* L.) germplasm resources in Chongqing, evidenced by morphological characterization. **Frontiers in Genetics**, v. 11, p. 1-9, 2020.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 5th ed. New York: John Wiley, 2012. 672 p.

NKHATA, W. *et al*. Population structure and genetic diversity analyses of common bean germplasm collections of East and Southern Africa using morphological traits and high-density SNP markers. **Plos One**, v. 15, n. 12, p. 1-23, 2020.

NOGUEIRA, A. F. *et al*. Morpho-agronomic, biochemical and molecular analysis of genetic diversity in the Mesoamerican common bean panel. **Plos One**, v. 16, n. 4, p. 1-14, 2021.

OLIVEIRA, L. S. G. *et al*. Genetic diversity of recombinant inbred lines of *Phaseolus vulgaris* L. **Revista Ciência Agronômica**, v. 52, n. 3, p. 1-9, 2021.

PEREIRA, H. S. *et al*. Genetic diversity among common bean cultivars based on agronomic traits and molecular markers and application to recommendation of parent lines. **Euphytica**, v. 215, n. 38, p. 1-16, 2019.

RESENDE, M. D. V.; ALVES, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal**, v. 2, n. 2, p. 1-31, 2020.

RIBEIRO, N. D.; MAZIERO, S. M. Number of experiments that should be considered in the cluster analysis of common bean genotypes for plant architecture and grain yield traits. **Euphytica**, v. 218, n. 89, p. 1-14, 2022.

RIBEIRO, N. D.; SANTOS, G. G.; ARGENTA, H. S. Genetic divergence and combined selection for traits that provide high agronomic performance in common bean lines. **Revista Ceres**, v. 69, n. 3, p. 274-282, 2022.

SANTOS, G. G. *et al*. Genetic diversity among bean landraces and cultivars for agronomy traits and selection of superior parents. **Revista Ciência Agronômica**, v. 53, p. 1-12, 2022.

SAVIC, A. *et al*. Genetic diversity of common bean (*Phaseous vulgaris* L.) breeding collection in Serbia. **Genetika**, v. 51, n. 1, p. 1-15, 2019.

SINGH, D. The relative importance of characters affecting genetic divergence. **The Indian Journal of Genetics & Plant Breeding**, v. 41, p. 237-245, 1981.

SUÁREZ-MARTÍNEZ, S. *et al*. Bean seeds: leading nutraceutical source for human health. **CyTA – Journal of Food**, v. 14, n. 1, p. 131-137, 2016.