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/10.1590/1983-21252025v3812587rc

Scientific Article

Grain technological quality traits and relationship to mineral concentration in common bean

Caracteres da qualidade tecnológica de grãos e relação com a concentração de minerais em feijão comum

Nerinéia D. Ribeiro[1]*, Fabricio F. de Andrade[1], Sandra M. Maziero[2]

[1]Department of Plant Science, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. [2]Universidade Federal da Fronteira Sul, Erechim, RS, Brazil.

ABSTRACT - This study proposes to investigate the correlations between technological traits and minerals under different degrees of multicollinearity; define the degree of multicollinearity to be used in Pearson's linear correlation analysis; and identify technological traits suitable for indirect selection aimed at biofortification in common bean (*Phaseolus vulgaris* L.). Twenty-five common bean cultivars of different grain types were evaluated regarding 17 technological traits and minerals across four experiments. Correlation analysis was carried out using three degrees of multicollinearity (severe, moderate to strong, and weak), achieved after excluding highly correlated traits. Most of the analyzed traits displayed genetic variability, enabling selection for both technological traits and minerals. The amplitude of variation observed for r values and the number of significant correlations varied when correlation analysis was performed under different degrees of multicollinearity. Under weak multicollinearity, all high r values (r ≥ 0.90) were excluded, completely eliminating the possible undesirable effects of multicollinearity on these estimates and thus enhancing the efficiency of indirect selection. Selecting based on the lowest values of lightness and mass of 100 grains is favorable for the indirect selection of mineral-biofortified common bean cultivars.

Keywords: *Phaseolus vulgaris*. Pearson's linear correlation analysis. Multicollinearity diagnostics.

RESUMO - Este estudo propôs investigar as correlações entre caracteres tecnológicos e minerais sob diferentes graus de multicolinearidade, definir o grau de multicolinearidade que deve ser usado na análise de correlação linear de Pearson e identificar caracteres tecnológicos úteis para a seleção indireta para a biofortificação do feijão comum. Para tanto, 25 cultivares de feijão comum de diferentes tipos de grãos foram avaliadas para 17 caracteres tecnológicos e minerais em quatro experimentos. A análise de correlação foi efetuada sob três graus de multicolinearidade (severa, moderada a forte e fraca) obtidos após a exclusão de caracteres altamente correlacionados. Variabilidade genética foi observada para a maioria dos caracteres analisados e isso possibilita a seleção para caracteres tecnológicos e minerais. A amplitude de variação observada para os valores de r e o número de correlações significativas foi variável quando a análise de correlação foi realizada sob diferentes graus de multicolinearidade. Com multicolinearidade fraca são excluídos todos os altos valores de r (r ≥ 0,90), removendo-se completamente os efeitos indesejáveis que a multicolinearidade pode causar sobre essas estimativas e isso propicia maior eficiência na seleção indireta. A seleção baseada nos menores valores de luminosidade e de massa de 100 grãos é favorável para a seleção indireta de cultivares de feijão comum biofortificadas para minerais.

Palavras-chave: *Phaseolus vulgaris*. Análise de correlação linear de Pearson. Diagnóstico de multicolinearidade.

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

Received for publication in: March 26, 2024.
Accepted in: November 7, 2024.

***Corresponding author:**
<nerineia.ribeiro@gmail.com>

INTRODUCTION

Current common bean (*Phaseolus vulgaris* L.) breeding programs have been directed towards the development of new cultivars with a focus on sustainability as well as food and nutritional security. In this scenario, the development of common bean cultivars with technological grain quality and high mineral concentration meets the demand of consumers who are aware of the nutritional quality of their diet.

To achieve this goal, a deeper understanding of the correlations between technological traits and mineral concentrations in common bean is essential, as it can empower breeding programs to optimize gains from selection. Recent studies have unveiled significant correlations between technological traits (BOROS; WAWER, 2018), between different minerals (MCCLEAN et al., 2017; KATUURAMU et al., 2018; DELFINI et al., 2020), and between technological traits and minerals (RIVERA et al., 2018) in common bean. This suggests that selecting for the enhancement of one trait can result a favorable change in the correlated trait. However, these studies did not mention whether Pearson's linear correlation analysis was implemented with or without addressing multicollinearity. Under severe multicollinearity, correlation coefficients may exhibit excessively high values that are incongruent with the biological phenomena studied in path analysis (CRUZ; CARNEIRO, 2003). The solution to this problem involves carrying out multicollinearity diagnostics on the data prior to undertaking path analysis. If multicollinearity is detected, it is necessary to identify its underlying causes and formulate strategies for its resolution.

The prior exclusion of highly correlated traits proved effective in overcoming multicollinearity in Pearson's linear correlation analysis between technological and/or nutritional traits in common bean genotypes (STECKLING et al., 2017; RIBEIRO et al., 2021a; KLÄSENER; RIBEIRO; ARGENTA, 2022; RIBEIRO; MAZIERO, 2023). However, it remains uncertain whether implementing Pearson's linear correlation analysis under different degrees of multicollinearity may lead to correlation coefficients with unexpected magnitude, sign, and/or significance or lacking biological significance. This unprecedented investigation holds significant importance for common bean breeding programs as it will greatly impact the efficiency of indirect selection.

The objectives of this study were to investigate correlations between technological traits and minerals under different degrees of multicollinearity; define the degree of multicollinearity to be used in Pearson's linear correlation analysis; and identify technological traits suitable for indirect selection aimed at biofortification in common bean.

MATERIAL AND METHODS

Database origin

The grains of the evaluated common bean cultivars were produced in four growing seasons, namely, the rainy seasons of 2019, 2020, and 2021, as well as the dry season of 2021. Rainy-season cultivation occurred from October to January, while dry-season cultivation spanned February to May, the recommended periods for common bean growing in Rio Grande do Sul (RS), Brazil. All experiments were conducted at the experimental field of the Federal University of Santa Maria, in Santa Maria, RS, Brazil (29º42′ S latitude, 53º49′ W longitude, and 95 m altitude), in soil classified as a typic alitic Argisol, Hapludalf.

The four experiments followed a randomized complete -block design with three replications. Each block comprised 25 experimental units (plots), and each plot consisted of four 4-m-long rows spaced 0.5 m apart, totaling an area of 8 m² per plot. The central two rows of each plot were considered the usable area (4 m²). A total of 25 common bean cultivars with different grain types were evaluated. The majority of these cultivars originate from the Mesoamerican gene pool, with 13 black-grain cultivars (IPR Uirapurú, IAC Netuno, BRS Esplendor, IPR Tuiuiú, BRS Valente, BRS Esteio, Fepagro 26, Guapo Brilhante, IPR Graúna, BRS Expedito, IPR Tiziu, Fepagro Triunfo, and BRS Campeiro) and ten carioca-grain (beige seed coat with brown streaks) cultivars: IPR Siriri, IPR Tangará, IAC Milênio, BRS MG Pioneiro, Fepagro Garapiá, Pérola, IAC Imperador, IPR Juriti, SCS 205 Riqueza, and BRS Estilo. Only two cultivars featured grains from the Andean gene pool of the cranberry type (cream seed coat with red streaks): BRS MG Realce and Iraí. All common bean cultivars evaluated are registered for cultivation in Rio Grande do Sul (MAPA, 2023), having been released by various breeding programs between 1980 and 2020. Therefore, the 25 common bean cultivars selected for this study represent advancements in the technological and nutritional quality of common bean breeding in Brazil over the past four decades.

For all experiments, the soil underwent preparation through the conventional cultivation system (one plowing and two harrowing operations). Soil samples were collected annually to determine the chemical composition. Fertilization was conducted based on the interpretation of the results from each year's soil analysis. Other management practices were uniformly applied and included: (1) seed treatment with the fungicide Maxim® (fludioxonil and metalaxil-M) and the insecticide Cruiser® 350 FS (thiamethoxam); (2) application of the insecticide Engeo™ Pleno (thiamethoxam and lambda-cyhalothrin); (3) mechanical control of weed plants; (4) application of the herbicides Dual Gold® (S-metolachlor) and Basagran® (bentazone); and (5) sprinkler irrigation during periods of observed water deficiency. The doses of fungicides, insecticides, and herbicides adhered to the recommendations provided by the respective manufacturers.

Harvesting occurred when 90% of the plants in the usable area exhibited dry pods (R9 stage). The plant harvesting and grain processing steps were executed manually to prevent mechanical damage to the grains. Subsequently, the grains were stored in a cold chamber (temperature of 5 ºC and relative humidity of 75%) until the start of analyses of technological and nutritional traits.

Determination of technological and nutritional traits

The grains from the 25 common bean cultivars obtained in the four experiments were analyzed for ten technological traits as well as the concentration of seven minerals. The quantitative assessment of grain color involved a random sample of 50 g of grains distributed inside a petri dish, with readings performed in triplicate for each replicate. This determination utilized a portable colorimeter, employing the three-dimensional scale L a b. On this scale, L indicated lightness (variation between black and white), a quantified chromaticity a* (variation between green and red), and b estimated chromaticity b* (variation between blue and yellow).

Grain dimensions (length, width, and thickness) were evaluated in ten randomly collected grains in each replicate using a digital caliper. Length was assessed parallel to the hilum, width was determined from the hilum to the opposite side of the grain, and thickness was measured perpendicular to the length and width of the grain. Mass of 100 grains was quantified in three samples of 100 grains collected randomly in each replicate, with moisture standardized to 13%.

The traits of normal grains, water absorption, and cooking time were analyzed in a sample of 25 grains randomly collected in each replicate. These grains were placed in Becker cups, and 50 mL of distilled water were added. After 8 h of soaking at room temperature (20 ± 2 ºC), the water was removed, and the grains were partially dried with paper towels. Normal grains were obtained by counting the grains that absorbed water (increased in size), and the result was expressed in %. Water absorption (%) was calculated using the expression:

water absorption = [(weight of grains after soaking – weight of grains before soaking)/weight of grains before soaking] × 100

The cooking time of the grains was quantified using a 25-plunger Mattson cooker. Common bean grains were placed in the holes in the device's support plate, and the cooker was positioned inside a pan with 3 L of boiling distilled water. Cooking was performed on a domestic stove, following a procedure similar to that described by Ribeiro et al. (2021a). The average dropping time of the first 13 plungers of the Mattson cooker constituted the cooking time for each sample.

The concentration of four macrominerals (potassium, phosphorus, calcium, and magnesium) and three microminerals (iron, zinc, and copper) was evaluated in a random sample of 30 g of raw grains per replicate. The grains were ground, and an aliquot of 0.5 g of common bean flour obtained was used for acid digestion, following the methodology described by Miyazawa et al. (2009). After completing the cold and hot digestion steps, the necessary dilutions for reading each mineral were carried out. The mineral concentration was determined using an atomic absorption spectrophotometer, except for potassium and phosphorus, which were quantified in a flame photometer and an optical emission spectrophotometer, respectively.

Statistical analyses

The traits of normal grains and water absorption, both expressed as percentages, underwent prior transformation using the Equation:

$$\sqrt{x + 0.5}$$

in which x represents the trait values.

Cooking time was converted from minutes to seconds. Subsequently, the database was subjected to the following statistical analyses: individual analysis of variance, combined analysis of variance, multicollinearity diagnostics, and Pearson's linear correlation analysis. All statistical procedures were conducted using Genes software (CRUZ, 2016), with a significance level set at 5% probability.

Individual analysis of variance was executed for data obtained in each growing season. The F test was employed to ascertain whether there was a significant effect of cultivars, while Hartley's maximum F test was used to assess the homogeneity of residual variances.

For combined analysis of variance, the cultivar and the mean were treated as fixed effects, whereas other effects were considered random. The significance of the main effects (cultivar and environment) and of the cultivar × environment interaction was evaluated using the F test.

Multicollinearity diagnostics utilized data from the phenotypic correlation matrix obtained from combined analysis of variance. The degree of multicollinearity was defined by the condition number (CN), categorized into three classes as proposed by Montgomery, Peck and Vining (2012): severe (CN ≥ 1000), moderate to strong (100 < CN < 1000), and weak (CN ≤ 100).

Pearson's linear correlation analysis was performed based on the phenotypic correlation matrix derived from combined analysis of variance. Correlations between technological traits and mineral concentration were examined under severe, moderate to strong, and weak multicollinearity. To achieve the latter two degrees of multicollinearity, highly correlated traits were excluded.

RESULTS AND DISCUSSION

Analysis of variance

In individual analyses of variance, the ratio between the highest and lowest residual mean square exceeded seven only for normal grains, mass of 100 grains, and iron concentration. For these three traits, residual variances were heterogeneous, requiring an adjustment of degrees of freedom of both the error and the genotype × environment interaction, as recommended by Cruz (2016). Hence, homogeneous residual variances were achieved for all 17 evaluated traits, which allowed combined analysis of variance to be performed without the necessity of trait exclusion.

In combined analysis of variance, a significant cultivar effect was observed for 13 traits (Table 1). This indicates genetic variability for most traits related to technological quality and mineral concentration in the grains among the common bean cultivars released for cultivation in RS over the last 40 years. Additionally, a significant cultivar × environment interaction was noted for 12 traits. This suggests that the cultivation of these common bean cultivars in different years and growing seasons results in variations in technological traits and macromineral concentration in the grains. Consequently, changes in the magnitude, sign, and/or significance of correlation coefficients between these traits are expected to occur depending on the growing environment, potentially compromising the efficiency of indirect selection.

A significant genotype × environment interaction was also reported for traits related to technological (SANTOS; RIBEIRO; MAZIERO, 2016; STECKLING et al., 2017; RIBEIRO; KLÄSENER, 2020; RIBEIRO et al., 2021a; KLÄSENER; RIBEIRO; ARGENTA, 2022) and nutritional (HOSSAIN et al., 2013; MCCLEAN et al., 2017; STECKLING et al., 2017; RIBEIRO; KLÄSENER, 2020; RIBEIRO et al., 2021a) quality in common bean grains. Given the significance of the genotype × environment interaction, it is important to base the choice of promising traits for indirect selection on experiments conducted in various environments. Using data from three experiments ensures a high percentage of coincidence in identifying significant correlations between technological traits and minerals in common bean grains through Pearson's linear correlation analysis (RIBEIRO; MAZIERO, 2023). The database of the present study covers evaluations from four experiments, allowing for a more accurate interpretation based on the correlation coefficients obtained between the analyzed traits.

Table 1. Combined analysis of variance containing the mean squares, degrees of freedom, mean, coefficient of experimental variation, and selective accuracy for the traits of lightness, chromaticity a*, chromaticity b*, grain length, grain width, grain thickness, mass of 100 grains, normal grains, water absorption, cooking time, and concentrations of potassium, phosphorus, calcium, magnesium, iron, zinc, and copper obtained in 25 common bean cultivars evaluated in four experiments carried out from 2019 to 2021.

Trait	Mean squares			Mean	CVE (%)	SA
	Cultivar (C) DF = 24	Environment (E) DF = 3	C x E DF = 72			
L	3387.53*	72.66*	7.28*	37.73	2.75	1.00
a	149.39*	22.23*	2.25*	4.90	9.54	0.99
b	1033.70*	20.79*	1.45*	7.23	6.01	1.00
Length (mm)	10.94*	1.53*	0.32*	10.89	3.06	0.98
Width (mm)	0.86*	3.55*	0.11*	6.79	2.63	0.93
Thickness (mm)	0.71*	1.06*	0.08*	4.84	3.59	0.94
Mass (g)	218.16*	242.11*	12.53*	24.68	5.63	0.97
NG (%)	0.17^{ns}	1.52*	0.20^{ns}	97.27	3.77	0.00
Absorption (%)	2.15^{ns}	27.86*	1.41*	92.84	7.36	0.59
CT (min:s)	98763.91*	970454.70*	40635.58*	17:48	13.46	0.77
K (g kg^{-1} DM)	4.84^{ns}	66.36*	3.57*	11.97	11.91	0.51
P (g kg^{-1} DM)	0.17^{ns}	15.71*	0.29*	4.51	9.55	0.00
Ca (g kg^{-1} DM)	0.29*	3.93*	0.06*	1.14	15.19	0.89
Mg (g kg^{-1} DM)	0.06*	16.28*	0.02^{ns}	1.43	8.80	0.85
Fe (mg kg^{-1} DM)	254.76*	2291.62*	114.50^{ns}	53.46	17.88	0.74
Zn (mg kg^{-1} DM)	26.92*	1117.03*	9.53^{ns}	24.55	11.08	0.80
Cu (mg kg^{-1} DM)	2.04*	127.11*	0.47^{ns}	7.01	10.41	0.88

*Significant by the F test at 0.05 probability; ^{ns}Not significant.
DF = degrees of freedom; CEV = coefficient of experimental variation; SA = selective accuracy; L = lightness; a = chromaticity a*; b = chromaticity b*; Length = grain length; Width = grain width; Thickness = grain thickness; Mass = mass of 100 grains; NG = normal grains; Absorption = water absorption; CT = cooking time; K = potassium; P = phosphorus; Ca = calcium; Mg = magnesium; Fe = iron; Zn = zinc; Cu = copper.

For normal grains, neither the cultivar nor the cultivar × environment interaction effects was significant. Thus, there is no genetic variability for this trait between the common bean cultivars evaluated. Consequently, this trait was excluded from multicollinearity diagnostics and Pearson's linear correlation analyses.

Experimental precision was assessed using two statistics: the coefficient of experimental variation (CEV) and selective accuracy (SA). We found CEV ≤ 17.88%, indicating high experimental precision for the determinations of different traits. Meanwhile, 15 traits showed 0.51 ≤ SA ≤ 1.00, characterizing moderate to very high experimental precision according to the classes proposed by Resende and Alves (2020). However, very low experimental precision was detected for normal grains and phosphorus concentration (SA ≤ 0.40). These results highlight that the SA statistics provide a better stratification of the experimental precision for the different technological traits and mineral concentrations evaluated in the common bean cultivars. The low experimental error, as indicated by the magnitude of the CEV and SA statistics for most technological and mineral traits evaluated in this study, contributes to greater efficiency in indirect selection.

Correlation between technological traits and minerals under severe multicollinearity

Multicollinearity diagnostics were executed for all traits, except for normal grains, which exhibited no significant differences for cultivar or cultivar × environment interaction (Table 1). In this scenario, a CN = 6,798.28 was obtained, placing it in the severe multicollinearity class according to the stratification proposed by Montgomery, Peck and Vining (2012).

Pearson's linear correlation analysis, implemented under severe multicollinearity, revealed correlation coefficients (r) ranging from -0.72 to 1.00 (Table 2). The absolute value of r does not exceed unity, concentrating in the range of -1 ≤ r ≤ 1 (CRUZ; REGAZZI, 1997).

Higher r values indicate that selection for one trait will have a greater impact on the progress of the other correlated trait. Therefore, in this study, only estimates of r ≥ 0.50 or r ≤ -0.50 were considered promising for use in indirect selection. Positive correlations imply that an increase in one trait leads to an increase in the other, while negative correlations suggest that selecting for an increase in one trait contributes to a decrease in the other correlated trait.

Table 2. Pearson's correlation coefficients obtained under severe multicollinearity between the traits of lightness, chromaticity a*, chromaticity b*, grain length, grain width, grain thickness, mass of 100 grains, water absorption, cooking time, and concentrations of potassium, phosphorus, calcium, magnesium, iron, zinc, and copper measured in 25 common bean cultivars evaluated in four experiments carried out from 2019 to 2021.

Trait	a	b	LEN	WI	THI	MAS	ABS	CT	K	P	Ca	Mg	Fe	Zn	Cu
L	0.97 *	1.00 *	0.55 *	0.42 *	0.56 *	0.53 *	0.33 ^{ns}	-0.20 ^{ns}	-0.27 ^{ns}	0.05 ^{ns}	-0.66 *	-0.20 ^{ns}	-0.55 *	0.33 ^{ns}	-0.50 *
a		0.97 *	0.69 *	0.51 *	0.62 *	0.66 *	0.30 ^{ns}	-0.12 ^{ns}	-0.32 ^{ns}	0.08 ^{ns}	-0.71 *	-0.29 ^{ns}	-0.55 *	0.39 *	-0.49 *
b			0.51 *	0.37 ^{ns}	0.54 *	0.50 *	0.34 ^{ns}	-0.23 ^{ns}	-0.22 ^{ns}	0.09 ^{ns}	-0.65 *	-0.20 ^{ns}	-0.54 *	0.28 ^{ns}	-0.49 *
LEN				0.76 *	0.83 *	0.98 *	0.22 ^{ns}	0.30 ^{ns}	-0.52 *	0.23 ^{ns}	-0.68 *	-0.44 *	-0.30 ^{ns}	0.44 *	-0.22 ^{ns}
WI					0.60 *	0.79 *	0.08 ^{ns}	0.29 ^{ns}	-0.44 *	0.31 ^{ns}	-0.50 *	-0.50 *	0.11 ^{ns}	0.53 *	0.01 ^{ns}
THI						0.90 *	0.31 ^{ns}	0.33 ^{ns}	-0.65 *	0.15 ^{ns}	-0.55 *	-0.28 ^{ns}	-0.35 ^{ns}	0.25 ^{ns}	-0.30 ^{ns}
MAS							0.24 ^{ns}	0.39 ^{ns}	-0.62 *	0.22 ^{ns}	-0.63 *	-0.41 *	-0.25 ^{ns}	0.41 *	-0.22 ^{ns}
ABS								-0.09 ^{ns}	-0.05 ^{ns}	0.19 ^{ns}	-0.11 ^{ns}	0.23 ^{ns}	-0.32 ^{ns}	-0.16 ^{ns}	-0.07 ^{ns}
CT									-0.51 *	-0.06 ^{ns}	0.19 ^{ns}	-0.06 ^{ns}	0.28 ^{ns}	-0.04 ^{ns}	0.07 ^{ns}
K										0.28 ^{ns}	0.31 ^{ns}	0.38 ^{ns}	0.02 ^{ns}	-0.17 ^{ns}	0.16 ^{ns}
P											-0.05 ^{ns}	-0.16 ^{ns}	0.00 ^{ns}	0.15 ^{ns}	0.25 ^{ns}
Ca												0.52 *	0.43 *	-0.33 ^{ns}	0.22 ^{ns}
Mg													-0.11 *	-0.40 *	-0.06 ^{ns}
Fe														0.19 ^{ns}	0.61 *
Zn															0.08 ^{ns}

*Significant by the t test at 0.05 probability; ^{ns}Not significant.
L = lightness; a = chromaticity a*; b = chromaticity b*; LEN = grain length; WI = grain width; THI = grain thickness; MAS = mass of 100 grains; ABS = water absorption; CT = cooking time; K = potassium; P = phosphorus; Ca = calcium; Mg = magnesium; Fe = iron; Zn = zinc; Cu = copper.

The application of Pearson's linear correlation analysis enhanced the understanding of associations between technological traits and mineral concentration in common bean (KAHRAMAN; ÖNDER, 2013; STECKLING et al., 2017; RIBEIRO et al., 2021a,b; RIBEIRO; MAZIERO, 2023). This approach is valuable for identifying promising traits for indirect selection, particularly for fast cooking and high mineral concentration. However, caution is warranted when performing path analysis under severe multicollinearity, as it may generate correlation coefficients with values that are excessively high and inconsistent with the biological phenomenon studied, leading to interpretation errors (CRUZ; CARNEIRO, 2003). In the case of Pearson's linear correlation analysis, is unclear whether the use of severe multicollinearity can result in correlation coefficients with unexpected magnitude, sign, and/or significance, or lacking biological significance.

In Pearson's linear correlation analysis under severe multicollinearity, 38 pairs of traits exhibited r ≥ 0.50 (Table 2). Among these, four positive correlations showed very high magnitude (r ≥ 0.97): L and a; L and b; a and b; and length and mass of 100 grains. Previous research also reported high correlations between L, a, and b in common bean when Pearson's linear correlation analysis was implemented with multicollinearity (KAHRAMAN; ÖNDER, 2013; RIBEIRO; KLÄSENER, 2020; RIBEIRO et al., 2021b). The existence of correlations between variables suggests interrelationship or multicollinearity. Under severe multicollinearity, correlation coefficients may be misinterpreted, potentially leading to incorrect conclusions (CRUZ; CARNEIRO, 2003).

In this study, four correlations (L and a; L and b; a and b; and length and mass of 100 grains) approached unity, indicating the presence of multicollinearity. Thus, evaluating the possibility of excluding one or more traits is crucial to eliminate unwanted effects of multicollinearity in Pearson's linear correction analysis. This strategy is expected to result in more reliable correlation estimates, enhancing the effectiveness of indirect selection for technological traits and minerals in common bean breeding programs.

Correlation between technological traits and minerals under moderate to strong multicollinearity

The initial multicollinearity diagnostics revealed a CN = 6,798.28, indicating severe multicollinearity (CN ≥ 1000), as per Montgomery, Peck and Vining (2012). In this analysis, the traits of L and b were highly correlated (r = 1.00). To reduce the effects of multicollinearity, the b trait was excluded, as it had the greatest contribution to the last eigenvectors and showed the highest variance inflation factor value. Multicollinearity diagnostics were then repeated, this time without b data. Despite this exclusion, this second analysis still demonstrated severe multicollinearity (CN = 2,116.10), with a high correlation between length and mass of 100 grains (r = 0.98). Employing the same criteria that led to the elimination of the b trait in the initial analysis, it was determined that the removal of the length trait was necessary.

In the third multicollinearity diagnostic analysis (excluding the values of b and length), CN = 1,267.91 and r = 0.97 were observed between the traits of L and a. The interpretation of the results from this analysis indicated the necessity to remove the a trait. In the fourth multicollinearity diagnostic analysis (excluding the values of b, length, and a), a CN = 125.11 was obtained, signifying moderate to strong multicollinearity.

When Pearson's linear correlation analysis was conducted under moderate to strong multicollinearity, r values displayed a smaller amplitude of variation, ranging from -0.66 to 0.90 (Table 3).

Table 3. Pearson's correlation coefficients obtained under moderate to strong multicollinearity between the traits of lightness, grain width, grain thickness, mass of 100 grains, water absorption, cooking time, and concentrations of potassium, phosphorus, calcium, magnesium, iron, zinc, and copper measured in 25 common bean cultivars evaluated in four experiments carried out from 2019 to 2021.

Trait	WI	THI	MAS	ABS	CT	K	P	Ca	Mg	Fe	Zn	Cu
L	0.42 *	0.56 *	0.53 *	0.33 ^{ns}	-0.20 ^{ns}	-0.27 ^{ns}	0.05 ^{ns}	-0.66 *	-0.20 ^{ns}	-0.55 *	0.33 ^{ns}	-0.50 *
WI		0.60 *	0.79 *	0.08 ^{ns}	0.29 ^{ns}	-0.44 *	0.31 ^{ns}	-0.51 *	-0.50 *	0.11 ^{ns}	0.53 *	0.01 ^{ns}
THI			0.90 *	0.31 ^{ns}	0.33 ^{ns}	-0.65 *	0.15 ^{ns}	-0.55 *	-0.28 ^{ns}	-0.35 ^{ns}	0.25 ^{ns}	-0.30 ^{ns}
MAS				0.24 ^{ns}	0.39 ^{ns}	-0.62 *	0.22 ^{ns}	-0.63 *	-0.41 *	-0.25 ^{ns}	0.41 *	-0.22 ^{ns}
ABS					-0.09 ^{ns}	-0.05 ^{ns}	0.19 ^{ns}	-0.11 ^{ns}	0.23 ^{ns}	-0.32 ^{ns}	-0.16 ^{ns}	-0.07 ^{ns}
CT						-0.51 *	-0.06 ^{ns}	0.19 ^{ns}	-0.06 ^{ns}	0.28 ^{ns}	-0.04 ^{ns}	0.07 ^{ns}
K							0.28 ^{ns}	0.31 ^{ns}	0.38 ^{ns}	0.02 ^{ns}	-0.17 ^{ns}	0.16 ^{ns}
P								-0.05 ^{ns}	-0.16 ^{ns}	-0.00 ^{ns}	0.15 ^{ns}	0.25 ^{ns}
Ca									0.52 *	0.43 *	-0.33 ^{ns}	0.22 ^{ns}
Mg										-0.11 ^{ns}	-0.40 *	-0.06 ^{ns}
Fe											0.19 ^{ns}	0.61 *
Zn												0.08 ^{ns}

*Significant by the t test at 0.05 probability; ^{ns}Not significant.
L = lightness; WI = grain width; THI = grain thickness; MAS = mass of 100 grains; ABS = water absorption; CT = cooking time; K = potassium; P = phosphorus; Ca = calcium; Mg = magnesium; Fe = iron; Zn = zinc; Cu = copper.

Therefore, the methodology of identifying the most correlated pair of traits and selecting one for elimination proved efficient in reducing multicollinearity. The exclusion of highly correlated traits efficiently addressed multicollinearity in path analysis for agronomic traits important for soybean breeding (DEL CONTE et al., 2020). In common bean, no previous studies were found examining the effect of different degrees of multicollinearity in Pearson's linear correlation analysis between technological and nutritional traits. This unprecedented investigation holds significant importance for common bean breeding programs as it will greatly impact the efficiency of indirect selection.

With moderate to strong multicollinearity, 18 pairs of traits exhibited r ≥ 0.50 (Table 3). This value marked a substantial decrease compared to the severe multicollinearity scenario (Table 2), reflecting a 47% reduction in the total number of pairs of traits with r ≥ 0.50 in Pearson's linear correlation analysis. Nonetheless, a high correlation persisted between thickness and mass of 100 grains (r = 0.90) (Table 3), indicating a partial resolution of the multicollinearity issues. When assessing numerous traits to identify promising traits for indirect selection aiming at fast cooking and high mineral concentration in common bean, focusing on a reduced number of significant correlations can enhance the efficiency of the breeding program. However, fully resolving multicollinearity problems will heighten the accuracy and efficiency of Pearson's linear correlation analysis in indirect selection.

Correlation between technological traits and minerals under weak multicollinearity

When Pearson's linear correlation analysis was carried out under weak multicollinearity (excluding the values of b, length, a, and thickness), a CN = 66.21 was recorded (Table 4). In this scenario, the r values ranged from -0.63 to 0.79, representing a narrower range of variation compared to analyses conducted under severe multicollinearity (Table 2) and moderate to strong multicollinearity (Table 3). By excluding highly correlated traits, which carry more weight in the last eigenvectors and have a higher variance inflation factor value, before implementing Pearson's linear correlation analysis, the amplitude of variation of r was reduced, and all high r values (r ≥ 0.90) were eliminated (Table 4). Consequently, this methodology completely removed the disruptive effects that multicollinearity can have on correlation estimates. Similarly, no pairs of traits with high r estimates were identified when Pearson's linear correlation analysis was employed without multicollinearity for technological and/or nutritional traits determined in common bean genotypes (STECKLING et al., 2017; RIBEIRO et al., 2021a; KLÄSENER; RIBEIRO; ARGENTA, 2022; RIBEIRO; MAZIERO, 2023). In these studies, the prior exclusion of highly correlated traits proved effective in overcoming multicollinearity, allowing for greater accuracy in the interpretation of r values.

Table 4. Pearson's correlation coefficients obtained under weak multicollinearity between the traits of lightness, grain width, mass of 100 grains, water absorption, cooking time, and concentrations of potassium, phosphorus, calcium, magnesium, iron, zinc, and copper measured in 25 common bean cultivars evaluated in four experiments carried out from 2019 to 2021.

Trait	WI	MAS	ABS	CT	K	P	Ca	Mg	Fe	Zn	Cu
L	0.42 *	0.53 *	0.33 ns	-0.20 ns	-0.27 ns	0.05 ns	-0.66 *	-0.20 ns	-0.55 *	0.33 ns	-0.50 *
WI		0.79 *	0.08 ns	0.29 ns	-0.44 *	0.31 ns	-0.51 *	-0.50 *	0.11 ns	0.53 *	0.01 ns
MAS			0.24 ns	0.39 ns	-0.62 *	0.22 ns	-0.63 *	-0.41 *	-0.25 ns	0.41 *	-0.22 ns
ABS				-0.09 ns	-0.05 ns	0.19 ns	-0.11 ns	0.23 ns	-0.32 ns	-0.16 ns	-0.07 ns
CT					-0.51 *	-0.06 ns	0.19 ns	-0.06 ns	0.28 ns	-0.04 ns	0.07 ns
K						0.28 ns	0.31 ns	0.38 ns	0.02 ns	-0.17 ns	0.16 ns
P							-0.05 ns	-0.16 ns	0.00 ns	0.15 ns	0.25 ns
Ca								0.52 *	0.43 *	-0.33 ns	0.22 ns
Mg									-0.11 ns	-0.40 *	-0.06 ns
Fe										0.19 ns	0.61 *
Zn											0.08 ns

*Significant by the t test at 0.05 probability; nsNot significant.
L = lightness; WI = grain width; MAS = mass of 100 grains; ABS = water absorption; CT = cooking time; K = potassium; P = phosphorus; Ca = calcium; Mg = magnesium; Fe = iron; Zn = zinc; Cu = copper.

Under weak multicollinearity, only 13 pairs of traits showed r values ≥ 0.50 (Table 4). The L value displayed a positive correlation with mass of 100 grains (r = 0.53) and a negative correlation with the concentrations of calcium (r = -0.66), iron (r = -0.55), and copper (r = -0.50). The highest L values were observed in Andean common bean cultivars BRS MG Realce and Iraí, which showcased the greatest lightness, i.e., a lighter grain color, and larger mass of 100 grains (≥ 40 g). Consequently, the greater grain lightness was associated with a higher mass of 100 grains and lower concentrations of calcium, iron, and copper, suggesting that Andean common bean cultivars had a lower nutritional value concerning these three minerals. However, the L value did not correlate with mass of 100 grains or mineral concentration for Mesoamerican and Andean common bean genotypes when Pearson's linear correlation analysis was conducted without multicollinearity (RIBEIRO et al., 2021a). Conversely, the magnitude, sign, and significance of r values between the L value and mineral concentration varied when correlation estimates were generated with multicollinearity, based on analyses performed with 121 Mesoamerican and 64 Andean common bean genotypes (RIBEIRO et al., 2021b). Therefore, the correlation estimates between technological and nutritional traits in common bean can be highly variable, potentially influenced by the genetic diversity of the analyzed germplasm and the absence or presence of multicollinearity in Pearson's linear correlation analysis.

Grain width exhibited a positive correlation with mass of 100 grains and zinc concentration, and a negative correlation with calcium and magnesium concentrations. However, no significant correlation was identified between width and mass of 100 grains in genotype selection experiments for carioca and black grains (SANTOS; RIBEIRO; MAZIERO, 2016) and black, carioca, red, and cranberry grains (KLÄSENER; RIBEIRO; ARGENTA, 2022) when analyses were conducted without multicollinearity. In the present study, we evaluated 25 common bean cultivars of different grain types, released for cultivation in Brazil over the past 40 years by various research institutions, contributing to greater genetic variability in germplasm concerning technological traits and mineral concentration. For common bean, no prior results of correlations between grain width and mineral concentration were found. In the current study, the selection for wider grains showed to be favorable for increasing mass of 100 grains and zinc concentration, while being unfavorable for increasing calcium and magnesium concentrations.

Mass of 100 grains exhibited a negative correlation with the concentrations of potassium (r = -0.62) and calcium (r = -0.63). Nevertheless, there was no significant correlation between mass of 100 grains and these two minerals (STECKLING et al., 2017) or between mass of 100 grains and calcium concentration (RIBEIRO; MAZIERO, 2023) when exclusively analyzing carioca- and black-grain genotypes, after overcoming issues related to multicollinearity. The present study involved the evaluation of cultivars with carioca, black, and cranberry grains, presenting greater genetic diversity compared to prior studies, which explains the observed differences. When a higher number of common bean cultivars of different grain types was included in the correlation study, it became evident that an increase in mass of 100 grains corresponded to a reduction in potassium and calcium concentrations (Table 4). Consistent with earlier studies on common bean lines featuring cranberry and red mottled grains, an inverse correlation between mass of 100 grains and the concentration of various nutrients was noted (RIBEIRO et al., 2021b). These findings demonstrate the association between common bean grain size and potassium and calcium concentrations. Specifically, in this study, cultivars with smaller grain sizes (mass of 100 grains ≤ 25 g) exhibited the highest concentrations of these macrominerals. Increased dietary potassium intake provides cardiovascular benefits (MCDONOUGH; YOUN, 2017), and raising the amount of calcium in the diet may prevent osteoporosis and bone loss (LI et al., 2018). Therefore, integrating common bean grains with a mass of 100 grains ≤ 25 g as well as rich in potassium and calcium into the diet could promote health benefits.

A negative correlation was established between cooking time and potassium concentration (r = -0.51). However, a positive correlation was observed between these two traits when Pearson's linear correlation analysis was performed under multicollinearity, based on data from experiments evaluating carioca- and black-grain common bean genotypes (RIBEIRO; KLÄSENER, 2020). These findings reinforce the importance of conducting multicollinearity diagnostics. In cases where moderate to severe multicollinearity is detected, it is essential to identify the underlying causes and implement strategies to eliminate the disruptive effects that multicollinearity may exert on correlation coefficients, as recommended by Cruz and Carneiro (2003). For the common bean cultivars evaluated, shorter cooking times were associated with increased potassium concentrations, suggesting that selecting for fast cooking also raises potassium concentration.

Positive correlation estimates (r ≥ 0.52) were found between calcium and magnesium and between iron and copper. Therefore, selecting for an increase in one of these minerals would result in an increase in the concentration of the other mineral. Similar correlation estimates in magnitude and sign between calcium and magnesium (RIBEIRO; MAZIERO, 2023) and between iron and copper (STECKLING et al., 2017) were reported for Mesoamerican common bean genotypes in analyses executed in the absence of multicollinearity. These findings support simultaneous biofortification for two minerals, representing advancements in improving the nutritional quality of common bean.

For other technological traits and mineral concentrations, correlation estimates of low magnitude or non -significance were observed. The lack of correlation between two traits indicates the absence of a linear relationship between them (CRUZ; REGAZZI, 1997). When there is no correlation between two traits, selecting for one trait does not induce a change in another, signifying their independence.

In this study, the elimination of highly correlated traits (r ≥ 0.90) before conducting Pearson's linear correlation analysis was efficient in fully resolving multicollinearity problems. This approach was valuable in removing redundant traits that provided similar information, allowing the breeding program to streamline the number of traits to be evaluated and the number of significant correlations to be analyzed. This contributes to a reduction in the time and costs associated with determining the technological and nutritional traits of common bean. Moreover, this methodology contributed to generating correlation estimates with magnitudes consistent with the biological phenomena studied, enhancing gains through indirect selection. Therefore, selecting for lower L* values is recommended for the indirect selection of common bean cultivars biofortified with calcium, iron, and copper, while choosing a lower mass of 100 grains is favorable for indirect selection aiming at high potassium and calcium concentrations in common bean grains.

CONCLUSIONS

The range of variation between correlation coefficients changes when Pearson's analysis is performed under different degrees of multicollinearity. Weak multicollinearity eliminates the disrupting effects multicollinearity can have on these estimates, increasing the efficiency of indirect selection. Selecting based on the lowest values for lightness and mass of 100 grains is favorable for the indirect selection of mineral-biofortified common bean cultivars.

ACKNOWLEDGMENTS

To the Coordination for the Improvement of Higher Education Personnel (CAPES) and to the National Council for Scientific and Technological and Development (CNPq) for financial support and scholarships.

REFERENCES

BOROS, L.; WAWER, A. Seeds quality characteristics of dry bean local populations (*Phaseolus vulgaris L.*) from National Center for Plant Genetic Resources in Radzików. **Legume Research**, 41: 669-674, 2018.

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

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 1. ed. Viçosa, MG: UFV, 2003. v. 2. 585 p.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa, MG: UFV, 1997. 390 p.

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

DELFINI, J. et al. Diversity of nutritional content in seeds of Brazilian common bean germplasm. **Plos One**, 15: 1-13, 2020.

HOSSAIN, K. G. et al. Interdependence of genotype and growing site on seed mineral compositions in common bean. **Asian Journal of Plant Sciences**, 12: 11-20, 2013.

KAHRAMAN, A.; ÖNDER, M. Correlations between seed color and nutritional composition of dry bean. **Ratarstvo i Povrtarstvo**, 50: 8-13, 2013.

KATUURAMU, D. N. et al. Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). **Molecular Breeding**, 38: 1-18, 2018.

KLÄSENER, G. R.; RIBEIRO, N. D.; ARGENTA, H. S.

Genetic divergence and selection of bean cultivars of different grain types based on physical traits. **Revista Ciência Agronômica**, 53: 1-12, 2022.

LI, K. et al. The good, the bad, and the ugly of calcium supplementation: a review of calcium intake on human health. **Clinical Interventions in Aging**, 13: 2443-2452, 2018.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. **Cultivares ou espécies registradas**. 2023. Available at: <https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/registro-nacional-de-cultivares-2013-rnc-1/cultivares-ou-especies-registradas>. Access on: Jul. 13, 2023.

MCCLEAN, P. E. et al. Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of Middle American ancestry. **Crop Science**, 57: 3129-3144, 2017.

MCDONOUGH, A. A.; YOUN, J. H. Potassium homeostasis: the knowns, the unknowns, and the health benefits. **Physiology**, 32: 100-111, 2017.

MIYAZAWA, M. et al. Análise química de tecido vegetal. In: SILVA, F. C. (Ed.). **Manual de análises químicas de solos, plantas e fertilizantes**. 2. ed. Brasília, DF: Embrapa Informação Tecnológica, 2009. v. 2, p. 191-233.

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

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**, 2: 1-31, 2020.

RIBEIRO, N. D. et al. Genetic diversity and selection of bean landraces and cultivars based on technological and nutritional traits. **Journal of Food Composition and Analysis**, 96: 1-10, 2021a.

RIBEIRO, N. D. et al. Technological-nutritional quality traits and relationship to bioactive compounds in Mesoamerican and Andean beans. **Revista Caatinga**, 34: 266-275, 2021b.

RIBEIRO, N. D.; KLÄSENER, G. Physical quality and mineral composition of new Mesoamerican bean lines developed for cultivation in Brazil. **Journal of Food Composition and Analysis**, 89: 1-8, 2020.

RIBEIRO, N. D.; MAZIERO, S. M. Indirect selection for culinary quality and minerals in beans based on genotype × environment interaction. **Revista Ciência Agronômica**, 54: 1-10, 2023.

RIVERA, A. et al. The Spanish core collection of common beans (*Phaseolus vulgaris* L.): an important source of variability for breeding chemical composition. **Frontiers in Plant Science**, 9: 1-11, 2018.

SANTOS, G. G.; RIBEIRO, N. D.; MAZIERO, S. M. Evaluation of common bean morphological traits identifies grain thickness directly correlated with cooking time. **Pesquisa Agropecuária Tropical**, 46: 35-42, 2016.

STECKLING, S. M. et al. Genetic diversity and selection of common bean lines based on technological quality and biofortification. **Genetics and Molecular Research**, 16: 1-13, 2017.