SANGIOVO, J. P.; CARVALHO, I. R.; PRADEBON, L. C.; LORO, M. V.; PORT, E. D.; SCARTON, V. D. B. Selection of soybean lines based on a nutraceutical ideotype. **Revista de Agricultura Neotropical**, Cassilândia-MS, v. 10, n. 3, e7356, July/Sep., 2023. ISSN 2358-6303. DOI: https://doi.org/10.32404/rean.v10i3.7356.

# Selection of soybean lines based on a nutraceutical ideotype

**Jaqueline Piesanti Sangiovo[1], Ivan Ricardo Carvalho[1], Leonardo Cesar Pradebon[1], Murilo Vieira Loro[2], Eduarda Donadel Port[1], Victor Delino Barasuol Scarton[1] Guilherme Hickembick Zuse[1], Eduardo Ely Foleto[1], Inâe Carolina Sfalcin[1]**

[1] Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, Rio Grande do Sul, Brasil. E-mail: jaqueline.sangiovo@sou.unijui.edu.br, carvalho.irc@gmail.com, pradebon22@gmail.com, port_duda@gmail.com, scartontt@gmail.com, guilhermezuse@gmail.com,eduardoely67@gmail.com, inae.sfalcin@hotmail.com.

[2] Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil. E-mail: muriloloro@gmail.com

Received: 20/01/2023; Accepted: 07/07/2023.

## ABSTRACT

This study aims to select soybean lines with desirable nutraceutical characteristics under specific meteorological conditions in the selection environments. The experiment was conducted during the 2021/2022 harvest at Escola Fazenda, Regional University of the Northwest of the State of Rio Grande do Sul, located in Augusto Pestana- RS, Brazil. The experimental design was augmented blocks, with 27 populations of F2 generation, 118 lines of F5 generation, 70 lines of F8 generation, and 15 soybean cultivars as interim checks, arranged in three replications. The chemical composition of soybeans was evaluated using Near Infrared Spectroscopy - NIRS method to analyze various nutritional components. The IRC050 genotype exhibited a nutraceutical profile characterized by an increment in stearic fatty acid (STEAR) and reduction in fiber (FB) and mineral material (MM) contents. Among the F5 generation, the following genotypes were selected based on their nutraceutical traits: L174F5, L307F5, L322F5, L205F5, L190F5, L196F5, L219F5, L183F5, L301F5, L172F5, L198F5, L182F5, L319F5, L21F5, L188F5, L217F5, L302F5, L315F5, L36F5, L318F5, L211F5, L165F5, L321F5, and L179F5. The indirect selection of genotypes with higher levels of linoleic acid and protein was facilitated using the enzyme peroxidase.

**Keywords:** *Glycine max,* MGIDI, Correlation.

## Seleção de linhagens de soja com base no ideótipo nutracêutico

## RESUMO

O objetivo deste trabalho foi selecionar linhagens de soja com ideótipo nutracêutico sob condições meteorológicas impostas nos ambientes de seleção. O experimento foi realziado na safra 2021/2022 na Escola Fazenda, Universidade Regional do Noroeste do Estado Rio Grande do Sul, localizada no município de Augusto Pestana - RS. O delineamento experimental utilizado foi o de blocos aumentados, com testemunhas intercalares sendo 27 populações geração F2, 118 linhagens geração F5, 70 linhagens geração F8 e 15 cultivares de soja, estes arranjados em three repetições. Avaliamos a composição química da soja pelo método de Espectroscopia no Infravermelho Próximo - NIRS. O genótipo IRC050 possui um ideótipo para incremento de ácido graxo esteárico (SEAR) e redução de fibra (FB) e material mineral (MM). Na geração F5 foram selecionados os genótipos L174F5, L307F5, L322F5, L205F5, L190F5, L196F5, L219F5, L183F5, L301F5, L172F5, L198F5, L182F5, L319F5, L21F5, L188F5, L217F5, L302F5, L315F5, L36F5, L318F5, L211F5, L165F5, L321F5 e L179F5. A seleção de genótipos com maiores teores de ácido linoleico e proteína pode ser realizada indiretamente por meio da enzima peroxidase.

**Palavras-chave**: *Glycine max,* MGIDI, Correlação.

## 1. Introduction

Soybean (*Glycine max* L.) is one of the main cultivated species in the world, holding significant agricultural, economic, and social importance. Brazil stands as the largest producer of this commodity, followed by the United States. According to data estimated by Conab (2022), the soybean crop achieved a production of 312.2 million tons in the 2021/22 harvest. The widespread cultivation and production of soybeans stem from their diverse applications across various industries.

This legume plays a crucial role in the Brazilian economy, being extensively used for animal protein production. Moreover, its use in human food consumption has been on the rise, offering a rich source of essential nutrients such as nitrogen, phosphorus, and potassium (Loro et al., 2021). The levels of proteins, oil, and other chemical compounds found in the grains are influenced by both genetic factors and the environment, with varying degrees of impact on different compounds.

Studies by Bakal et al. (2017) have revealed that the growth and development of soybean crops are significantly affected by the cultivation environment, which, in turn, affects the quality of the seeds in terms of chemical composition, germination, and vigor (Xavier et al., 2015). Additionally, temperature influences the composition of fatty acids, protein, and oil in the seeds (Marcos Filho, 2015). Germination and vigor processes are influenced by protein, lipid, starch, and sugar contents, with seeds exhibiting low vigor resulting in reduced emergence speed and subsequently affecting the crop's establishment and long-term performance (Ferraresi et al., 2014).

Soybeans possess nutraceutical properties, characterized as a source of proteins, carbohydrates, lipids, vitamins, saturated and unsaturated fatty acids, and phytochemicals. These compounds participate in various metabolic activities, including isoflavones, and are rich in essential minerals such as iron, potassium, magnesium, zinc, copper, phosphorus, manganese, and B vitamins (Silva et al., 2012). The plant is regarded as a functional food, offering several health benefits that aid in reducing the risk of chronic and degenerative diseases. The typical composition of soybeans on a dry basis comprises an average of 34% protein, 21% oil, 34% carbohydrate, and 4.9% ash (Souza et al., 2020). The variation in protein and oil content is primarily influenced by genetic factors (Embrapa, 2015).

Although soybean breeding programs focus on selecting genotypes with characteristics that result in higher grain yields, traits related to seed quality, performance, and grain biofortification has received less emphasis (Carvalho and Nakagawa, 2012). Thus, this study aimed to select soybean lines with desirable nutraceutical traits under specific meteorological conditions within the selection environments.

## 2. Material and Methods

The experiment was conducted during the 2021/2022 harvest at the Farm School of the Regional University of the Northwest of the State of Rio Grande do Sul. It is situated in the city of Augusto Pestana- RS, Brazil (28º23'16" S latitude and 53º54'53" W longitude). The experimental design used an augmented block approach, comprising 15 soybean cultivars (Table 1), 27 populations of F2 generation, 118 lines of F5 generation, and 70 lines of F8 generation as interim checks (Table 2). The treatments were organized into three replications. The chemical composition of harvested grains was evaluated using Near Infrared Spectroscopy - NIRS (Near-Infrared Spectroscopy) as the analytical method.

**Table 1.** Characteristics of the commercial cultivars evaluated in the study

| Cultivar | RMG[1] | Growth habit | Flower color |
|---|---|---|---|
| DM 57i52 | 5.7 | Indeterminate | Purple |
| DM 5958 RSF IPRO® | 5.8 | Indeterminate | Purple |
| TMG 7363 RR® | 6.3 | Indeterminate | White |
| BRS 525 | 5.6 | Indeterminate | Purple |
| BRS 537 | 6.0 | Indeterminate | Purple |
| NS 4823 RR® | 4.8 | Indeterminate | Purple |
| NA 5909 RG RR® | 6.7 | Indeterminate | Purple |
| M 5947 IPRO | 5.9 | Indeterminate | Purple |
| 95R51 IPRO | 5.1 | Indeterminate | Purple |
| BMX LANÇA IPRO | 5.8 | Indeterminate | White |
| BMX LOTUS IPRO | 6.1 | Indeterminate | White |
| BMX CROMO TF IPRO | 5.7 | Indeterminate | White |
| VTOP RR® | 5.9 | Indeterminate | White |
| AS 3590 IPRO | 5.9 | Indeterminate | White |
| BMX COMPACTA IPRO | 6.5 | Indeterminate | Purple |

[1]Relative maturity group

**Table 2.** Soybean lines derived from $F_2$, $F_5$, and $F_8$ generations used in the study

| | |
|---|---|
| F2 | IRC 047; IRC023;IRC038;IRC035;IRC037;IRC017;IRC025;IRC013;IRC030; IRCO36; IRC004;IRC011;IRC028;IRC039;IRC026;IRC019;IRC003;IRC001; IRC032;IRC041;IRC002;IRC034;IRC040;IRC012;IRC050; IRC033; IRC008 |
| F5 | L174F5;L307F5;L322F5;L205F5;L190F5;L196F5;L219F5;L183F5;L301F5; L172F5;L198F5;L182F5;L319F5;L21F5;L302F5;L315F5;L36F5;L318F5; L211F5;L165F5;L321F5;L179F5;L208F5;L207F5;L310F5;L33F5;L154F5; L185F5;L1F5;L178F5;L304F5;L156F5;L56F5;L308F5;L163F5;L35F5; L8F5;L51F5;L313F5;L184F5;L305F5;L57F5;L320F5;L306F5;L329F5; L61F5;L327F5;L7F5;L209F5;L200F5;L15F5;L27F5;L303F5;L69F5;L199F5; L34F5;L180F5;L311F5;L326F5;L328F5;L9F5;L212F5;L46F5;L72F5;L55F5; L11F5;L4F5;L59F5;L325F5;L68F5;L17F5;L19F5;L23F5;L317F5;L168F5; L213F5;L41F5;L160F5;L48F5;L16F5;L28F5;L44F5;L53F5;L3132F5;L203F5; L2F5;L70F5;L22F5;L63F5;L45F5;L37F5;L50F5;L75F5;L330F5;L43F5;L192F5; L24F5;L202F5;L6F5;L26F5;L316F5;L54F5;L47F5;L20F5;L30F5;L25F5;L67F5; L71F5;L186F5;L186F5;L10F5;L31F5;L49F5;L66F5;L42F5;L74F5;L65F5;L5F5; L38F5;L323F5;L12F5;L13F5 |
| F8 | L344F8;L142F8;L142F8;L145F8;L333F8;L128F8;L338F8;L343F8;L334F8; L136F8;L337F8;L345F8;L99F8;L331F8;L112F8;L80F8;L144F8;L143F8; L115F8;L114F8;L348F8;L119F8;L113F8;L85F8;L87F8;L141F8;L339F8; L147F8;L97F8;L336F8;L81F8;L102F8;L124F8;L138F8;L135F8;L76F8;L139F8; L340F8;L77F8;L117F8;L129F8;L341F8;L125F8;L91F8;L132F8;L148F8;L78F8; L137F8;L95F8;L96F8;L101F8;L121F8;L84F8;L130F8;L146F8;L332F8;L150F8; L153F8;L82F8;L104F8;L131F8;L90F8;L98F8;L126F8;L86F8;L106F8; L140F8;L88F8;L133F8;L120F8;L103F8;L107F8;L151F8 |

Sowing took place in the second half of December 2020, with each plot consisting of three rows spaced at 0.45 meters and extending five meters in length, resulting in a total usable area of 4.5 square meters. Management practices were implemented to minimize the impact of biotic factors. The lines were sourced from the municipality of Campos Borges. A descriptive analysis of grain chemical composition was conducted, stratified by commercial cultivars and lines F2, F5, and F8, to discern trends and variability in the evaluated genotypes. Measures of central tendency, such as mean and median, and measures of dispersion, including maximum and minimum values and the coefficient of variation, were utilized for this purpose.

Subsequently, the multi-trait genotype-ideotype distance index (MGIDI) was employed to select genotypes and lines based on multiple traits. An ideotype was formed to enhance (positive gains) the indices of stearic fatty acid (STEAR), linoleic acid (LLEI), linolenic acid (LLEN), 100-grain weight (100W), oleic fatty acid (OLE), oil (OIL), and protein (PTN), while reducing (negative sense) the contents of fiber (FB), palmitic fatty acid (PAL), and mineral material (MM).

To explore the relationships between the evaluated characters, a Pearson linear correlation analysis was conducted, with a significance level set at a 5% error probability, using the t-test. All analyses were performed using the R software. The “metan” package was employed for applying the MGIDI and correlation, while the ggplot2 package was utilized to create the graphs (R Core Team 2022).

## 3. Results and Discussion

All evaluated variables had low coefficients of variation (CV), ranging from 1.59% to 10.68%, except for linolenic acid, which showed a significantly higher coefficient of 63.81% (Table 3). These results indicate a prominent level of experimental precision and data homogeneity. However, the higher variability observed in linolenic acid suggests that its expression is strongly influenced by environmental factors. As highlighted by Marcos Filho (2015), the chemical composition of soybean seeds can vary based on factors such as the environmental conditions, genotype of the plant, position of the seed within the plant, stage of maturation, and cultural practices, among other relevant aspects.

In our study, the F2 population showed average percentages of stearic, linoleic, linolenic, oleic, and palmitic fatty acids of 4.13%, 57.44%, 2.14%, 24.99%, and 9.11%, respectively (Table 3). According to Marcos Filho (2015), the corresponding averages for these fatty acids were 3%, 54%, 8%, 22%, and 11%, respectively. Thus, overall, our findings for the F2 population closely align with the literature findings.

The F5 generation exhibited low CV values, ranging from 1.11% to 9.81%, except for linolenic acid, which had a high CV of 63.88% (Table 3). The average percentages were 4.16%, 5.87%, 59.69%, 2.42%, 15.72%, 5.14%, 22.21%, 21.98%, 10.36%, and 38.52% for stearic acid, fiber, linoleic acid, linolenic acid, 100-grain weight, mineral material, oleic acid, oil, palmitic acid, and protein, respectively. These averages were similar to those observed in the seeds of the F2 population.

**Table 3.** Chemical composition of seeds from populations F2, F5, F8, and commercial cultivars

| VAR[15] | F2 population | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | STE[5] | FB[6] | LLEI[7] | LLE[8] | HGW[9] | MM[9] | OLE[10] | OIL[11] | PAL[12] | PTN[13] |
| CV[1] | 3.30 | 3.90 | 4.31 | 63.81 | 9.22 | 1.59 | 9.02 | 3.66 | 10.68 | 4.19 |
| MAX[2] | 4.40 | 6.27 | 60.98 | 4.66 | 19.32 | 5.39 | 28.50 | 22.92 | 11.40 | 41.38 |
| MEAN | 4.13 | 5.92 | 57.44 | 2.14 | 14.32 | 5.16 | 24.99 | 21.49 | 9.11 | 38.17 |
| MED[4] | 4.13 | 5.97 | 57.81 | 2.21 | 14.22 | 5.16 | 25.53 | 21.43 | 9.11 | 37.86 |
| MIN[5] | 3.80 | 5.15 | 46.90 | 0.00 | 12.33 | 4.99 | 18.16 | 20.04 | 7.46 | 35.37 |
| VAR[15] | F5 population | | | | | | | | | |
| | STE[5] | FB[6] | LLEI[7] | LLE[8] | HGW[9] | MM[9] | OLE[10] | OIL[11] | PAL[12] | PTN[13] |
| CV[1] | 3.36 | 2.99 | 2.89 | 63.80 | 9.81 | 1.11 | 9.00 | 3.75 | 9.78 | 4.03 |
| MAX[2] | 4.51 | 6.32 | 63.98 | 6.41 | 19.86 | 5.43 | 28.06 | 24.09 | 12.99 | 41.85 |
| MEAN | 4.16 | 5.87 | 59.69 | 2.42 | 15.72 | 5.14 | 22.21 | 21.98 | 10.36 | 38.52 |
| MED[3] | 4.17 | 5.90 | 59.57 | 2.45 | 15.46 | 5.14 | 22.02 | 22.02 | 10.41 | 38.63 |
| MIN[4] | 3.78 | 5.28 | 54.77 | 0.00 | 12.58 | 4.99 | 17.75 | 19.72 | 7.42 | 34.27 |
| VAR[15] | F8 population | | | | | | | | | |
| | STE[5] | FB[6] | LLEI[7] | LLE[8] | HGW[9] | MM[9] | OLE[10] | OIL[11] | PAL[12] | PTN[13] |
| CV[1] | 3.73 | 2.95 | 2.72 | 46.70 | 9.26 | 1.03 | 8.25 | 4.10 | 10.38 | 4.30 |
| MAX[2] | 4.48 | 6.21 | 63.98 | 5.86 | 18.82 | 5.31 | 27.61 | 24.00 | 12.00 | 41.70 |
| MEAN | 4.14 | 5.88 | 59.58 | 2.55 | 15.44 | 5.15 | 22.68 | 21.95 | 9.81 | 38.45 |
| MED[3] | 4.15 | 5.91 | 59.71 | 2.63 | 15.36 | 5.16 | 22.54 | 21.87 | 10.00 | 38.76 |
| MIN[4] | 3.79 | 5.46 | 55.80 | 0.00 | 12.44 | 5.03 | 17.11 | 20.31 | 7.13 | 33.91 |
| VAR[15] | Commercial cultivars | | | | | | | | | |
| | STE[5] | FB[6] | LLEI[7] | LLE[8] | HGW[9] | MM[10] | OLE[11] | OIL[12] | PAL[13] | PTN[14] |
| CV[1] | 2.48 | 3.64 | 2.77 | 75.69 | 7.25 | 1.37 | 9.58 | 4.27 | 7.78 | 4.37 |
| MAX[2] | 4.36 | 6.16 | 64.08 | 3.87 | 19.42 | 5.29 | 26.38 | 24.09 | 11.51 | 41.12 |
| MEAN | 4.12 | 5.81 | 60.42 | 1.50 | 17.20 | 5.12 | 22.45 | 22.34 | 10.23 | 38.10 |
| MED[3] | 4.12 | 5.85 | 60.46 | 1.16 | 17.09 | 5.13 | 22.00 | 22.30 | 10.13 | 38.15 |
| MIN[4] | 3.97 | 5.48 | 56.68 | 0.00 | 14.97 | 5.00 | 18.50 | 20.54 | 8.71 | 34.44 |

[1]Coefficient of variation (%); [2]maximum, [3]median, and [4]minimum (%); [5]stearic fatty acid; [6]fiber; [7]linoleic fatty acid; [8]linolenic fatty acid; [9]hundred-grain weight; [10]mineral material; [11]oleic fatty acid; [12]oil; [13]palmitic fatty acid; [14]protein; and [15]variables.

For the F8 generation lines, CV values were low, ranging from 1.03% to 10.38%, except for linolenic acid, which exhibited a higher coefficient of 46.70% (Table 3). The average composition of the variables in the F8 generation seeds were as follows: stearic acid 4.14%, fiber 5.88%, linoleic acid 59.58%, linolenic acid 2.55%, 100-grain weight 15.44%, mineral material 5.15%, oleic acid 22.68%, oil 21.95%, palmitic acid 9.81%, and protein 38.45%.

Table 4 details the chemical composition of commercial cultivars, including coefficient of variation (CV) for several variables: stearic acid, fiber, linoleic acid, linolenic acid, 100-grain weight, mineral material, oleic acid, oil, palmitic acid, and protein. The observed CV values were 2.48%, 3.63%, 2.76%, 75.69%, 7.25%, 1.36%, 9.58%, 4.27%, 7.77%, and 4.37%, respectively (Table 3). Notably, linolenic acid exhibited the highest CV at 75.69%. The average compositions for these variables were: 4.11% (stearic acid), 5.81% (fiber), 60.42% (linoleic acid), 1.49% (linolenic acid), 17.19% (100-grain weight), 5.11% (mineral material), 22.45% (oleic acid), 22.33% (oil), 10.22% (palmitic acid), and 38.10% (protein).

The multi-trait selection analysis of the F2 population, when compared to commercial cultivars, highlighted the primary characteristics for selecting a progeny with an optimal nutraceutical ideotype. This approach aims to combine these characteristics for the best outcome, considering variations in the variables and analyzing potential increases and decreases (Table 4). Four factorial groups or latent variables were discerned. Factor 1 comprised 100-grain weight and oleic acid. The goal was to increase levels in the grains (positive selection) and reduce palmitic acid levels (negative gains). Factor 2 included mineral material, fiber, and stearic acid (Table 4).

The aim was to reduce the content of fiber and mineral material, as they are not significant for soybean's nutritional quality, while stearic acid was targeted for positive increases. Factor 3 contained protein and oil variables, both aimed for augmentation. Factor 4 showcased linoleic and linolenic fatty acids. The intent was to elevate both, given their status as essential unsaturated fatty acids in soybean oil composition.

The selection differential in percentage (SDperc) denotes the disparity between current averages and target values. For the 100-grain weight and palmitic fatty acid, values of 8.82% and 7.01% respectively indicated challenges in achieving the desired levels. In contrast, oleic fatty acid (for increase) and fiber (for reduction) exhibited SDperc values of -3.05% and -2.79%, suggesting easier attainment of target values. Mineral material (targeted for reduction) and stearic fatty acid (for increase) presented SDperc values of 0.147% and 0.479%, indicating a favorable trajectory towards desired levels. Protein had an SDperc of -0.548% and linolenic fatty acid -22.6%, both suggesting that current levels exceed the target. Finally, oil and linoleic fatty acid showcased SDperc values of 3.23% and 1.73%, implying the feasibility of reaching optimal levels.

Follmann et al. (2019) reported a broad-sense heritability ($H^2$) of 0.93 for the 100-grain weight, while Barbosa et al. (2021) found a narrow-sense heritability ($h^2$) of 0.006, indicating a strong genetic influence. For the variables of crude fiber, mineral material, and palmitic acid, Bellinasso et al. (2021) observed mean genotype heritability ($h^2_g$) values of 0.26, 0.96, and 0.44, respectively. As for stearic fatty acid, oleic fatty acid, linoleic fatty acid, and linolenic fatty acid, this author noted broad-sense heritability ($h^2_g$) values for total genotypic effects, with values of 0.82, 0.75, 0.78, and 0.53, respectively. This further underscores the strong genetic influence on these variables. The multi-trait selection analysis, as depicted in Figure 1, showcases the alignment of commercial cultivars and genotypes in relation to the soybean nutraceutical ideotype compared to the F2 population.

**Table 4.** Multi-trait genotype-ideotype distance (MGIDI) analysis for multi-trait selection of $F_2$ × Cultivars

| VAR[1] | FAC[2] | $X_o$[3] | $X_s$[4] | SDperc[5] | Goal |
|---|---|---|---|---|---|
| HGW[6] | FA1 | 15.2 | 16.6 | 8.82 | Increase |
| PAL[7] | FA1 | 9.54 | 10.2 | 7.01 | Decrease |
| OLE[8] | FA1 | 24 | 23.2 | -3.05 | Increase |
| FB[9] | FA2 | 5.88 | 5.72 | -2.79 | Decrease |
| MM[10] | FA2 | 5.15 | 5.16 | 0.147 | Decrease |
| STEAR[11] | FA2 | 4.13 | 4.15 | 0.479 | Increase |
| PTN[12] | FA3 | 38.2 | 37.9 | -0.548 | Increase |
| OIL[13] | FA3 | 21.8 | 22.5 | 3.23 | Increase |
| LLEI[14] | FA4 | 58.6 | 59.7 | 1.73 | Increase |
| LLEN[15] | FA4 | 1.9 | 1.47 | -22.6 | Increase |

[1]Variables; [2]Factor; [3]observed mean; [4]selected lines mean; [5]selection differential in percentage; [6]hundred-grain weight (in g); [7]palmitic fatty acid (in %); [8]oleic fatty acid (in %); [9]fiber; [10]mineral material (in %); [11]stearic fatty acid (in %); [12]protein (in %); [13]oil (in %); [14]linoleic fatty acid (in %); [15]linolenic fatty acid (in %).

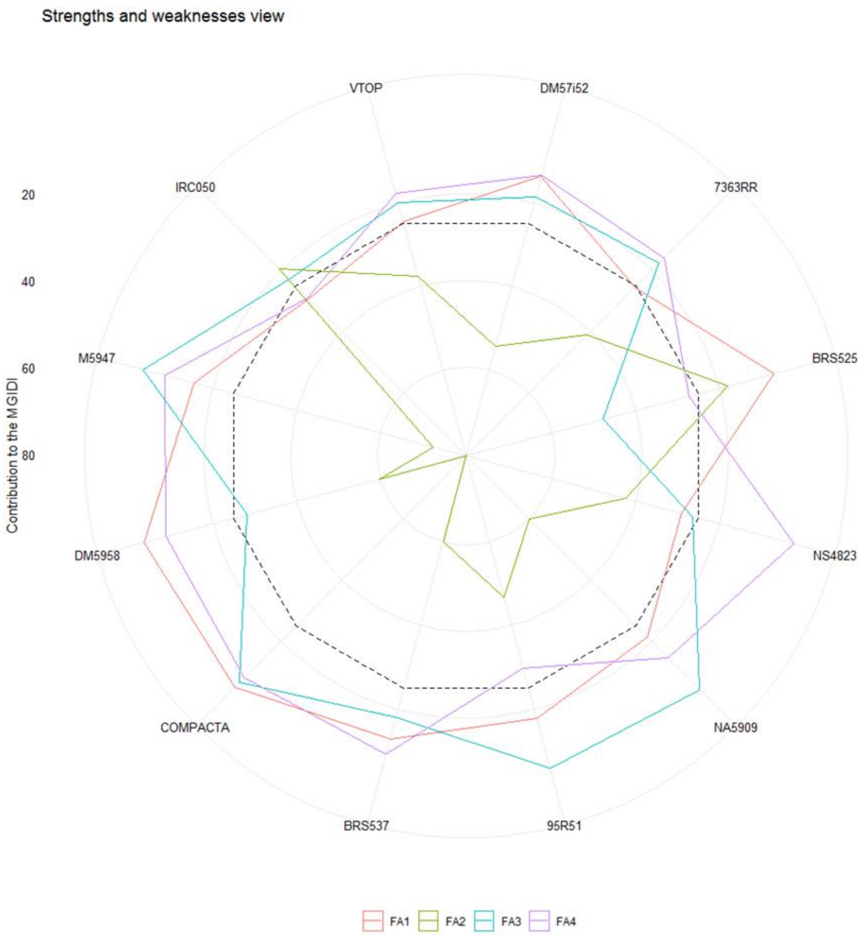


**Figure 1.** Proportions of each factor in the multi-trait genotype-ideotype distance index (MGIDI). Smaller proportions (nearer to the outer edge) indicate that characteristics of that factor closely aligning with the ideotype. The dashed line represents equal contribution from all factors.

Commercial cultivars, including DM57i52, 7363RR, BRS525, NS4823, NA4823, NA5909, 95R51, BRS537, COMPACTA, DM5958, M5947, and VTOP, exhibited strong similarities to the nutraceutical ideotype. Additionally, the genotype IRC 050 notably aligned with the ideotype, distinguishing itself from other studied genotypes. As such, these genotypes are prime candidates for germplasm banks to enhance nutraceutical compounds in grains.

Figure 1 further pinpoints the cultivars or commercial genotypes that align with each variable from the multi-trait selection analysis: Factor 1, emphasizing 100-grain weight and oleic fatty acid for enhancement and palmitic acid for reduction, pinpointed DM5958, COMPACTA, and BRS525 as cultivars closely aligned with the ideotype for these three variables; Factor 2 highlighted the variables of fiber and mineral material for reduction and stearic fatty acid for an increase. Here, genotype IRC050 and the commercial cultivar BRS525 matched the ideotype for these traits; Factor 3, which encompasses protein and oil for increment, found 7363RR, NA5909, 95R51, COMPACTA, M5947, and VTOP as cultivars in line with the desired ideotype; and Factor 4, focusing on elevating linoleic and linolenic acid, showed DM57i52, 7363RR, NS4823, BRS537, M5947, and VTOP as standout cultivars.

In the multi-trait selection analysis of F5 lines, five distinct clusters (factors) emerged that depict the optimal nutraceutical ideotype, defined by traits designated for either increase or reduction. Table 5 presents the differential selection percentages for each variable in the F5 generation lines, emphasizing the disparity between existing and desired values for the nutraceutical ideotype. In examining the variables, the data for palmitic fatty acid suggests a need for reduction, with a selection differential percentage (SDperc) of -1.21%, indicating that the desired value surpasses the current one. Oleic and stearic fatty acids, as well as the oil variable, registered SDperc values of 5.6%, 1.17%, and 2.14%, respectively, reflecting the need to increase these variables. The factor analysis for the F5 lines identifies genotypes aligning closely with the nutraceutical ideotype. For Factor 1, which aims to diminish palmitic acid and elevate oleic acid, the lines L182F5, LI83F5, L205F5, L321F5, L318F5, and L302F5 demonstrated a predisposition, exhibiting elevated values. In addition, the cultivars COMPACTA, DM57I52, and BRS 525 aligned closely with this factor, as illustrated in Figure 2.

In Group 2, with the goal of decreasing mineral material and increasing linoleic and linolenic acid, the lines L30F5, L20F5, L322F5, L179F5, L36F5, and L188F5 exhibited superior performance, complemented by the cultivars BRS537 and COMPACTA. For Group 3, which focused on increasing protein and oil, standout lines included L198F5, LI83F5, L307F5, L208F5, L318F5, and L21F5. Among the cultivars, M5947 and DM57i52 showed superior attributes. In Group 4, dedicated to enhancing stearic fatty acid, lines such as L217F5, L219F5, L190F5, L179F5, L211F5, and L36F5 were preeminent, with TMG 7363RR and BRS525 emerging as the superior cultivars. However, Group 5, which targeted an increase in the 100-grain weight and a reduction in fiber, was below the average, with no cultivar nor line showing superiority for these characteristics.

The multi-trait selection analysis highlighted the proximity of F5 generation lines and selected cultivars to the agronomic ideotype concerning nutraceutical properties and grain composition. Noteworthy lines included L174F5, L307F5, L322F5, L205F5, L190F5, L196F5, L219F5, L183F5, L301F5, L172F5, L198F5, L182F5, L319F5, L21F5, L188F5, L217F5, L302F5, L315F5, L36F5, L318F5, L211F5, L165F5, L321F5, and L179F5. As for the cultivars, BRS 525, TMG 7363RR, M5947, DM57i52, BRS537, DM5958, BRS5804RR, and BMX COMPACTA IPRO stood out (Figure 2).

**Table 5.** Multi-trait genotype-ideotype distance (MGIDI) analysis for multi-trait selection of F5 × cultivars.

| VAR[1] | FAC[2] | $X_o$[3] | $X_s$[3] | SDperc[5] | Goal |
|---|---|---|---|---|---|
| PAL[6] | FA1 | 10.3 | 10.2 | -1.21 | Decrease |
| OLE[7] | FA1 | 22.2 | 23.5 | 5.6 | Increase |
| MM[8] | FA2 | 5.14 | 5.15 | 0.0918 | Decrease |
| LLEI[9] | FA2 | 59.8 | 59.6 | -0.348 | Increase |
| LLEN[10] | FA2 | 2.3 | 1.68 | -26.7 | Increase |
| PTN[11] | FA3 | 38.5 | 38.6 | 0.438 | Increase |
| OIL[12] | FA3 | 22 | 22.5 | 2.14 | Increase |
| STEAR[13] | FA4 | 4.16 | 4.21 | 1.17 | Increase |
| HGW[14] | FA5 | 15.9 | 17.4 | 9.45 | Increase |
| FB[15] | FA5 | 5.87 | 5.71 | -2.75 | Decrease |

[1]Variables; [2]Factor; [3]observed mean; [4]selected lines mean; [5]selection differential in percentage; [6]palmitic fatty acid; [7]oleic fatty acid; [8]mineral material; [9]linoleic fatty acid; [10]linolenic fatty acid; [11]protein; [12]oil; [13]stearic fatty acid; [14]hundred-grain weight; [15]fiber.

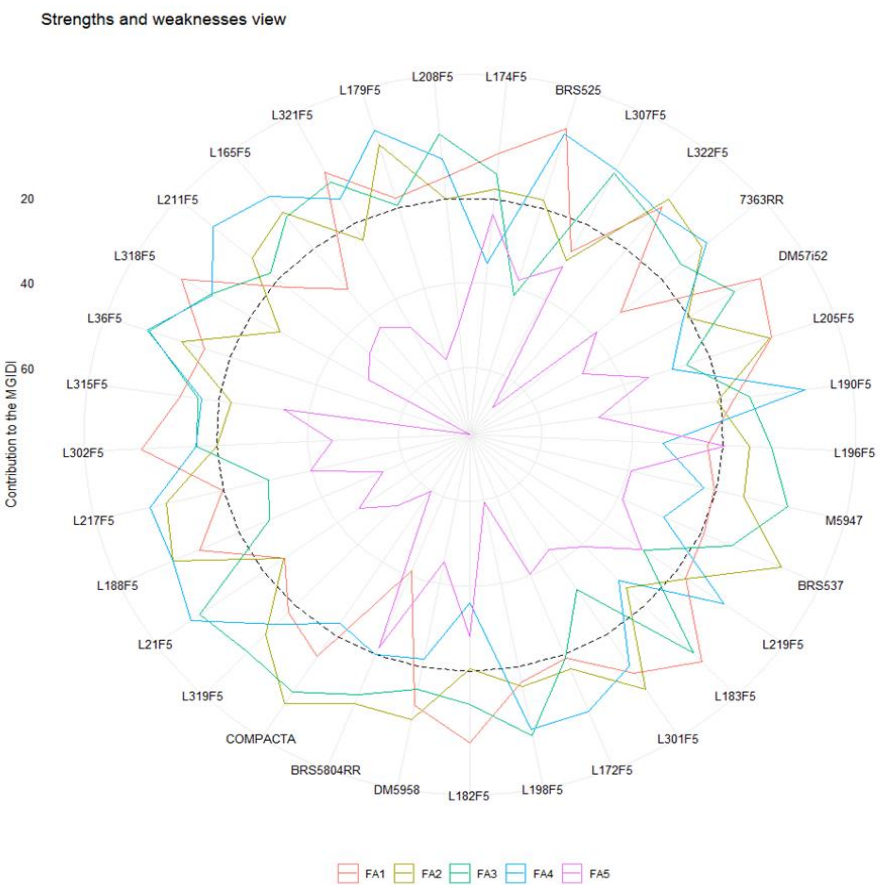


**Figure 2.** Proportions of each factor in the calculated multi-trait genotype-ideotype distance indexes (MGIDI).

Examining the selection differential in percentage (SDperc), mineral material showed a value of 0.146% signaling a need for reduction. For stearic acid, linolenic acid, linoleic acid, and protein, values stood at -0.426%, -0.28%, -21.9%, and -0.0219% respectively. These negative values emphasize that the current concentrations surpass the desired levels (Table 6). Oil, oleic fatty acid, and 100-grain weight exhibited SDperc values of 1.34%, 6.7%, and 7.53%, respectively, and must increase to match the ideotype. Conversely, palmitic acid and fiber showed SDperc values of respectively -6.76% and -2.76, thus requiring a reduction.

The factor analysis conducted on F8 cultivars, in comparison to elite commercial cultivars, identified four distinct groups (factors). Factor 1 included mineral material targeted for reduction, as well as linoleic acid, which was intended for an increase (refer to Figure 3). Factor 2 encompassed protein, oil, and stearic acid, all of which needed to rise to achieve the optimal nutraceutical potential. Factor 3 was characterized by palmitic acid, which was aimed to be reduced, and both oleic and linolenic acids, marked for an increase. Factor 4 centered around the 100-grain weight, which had to be increased, together with fiber, designated for decrement.

In the factor analysis, standout progenies included L344F8, L142F8, L111F8, L145F8, L333F8, L128F8, L338F8, L343F8, L122F8, L94F8, L79F8, L108F8, L134F8, L346F8 and L93F8. The commercial cultivars pinpointed were DM57I52, BRS 525, TMG 7363 RR, DM5958, COMPACTA, and BRS 537. Upon dissecting the factor analysis and the multi-trait selection assessment, specific genotypes and elite cultivars emerged triumphant across the four aforementioned groups. For instance, Group 1, distinguished by variables like mineral material and linolenic fatty acid, showcased lines such as L93F8, L344F8, L333F8, L128F8, L338F8, L343F8, L122F8, L94F8 and L134F8 (Figure 3). In the factor analysis, the elite cultivars notable for both variables were DM57i52 and COMPACTA.

The second group, which highlighted protein, oil, and stearic fatty acid—all of which need to be increased in line with the nutraceutical ideotype—selected lines L346F8, L344F8, L111F8, and L79F8. The elite cultivars chosen based on these three variables included DM57i52, DM5958, and COMPACTA. Group 3 emphasized the reduction of palmitic fatty acid and the increase of oleic and linolenic fatty acids. This group featured lines L93F8, L145F8, L128F8, L334F8, and L94F8, with BRS525 being the distinct elite cultivar of choice. The fourth and final group, which focused on the increase of the 100-grain weight and the reduction of fiber, singled out BRS 525 as the only commercial cultivar distinguishing itself from other genotypes.

**Table 6.** Multi-trait genotype-ideotype distance (MGIDI) analysis for multi-trait selection of $F_8$ × Cultivars.

| VAR[1] | FAC[2] | $X_o$[3] | $X_s$[4] | SDperc[5] | Goal |
|---|---|---|---|---|---|
| MM[6] | FA1 | 5.15 | 5.16 | 0.146 | Decrease |
| LLEI[7] | FA1 | 59.8 | 59.5 | -0.426 | Increase |
| PTN[8] | FA2 | 38.4 | 38.4 | -0.0219 | Increase |
| OIL[9] | FA2 | 22 | 22.3 | 1.34 | Increase |
| STEAR[10] | FA2 | 4.14 | 4.13 | -0.28 | Increase |
| PAL[11] | FA3 | 9.9 | 9.23 | -6.76 | Decrease |
| OLE[12] | FA3 | 22.6 | 24.2 | 6.7 | Increase |
| LLEN[13] | FA3 | 2.35 | 1.84 | -21.9 | Increase |
| HGW[14] | FA4 | 15.7 | 16.9 | 7.53 | Increase |
| FB[15] | FA4 | 5.87 | 5.71 | -2.76 | Decrease |

[1]Variation; [2]Factor; [3]observed mean; [4]selected lines mean; [5]selection differential in percentage; [6]mineral material; [7]linoleic fatty acid; [8]protein; [9]oil; [10]stearic fatty acid; [11]palmitic fatty acid; [12]oleic fatty acid; [13]linolenic fatty acid; [14]hundred-grain weight; [15]fiber.

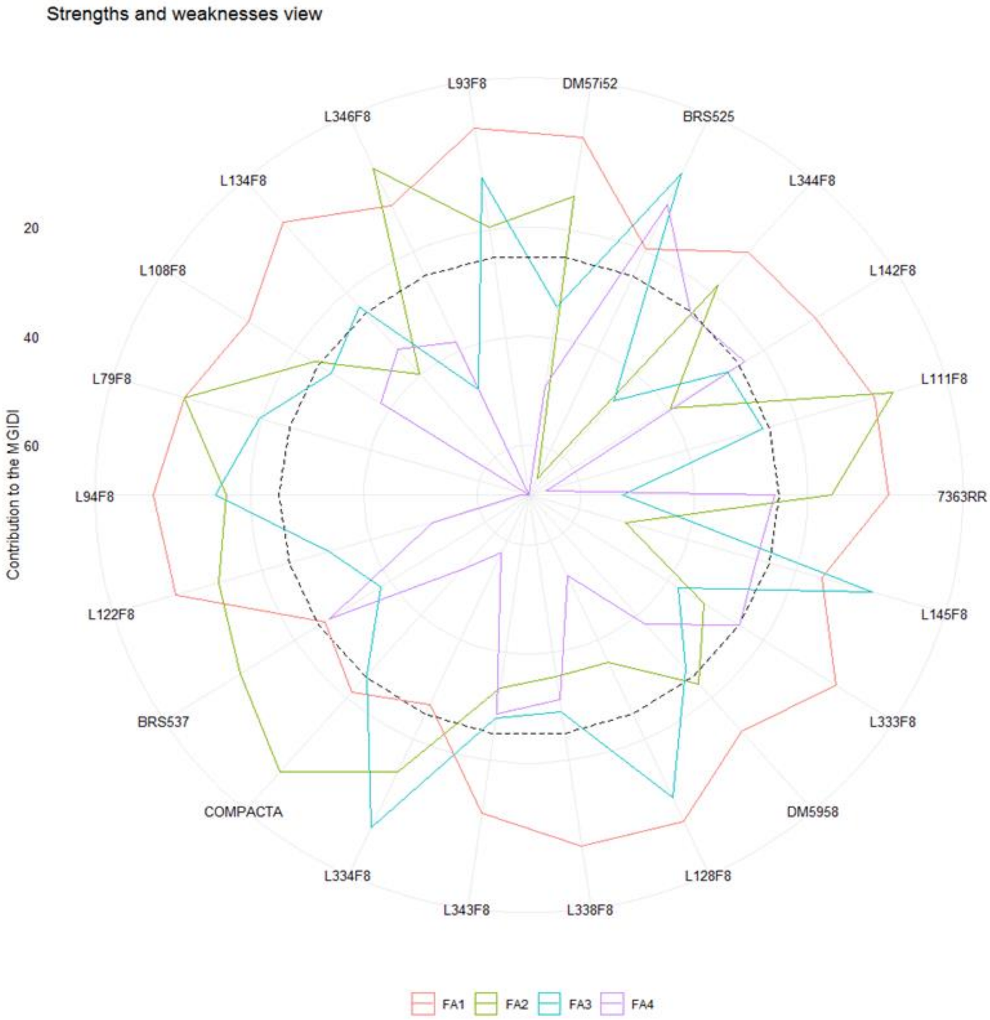


**Figure 3.** Proportions of each factor in the multi-trait genotype-ideotype distance index (MGIDI). Smaller proportions (nearer to the outer edge) indicate that characteristics of that factor closely aligning with the ideotype. The dashed circle represents the theoretical value when all factors contribute equally.

The comprehensive multi-trait analysis for F8 lines, when juxtaposed with elite commercial cultivars, clarified the distance from each genotype to the nutraceutical ideotype. Lines such as L344F8, L142F8, L11F8, and L145F8, L33F8, L128F8, L343F8, L334F8, L122F8, L94F8, L79F8, L108F8, L134F8 were among the prominent selections. The most distinguished cultivars identified were DM57i52, BRS525, TMG 7363RR, DM5958, and BMX COMPACTA IPRO. Pearson's linear correlation served to correlate the variables under test. The correlation values can range from -1 (perfect negative correlation) to +1 (perfect positive correlation), with zero signifying no correlation (Loro et al., 2021). Out of the 23 variables introduced, there were 276 associations, 99 of which were deemed significant, with their values oscillating between 0 and 0.70 (Figure 4).

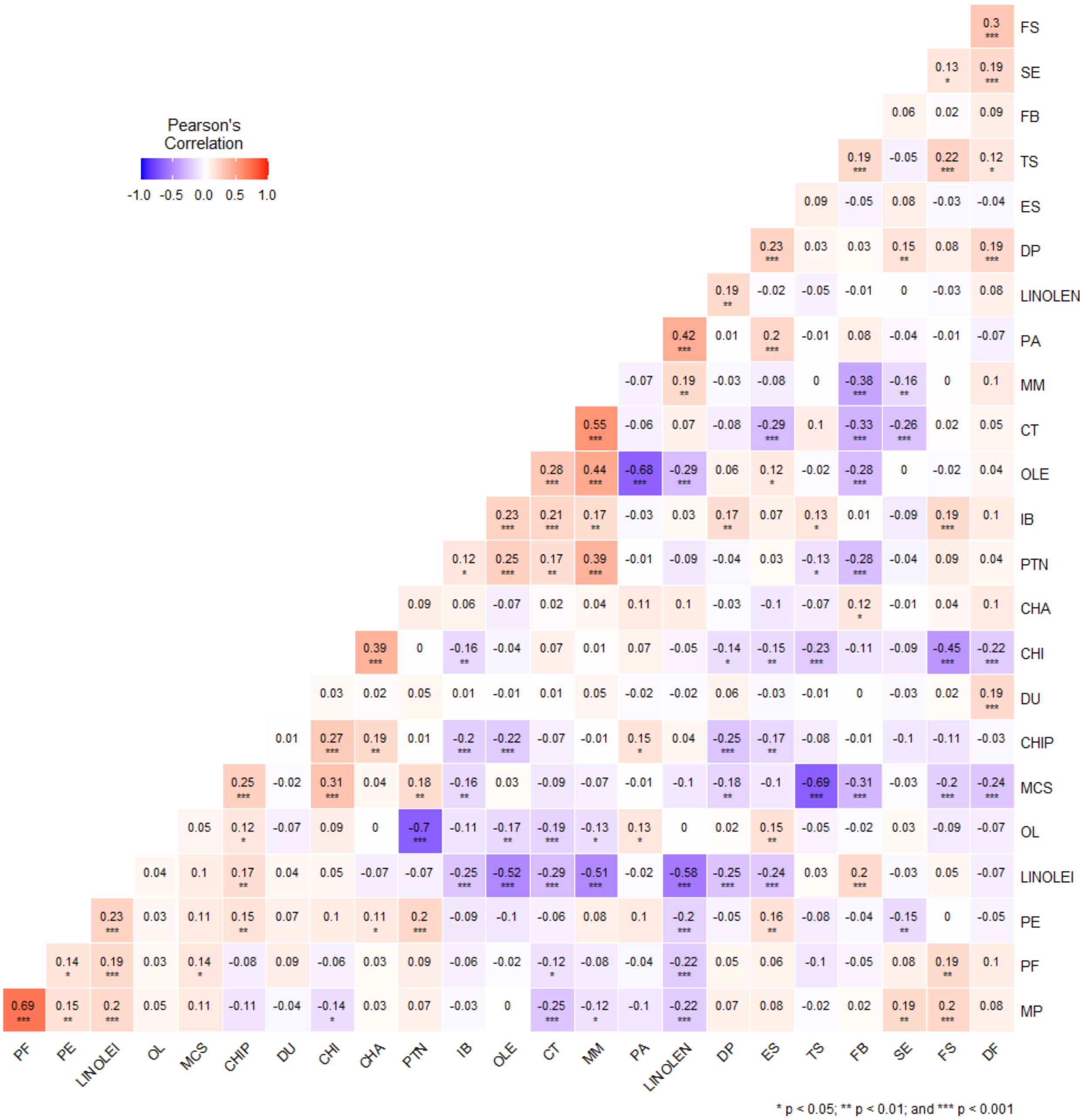


**Figure** 4. Pearson's linear correlation values for seed chemical composition, diseases, and physiological quality. Pearson's linear correlation coefficient significant at 5% probability by the t-test. Presence of fungus (PF), Peroxidase (PE), linoleic fatty acid (LLEI), oil (OIL), hundred seed weight (HSW), hypocotyl color (CHIP), humidity damage (HD), hilum color (CHI), halo color (CHA), protein (PTN), gloss index (GI), oleic fatty acid (OLE), tegument color (TC), mineral material (MM), palmitic fatty acid (PA), linolenic fatty acid (LLEN), bedbug damage (BD), stearic fatty acid (ST), seed size (SSZ) fiber (FB), greenish seed (GS), seed shape (SSH), and physical damage (PD).

The correlation analysis highlighted a strong positive relationship (r= 0.69) between the occurrence of fungus and the appearance of purple-bordered spots. As the presence of the fungus escalates, there was a rising incidence of these spots. Additionally, there was a modest correlation (r= 0.15) between peroxidase levels and the purple-bordered spots. As pointed out by Beninca et al. (2008), peroxidase plays multiple roles in physiological processes, importantly bolstering plant resistance against pathogens and establishing biochemical defenses against fungal growth.

On the other hand, there was a notable inverse relationship between oil and protein content (r= -0.7). Another significant inverse correlation (r= -0.69) was observed between the 100-grain weight and grain size, as well as between oleic acid and linolenic acid. Focusing on the peroxidase enzyme, the positive correlations identified were with linoleic fatty acid (r= 0.23), hypocotyl color (r= 0.15), protein content (r= 0.2), and stearic fatty acid (r= 0.16). Conversely, peroxidase exhibited a moderate negative correlation with linolenic fatty acid (r= -0.58) and a weaker negative association with the presence of greenish seeds (r= -0.15).

## 4. Conclusions

In the analysis of seed chemical composition across the F2, F5, and F8 generations, linolenic acid displayed a high coefficient of variation in all three generations. Using the Multi-trait Genotype-Ideotype Distance Index (MGIDI) and examining the proportion of each factor, the IRC050 genotype showed to be an ideotype ideal for increasing stearic fatty acid (STEAR) and reducing fiber (FB) and mineral material (MM). This makes it a favorable selection for nutraceutical characteristics.

For the F5 generation, the selected genotypes were L174F5, L307F5, L322F5, L205F5, L190F5, L196F5, L219F5, L183F5, L301F5, L172F5, L198F5, L182F5, L319F5, L21F5, L188F5, L217F5, L302F5, L315F5, L36F5, L318F5, L211F5, L165F5, L321F5, and L179F5. Lastly, to target genotypes with elevated levels of linoleic acid and protein, an indirect approach can be employed using the enzyme peroxidase.

## Authors' Contribution

Jaqueline Piesanti Sangiovo responsible for writing and conducting the essays, Ivan Ricardo Carvalho carried out the writing, advice and data analysis, Leonardo Cesar Pradebon and Murilo Vieira Loro responsible for carrying out the evaluations, Eduarda Donadel Port, Victor Delino Barasuol Scarton conducted the corrections and revisions, Guilherme Hickembick Zuse, Eduardo Ely Foleto, Inâe Carolina Sfalcin evaluations and data measurements.

## Bibliographic References


Bakal, H., Gulluoglu, L., Onat, B.Z., Arioglu, H. 2017. The effect of growing seasons on some agronomic and quality characteristics of soybean varieties in Mediterranean region in Turkey. Turkish Journal. Field Crop. 22(2), 187-196. https://doi.org/10.17557/tjfc.356213

Barbosa, M.H., Carvalho, I.R, Silva, J.A.G., Magano, D.A., Souza, V.Q., Szareski, V.J., Lautenchleger, F., Hutra, D.J., Moura, N.B., Loro, M.V. 2021. Contribution of the additive genetic effects in soybean breeding aiming at the agronomic ideotype. Functional Plant Breeding Journal. 3(1), 1-9. http://dx.doi.org/10.35418/2526-4117/v3n1a1

Bellinasso, A.V., Carvalho, I.R., Silva, J.A.G., Moura, N.B., Hutra, D.J., Loro, M.V., Bubans, V.E., Lautenchleger, F. 2021. Cover crops and their relationship with the qualitative and quantitative attributes of soybeans. Revista Brasileira de Agricultura, 96(1), 294-313. https://doi.org/10.37856/bja.v96i1.4260

Beninca, C.P., Franzener, G., Assi, L., Iurkiv, L.E.B., Costa, V.C., Nogueira, M.A., Stangarlin, J.R., Estrada, K.R.F.S.2008. Indução de fitoalexinas e atividade de peroxidases em sorgo e soja tratados com extratos de basidiocarpos de *Pycnoporus sanguineus*. Revista Arquivos do Instituto Biológico. 75(3), 285-292. https://doi.org/10.1590/1808-1657v75p2852008

Carvalho, N.M., Nakagawa, J. 2012. Sementes: ciência, tecnologia e produção. 5. ed. FUNEP, Jaboticabal.

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. 2017. A produtividade da soja: análise e perspectivas. Safra 2017. v. 10. CONAB, Brasília. http://www.conab.gov.br. (accessed April 15, 2022).

EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. 2015. Teores de óleo e proteína em soja: fatores envolvidos e qualidade para a indústria. Comunicado Técnico, Embrapa, Londrina.

Ferraresi, L.M.D., Villela, F.A., Aumonde, T.Z. 2014. Desempenho fisiológico e composição química de sementes de soja. Revista Brasileira de Ciências Agrárias. 9(1), 14-18. https://doi.org/10.5039/agraria.v9i1a2864

Follmann, D.N., Souza, V.Q., Cargnelutti Filho, A., Demari, G.H., Nardino, M., Olivoto, T., Carvalho, I.R., Silva, A.D.B., Meira, D., Meier, C. 2019. Agronomic performance and genetic dissimilarity of second-harvest soybean cultivars using REML/BLUP and Gower's algorithm. Bragantia. 78(2), 197-207. https://doi.org/10.1590/1678-4499.20180194

Loro, M.V., Carvalho, I.R., Silva, J.A.G., Moura, N.B., Hutra, D.J., Lautenchleger, F. 2021. Artificial intelligence and multiple models applied to phytosanitary and nutritional aspects that interfer in the physiological potential of soybean seeds. Revista Brasileira de Agricultura. 96(1), 324-338. https://doi.org/10.37856/bja.v96i1.4258

Marcos Filho, J. 2015. Fisiologia de sementes de plantas cultivadas. Editora ABRATES - Associação Brasileira de Tecnologia de Sementes, Londrina.

R Core Team. 2022 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Silva, C.E., Carrão-Panizzi, M.C., Mandarino, J.M.G., Leite, R.S., Mônaco, A.P.A. 2012. Teores de isoflavonas em grãos inteiros e nos componentes dos grãos de diferentes cultivares de soja (*Glycine max* L.) Merrill). Brazilian Journal of Food Technology, 15(2), 150-156. https://doi.org/10.1590/S1981-67232012005000008

Souza, S.M., Morais, R.A., Gualberto, L.S., Sousa, H.M.S., Martins, G.A.D., Peluzio, J.M. 2020. Teor de óleo em cultivares de soja visando a produção de biocombustível no estado do Tocantins. Desafios - Revista Interdisciplinar da Universidade Federal do Tocantins. 7(1), 82-86. http://dx.doi.org/10.1590/1517-869220162203142486

Xavier, T.S., Daronch, D., Peluzio, J.M., Afféri, F.S., Carvalho, E.V., Santos, W.F. 2015. Época de colheita na qualidade de sementes de genótipos de soja. Comunicata Scientiae, 6(2), 241–245.