SANGIOVO, P. S.; CARVALHO, I. R.; ZUSE, G. H.; SILVA, T.; PRADEBON, L. C.; LORO, M. V. Definition of soybean gene micropools based on genes with non-complex genetic actions. **Revista de Agricultura Neotropical**, Cassilândia-MS, v. 13, n. 1, e9295, 2026. ISSN 2358-6303. DOI: https://doi.org/10.32404/rean.v13i1.9295.

Definition of soybean gene micropools based on genes with non-complex genetic actions

Jaqueline Piesanti Sangiovo[1], Ivan Ricardo Carvalho[2], Guilherme Hickembick Zuse[2], Thayane Silva[2], Leonardo Cesar Pradebon[1], Murilo Vieira Loro[1]

[1] Universidade Federal de Santa Maria, Santa Maria - RS, Brasil. E-mail: jaqueline.sangiovo@sou.unijui.edu.br, leonardopradebon@gmail.com, loro@gmail.com
[2] Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí - RS, Brasil. E-mail: carvalho.irc@gmail.com, zuseeee@gmail.com, silva@gmail.com

Received: 21/01/2025; Accepted: 04/03/2026.

ABSTRACT

The aim of this study was to define soybean gene micropools using genes with non-complex actions and to guide future selection strategies. The study was conducted in the 2022/2023 growing season in Ijuí, Rio Grande do Sul, Brazil, at the experimental area of the Regional University of the Northwest of the State of Rio Grande do Sul (UNIJUÍ). The experimental design used was an augmented block design with interspersed controls, with four blocks distributed across the experiment area of 1500 m^2. 25 soybean gene micropools were formed based on agronomic traits and morphological and biochemical genetic descriptors. High genetic variability was associated with genes conferring tolerance to downy mildew (*Peronospora manshurica*) and frogeye leaf spot (*Cercospora sojina*). The gene micropool (C3) was identified to obtain early lines and cultivars, lanceolate leaflets and tolerance to target spot (*Corynespora cassicola*). 42 lines were selected for advancement to the next generation due to the absence of the major foliar diseases, which are intended for environments with reduced agrochemical use and increased sustainability.

Keywords: *Glycine max***,** Lines, Descriptors.

Definição de micropools de gene de soja através de genes com ações não complexas

RESUMO

O objetivo deste estudo foi definir micropools gênicos de soja utilizando genes com ações não complexas e orientar futuras estratégias de seleção. O estudo foi realizado na safra 2022/2023 no município de Ijuí localizado no estado do Rio Grande do Sul, na área da Universidade Regional do Noroeste do Estado do Rio Grande do Sul – UNIJUÍ. O delineamento experimental utilizado foi o de blocos aumentados com controles intercalados, com quatro blocos distribuídos ao longo do experimento em uma área de 1500 m^2. Foram semeadas 1449 linhagens de soja (L) da geração segregante F3 das seguintes populações. Foram formados 25 micropools gênicos de soja com base em características morfológicas, agronômicas, descritores genéticos morfológicos e bioquímicos. Alta variabilidade genética é atribuída aos genes de tolerância ao míldio (*Peronospora manshurica*) e à mancha-olho-de-rã (*Cercospora sojina*). O micropool genético (C3) é identificado para obter linhagens e cultivares precoces, folíolos lanceolados e tolerância à mancha-alvo (*Corynespora cassicola*). Foram selecionadas 42 linhagens para avanço geracional devido à ausência das principais doenças foliares, as quais são destinadas a ambientes com menor uso de agroquímicos e maximização da sustentabilidade.

Palavras-chave: *Glycine max*, Linhagens, Descritores.

1. Introduction

Soybean is one of the most important agricultural commodities worldwide due to its versatility in oil and protein production. Its production in the 2022/2023 growing season was evaluated at 370.24 million tons of grains, higher than in 2021/2022 (USDA, 2023). Soybean grain yield has increased significantly in recent years, mainly due to genetic improvement and the creation of cultivars (Hamawaki et al., 2019). For the genotype selection, the characterization of morphoagronomic components, quantitative and qualitative variables that are easy to identify, is used. These, when qualitative, are called morphological genetic descriptors or markers. These descriptions are used for characterization and classification of genotypes by the National Cultivar Protection Service (SNPC) of the Brazilian Ministry of Agriculture (SNPC 2008; Machado et al., 2017).

The phenotypic manifestation in soybean is dependent on the genetic constitution of the genotype, environmental effects, gene interactions and gene actions, however, the degree of expression is determined by the number of genes, gene or allelic actions involved, heritability and biotic and abiotic pressure, influenced by the cultivation environment (Viana et al., 2012). In this way, available germplasm gene micropools, the degree of variability and the alleles available in the germplasm become crucial to the success of the breeding program (Singh et al., 2021).

The combination of several interconnected genes is called a gene micropool; however, simply combining high genetic variability is not effective, making it necessary to characterize, discriminate gene micropools and understand the variation components intrinsically involved in each set of lines. Thus, effective selection within a soybean genetic improvement program is based on the combination of genes with important characteristics that influence the productive potential of soybean and its behavior in the face of the main diseases that affect this crop. In view of this, this study aimed to define soybean gene micropools through genes with non-complex actions and guide future selection strategies.

2. Material and Methods

The study was carried out in the 2022/2023 growing season in the municipality of Ijuí, located in the state of Rio Grande do Sul, in the area of the Regional University of the Northwest of the State of Rio Grande do Sul - UNIJUÍ (28°23'37.1"S and 53°56'40.5"W). The soil in the experimental area is classified as a Typical Dystroferric Red Oxisol. The climate is classified *Cfa* according to the Köppen climate classification, that is, characterized as a temperate climate (Dubreuil et al., 2018).

Sowing took place in the first fortnight of October 2022. The experimental design used was that of augmented blocks with interspersed controls, with four blocks distributed throughout the experiment in an area of 1500 m^2. 1449 soybean lines (L), no replication, of the F3 segregating generation were sown from the following populations: IRC_{001} (112_L), IRC_{002} (11_L), IRC_{003} (6_L), IRC_{005} (55_L), IRC_{007} (17_L), IRC_{008} (2_L), IRC_{009} (1_L), IRC_{011} (20_L), IRC_{012} (70_L), IRC_{013} (63_L), IRC_{017} (54_L), IRC_{020} (2_L), IRC_{021} (200_L), IRC_{022} (9_L), IRC_{025} (34_L), $_{IRC026}$ (1_L), IRC_{028} (3_L), IRC_{029} (5_L), IRC_{031} (71_L), IRC_{033} (11_L), IRC_{034} (95_L), IRC_{035} (8_L), IRC_{036} (1_L), IRC_{037} (67_L), IRC_{038} (2_L), IRC_{039} (15_L), IRC_{040} (27_L), IRC_{041} (3_L), IRC_{042} (7_L), IRC_{043} (5_L), IRC_{044} (106_L), IRC_{045} (7_L), IRC_{046} (50_L), IRC_{047} (2_L), IRC_{048} (1_L), IRC_{049} (88_L), IRC_{050} (7_L), IRC_{051} (1_L), IRC_{052} (24_L), IRC_{053} (7_L), IRC_{054} (11_L), IRC_{055} (37_L), IRC_{056} (17_L), IRC_{057} (13_L), IRC_{058} (71_L), IRC_{059} (4_L), IRC_{060} (67_L) and IRC_{061} (1_L). Four commercial genotypes were used as controls, namely VTOP SYN 1059 RR, BS 2607 IPRO, BMX ZEUS 55I57 RSF IPRO and NS 4823 RR, with four replications each, being randomly distributed throughout the blocks in a regular manner.

Each line was manually sown in a 2-m row at a density of 10 seeds m^{-1}. Fertilization took place with NPK (03-21-21) at a dose of 300 kg ha^{-1}. Phytosanitary management took place preventively to avoid the effect of biotic factors. The assessments took place as recommended by the Ministry of Agriculture (SNPC, 2008).

After seedling emergence, the following were carried out: Initial establishment assessments:

-Initial emergence at ten days (S10D): assessment carried out 10 days after sowing considering expanded cotyledons (0- slow; 1- fast).

-Initial emergence at 20 days (S20D): evaluation carried out at 20 days after sowing considering the expanded cotyledons (0- slow; 1- fast).

-Cotyledon damage at 20 days (CD20D): assessment carried out on seedlings emerged up to 20 days (0: no damage; 1: with damage).

-Cotyledon retention at 30 days (CR30D): assessment carried out 30 days after sowing considering expanded cotyledons (0- absence; 1- presence).

-Tipping (TI): assessment carried out after seedling emergence (0- absent; 1- presence)

Assessments carried out at the beginning of the reproductive period:

-Flower color (FC): evaluations carried out at the R2 stage (1-white; 2-purple).

-Leaflet shape (LSH): assessments performed on the leaflet (1-narrow lanceolate, 2- lanceolate, 3- triangular, 4- oval-pointed and 5- rounded oval);

-Leaflet size (LSI): assessment performed on the leaflet (1-small, 2-medium and 3-large).

Assessments carried out at full grain filling (R5):

-Mildew (MFC): (0- absence; 1- presence).

-Target spot (TSFC): (0- absence; 1- presence).

-Asian rust (ARFC): (0- absence; 1- presence).

-Powdery mildew (PMFC): (0 - absence; 1 - presence)

-Frogeye leaf spot (FSFC): (0- absence; 1- presence).

-Nodulation (ND): (3- high; 2- medium; 1- low).

-Growth habit (GH): (3-horizontal; 1- semi-erect, 2- erect).

-Entrapment (ENT): (2 -high, 1- low and 0- absent).

-Color of pubescence in the legume (CPL): (1- light gray, 2- dark gray, 3- light brown, 4- medium brown and 5- dark brown).

-Color of pubescence in the stem (CPS): (1-gray, 2- light brown and 3-medium brown).

Assessment carried out post-growing season:

-Hilum color (HC): (1-gray, 2- yellow, 3- light brown, 4- medium brown, 5- imperfect black and 6- black).

-Physical seed damage (SD): (0: absence; 1: presence).

-Green seeds (GS): (0 - absence; 1 - presence).

-Purple spot (PS): (0 - absence; 1 - presence).

-Stink bug seed damage (SSD): (0 - absence; 1- presence).

-Integument opening damage (ID): (0- absence; 1- presence).

-Brightness (BR): (1- low, 2- medium and 3- high).

-Halo color (HAC): (divided into 1- gray, 2- brown, 3- yellow and 4- black).

-Moisture Damage (MD): (0: absence; 1: presence).

-Peroxidase (PX): (1: positive and 0: negative.

-Color of integument (CI): (1- yellow, 2- greenish yellow).

- Seed shape (SSH): (spherical, flattened spherical).

The variables mean air temperature (°C) and precipitation (mm) were obtained through the Nasa Power platform (Nasa Power, 2023), compiled during the soybean cultivation cycle that extended from October 2022 to April 2023 totaling 200 days, divided into the vegetative period (VE - emergence) to Vn - nth node), reproductive period (R1 - beginning of flowering) to R8 (full maturation) and physiological maturation.

From the data, a descriptive analysis took place for qualitative characters based on the frequency of each observed variable. Afterwards, Kendal's linear correlation was performed. The unsupervised machine learning method was used to build lineage profiles through their agronomic attributes, using the Kohonen map, which defines the central tendencies and gathers them into centroids, based on the information expressed in each neuron. Afterwards, the main components analysis was carried out in order to establish multivariate associations between the variables and the soybean segregating populations, with these inferences being constructed through the average Euclidean distance matrix, which also contributed to the creation of the dendrogram with groupings obtained by UPGMA. Based on the analysis of principal components on the targeted agronomic ideotype, the multitrait index of the distance between the genotype and the intended agronomic ideotype (MGDI) was used (Olivoto and Nardino, 2021), the objective of the selection was to obtain lines that present a low incidence of diseases such as powdery mildew (PMFC), target spot (TSFC), mildew (MFC), Asian rust (ARFC), frogeye leaf spot (FSFC), purple spot (PS) and fungal damage (FD). The MGIDI method allows estimating the parameters: observed mean (Xo), mean of selected lines (Xs), selection differential (Sd, %) and direction of selection. The multivariate index is based on the model:

$$MGIDI_i = \left[\sum_{j=i}^{f} (\Upsilon ij - \Upsilon j)^2\right] 0,5$$

Where: MGIDI means the index of the distance between the genotype and the intended ideotype in the progeny, 'Yij represents the score of the i-th progeny in the j-th factor, 'Yj is the j-th score of the intended ideotype. Therefore, the progeny, which has a lower MGIDI index, is closer to the expected ideotype. The i-th line of the progeny, explained by the j-th factor (ω ij) is used to demonstrate the potentialities and shortcomings of the progenies (Olivoto and Nardino, 2021).

The analyzes were carried out using the R software using the packages agricolae, metan, ggplot2, readxl, rio, cowplot, dplyr, pvclust, ape, kohonen (R core team, 2022).

3. Results and Discussion

Mean temperatures did not exceed 19 °C between October and November, which coincided with the vegetative period of soybean crops, indicating favorable temperature conditions for the development of soybean plants (Pilau et al., 2022). In the soybean reproductive period, which occurred between November and February, there were mean temperatures between 19 °C to 25 °C, as well as at physiological maturity (Figure 1). It is noted that there were favorable mean temperature settings for the development of soybean plants, ranging from 13 °C to 40 °C (Tagliapietra et al., 2022).

The water demand of soybeans ranges from 450 to 800 mm throughout its cycle, with the critical periods being the germination, flowering and grain filling phase

(Carvalho et al., 2013). Moderate precipitation levels were observed, with an accumulation of 637 mm during the crop cycle, when stratifying the water supply in the months of study, an accumulation for October of 12.87%, November of 6.90%, December of 21% was noted. , January 9.41%, February 4.7%, March 21% and April 23%, revealing water disparity. To ensure the full progress of the study, maintenance irrigation took place throughout the soybean cycle.

Of the 1449 lines, it was observed that 18.56% of these did not present physiological potential and viability, being discarded from the test. At the final cycle, 15.94% of the remaining lines were discarded due to sensitivity to high air temperatures.

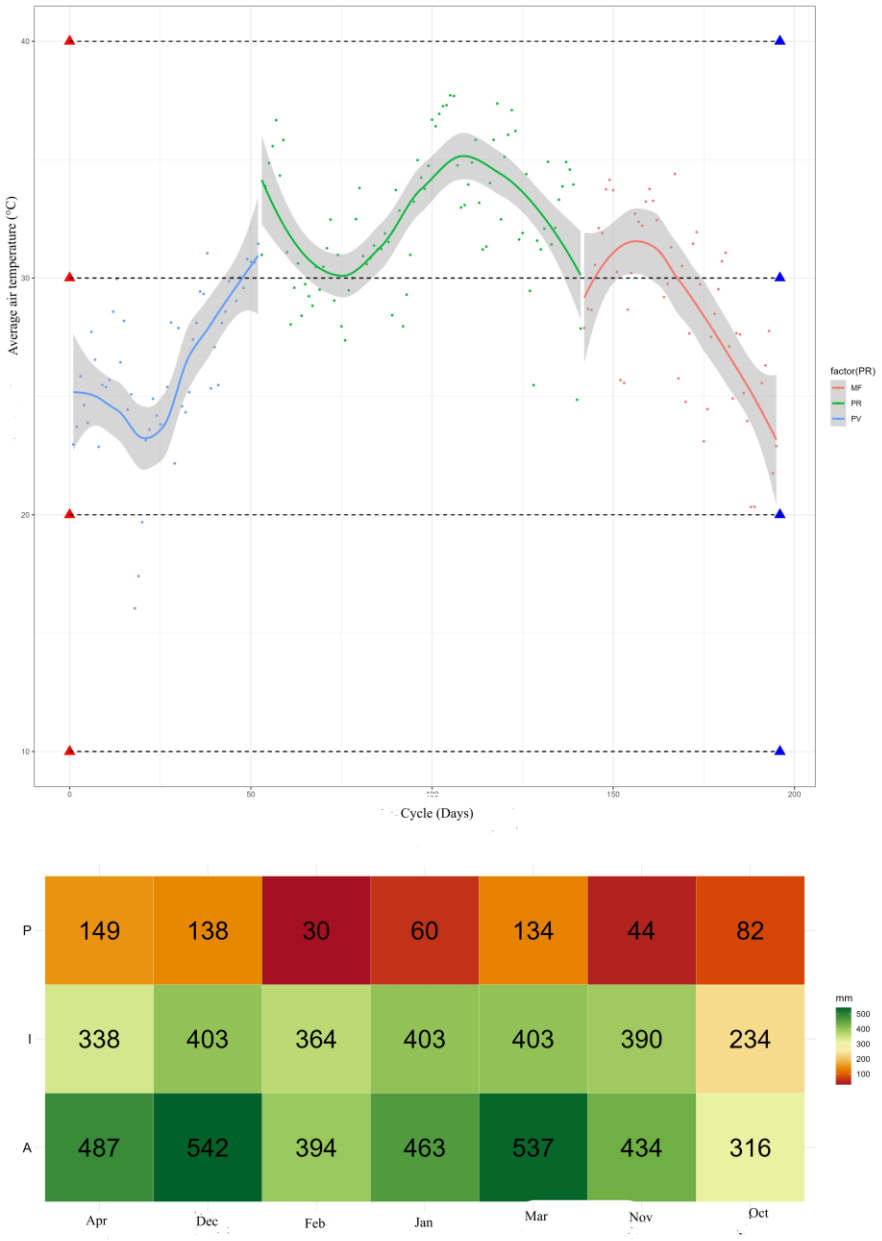

Figure 1. Mean air temperature and precipitation during October, November, December, January, February, March and April, UNIJUI, Ijuí-RS. PM: physiological maturity (days); RP: reproductive period (days); VP: vegetative period (days); Precipitation (P); Irrigation (I) and accumulated (A).

Initial plant stand assessments at 10 days (S10D) revealed viability of 1007 lines with plant stand emergence. At 20 days (S20D) it was found that 1194 lines had established themselves (Figure 2), resulting in the formation of two distinct initial plant stand scenarios. According to Farias et al. (2007), complete emergence is only considered when the cotyledons are 1.5 cm above the soil surface. Regarding the cotyledons damage at 20 days (S20D), it revealed that 1078 lines showed rupture of the skin caused by incidence of insect pests. The establishment of the lines was affected by physiological stress, mainly due to damping-off.

In order to improve the description of the lines, morphoagronomic descriptors were used, which are

responsible for the characterization of new cultivars recognized by the Ministry of Agriculture, Livestock and Supply (Mapa). For flower color (FC), 471 lines with white color and 683 with purple petals were identified. The W gene has a pleiotropic effect to control flower color and hypocotyl color, with the allele with dominant action (W1W1) determining purple color, while the recessive allele originates white flowers (w1w1) (Dellagostin et al., 2011). Regarding the presence of Asian rust at the beginning of flowering, it was identified that 931 lines are resilient and did not express the initial effects of the pathogen. For Nascimento et al. (2021), the first signs of the presence of the pathogen can be observed in the lower third of the plant, considering that they occur after flowering and the closure of the rows. The leaflet shape (LSH) is of great importance mainly for the differentiation of lines in the field. Narrow lanceolate conformation was expressed only in 14 lines, lanceolate in 701 lines, triangular in 275 lines, pointed oval in 168 lines and rounded in 9 lines. The gene responsible for controlling the leaflet shape (wide or narrow) is called Ln (wide leaflet) and ln (narrow leaflet) (Frang et al., 2023). In this context, there is a predominance of the gene in the recessive form that results narrower leaflets.

Figure 2. Qualitative morphoagronomic descriptors S10D: initial emergence at 10 days (a); S20D: initial emergence at 20 days (b); CD20D: cotyledon damage at 20 days (c); CR30D: cotyledon retention at 30 days (d); TI: Tipping (e); FC: flower color (f), PF: presence of fungus (g), LSH: leaflet shape (h).

Assessments aimed at the presence or absence of diseases (Figure 3) carried out in the reproductive period, for powdery mildew (PMFC) (*Microsphaera diffusa*) revealed that 478 lines are tolerant to the disease. For target spot (TSFC) it was identified that 975 lines were tolerant. For the incidence of mildew (MFC) (*Peronospora manshurica*) there were 630 tolerant lines. The incidence of Asian rust (ARFC) (*Phakopsora pachyrhizi*) is listed under various pressures, and it is evident that only 313 lines are tolerant to this pathogen, the remaining lines have a certain level of incidence ranging from 1% to 80%. Frogeye leaf spot (FSFC) (*Cercospora sojina*) exhibited lower levels of damage from this pathogen, with only 34 lines being affected. Overall, 41% of lines were identified as carrying genes conferring disease tolerance.

For the hilum color (HC), it was observed the existence of 355 lines with a light brown hilum, 226 with a yellow hilum, 142 with a black hilum, 91 with a medium brown hilum, 72 lines with an imperfect black hilum and 63 lines with a gray hilum, as observed in the literature, genetic control of the hilum is simple, but the variations in tone are affected by certain genetic peculiarities and environmental conditions (Pípolo et al., 2007). To characterize the presence or absence of the peroxidase enzyme (PX), it was observed that 411 lines are considered non-reactive, 420 are reactive and 119 lines are intermediate, a fact that expresses the great importance of using biochemical genetic markers to differentiate lines, being resulting from the action of the Ep gene (Valário et al., 2014).

For the halo color (HAC), called a proposal for a new morphological genetic marker for soybean, high

variability and independence of the hilum color is identified, in this way, 450 lines express a brown halo, 370 lines with a yellow halo, 125 lines with black halo and 6 lines with gray color. In this context, the halo color can be a very important trait, verifying the need for implementation together with the other evaluated traits (Figure 4). The seed shape (SSH) revealed the predominance of 922 lines with flattened spherical seeds, regarding the seed size (SSI), 937 lines showed medium seeds. For green seeds (GS), 553 absent lines were identified, making it imperative that these were considered tolerant to thermal adverse conditions, the main causes of this physiological disorder.

The purple spot (PS) indicates that 835 lines are free from the latent effects of this pathogen, showing low transmissivity throughout the breeding generations. 736 lines showed tolerance to fungal damage (FD) in the seed coat. Regarding data on integument opening damage (ID) directly linked to the degree of humidity at physiological maturity, field maturation, latent damage and physiological potential of the seeds, it is revealed that 53 lines present this injury. Stink bug seed damage (SSD) is expressed in only 45 lines and 910 lines showed some degree of moisture damage and wrinkling. In this way, it is concluded that 44% of the lines present tolerance to the main damages caused by environmental changes.

Figure 3. Assessments of diseases and morphoagronomic components in soybean crops PMFC: powdery mildew 0 no - 1 yes (a), TSFC: target spot 0 no - 1 yes (b), MFC: mildew 0 no - 1 yes (c), ARFC: Asian rust 0 no - 1 yes (d), FSFC: frogeye leaf spot 0 no - 1 yes (e), HC: hilum color 0 no - 1 yes (f), PX: peroxidase 0 no - 1 yes (g).

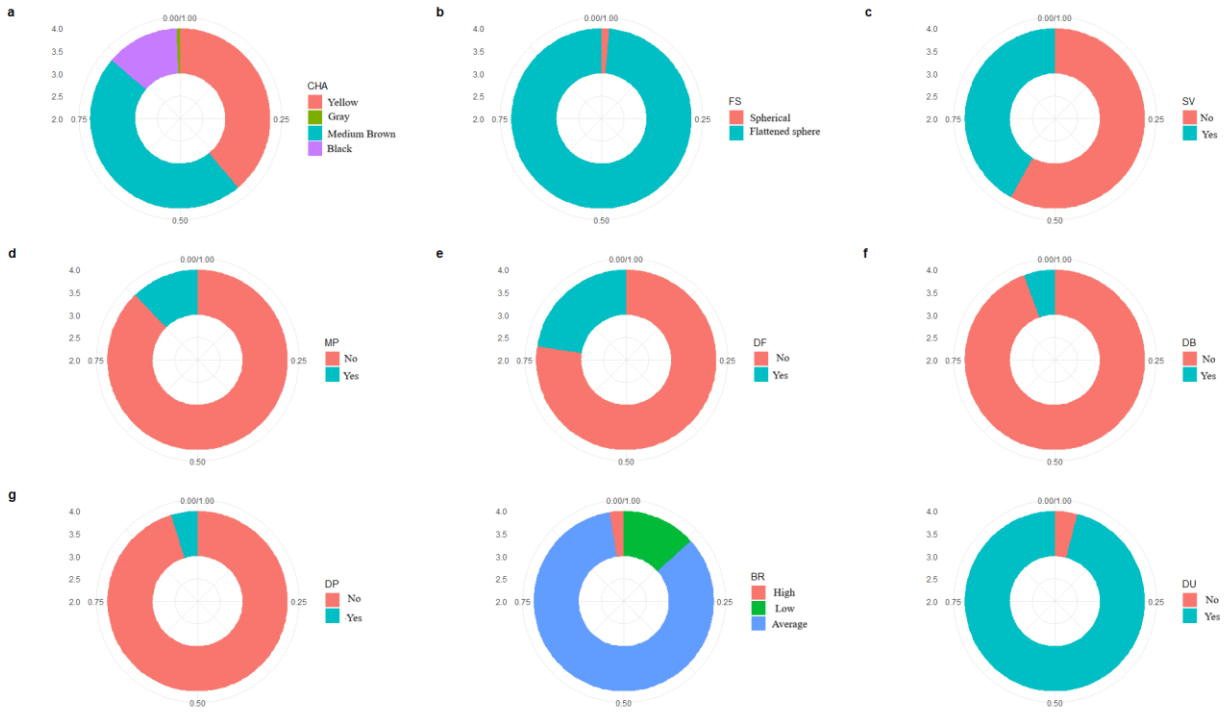

Figure 4. Assessments of diseases and morphoagronomic components in soybean crops HAC: halo color -1 black; 2 yellow; 3 brown; 4 gray (a), SSH: seed shape - 1 spherical; 2 flattened spherical (b), GS: green seeds 0 absence 1 presence (c), PS: purple spot 0 absence 1 presence (d), FD: fungal damage 0 absence 1 presence (e); ID: integument damage 0 absence 1 presence (f), SSD: stink bug seed damage 0 absence 1 presence (g), BR: brightness 1 low; 2 medium; 3 high (h) MD: moisture damage 0 absence 1 presence (i).

For nodulation (ND) it was observed that only 13 lines express a high capacity to nodulate, this can be explained by the presence of genes linked to the capacity for biological nitrogen fixation, as well as the biological conditions of the cultivation environment (Table 1).

Leaflet size (LSI) revealed that 1163 lines express medium leaflets. The color of pubescence in the stem (CPS) is controlled by two genes (TT and Td), and TTTd allelic combinations determine medium brown pubescence, TTtd light brown and tt Td have gray pubescence (Iwashina et al., 2006). Among the lines evaluated, it was observed that 1008 lines have gray pubescence, 165 lines are light brown and only two are medium brown. For the color of pubescence in the legume (CPL), 252 lines are light brown, three lines are dark gray and most lines express light gray tone. The growth habit (GH) expresses that 28 lines have a horizontal habit and must be eliminated in the selection, 534 semi-erect lines that result in care in future selections and only 612 lines have a completely erect habit and are conditioned to the agronomic ideotype.

The correlation between qualitative characters represents important associations that help in the selection of lines, moisture damage to the seed showed a negative correlation (r = -0.72) with seed brightness, which proves that excess moisture causes damage due to the integumentary wrinkling of the seed (Brandelero et al., 2019). The halo color was negatively associated with the color of pubescence in the legume (r = - 0.42), the color of pubescence in the stem (r = - 0.60) and peroxidase (r = -0.34), information linked to halo coloration and its possible modifications in important traits are scarce, therefore more studies are necessary, considering the importance of adding items in the description of lines, facilitating the breeder's activities. Leaflet shape was negatively correlated with flower color (r = - 0.33) and color of pubescence in the stem (r = -0.22). The flower color was positively associated with the color of pubescence in the legume (r = 0.30) and the color of pubescence in the stem (r = 0.34), there is a proven relationship that the high frequency of lines with purple flowers lead to the assembly of genes that determine brown pubescence. The color of pubescence in the stem was positively related to the color of pubescence in the legume (r = 0.69), considering that the same genes and alleles are responsible for the expression of both traits. Soybean seeds are sensitive to the actions of environmental factors and this is due to the morphological and chemical traits of the seed (Forti et al., 2010). Integument opening damage correlates with stink bug seed damage (r= 0.40) and fungal damage (r = 0.37), purple spot (r= 0.28) and green seeds (r= 0. 19) considering that the softness and ease of breaking the seed integument in the field, indicates that the lines will be more susceptible to other biotic and abiotic stresses.

Table 1. Kendal's linear correlation for the variables: DU: moisture damage (0 absence 1 presence); FAFC: end-of-cycle rust (0 little 100 very); CHA: halo color (1 black; 2 yellow; 3 light brown; 4 medium brown; 5 imperfect black; 6 black); MAFC: end-of-cycle target spot (0 absence 1 presence); MFC: end-of-cycle downy mildew (0 absence 1 presence); TF: leaflet size (1 small 2 medium 3 large); BR: seed brightness (1 low 2 medium 3 high); FOF: leaflet shape (1 narrow lanceolate 2 lanceolate 3 triangular 4 pointed oval); RC30D: cotyledon retention at 30 days (0 no 1 yes); PX: peroxidase (0 non-reactive 1 reactive and 2 medium reactive); MRFC: end-of-cycle frog-eye spot (0 absence 1 presence); CF: flower color (1 white 2 purple); CPH: color of the pubescence of the pod (1-gray, 2- light brown and 3-medium brown); CL: Color of the pubescence on the pod (1- light gray, 2- dark gray, 3- light brown, 4- medium brown and 5- dark brown); BD: damage by banding (0 absence 1 presence); DPS: damage perceived in the seed (0 absence 1 presence); MP: purple spot (0 absence 1 presence); SV: green seed (0 absence 1 presence); DF: fungal damage (0 absence 1 presence); CH: hilum color; DC20D: damage to the cotyledon at 20 days (0 absence 1 presence); A10D: initial sprouting at 10 days (0 absence 1 presence); A20D: initial sprouting at 20 days (0 absence 1 presence).

Associations	r	Associations	r
S10D X S20D	0.40*	PX X CPS	0.22*
CD20D X S20D	0.32*	PX X FC	-0.19*
CD20D X S10D	0.09*	LSH X S20D	-0.17*
GS X FD	0.53*	LSH X S10D	-0.11*
PS X FD	0.49*	LSH X GS	-0.11*
PS X GS	0.31*	LSH X SSD	-0.09*
SSD X FD	0.22*	LSH X CPL	-0.16*
SSD X GS	0.13*	LSH X CPS	-0.22*
SSD X PS	0.20*	LSH X FC	-0.33*
ID X FD	0.37*	BR X S20D	0.14*
ID X GS	0.19*	BR X LSH	0.17*
ID X PS	0.26*	LSI X S20D	-0.15*
ID X SSD	0.40*	LSI X BR	0.14*
CPL X HC	0.14*	MFC X GS	-0.14*
CPS X CPL	0.69*	MFC X CPL	-0.12*
FC X CPL	0.30*	MFC X CPS	-0.11*
FC X CPS	0.34*	TSFC X S20D	-0.11*
FSFC X CPL	0.14*	TSFC X GS	-0.12*
FSFC X CPS	0.18*	TSFC X MFC	0.29*
PX X CPL	0.12*	HAC X CPL	-0.42*
HAC X LSH	0.22*	HAC X CPS	-0.60*
LSH X MFC	-0.19*	HAC X FSFC	-0.17*
ARFC X CR30D	-0.16*	HAC X PX	-0.34*
ARFC X LSH	0.15*	MD X LSH	-0.18*
ARFC X TSFC	0.22*	MD X BR	-0.72*
MD X LSI	-0.13*	PMFC X TSFC	0.16*
PMFC X FC	-0.17*	PMFC X MD	0.09*

The Kohonen map revealed the grouping of 1449 lines into 25 groups (clusters) that act as a reference for the creation of gene micropools, taking into account the proximity between the highlighted variables and the equidistance between the groups (Figure 5). Among the main groups that presented connection points of great magnitude is the C1 group with 23 lines where the determining variable is the emergence at 10 days, this group being responsible for bringing together the earliest lines. The C3 centroid grouped 7 lines, which was constructed by the action of the variable initial emergence at 10 days, 20 days and cotyledon damage. The C4 group brought together only two lines that were highly susceptible to purple spot. The C7 centroid brought together five lines where the determining variables were the color of pubescence in the stem.

The C14 centroid brought together eight lines with seed brightness being the determining factor, the C15 and C18 groups supported 35 lines resulting from susceptibility to integument opening damage. The use of artificial intelligence grouping methods facilitates the distinction of lines clearly and effectively, with an emphasis on the selection of lines with greater tolerance to biotic and abiotic effects.

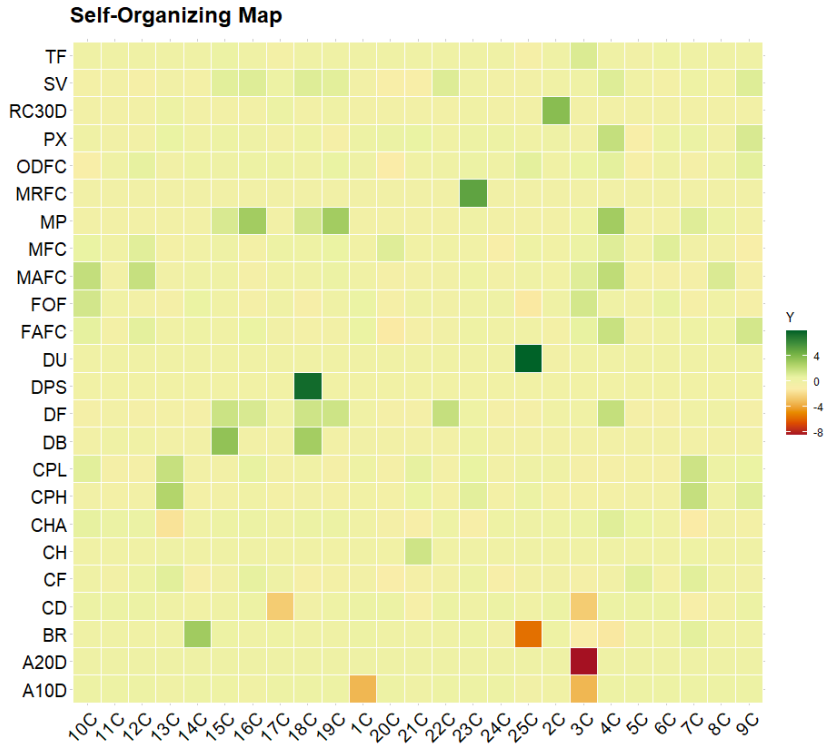

Figure 5. Kohonen self-organizing map based on 25 centroids based on the variables LSI: leaflet size, GS: green seed; CR30D: cotyledon retention at 30 days; PX: peroxidase; PMFC: powdery mildew final cycle, FSFC: frogeye leaf spot final cycle, PS: purple spot; MFC: mildew final cycle; TSFC: target spot final cycle; LSH: leaflet shape; ARFC: Asian rust final cycle, MD: moisture damage; SSD: stink bug seed damage; FD: fungal damage; ID: integument damage; CPL: color of pubescence in the legume; HAC: halo color, CPS: color of pubescence in the stem, HC: hilum color; FC: flower color; CD: cause of damage; BR: seed brightness; S10D: initial emergence at 10 days S20D: initial emergence at 20 days.

The principal components analysis (PCA) identified the affinity of the segregating populations with the analyzed variables, with an explained variance of 27.82% (Figure 6). The first quadrant grouped the IRC $_{033}$; IRC $_{043}$; IRC $_{055}$; IRC 0_{46}; IRC $_{007}$; IRC $_{009}$; IRC $_{039}$; IRC $_{035}$; IRC $_{017}$; IRC $_{060}$; IRC $_{003}$; IRC $_{011}$; IRC $_{034}$; IRC $_{013}$; IRC $_{059}$; IRC $_{050}$; IRC $_{044}$; IRC $_{031}$; IRC $_{005}$; IRC $_{025}$; IRC $_{058}$; IRC $_{042}$; IRC $_{037}$; IRC $_{012}$; IRC $_{008}$; IRC $_{042}$; IRC $_{029}$; IRC $_{054}$; IRC $_{053}$; IRC $_{056}$. These populations showed affinity for the variables fungal damage (FD), revealing that there was no presence of damage in the lines from the observed populations, leaflet shape (LSH) with prevalence of pointed oval and lanceolate forms, absence of integument opening damage (ID), presence of moisture damage (MD), presence of green seeds (GS) and absence of target spot (TSFC).

For the second quadrant, the presence of only one IRC $_{022}$ population was observed, which demonstrated affinity for the variables absence of purple spot (PS), absence of cotyledon retention at 30 days (CR30D) and absence of frogeye leaf spot (FSFC). Quadrant three grouped the populations IRC $_{021}$, IRC $_{020}$ and IRC $_{036}$ with the presence of mildew (MFC), purple flowers (PF), peroxidase reactivity (PX), average seed brightness intensity (BR), black hilum color (HC), color of pubescence in the stem (CPS) with variation between light brown and gray and color of pubescence in the legume (CPL) with light brown color dominance. The fourth quadrant grouped the populations IRC $_{051}$, IRC $_{047}$, IRC $_{057}$, IRC $_{040}$, IRC $_{041}$, IRC $_{026}$, IRC $_{009}$ and IRC $_{068}$ for the variables powdery mildew (PMFC), halo color (HAC), Asian rust (ARFC), cotyledon damage at 20 days (CD20D), initial rust (INI_RUS), initial emergence at 10 days (S10D) and initial emergence at 20 days (S20D).

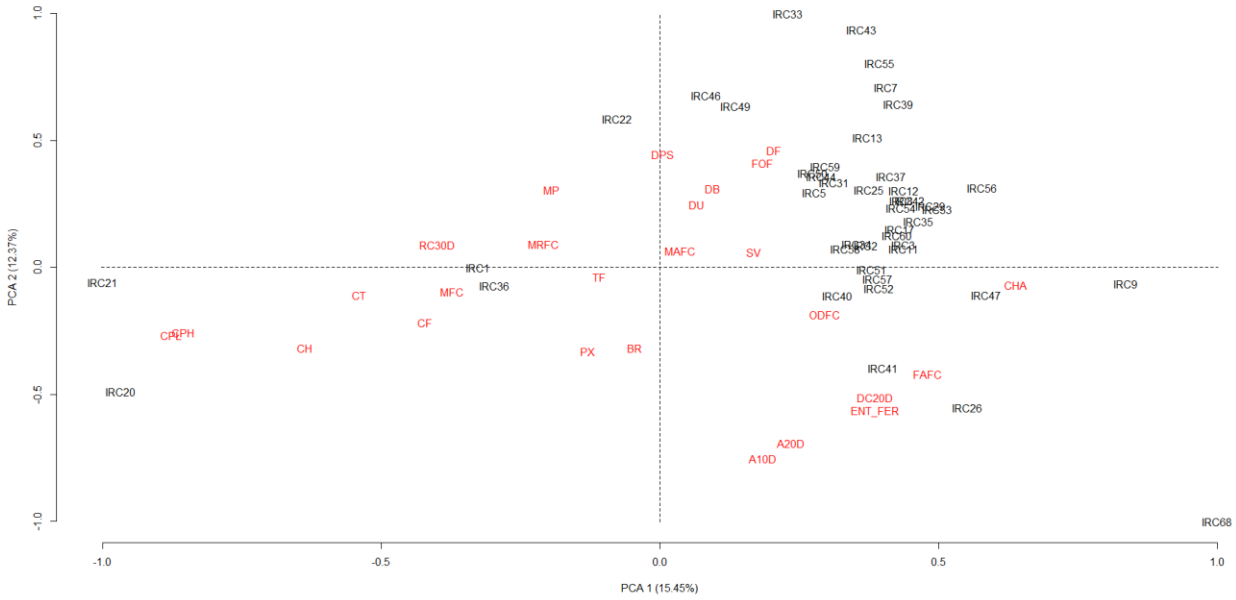

Figure 6. Biplot Principal Component Analysis (PCA) MD: moisture damage; ARFC: Asian rust final cycle; HAC: halo color; TSFC: target spot final cycle; MFC: mildew final cycle; LSI: leaflet size; BR: seed brightness; LSH: leaflet shape; CR30D: cotyledon retention at 30 days; PX: peroxidase; FSFC: frogeye leaf spot final cycle; FC: flower color; CPS: color of pubescence in the stem; CPL: color of pubescence in the legume; ID: integument damage; SSD: stink bug seed damage; PS: purple spot; GS: green seed; FD: fungal damage; HC: hilum color; CD20D: cotyledon damage at 20 days; S10D: initial emergence at 10 days; S20D: initial emergence at 20 days.

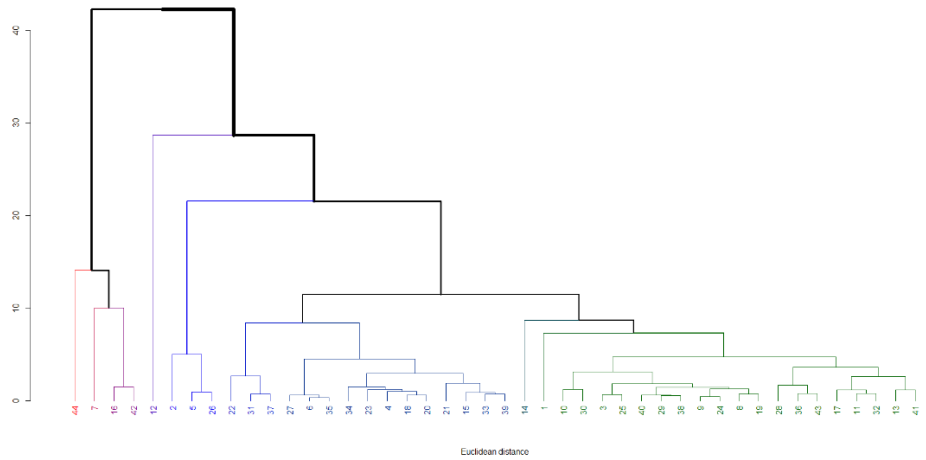

Figure 7. Genetic dissimilarity dendrogram constructed by Euclidean distance and UPGMA clustering for populations IRC044, IRC007, IRC042, IRC012, IRC002, IRC005, IRC026, IRC022, IRC031, IRC037, IRC027, IRC006, IRC035, IRC034, IRC023, IRC004, IRC018, IRC020, IRC021, IRC015, IRC033, IRC039, IRC014, IRC001, IRC010, IRC030, IRC003, IRC025, IRC040, IRC029, IRC038, IRC009, IRC024, IRC008, IRC019, IRC028, IRC036, IRC043, IRC017, IRC011, IRC032, IRC013, IRC041.

The genetic dissimilarity dendrogram revealed the formation of 7 distinct groups in which the populations were grouped according to the expression of their traits (Figure 7). The first group was formed by the IRC 44 population, the second group by the IRC 7 population. The third group by the IRC 16 and IRC 42 populations, these being earlier in terms of their establishment and flowering between 65 and 70 days, with a cycle between 143 and 167 days indicating greater precocity, pubescence of the legume and gray stem. Group four formed only by the IRC 12 population, group five brought together the IRC 2, IRC 5 and IRC 26 populations similar to flowering between 70 and 95 days, rust at final cycle, which are considered late. Group six comprised the largest number of populations, being IRC 39, IRC 33, IRC 15, IRC 21, IRC 20, IRC 18, IRC 4, IRC 23, IRC 34, IRC 35, IRC 6, IRC 27, IRC 37, IRC 31 and IRC 22 considered late flowering (67 and 103 days), and cycle of up to 191 days.

Gains in the selection process occur mainly through the genotype selection in the initial generations, with the help of estimates of genetic parameters (Santos et al.,

2019). Considered an extremely important parameter in genetic improvement, heritability allows estimating part of the phenotypic variance that will be heritable. Broad-sense heritability (H^2) can vary from 0 to 1, the closer to 1 the greater the genetic contributions to the phenotypic manifestation of the character (Hamawaki et al., 2019). Given the broad sense heritability, the mildew revealed high genetic variability compared to the other variables (H^2:0.94), frogeye leaf spot (H^2:0.50), onset at 20 days (H^2:0.49), stink bug seed damage (H^2: 0.45), cotyledon retention, initial emergence at 10 days and flower color revealed H^2:0.41, H^2:0.33, and H^2:0.3.

Among the complex of diseases that cause the most losses in soybean crops, those caused by fungi stand out (Almeida et al., 2013). The selection analysis for the ideal ideotype aimed at foliar diseases (Table 2) presented the combination of three factors (FA1, FA2 and FA3), which target the directions, pressures and variables that should be computed in the multi-trait selection. The first factor (FA1) encompassed the reduction in the incidence of powdery mildew and target spot (reduction of – 0.126% and - 0.234%). For factor two, the objective was to reduce mildew, Asian rust final cycle, frogeye leaf spot and purple spot (reductions of -0.674%, -0.227%, -0.537% and -26.9%), respectively. Factor three was responsible for reducing fungal damage to seeds with a reduction of -0.0365%.

Table 2. Multitrait genotype-ideotype distance analysis, multi-trait selection of soybean lines (MGDI) and broad-sense heritability parameters

VAR	FACTOR	Xo	SD	Sense
Powdery mildew	FA1	0.126	-0.126	Decrease
Target spot	FA1	0.234	-0.234	Decrease
Mildew	FA2	0.674	-0.674	Decrease
Asian Rust	FA2	0.227	-0.227	Decrease
Frogeye leaf spot	FA2	0.537	-0.537	Decrease
Purple spot	FA2	26.9	-26.9	Decrease
Fungal damage	FA3	0.0365	-0.0365	Decrease
Heritability (H^2)				
Initial emergence at 10 days	0.3360654			
Initial emergence at 20 days	0.4941751			
Cotyledon damage at 20 days	0.08818893			
Cotyledon retention at 30 days	0.4124855			
Flower color	0.3167257			
Powdery mildew	0.01281377			
Target spot	0.004095815			
Mildew	0.9421089			
Frogeye leaf spot	0.5091603			
Integument damage	0.03731396			
Stink bug seed damage	0.458424			

When ranking, grouping lines and populations, it is determined which main genes are gathered or separated into gene micropools, which are responsible for phenotypically bringing together traits of agronomic importance that meet the requirements of new cultivars. Without a doubt, this completed stage facilitates the work of the breeder, reduces time and financial expenses. Among the traits evaluated, the separation of 1449 lines into 25 groups was observed, assisting in the selection of these in different profiles. Considering the agronomic ideotype characterized by absence of foliar diseases, the lines L101, L1012, L1073, L1097, L1148, L1234, L1280, L1305, L1306, L138, L1403, L1406, L1413, L1414, L1424, L1429, L1446, L1 were selected. 448, L1449, L295, L298, L299, L36, L86, L897, L95, L237, L247, L1102, L1354, L1376, L718, L923, L1105, L296, L1288, L327, L496, L498. Once selected, they could be a source of tolerance genes, as well as candidates for a cultivar intended for environments where the aim is to reduce the use of agrochemicals, bringing agronomic, environmental and economic sustainability.

4. Conclusions

25 soybean gene micropools were formed based on agronomic traits, morphological and biochemical genetic descriptors.

High genetic variability was associated with tolerance to mildew (*Peronospora manshurica*) and frogeye leaf spot (*Cercospora sojina*) tolerance genes.

The gene micropool (C3) was identified to obtain early lines and cultivars, lanceolate leaflets and tolerance to target spot (*Corynespora cassicola*).

42 lines were selected for generation advancement due to the absence of the main foliar diseases, which are intended for environments with reduced use of agrochemicals and maximizing sustainability.

These findings reinforce the potential of this approach to accelerate the development of soybean cultivars suited for low-input systems, promoting more sustainable and resilient agricultural practices.

Authors' Contribution

J. P. Sangiovo: study conception, data collection, statistical analysis, interpretation of results, and manuscript writing.

I. R. Carvalho: study conception, supervision, statistical analysis, critical revision, and final approval of the manuscript.

G. H. Zuse: support in data analysis and manuscript revision.

T. B. Silva: contribution to data collection and manuscript revision.

L. C. Pradebon: assistance in results interpretation and manuscript revision.

M. V. Loro: supervision, critical content revision, and final approval of the manuscript.

Bibliographic References

Almeida, F.A., Bruscke, E.L., Polizel, A.C., Petter, F.A., Hamawaki, O.T., Neto, F.A. 2013. Agronomic performance of soybean lines and cultivars against foliar diseases. Revista de Ciências Agrárias, 26 (2), 88-94. DOI: http://dx.doi.org/10.4322/rca.2013.014.

Brandelero, W., Barbacovi, A., Rosbach, M.G.O., Viebrantz, C., Girardi, L.B., Mayer, A.R., Casassola, A. 2019. Vigor and viability of soybean seeds in response to humidity during the storage process. Brazilian Journal of Development, Curitiba, 5 (1), 342–350. DOI: https://doi.org/10.34117/bjdv5n1-924

Carvalho, I.R., Silva, J.A.S., Moura, N.B., Ferreira, L.L., Lautenchleger, F., Souza, V.Q. 2022. Methods for estimation of genetic parameters in soybeans: an alternative to adjust residual variability. Acta Scientiarum. Agronomy, 45(1), 1-10. DOI: https://doi.org/10.4025/actasciagron.v45i1.56156

Dellagostin, M., Henning, F.A., Mertz, L.M., Knopp, M.M., Crestani, M., Schuster, I., Zimmer, P.D. 2011. Genetic dissimilarity in a segregating soybean population with variability in seed morphological characters. Revista Brasileira de Sementes, 33 (4), 689 – 698. DOI: 10.1590/S0101-31222011000400011

Dubreuil, V., Fante, K.P., Planchon, O., Sant'anna Neto, J.L. 2018. The types of annual climates in Brazil: an application of the Köppen classification from 1961 to 2015. Franco-Brazilian geography Journal, 37. DOI: https://doi.org/10.4000/echogeo.15017.

Farias, J.R.B., Nepomuceno, A.L., Neumaier, N. 2007. Soybean ecophysiology. Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA. Londrina- PR.

Forti, V.A., Cicero, S.M., Pinto, T.L.F. 2010. Assessment of the evolution of "humidity" damage and reduced vigor in soybean seeds, cultivar TMG113-RR, during storage, using x-ray images and physiological potential tests. Revista Brasileira de sementes, 32 (3), 123-133. DOI: https://doi.org/10.1590/S0101-31222010000300014.

Frang, C., Du, H., Wang, L., Liu, B., Kong, F. 2024. Mechanisms underlying key agronomic traits and implications for molecular breeding in soybean. Journal of Genetics and Genomics, 51 (4), 379-393. DOI: https://doi.org/10.1016/j.jgg.2023.09.004.

Hamawaki, O.T., Hamawaki, R.L., Nogueira, A.P.O., Glasenapp, J.S., Hamawaki, C.D.L., Silva, C.O. 2019. Evaluation of soybean breeding lines to new sources of root-knot nematode resistance. Ciência e Agrotecnologia, 43, 1-8. DOI: https://doi.org/10.1590/1413-7054201943009519.

Iwashina, T., Benites, E.R., Takahashi, R. 2006. Analysis of flavonoids in pubescence of soybean near-isogenic lines for color of pubescence loci. Journal of Heredity 97(5), 438–443. DOI: doi:10.1093/jhered/esl02.

Kohonen, T. 1990. The self-organizing map. Proceedings of the IEEE, 78, 1464–1480. https://doi.org/10.1109/5.58325.

Machado, B.Q.V., Nogueira, A.P.O., Hamawaki, O.T., Rezende, G.F., Jorge, G.L., Silveira, I.C., Medeiros, L.A., Hamawaki, R.L., Hamawaki, C.D.L. 2017. Phenotypic and genotypic correlations between soybean agronomic traits and path analysis. Genetics and Molecular Research, 16(2), 1-11. DOI: 10.4238/gmr16029696.

NASA. National Aeronautics and Space Administration. 2023. NASA Prediction of Worldwide Energy Resources. Disponível em: https://power.larc.nasa.gov/.

Nascimento, R.S.M., Ferreira, L.R., Zambolim, L., Parreira, D.F., Costa, Y.K.S., Damascena, J.E., Cecon, P.R. 2021. Spray mixture volume in the control of Asian soybean rust. Crop Protection, 146. DOI: https://doi.org/10.1016/j.cropro.2021.105662.

Olivoto, T. and Nardino, M. 2021. Mgidi: Toward an effective multivariate selection in biological experiments. Bioinformatics, 37 (10), 1383-1389.

Pilau, F.G., Marin, F.R., Grubert, D.A.V., Dalmago, G.A., Romanelli, T.L. 2022. Impact of ENSO-related rainfall variability on soybean yield in the state of Rio Grande do Sul, Brazil. Agrometeoros, Passo Fundo, 30. DOI: http://dx.doi.org/10.31062/agrom.v30.e027115.

Pípolo, V.C., Souza, A., Almeida, L.A., Kiihl, R.A.S, Pípolo, A.E. 2007. Inheritance of scattered hilum in soybean seeds. Pesquisa Agropecuária Brasileira. Brasilia, 42 (1), 127-129. DOI: https://doi.org/10.1590/S0100-204X2007000100017

R core team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2022.

Santos, E.R., Spehar, C.R., Pereira, P.R., Capone, A., Barros, H.B. 2019. Genetic parameters and agronomic evaluation in F2 progenies in the Federal District, Brazil. Brazilian Journal of Agricultural Sciences, Recife, 14 (1). DOI: https://doi.org/10.5039/agraria.v14i1a5625

SNPC –Serviço Nacional de Proteção de Cultivares. MAPA – Ministério da Agricultura, Pecuária e Abastecimento. 2008.(accessed on March 18, 2024).

Singh, D.P., Singh, A.K., Singh, A. 2021. Plant breeding and cultivar development. Academic Press. 2021. ISBN: 978-0-12-817563-7.

Tagliapietra, E.L., Zanon, A.L.Z. 2022. Ecofiosiologia da soja: Visando altas produtividades. 2, 1-432.

USDA. United States Department of Agriculture. 2023.World Agricultural Production – Circular Series, https://apps.fas.usda.gov/psdonline/circulars/production.pdf. (accessed on May 14, 2024).

Valário, B.P., Cavariani, C., Neto, J.B.F., Vieira, E.S.N., Bravo, J.P., Silva, E.A.A. 2014. Use of "EP"(Peroxidase) allele in soybean varietal characterization. Revista Brasileira de Sementes, 36 (4) 465-470.

Viana, J.M.S., Cruz, C.D., Barros, E.G. 2003. Genetics, volume 1, Fundamentals. 2 editions. Federal University of Viçosa - UFV.