ARTICLE

Protein selection gain in soybean grains based on segregating generations

João Pedro Dalla Roza[1], Ivan Ricardo Carvalho[1]✉, Christiane de Fátima Colet[1], Eduardo E. Foleto[1], Guilherme H. Zuse[1] and Jaqueline P. Sangiovo[2]

Crop Breeding and Applied Biotechnology
26(1): e54192624, 2026
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332026v26n2a18

Abstract: *Soybean (Glycine max L.) is a major global crop due to its diverse uses, high nutritional value, and strong production potential. This study quantified selection gains for grain protein content across segregating generations and proposed breeding strategies adjusted to heterozygosity levels. The experiment was conducted in Ijuí, Rio Grande do Sul, using an augmented block design with common controls, including 170 progenies and 44 cultivars. Genetic effects, heritability, and selection gains were estimated through Bayesian inference with Markov Chain Monte Carlo. Broad-sense heritability ($H^2 = 0.451$) indicated moderate genetic control of protein content. Greater selection intensity was suitable for the F_{10} generation due to low heterozygosity (0.19%), whereas earlier generations (F_5, F_7, and F_8) required milder selection to preserve genetic variability. Adjusting selection intensity across generations ensures reliable parameter estimation and consistent genetic gain, highlighting the usefulness of Bayesian approaches in protein-oriented soybean breeding.*

Keywords: *Yield, biochemical composition, heterozygosity*

INTRODUCTION

Soybean (*Glycine max* L.) is one of the most important commodities in the global agricultural landscape. Its prominence is mainly due to its wide range of uses, nutritional quality, and production potential. Considered the main source of vegetable protein, soybeans play a fundamental role in both human and animal nutrition, in addition to being used in various industries, such as pharmaceuticals and cosmetics (Kim et al. 2021). Furthermore, the biochemical composition of soy protein includes the amino acids leucine, isoleusine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine (Kudelka et al. 2021). These attributes make soy an important ingredient in the formulation of products such as protein beverages, flour, and nutritional supplements.

Historically, breeding programs have prioritized increasing productivity, which has reduced the focus on quality traits. Thus, selection strategies for protein should be conducted carefully, especially over segregating generations, where heterozygosity is greater and genetic variability can be exploited more efficiently.

In early segregating generations, elevated heterozygosity not only increases genetic variability but also affects the expression of genotype × environment (G×E) interaction. Heterozygous loci may exhibit dominance or partial dominance effects that are differentially expressed across environments, thereby amplifying

***Corresponding author:**
E-mail: carvalho.irc@gmail.com

Scientific Editor:
Luiz Antônio dos Santos Dias

Received: 10 October 2025
Accepted: 18 February 2026
Published: 24 February 2026

[1] Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Rua do Comércio, 3000, 98700000, Ijuí, RS, Brazil
[2] Universidade Federal de Santa Maria, Avenida Roraima, 1000, 97105-900, Santa Maria, RS, Brazil

phenotypic plasticity and inflating the apparent G×E interaction. As a result, phenotypic performance in these generations may reflect environment-specific dominance effects rather than stable additive genetic values, reducing the accuracy of selection for protein content. As homozygosity increases in later generations, dominance effects are progressively reduced, and the additive genetic signal becomes more consistent across environments, allowing a clearer discrimination of genotypic performance.

Previous studies on protein selection in soybeans demonstrate modest gains, especially when selection is applied late or without adjusting the intensity to the level of available genetic variability (Assefa et al. 2019). However, few studies have evaluated how selection intensity should vary between generations (F_2, F_3, F_4...), nor how different strategies can maximize additive gain in the face of progressively decreasing heterozygosity.

Furthermore, selection strategies applied to each generation still lack comparative studies, especially considering that the heritability of protein content is influenced by genotype × environment interaction and the polygenic nature of the trait (Pathan et al. 2013). This gap hinders the definition of optimized protocols for segregating populations.

Bayesian methods have gained prominence in plant breeding for offering greater stability in scenarios of small samples, hierarchical structure, and high uncertainty typical conditions of segregating populations. Unlike frequentist approaches, these methods allow the incorporation of prior information, estimation of the complete distribution of parameters, and better handling of variability between generations, making them a promising alternative for predicting selection gains.

Given these limitations and opportunities, this study aimed to quantify selection gains for protein content across different segregating generations and propose selection strategies adjusted to the heterozygosity level of each generation, using Bayesian modeling to improve the accuracy and robustness of the estimates.

MATERIAL AND METHODS

The study was conducted in Ijuí (lat 28° 53' 10" S and long 52° 59' 55" W, alt 300 m asl), in the state of Rio Grande do Sul, by the genetic improvement program of the Regional University of Northwest of the State of Rio Grande do Sul. The soil is classified as a Typical Dystrophic Red Latosol, and the climate is characterized as humid subtropical Cfa. The progenies were sown in the first half of November 2023 and harvested in the first half of April 2024.

The experimental design was an augmented block design with common controls. The regular treatments consisted of 170 soybean progenies (Figure 1), and the common treatments consisted of 44 commercial soybean cultivars (Tables 1 and 2), arranged in four blocks. The experimental units consisted of a ten-meter-long seeding row spaced 0.5 m apart. For all genotypes, a seeding density of 18 seeds per linear meter was used, with a base fertilizer of 250 kg ha^{-1} of NPK (05-20-20). Phytosanitary management was carried out preventively to mitigate biotic effects (insect pests, diseases, and invasive plants) in the experiment.

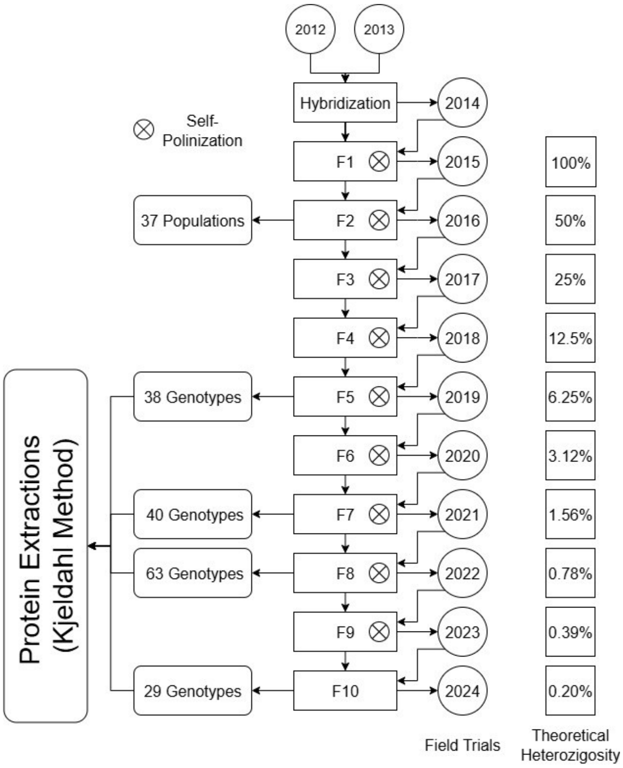

Figure 1. Chronological diagram of the genetic improvement process using hybridizations, F_1 to F_{10}.

The progenies were stratified into 38 F_5 lines (87.5% homozygosity and 6.25% heterozygosity), 40 F_7 lines (96.88% homozygosity and 1.56% heterozygosity), 63 F_8 lines (98.44% homozygosity and 0.78% heterozygosity) and 29 F_{10} lines (99.61% homozygosity and 0.19% heterozygosity) (Tables 1 and 2). The main variable evaluated after harvest was grain protein, obtained by assessing 50 representative plants

Table 1. Soybean control cultivars and the relative maturity group

Cultivars	GMR*	Cultivars	GMR
95R51	5.0	FPS JUPTER RR	5.9
35Y02 IPRO		FPS PARANAPANEMA RR	5.6
95Y92	5.0	FPS SOLIMÕES RR	
AS 3590 IPRO	5.9	IPR BASALTO	6.4
BMX COMPACTA 65I65 RSF IPRO	6.5	IPR PÉ VERMELHO	6.5
BMX CROMO TF IPRO 57I59 RSF IPRO	5.7	IPR PETROVITA	6.5
BMX LANÇA 58I60 RSF IPRO	5.8	M 5705 IPRO	5.7
BMX LÓTUS 61I63 RSF IPRO	6.1	M 5838 IPRO	5.8
BMX RAIO 50I52 RSF IPRO	5.0	M 5917 IPRO	5.9
BMX ZEUS 55I57 RSF IPRO	5.5	M 5947 IPRO	5.9
BRS 10001 IPRO		M 6210 IPRO	6.2
BRS 511	6.4	M 6410 IPRO	6.4
BRS 525	5.6	NA 5909 RG RR	6.2
BRS 537	5.7	NS 4823 RR	4.8
BRS 539	6.1	NS 5959 RR	5.9
BRS 546	6.0	NS 6006 IPRO	6.0
BRS 580 4RR	5.8	NS 6601 IPRO	6.6
BRS 6013	6.0	SYN 1059 RR	5.9
C 2570 RR	5.7	TMG 7061 IPRO	6.1
DM 57I52 RSF IPRO	5.7	TMG 7262 RR	6.2
DM 5958 RSF IPRO	5.8	TMG 7363 RR	6.3
FPS1859 RR	5.9	URNRS23	6.0

* Relative Maturity Group (GMR)

from each experimental unit, where the grains were manually threshed and grouped, and a 100 g sample of grains was separated and subjected to a drying process until it reached 13% moisture.

Subsequently, the percentage of total crude protein (%) was determined using the acid-catalytic digestion method, followed by distillation and titration using the Kjeldahl method, employing a conversion factor of 6.25, indicated for soybean cultivation (Kjeldahl 1883). The data obtained were subsequently subjected to analysis of the assumptions of the statistical model, normality and homogeneity of residual variances, and model additivity. Afterwards, the selection differentials were calculated using the overall mean of the commercial controls (C) and the segregating generations (F_5, F_7, F_8, and F_{10}). Thus, the overall standard deviation of the experiment was calculated for the next steps (Table 3).

Subsequently, the Bayesian inference model based on the Markov Chain Monte Carlo (MCMC) algorithm was used, through the MCMC*glmm* function, to estimate the genetic effects. Thus, informative prior matrix, random effects for the genotypes, 100,000 iterations and a burn-in of 10,000 were considered. The model used was based on Bandeira et al. (2025a), as follows:

$$Y = X\beta + Z_1\delta_1 + Z_2\delta_2 + e$$

Where: y is the vector of phenotypic values, β is the vector of the incidence matrix and corresponding to the vector of systematic effects (general mean), Z_1 and Z_2 are the incidence matrices of random effects, δ_1 is the vector of block effects and δ_2 is the vector of genetic values, e is the residual vector.

In this model, the mean parameters of the protein percentage referring to the posterior distribution (post mean), 95% credible interval (UP-95% CrI), significance of the probabilistic model by the Markov chain Monte Carlo method (pMCMC) at 5% probability and broad-sense heritability (H^2) were estimated. To understand the transgressive profile of the lines and promote selection, the method of identifying transgressives based on selection differentials graphically, segregated by generation, was used, using the *transg* function through the *EstimateBreed* package (Bandeira et al. 2025b). All analyses were performed using R software (R Core Team 2023).

Table 2. Genealogy of soybean segregating lines and populations

Population	MP	PP	*NL*	*F_5*	*F_7*	*F_8*	*F_{10}*
IRC 001	DM 7.0 BMX MAGNA RR	FUNDACEP 66 RR	13		3	7	3
IRC 002	DM 5.8 BMX APOLO RR	FUNDACEP 66 RR	8		4	3	1
IRC 003	DM7.OBMXMAGNA RR	MONASCA RR	4		3		1
IRC 011	BMX POTÊNCIA RR	MAR.M4 CONDOR BRANQ	2			1	1
IRC 012	FPS NETUNO RR	DM 5.8 BMX APOLO RR	15	4	4	3	4
IRC 013	DM 7.0 BMX MAGNA RR	DM 5.8 BMX APOLO RR	5		2		3
IRC 016	ROOS CAMINO RR	FPS URANO RR	1				1
IRC 017	DM 7.0 BMX MAGNA RR	FUNDACEP 66 RR	4			3	1
IRC 021	TMG 7062 RR	6700 RR	7	3			4
IRC 023	6700 RR	5601 IPRO	1				1
IRC 028	MASSAL	COLETA TP	2				2
IRC 031	NS 5958 RR	MAR.M2 CHAPADA	6		3	1	2
IRC 032	DM 7.0 BMX MAGNA RR	NS 5909 RR	2				2
IRC 035	BMX FORÇA RR	MAR.M4 CONDOR BRANQ	3		1	1	1
IRC 038	6700 RR	MAR M3 VICENTE DUTRA	1				1
IRC 039	TMG 7166 RR	NS 5909 RR	6			5	1
IRC 005	ROOS CAMINO RR	FPS PARANAPANEMA RR	3	1	1	1	
IRC 007	DM 5.8 BMX APOLO RR	MAR.M4 CONDOR BRANQ	2	2			
IRC 034	NS 5958 RR	6700 RR	14	6	3	5	
IRC 037	NS 5909 RR	MAR M4 CONDOR BRANQ	3	3			
IRC 040	BMX TURBO RR	TMG 7166 RR	2	2			
IRC 044	M5710	TMG2757	7	2	3	2	
IRC 048	DM 7.0 BMX MAGNA RR	FPS PARANAPANEMA RR	2	1		1	
IRC 056	HO PURICA IPRO	HO AMABOY IPRO	2	2			
IRC 058	15b70 IPRO	6563 IPRO	7	7			
IRC 060	HO JACUI IPRO	HO PURICA IPRO	4	4			
IRC 025	MASSAL	COLETA DP	11	1	2	8	
IRC 022	6700 RR	BMX ATIVA RR	3		1	2	
IRC 033	MAR.M4 CONDOR BRANQ	NS 5909 RR	6		1	5	
IRC 042	DM 7.0 BMX MAGNA RR	DM 5.9 BMX ALVO RR	2		1	1	
IRC 046	DM 7.0 BMX MAGNA RR	BMX ATIVA RR	2		2		
IRC 049	6700 RR	DM 6160 IPRO	4		2	2	
IRC 050	BMX POTENCIA RR	5909	9		4	5	
IRC 030	4823RR	ROTA54	1			1	
IRC 047	TMG 7262 RR	NS 5958 RR	1			1	
IRC 045	FPS JÚPITER RR	DM 7.0 BMX RR	2			2	
IRC 043	DM 6563 IPRO	NS 5958 RR	3			3	

Maternal Parent (MP), Paternal Parent (PP), Number of Lines (NL).

Table 3. Selection differentials (SD) used for line selection

SD	Definition	Nr. sample	**Reference** (% protein)
SD_1	Overall mean for the segregating generation – F_5	38	36.47
SD_2	Overall mean for the segregating generation – F_7	40	36.29
SD_3	Overall mean for the segregating generation – F_8	63	36.13
SD_4	Overall mean for the segregating generation – F_{10}	29	35.82
SD_5	Overall mean of 44 commercial controls – C	44	35.66
SD_6	Mean of the best commercial control – BC	-	44.29
SD_7	Reference mean - (Breeding)	-	40%
SD_8	Overall mean of controls plus one standard deviation (3.62g) - (C + 1S)	-	39.24
SD_9	Overall mean of controls plus two standard deviations (2 x 3.62g) - (C + 2S)	-	42.29
SD_{10}	Overall mean of controls plus three standard deviations (3 x 3.62g) - (C + 3S)	-	45.34

Table 4. Estimated genetic parameters for the categories and segregating generations of the studied genotypes

Categories	**Mean *a posteriori***	**CrI: 95%**		**N**	***Pmcmc***	H^2
		LOW	**UP**			
Overall	36.09	35.68	36.49	9,000	0.004*	0.451
F_5	36.47	35.55	37.37	9,000	0.004*	0.471
F_7	36.29	35.22	37.29	9,000	0.004*	0.532
F_8	36.13	35.44	36.82	8,361	0.004*	0.469
F_{10}	35.82	34.4	37.24	9,000	0.004*	0.517
Controls	35.67	34.85	36.6	9,000	0.004*	0.447

CrI: 95% Credibal Interval; N: Number of Observations; *Pmcmc*: Probability; H^2: Heritability

RESULTS AND DISCUSSION

The use of selection differentials enabled the evaluation of the productive performance of progenies in relation to the objectives established by the breeding program, allowing the estimation of potential genetic gains for grain protein content. This methodology complements the selection strategies applied across generations (Table 4), including F_5, F_7, F_8, and F_{10}, and enables comparisons with the overall mean of the controls (C = 35.66%), the best-performing control (BC), and reference thresholds based on the standard deviation of the controls (C + 1S, C + 2S, C + 3S).

The analysis of genetic parameters revealed that the F_5 generation presented the highest posterior mean for protein content (up to 36.47%), whereas F_{10} showed the lowest mean (35.82%). The overall category presented the narrowest credible interval (0.81%), while segregating generations particularly F_7 (2.07%) and F_{10} (2.87%) displayed broader intervals. This pattern is consistent with the increased homozygosity and the expression of additive genetic variance as lines become more genetically fixed (Sun et al. 2025).

Although the F_5 generation exhibited the highest mean protein content and F_{10} the lowest, these differences should not be interpreted as being driven exclusively by heterozygosity. The segregating generations evaluated in this study originated from different parental combinations, and variation in parental genetic background may have contributed to the observed contrasts among generations. Therefore, the higher protein concentration observed in F_5 may reflect, at least in part, the superior genetic potential of specific crosses rather than an intrinsic advantage of higher heterozygosity. In this context, heterozygosity should be interpreted as a contributing factor rather than a causal determinant, and the results represent generation-level patterns within a heterogeneous breeding population rather than direct within-cross comparisons.

Broad-sense heritability (H^2 = 0.451) demonstrated moderate genetic control of protein content when considering the 170 progenies and 44 controls. Although lower than the values reported by Zhang et al. (2025) (H^2= 0.7 – 0.9), the magnitude still indicates the presence of exploitable genetic variability. In evaluations conducted in a single year and environment, it is expected that the reduction in environmental variance may inflate the H^2 estimates due to confounding between genetic effects and genotype × environment interaction. However, the moderate value observed in this study suggests that the genetic variance for protein content in this specific population is relatively limited and/or that the residual variance remains significant, even under controlled environmental conditions. This behavior may be associated with the high level of homozygosity of the evaluated lines, the relatively narrow genetic base of the material, and the highly polygenic nature of protein content. Previous studies also highlight that heterogeneity between environments is substantial for this trait, due to the effects of temperature and nitrogen availability during grain filling (Carneiro et al. 2019). Therefore, validation in multiple environments and growing seasons is recommended to confirm the stability of the lines and ensure more robust and generalizable heritability estimates.

Table 5. Selection gain, stratified by segregating generations

Heterozygosity (%)	**6.25**	**1.56**	**0.78**	**0.19**	
Pressure (%)	F_5	F_7	F_8	F_{10}	**Mean gain**
1	3.46	4.72	3.52	5.21	4.23
2	3.13	4.26	3.18	4.71	3.82
3	2.91	3.97	2.96	4.38	3.55
4	2.77	3.77	2.82	4.17	3.38
5	2.66	3.63	2.71	4.01	3.26
10	2.25	3.07	2.30	3.40	2.75
20	1.79	2.44	1.83	2.70	2.19
40	1.24	1.69	1.26	1.87	1.52
50	1.02	1.39	1.04	1.54	1.25
70	0.64	0.87	0.65	0.96	0.78
80	0.44	0.61	0.45	0.67	0.54
90	0.25	0.34	0.26	0.38	0.31

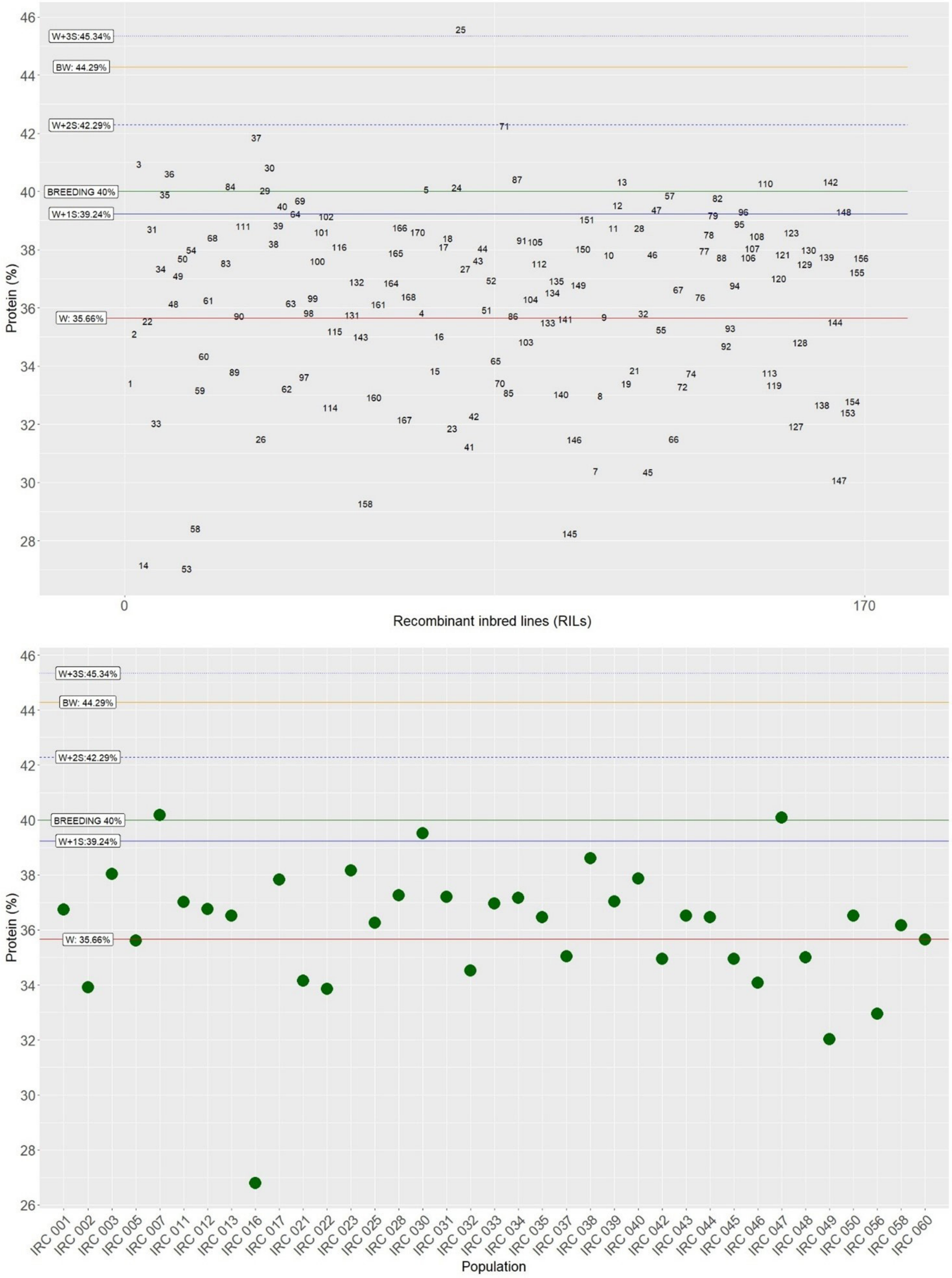

Figure 2. a. Transgressive selection of populations, subjected to selection differentials (SD): SD 5: Overall mean of 44 commercial controls – C; SD 6: Mean of the best commercial control – BC; SD 7: Reference mean to obtain 40% protein – Breeding; SD 8: Overall mean of controls plus one standard deviation (3.62%) - (C + 1S); SD 9: Overall mean of controls plus two standard deviations (2 x 3.62%) - (C + 2S); SD 10: Overall mean of controls plus three standard deviations (3 x 3.62%) - (C + 3S); b. Transgressive selection of 170 segregating lines, F$_5$ (93.75% homozygosity and 6.25% heterozygosity), 40 lines F$_7$ (98.44% homozygosity and 1,56% heterozygosity), 63 lines F$_8$ (99.22% homozygosity and 0.78% heterozygosity) and 29 lines F$_{10}$ (99.80% homozygosity and 0.20% heterozygosity), subjected to the selection differentials (SD) SD 5: General mean of the 44 commercial controls – C; SD 6: Mean of the best commercial control – BC; SD 7: Reference mean to obtain 40% protein – Breeding); SD 8: General mean of the controls plus one standard deviation (3.62%) - (C + 1S); SD 9: General mean of the controls plus two standard deviations (2 x 3.62%) - (C + 2S): SD 10: General mean of the controls plus three standard deviations (3 x 3.62%) - (C + 3S).

Selection gains differed across generations and intensities (Table 5). The highest gains (up to 4.23%) were obtained under strong selection pressure (1%), especially in F_{10} and F_7, which presented greater fixation and narrower phenotypic variance. In contrast, early selection in F_5 requires lower selection pressure (80–90%) to maintain genetic diversity, followed by progressive intensification in F_7 and rigorous selection in F_8 and F_{10}. This strategy aligns with the dynamics of heterozygosity (6.25% in F_5 to 0.19% in F_{10}) and ensures the retention of favorable recombinants. Similar patterns were described by Carvalho et al. (2024), who observed meaningful genetic gains for protein content under indirect and stringent selection strategies.

The protein data also highlighted contrasts between population-level and individual selection (Figure 2). The use of C + 1S (39.24%), C + 2S (42.29%), and C + 3S (45.34%) as thresholds enabled identification of high-performing populations (e.g., IRC 007 and IRC 047) and detection of transgressive recombinant inbred lines. Notably, line L25 exceeded the C + 3S threshold and achieved values above 45%, originating from DM 7.0 BMX Magna RR × Monasca RR, indicating strong complementarity between parental alleles affecting nitrogen assimilation and storage protein pathways (Port et al. 2024). Although values above 45% are uncommon in commercial soybean, typically ranging from 38–42%, high-protein breeding lines have occasionally surpassed these levels in controlled environments, particularly when derived from crosses specifically targeted for seed quality traits. Thus, the value for L25 appears plausible rather than resulting from measurement error, though replication in multi-environment trials is advisable.

Comparison with commercial controls (Figure 3) showed that only two cultivars (DM5958 RSF IPRO and M6410 IPRO) approached the C + 2S threshold, confirming their potential for use in quality-oriented germplasm banks. The relatively

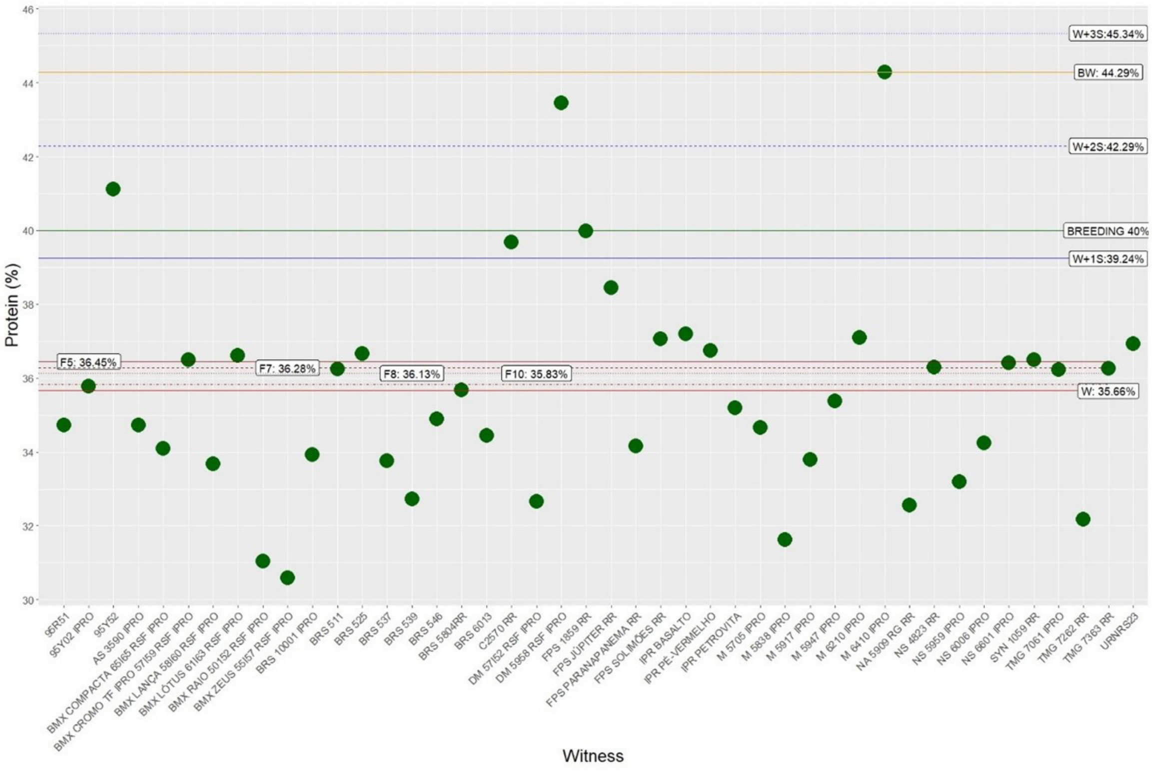

Figure 3. Descriptive analysis of controls according to selection differences (SD) SD 1: General mean for the segregating generation F_5; SD 2: General mean for the segregating generation F_7; SD 3: General mean for the segregating generation F_8; SD 4: General mean for the segregating generation F_{10}; SD 5: General mean of the 44 commercial controls - C; SD 6: Mean of the best commercial control - BC; SD 7: Reference mean to obtain 40% protein - Breeding); SD 8: General mean of the controls plus one standard deviation (3.62%) - (C + 1S); SD 9: General mean of the controls plus two standard deviations (2 x 3.62%) - (C + 2S): SD 10: General mean of the controls plus three standard deviations (3 x 3.62%) - (C + 3S).

low mean of the control group (35.66%) below typical commercial standards likely reflects the specific environmental conditions during grain filling, especially temperature and water availability, which are known to reduce protein levels (Carneiro et al. 2019). This reinforces the role of G×E in the expression of this trait and highlights the need for expanded trials to validate the superiority of high-protein lines.

Overall, the results confirm the presence of exploitable genetic variability for seed protein content, support moderate heritability, and highlight the identification of superior transgressive lines such as L25. These findings demonstrate potential for genetic progress, though future evaluations must incorporate multi-environment trials to assess stability, quantify G×E interactions, and reconcile the documented negative genetic correlation between protein and oil content an important factor for breeding decisions.

CONCLUSION

The selection gain based on the F_{10} generation with 0.19% heterozygosity appears promising and supports the use of higher selection intensity. However, the segregating generations F_5, F_7, and F_8 should be cautious regarding soybean protein selection and employ milder pressures to ensure genetic gain.

CREDIT STATEMENT

JPD Roza: Conceptualization; IR Carvalho: Data curation and methodology; CF Colet: Formal analysis; EE Foleto: Funding acquisition; GH Zuse: Investigation; JP Sangiovo: Project administration.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current research are available from the corresponding author upon reasonable request.

REFERENCES

Assefa ST, Yang EY, Chae SY, Song M, Lee J, Cho MC and Jang S (2019) Alpha glucosidase inhibitory activities of plants with focus on common vegetables. **Plants 9**: 2.

Bandeira WJA, Carvalho IR, Pradebon LC, Sausen NH, Sangiovo JP, Bruinsma GMW and Loro MV (2025a) Estimation of genetic parameters and selection pressure in canary seed (*Phalaris canariensis* L.). **Pesquisa Agropecuária Brasileira 60**: e03781.

Bandeira WJA, Carvalho IR, Loro MV, Pradebon LC and Silva JAG (2025b) EstimateBreed: Estimation of environmental variables and genetic parameters. Available at: https://cran.r-project.org/web/packages/EstimateBreed/index.html. Accessed on January 30, 2025.

Carneiro AK, Bruzi AT, Pereira JLAR and Zambiazzi EV (2019) Stability analysis of pure lines and a multiline of soybean in different locations. **Crop Breeding and Applied Biotechnology 19**: 395-401.

Carvalho IR, Silva JAG, Loro MV, Sarturi MVDR, Hutra DJ and Lautenchleger F (2021) Soybean canonical nutraceutical interrelations and their reflections on breeding. **Agropecuária Catarinense 34**: 67-75.

Kjeldahl J (1883) Neue methode zur bestimmung des stickstoffs in organischen Körpern. *Fresenius, Zeitschrift f. anal.* **Chemie 22**: 366-382.

Kim SH, Tripathi P, Yu S, Park J-M, Lee JD, Chung YS, Chung G and Kim Y (2021) Selection of tolerant and susceptible wild soybean (*Glycine soja* Siebold & Zucc.) accessions under waterlogging condition using vegetation indices. **Polish Journal of Environmental Studies 30**: 3659-3675.

Kudełka W, Kowalska M and Popis M (2021) Quality of soybean products in terms of essential amino acids composition. **Molecules 26**: 5071.

Pathan SM, Vuong T, Clark K, Lee J, Shannon JG, Roberts CA, Ellersieck MR, Burton JW, Cregan PB, Hyten DL, Nguyen HT and Sleper DA (2013) Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. **Crop Science 53**: 765-774.

Port ED, Carvalho IR, Pradebon LC, Loro MV, Colet CF, Silva JAG and Sausen NH (2024) Early selection of resilient progenies to seed yield in soybean (*Glycine max* L.) populations. **Ciência Rural 54**: e20230287.

R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Sun X, Hu B, Li WX and Ning HL (2025) Genetic basis and simulated breeding strategies for enhancing soybean seed protein content across multiple environments. **Plants 14**: 2117.

Zhang D, Sun X, Hu B, Li WX and Ning H (2025) QTN Mapping, gene prediction and molecular design breeding of seed protein content in soybean. **The Crop Journal 13***: 1116-1126.*