ARTICLE

Early selection of soybean yield via agronomic trait phenotyping under Asian soybean rust pressure

Crop Breeding and Applied Biotechnology
26(1): e540726110, 2026
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332026v26n1a10

Claudio Guilherme Portela de Carvalho[1*], Cosme Damião Cruz[2], Carlos Alberto Arrabal Arias[1] and Aaron J. Lorenz[3]

Abstract: *Asian soybean rust (ASR) is a major threat due to its aggressiveness, fungicide tolerance, and ability to overcome resistance genes. Breeding high-yielding cultivars under ASR pressure presents challenges, including low early selection accuracy and managing several lines. This study evaluated whether early agronomic traits could predict yield in later generations under ASR. Traits assessed included seed yield, days to flowering and maturity, plant height, and 50-seed weight.* F_2 *or* $F_{2:3}$ *data were used in regression and machine learning models to predict yield in* $F_{2:5}$ *progenies. Similar* R^2 *values across approaches suggested a mainly linear relationship among predictors. Using* $F_{2:3}$ *data improved* R^2*, especially for flowering, maturity, and height. Univariate models with these traits performed best, reaching* R^2 *values up to 52.12%. These models can improve early selection and reduce the breeding workload.*

Keywords: *Agronomic trait, indirect selection, linear regression, machine learning,* Phakopsora pachyrhizi

INTRODUCTION

Asian soybean rust (ASR) is caused by *Phakopsora pachyrhizi*, and poses a severe threat to soybean production, leading to near-total yield losses in affected areas (Santos et al. 2018, Chicowski et al. 2023). The pathogen exhibits low sensitivity to fungicides and quickly overcomes plant resistance genes, making the development of high-yielding cultivars under disease pressure a crucial goal for breeding programs, even if such cultivars display susceptibility symptoms (Godoy et al. 2016, Zambolim et al. 2022).

Phenotypic selection for yield is often inaccurate in the early stages of breeding programs due to the difficulty of evaluating individual plant performance. Due to this low accuracy, programs tend to advance a large number of segregating lines, increasing both costs and logistical complexity. Predicting yield in later generations based on early-stage genotyping could enable early elimination of low-potential lines, thereby reducing the number of lines and the experimental area required. This approach is especially valuable when working under the specific environmental conditions necessary for ASR development (Bock et al. 2022), which are hard to achieve uniformly across large field trials.

Carvalho et al. (2025) predicted mean soybean agronomic traits in $F_{2:5}$ progenies under ASR pressure based on genotyping of an F_2 population.

***Corresponding author:**
E-mail: portela.carvalho@embrapa.br

Received: 23 May 2025
Accepted: 12 January 2026
Published: 17 January 2026

[1] Embrapa Soja, Rodovia Carlos João Strass, Acesso Orlando Amaral, Distrito de Warta, s/n, 86085-981, Londrina, PR, Brazil
[2] Universidade Federal de Viçosa, Departamento de Biologia Geral, Avenida Peter Henry Rolfs, s/n, 36570-000, Viçosa, MG, Brazil
[3] University of Minnesota, 100 Church Street SE, s/n, 55455, Minneapolis, Minnesota, United States

However, the number of plants genotyped can be an economically limiting factor for genomic selection, especially in public-sector breeding programs (Wartha and Lorenz 2021). An alternative may be to predict yield in later generations using early-stage phenotyping of agronomic traits. Prediction models based on multiple linear regression (MLR) and machine learning methods such as bagging (BA) and random forests (RF) (Yoosefzadeh-Najafabadi et al. 2021, Sousa et al. 2021, Piekutowska et al. 2021, Pérez et al. 2024) support informed selection decisions.

This study assessed whether phenotyping agronomic traits in early generations can predict seed yield in later generations of the population examined by Carvalho et al. (2025). We obtained models for phenotypic prediction which exhibited higher coefficients of determination (R^2) than those for genomic prediction. These results suggest that phenotypic prediction could improve the efficiency of the selection process by enabling earlier decisions in the breeding cycle. Incorporating such models into breeding programs may optimize resource use, reduce the scale of field trials, and support the development of high-yielding cultivars under ASR pressure.

MATERIAL AND METHODS

Field trials

Field experiments were conducted at Embrapa Soja (lat 23° 11’ 37” S, long 51° 11’ 03” W, and alt 630 m asl), Londrina, PR, Brazil; to evaluate F_2, $F_{2:3}$, and $F_{2:5}$ generations derived from a cross between the ‘BRQ16-5409’ and ‘BR13-9499’ soybean lines (Supplementary Tables and Figures). These elite lines were selected for seed yield, with only the ‘BRQ16-5409’ line being previously exposed to Asian soybean rust (ASR). Neither parental line carries Rpp1 to Rpp7 resistance genes against *Phakopsora pachyrhizi* (Childs et al. 2018).

The F_2 population (230 plants) and the two parental lines (50 replicates each) were grown in November 2019 in 330 randomly single-plant hill plots (one hill plot = one plant) (Lima et al. 2012). Then, we evaluated 230 $F_{2:3}$ progenies and the two parental lines in November 2020 in a completely randomized design with 30 replicates for each parental line and six replicates (plants) for each progeny, totaling 1,440 hill plots (Figure S1b). Plots were spaced 20 cm apart within rows and 1.5 m between rows in both 2019 and 2020. Two rows of the susceptible ‘BRS Conquista’ cultivar were planted as borders within this 1.5 m distance and around the trial area.

In turn, the experimental design in November 2022 involved segregating progenies with intercalary lines comprising 230 $F_{2:5}$ progenies, each with six lines (totaling 1,380 $F_{2:5}$ lines), along with the two parental lines and two control cultivars ‘BRS 531’ (resistant) and ‘BRS 523’ (susceptible) (Figure S1c). The 1,380 $F_{2:3}$ plants gave rise to $F_{2:4}$ plants, which subsequently produced seeds for the F2:5 lines under greenhouse conditions. Each parental line and control cultivar was replicated 30 times. One row of each parental line and control cultivar were sown for every 50 $F_{2:5}$ lines, resulting in 1,500 single-row plots, all 3 m long and spaced 0.5 m apart. The ‘BRS Conquista’ cultivar was also sown as a border around the trial area.

Crop management followed standard soybean production protocols. ASR inoculum was prepared and applied to border plants at the V3 stage, following the procedures described by Ribeiro et al. (2008) and Lima et al. (2012). Phenotypic evaluations included seed yield per plant or progeny (SY), days to flowering (DF) and maturity (DM), plant height (PH), 50-seed weight (50SW), ASR severity (ASRS), and area under the disease progress curve (AUDPC). The agronomic traits and ASRS were recorded in the F_2, $F_{2:3}$, and $F_{2:5}$ generations. ASRS was assessed as the percentage of infected leaf area using a graphical scale (Franceschi et al. 2020). Four evaluations were performed at approximately 7-day intervals starting 80 day post-emergence when the crop canopy had closed and a more favorable microclimate for infection was established (Isard et al. 2006). AUDPC was calculated by trapezoidal integration (Silveira et al. 2003). Lesion types were classified as reddish-brown (RB, resistant) or tan-colored (TAN, susceptible) (Goellner et al. 2010). Plants reached physiological maturity at the R7 stage (Fehr and Caviness 1977), after which they were dried, threshed, and weighed individually for F_2 and $F_{2:3}$ generations, and by row for $F_{2:5}$.

Heritability and correlation

Narrow-sense heritabilities of the agronomic traits were estimated at the individual level in the F_2 generation (h^2_{F2}) and at the progeny mean level in the $F_{2:3}$ generation ($h^2_{F2:3}$). Genetic variance components were calculated using the weighted least squares method (Mather and Jinks 1984). Both parental lines were cultivated in the same area along with each F_2, $F_{2:3}$ and $F_{2:5}$ generations (in 2019, 2020 and 2022) (Figure S1b, c), and were used as checks with 50, 30 and 30 replicates, respectively, obtaining variance among plants in each year to estimate the environmental variance used to estimate heritability. Phenotypic correlations among traits were estimated within each generation, as well as the associations between traits evaluated in F_2 and $F_{2:3}$ with SY in $F_{2:5}$. The statistical significance of the correlations was assessed using the t-test.

Phenotypic prediction

The predict phenotypic values (PPVs) for SY in F2:5 were calculated using models constructed with MLR (Piekutowska et al. 2021), BA (Yoosefzadeh-Najafabadi et al. 2021, Sousa et al. 2021), and RF (Prasad et al. 2006, Yoosefzadeh-Najafabadi et al. 2021, Sousa et al. 2021, Costa et al. 2022, Pérez et al. 2024). Multicollinearity for the MLR model was assessed via the condition number (CN) of the correlation matrix (Montgomery and Peck 1981). CN values <100 indicated weak, 100–1000 moderate to strong, and >1000 severe multicollinearity. Both BA and RF utilized 500 trees (Costa et al. 2022), with all variables included in BA and one-third of the variables in RF.

Models used the phenotypic value of individual plants in the F_2 generation, the mean phenotypic value of six plants from each $F_{2:3}$ progeny, and the mean phenotypic value of six lines from each $F_{2:5}$ progeny. Predictions were based on combinations (response variable vs. predictor variable): $F_{2:5}$ vs. F_2 and $F_{2:5}$ vs. $F_{2:3}$.

The population (230 F_2 plants or 230 $F_{2:3}$ progenies and their respective $F_{2:5}$ progenies) was randomly divided into five groups (folds) of 46 genotypes (F_2 plants or $F_{2:3}$ progenies and their respective $F_{2:5}$ progenies) each to obtain the models. In turn, four folds (184 genotypes) were used in each iteration as the training population and the remaining fold (46 genotypes) as the validation population. This five-fold cross-validation was repeated four additional times, resulting in five iterations and ensuring that each genotype participated in four training and one validation set. We built genomic prediction models using phenotypic data of agronomic traits from each training population. SY values in $F_{2:5}$ served as response variables, while traits from F_2 and $F_{2:3}$ acted as predictors. We then applied these models to estimate the PPVs of $F_{2:5}$ progenies in the validation sets for each iteration. Predictive ability was assessed by calculating the correlation between PPVs and observed mean phenotypic values of $F_{2:5}$ progenies in the validation population. The mean predictive ability of the models obtained in the five iterations was defined as the predictive ability of the first validation. This procedure was repeated nine more times, yielding predictive abilities for 10 validations. Finally, we calculated the mean and the respective standard deviation of the predictive abilities of the 10 validations.

The Tukey's test compared the predictive abilities of models obtained using different approaches and agronomic traits evaluated in the F_2 and $F_{2:5}$ generations. Comparisons between approaches within the same trait and between traits within the same approach were performed when there was a significant Trait × Approach interaction ($p < 0.05$). When the interaction was not significant, comparisons were made among traits and among approaches. Prediction models based on multiple linear regression and machine learning approaches were compared to evaluate the importance of additive and non-additive genetic effects on SY prediction.

Regression analyses

Stepwise regression analysis was conducted to identify the independent variables which most strongly explained the dependent variable. These variables were subsequently incorporated into prediction models constructed using MLR, BA, and RF. Furthermore, variables that contributed to the prediction of the response variable in the multivariate analysis were also included. The contribution of each variable was estimated using R^2 values, while accounting for the effects of the remaining variables in the MLR model.

Computational analyses

Heritability was calculated using GENFIT (Toledo 1991). Analyses for multicollinearity, correlations, PPVs, regression analyses, predictive ability, and R^2 values of the models were conducted using GENES (Cruz 2016).

RESULTS AND DISCUSSION

Disease progression and symptoms

The 95% confidence intervals (95%CI) for the AUDPC trait in the F_2 population and the 'BRQ16-5409' and 'BR13-9499' parental lines were 1261.4 ± 76.5, 1280.2 ± 69.5, and 1345.5 ± 72.4, respectively (Figure 1). All plants exhibited AUDPC values above 1000 in 2019, which were higher than those observed in the $F_{2:3}$ and $F_{2:5}$ populations and parental lines in 2020 and 2022 with the values ranging from 737.6 ± 80.8 to 909.8 ± 69.8. The susceptible 'BRS 523' cultivar (875.5 ± 62.3) showed similar AUDPC values to the $F_{2:5}$ population in 2022, whereas the resistant 'BRS 531' cultivar exhibited significantly lower values (255.8 ± 48.2). These differences in AUDPC between resistant and susceptible genotypes, as illustrated in Figure 1S, were comparable to or greater than those reported in other studies involving soybean genotypes with and without *Rpp* genes (Vuong et al. 2016, Sacon et al. 2020).

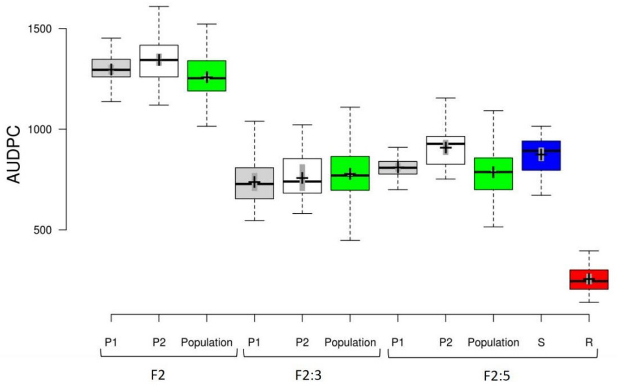

Figure 1. Area under disease progress curve (AUDPC) of a soybean population from the cross between 'BRQ16-5409' (P1) and 'BR13-9499' (P2), under pressure from Asian rust. The resistant 'BRS 531' cultivar (R) and the susceptible 'BRS 523' cultivar (S) were also evaluated in the $F_{2:5}$ generation.

All plants from the parental lines, the susceptible control 'BRS 523' and the F_2, $F_{2:3}$, and $F_{2:5}$ populations developed TAN-type lesions (tan-colored) with abundant sporulation (Figure S1d). In contrast 'BRS 531' developed RB-type lesions (reddish brown) characterized by limited sporulation and extensive necrosis. The TAN-type lesions indicate the absence of *Rpp* genes in the population derived from the cross between 'BRQ16-5409' and 'BR13-9499'. The AUDPC values obtained for the F_2, $F_{2:3}$, and $F_{2:5}$ generations confirmed that environmental conditions were conducive to disease development in the field (Lima et al. 2012), which provided an opportunity to construct prediction models for PPVs of soybean genotypes for agronomic traits under Asian soybean rust pressure.

Regression analyses

Models including DM and 50SW in F_2 or $F_{2:3}$ as explanatory variables were estimated using stepwise regression, yielding R^2 values of 46.2% and 56.06%, respectively (Table S1). However, these values were similar to those obtained from univariate analysis including DM (45.12% in F_2 and 54.36% on $F_{2:3}$). The regression coefficient values for the univariate analysis of DM were negative in F_2 (-5.89) and $F_{2:3}$ (-9.40).

When examining the contribution (R^2) of each explanatory variable to the response variable prediction in the multivariate analysis, it can be noted that the greatest contribution of a variable in the model including the other variables was from DM (R^2 = 5.10%) in F_2 and from 50SW (R^2 = 3.44%) in $F_{2:3}$ (Table S2). However, as the other agronomic variables exhibited non-zero R^2 values (Table S2), their univariate models were also evaluated in the prediction analyses, in addition to the multivariate model.

Phenotypic prediction

The prediction models were obtained using the phenotypic value of SY in $F_{2:5}$ as the response variable and phenotypic values of agronomic traits in F_2 or $F_{2:3}$ as predictor variables. We used the MLR, BA, and RF methods to construct univariate and multivariate models from the training populations. All traits evaluated in F_2 or $F_{2:3}$, including SY, were considered in the multivariate models. These models were then used to predict SY in $F_{2:5}$ in the validation populations, and their R^2 values are shown in Figures 2 and 3.

No significant trait x approach interaction was observed for the coefficient of determination of the models (Table S3). Tukey's test was performed to compare traits or approaches (Figures 2 and 3). Models based on MLR showed higher R^2 values than those obtained using machine learning methods (Figures 2a and 3a). Machine learning algorithms

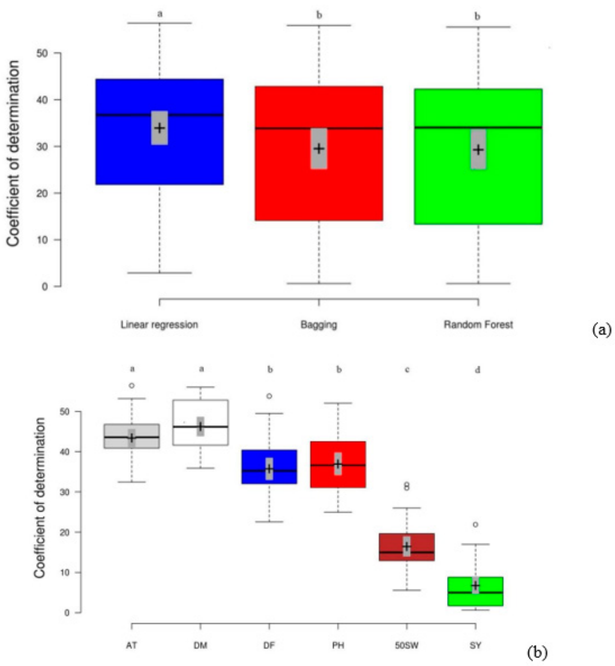

Figure 2. Coefficients of determination (%) of models obtained from different approaches (linear regression, bagging, and random forest) (a) and agronomic traits (AT: all traits, DM: days to maturity, DF: days to flowering, PH: plant height, 50SW: 50-seed weight and SY: seed yield per plant) (b) evaluated in F_2 and $F_{2:5}$ soybean populations under Asian rust pressure. The models' dependent variable was SY in $F_{2:5}$ and the independent variables were agronomic traits in F_2. Models with the same lowercase letter do not differ by Tukey's test at 5% probability.

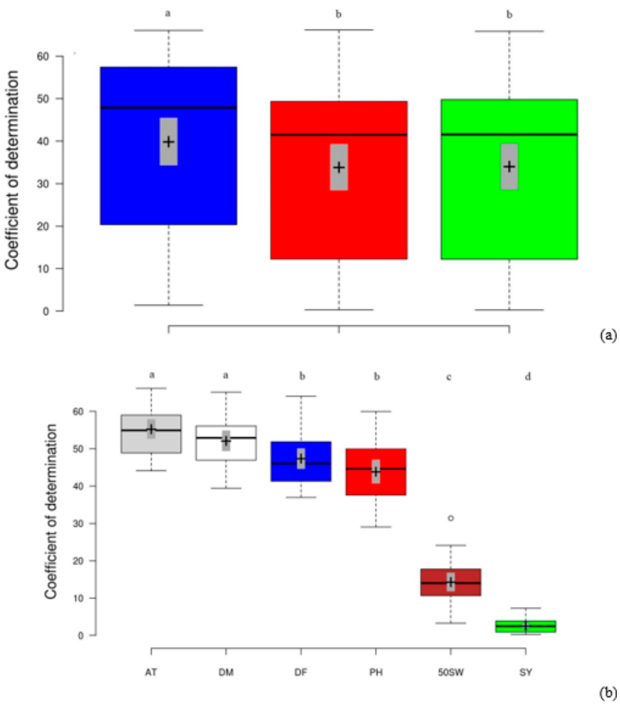

Figure 3. Coefficients of determination (%) of models obtained different approaches (linear regression, bagging, and random forest) (a) and from agronomic traits (AT: all traits, DM: days to maturity, DF: days to flowering, PH: plant height, 50SW: 50-seed weight and SY: seed yield per plant) (b) evaluated in $F_{2:3}$ and $F_{2:5}$ soybean populations under Asian rust pressure. The models' dependent variable was SY in $F_{2:5}$ and the independent variables were agronomic traits in $F_{2:3}$. Models with the same lowercase letter do not differ by Tukey's test at 5% probability.

are generally effective at capturing nonlinear patterns in predictive modeling (Parmley et al. 2019, Sousa et al. 2021, Silva Junior et al. 2023). These algorithms make no assumptions about the underlying model, enabling identification of complex genetic interactions, such as epistasis and dominance, within predictive models (Sousa et al. 2021). Thus, the **superiority of the** R^2 values obtained **from MLR-based models indicates a limited contribution of non-additive genetic effects to SY prediction and** suggests a predominantly linear relationship among the predictor variables and consequently greater importance of additive effects in the expression of SY. Ribeiro et al. (2008) also reported predominantly additive effects of polygenes controlling soybean yield in the presence and absence of *P. pachyrhizi*. When additive effects predominate, selecting segregating populations under severe rust pressure (Figure 1) may be an appropriate strategy for cultivar development.

Multicollinearity among traits can adversely affect the estimation of coefficients in a regression analysis (Del Conte et al. 2020). Multicollinearity occurs when the sample observations of the explanatory or predictor variables, or their linear combinations, are correlated. The variances associated with the estimators of regression coefficients in the presence of multicollinearity can reach excessively high values, making them unreliable (Bizeti et al. 2004). Furthermore, parameter estimates may assume absurd values or values that lack coherence with the biological phenomenon under study. In the present study, the CN of the correlation matrices in F_2 and $F_{2:3}$ populations was 63.3 and 67.7, respectively, indicating weak multicollinearity, according to Montgomery and Peck (1981). Thus, we did not perform ridge regression or trait culling to overcome collinearity problems and estimate univariate and multivariate models (Bizeti et al. 2004).

The mean R^2 values (46.25% in F_2 and 52.12% in $F_{2:3}$) for the univariate analyses of DM did not differ statistically ($p < 0.05$) from those obtained in the multivariate analyses (43.39% in F_2 and 55.30% in $F_{2:3}$), and were higher than those of the other univariate models (Figures 2b and 3b). Along with DM, univariate models using DF or PH exhibited R² values above 35% when using phenotypic data from the $F_{2:3}$ and $F_{2:5}$ generations.

The narrow-sense heritabilities at the individual level in F_2 were 0.59 (SY), 0.79 (DM), 0.86 (DF), 0.84 (PH), and 0.53 (50SW). Heritabilities at the progeny mean level in $F_{2:3}$ were 0.43 (SY), 0.95 (DM), 0.85 (DF), 0.93 (PH), and 0.84 (50SW). These heritability values can influence predictive accuracy. Traits with higher heritability specifically tended to produce higher R² values in the prediction models. For example, traits with heritabilities around 0.80 in F_2 resulted in models with R² values above 35.74% (Figures 2b and 3b), whereas traits with heritabilities around 0.55 produced models with R² values below this threshold. Supporting this pattern, Carvalho et al. (2025) also found that traits with greater heritabilities showed higher R² values in genomic prediction using the population of the present study.

Breeding program

Breeding programs aiming to develop high-yielding soybean cultivars under Asian soybean rust pressure need to devise strategies to enhance selection accuracy in the early stages and reduce the number of lines under selection in later stages to increase logistical feasibility. In this regard, MRL and machine learning were performed for indirect selection, and models were constructed using agronomic trait phenotyping in the early stages of the breeding program (F_2 or $F_{2:3}$ generation) and yield phenotyping in the later stages ($F_{2:5}$).

Using phenotypic values from $F_{2:3}$ to enhance early-stage selection accuracy, as opposed to F_2, increased the R² values of multivariate models or univariate models for DM, DF, or PH (Figures 2b and 3b). The highest R² with $F_{2:3}$ data reached 55.30%, while the maximum with F_2 data was 46.25%. Replacing F_2 phenotypic values with the mean of $F_{2:3}$ progenies enabled improved model fit. However, although $F_{2:3}$ offers greater accuracy and potential genetic gains, breeders must weigh this against the increased mechanical and financial costs required to advance this generation. Therefore, the decision to select in $F_{2:3}$ should consider both the improved prediction accuracy and the resource demands associated with evaluating progeny rows.

As prediction models showed R² values near 50%, evaluating six plants per $F_{2:3}$ progeny was adequate, as supported by previous inheritance studies (Ribeiro et al. 2008, Lima et al. 2012). The F_2, $F_{2:3}$, and $F_{2:5}$ generations were evaluated in the field, while $F_{2:4}$ was assessed in a greenhouse. However, if agronomic traits are measured in F_2, $F_{2:3}$ can be conducted in a greenhouse, and vice versa. Additionally, lines can be evaluated in $F_{2:4}$ instead of $F_{2:5}$, shortening the breeding cycle by one year.

Phenotypic prediction enables reducing the number of lines in later selection stages, improving logistical feasibility. For instance, selecting 2% of 10,000 F_2 plants yields 200 plants. If each produces 50 $F_{2:5}$ lines, 10,000 lines are generated. Without prior modeling in the training/validation population, all 10,000 F_2 plants would result in 500,000 field lines. This volume hinders maintaining optimal conditions for pathogen development (Bock et al. 2022). Reducing the number of lines helps preserve environmental suitability, enhancing the effectiveness of selecting productive lines under disease pressure in advanced breeding stages.

Managing ASR in Brazil involves a soybean-free period, during which planting is prohibited in certain areas to reduce inoculum pressure for the next season (Godoy et al. 2016). Additionally, growing early-maturing cultivars is another strategy, as these cultivars spend less time in the field and can escape late-season infections or reduce disease severity (Zambolim et al. 2022). In this study, DM in F_2 and $F_{2:3}$ was negatively correlated with SY in $F_{2:5}$ (-0.67 and -0.74, respectively) (Table S4). Furthermore, univariate models using DM as a predictor had the highest R² values (Figures 2b and 3b), with negative regression coefficients, suggesting that early maturity in early generations contributed to selecting the most productive $F_{2:5}$ progenies. However, unlike typical early sowing conditions in Brazil, the inoculum pressure in our study differed. While soybeans in Londrina are usually sown from September to December, the trials reported herein were planted in November and subsequently inoculated with *P. pachyrhizi* to create high rust pressure, resulting in elevated ASRS and AUDPC values for all plants (Figure 1). Despite these late-sowing and high ASR pressure conditions, Carvalho et al. (2025) reported that earlier plants experienced shorter exposure to rust, and this may reduce physiological damage. However, in this study, positive correlations between grain yield and days to maturity were observed in the F_2 and $F_{2:3}$ generations (Table S5).

The accuracy of predictive models in soybean for grain yield has varied according to the objectives of the breeding program or selection methods (Yoosefzadeh-Najafabadi et al. 2021, Pérez et al. 2024). For example, the genomic prediction of $F_{2:5}$ progenies conducted by Carvalho et al. (2025) based on the population under study generated models with an R^2 of up to 36.00% (predictive ability = 0.60). On the other hand, the coefficients obtained from phenotypic prediction, as shown in Figures 2 and 3, reached 55.30%. In addition to the difference in accuracy of the predictive models, the cost of genotyping F_2 plants can be unfeasible for breeding programs in public institutions, depending on the number of plants genotyped (Wartha and Lorenz 2021). In contrast, evaluating DM, DF, or HP (for example) is a routine activity in these programs.

In addition to DM, the DF and HP traits in F_2 (-0.59 and -0.63, respectively) and $F_{2:3}$ (-0.81 and -0.72, respectively) showed negative correlations with SY in $F_{2:5}$. These correlations are consistent with the negative regression coefficients in the univariate models for these traits. Moreover, the development of early, shorter, and high-yielding cultivars is not only important for improved rust control, but also for better adaptation to the Brazilian production system.

The possibility of selecting agronomic traits such as DM, DF, or PH in the early breeding stages enables reducing the segregating lines, enhances the ability to maintain favorable climatic conditions for optimal pathogen development, and minimizes the need for a larger logistical structure to manage and harvest the lines.

CONCLUSION

Phenotyping of agronomic traits in the early stages of a breeding program can enable selecting seed yield in later stages of a soybean population under Asian soybean rust pressure. This strategy contributes to cost reduction, facilitates maintaining favorable climatic conditions for optimal pathogen development, allows for early-generation selection, reduces the need for extensive physical infrastructure and labor for harvesting and other cultural practices, and increases genetic gains.

ACKNOWLEDGMENTS

We thank the Brazilian National Council for Scientific and Technological Development (CNPq) for granting a scholarship to the first author.

DATA AVAILABILITY

The datasets generated and/or analyzed in this study, as well as the supplementary tables and figures, are available from the corresponding author upon reasonable request.

REFERENCES

Bizeti HS, Carvalho CGP, Souza JRP and Destro D (2004) Path analysis under multicollinearity in soybean. **Brazilian Archives of Biology and Technology 47**: 669-676.

Bock CH, Chiang K and Del Ponte EM (2022) Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. **Tropical Plant Pathology 47**: 25-42.

Carvalho CGP, Cruz CD, Arias CAA and Lorenz AJ (2025) Genomic prediction of yield components in soybean under Asian soybean rust pressure for early-generation selection. **Australian Journal of Crop Science 19**: 351-360.

Chicowski AS, Bredow M, Utiyama AS, Marcelino-Guimarães FC and Whitham AS (2023) Soybean-*Phakopsora pachyrhizi* interactions: towards the development of next-generation disease-resistant plants. **Plant Biothecnology Journal 22**: 296-315.

Childs SP, Buck JW and Li Z (2018) Adult-plant resistance to net form net blotch in barley (*Hordeum vulgare* L.): Durability and genetic control. **Plant Breeding 137**: 250-261.

Costa WG, Celeri MO, Barbosa IP, Silva GN, Azevedo CF, Borem A, Nascimento M and Cruz CD (2022) Genomic prediction through machine learning and neural networks for traits with epistasis. **Computational and Structural Biotechnology Journal 20**: 5490-5499.

Cruz CD (2016) Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum-Agronomy 38**: 547-552.

Del Conte MV, Carneiro PCS, Resende MDV, Silva FL and Peternelli LA (2020) Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. **Plos ONE 15**: e0233290.

Fehr WR and Caviness CE (1977) **Stages of soybean development**. Iowa State University of Science and Tecnology, Ames, 11p.

Franceschi VT, Alves KS, Mazaro SM, Godoy CV, Duarte HSS and Del Ponte EM (2020) A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. **Plant Pathology 69**: 495-505.

Godoy CV, Seixas CDS, Soares RM, Guimarães FCM, Meyer MC and Costamilan LM (2016) Asian soybean rust in Brazil: past, present, and future. **Pesquisa Agropecuária Brasileira 51**: 407-421.

Goellner K, Loehrer M, Langenbach C, Conrath U, Koch E and Schaffrath U (2010) *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. **Molecular Plant Pathology 11**: 169-177.

Isard SA, Dufault NS, Miles MR, Hartman GL, Russo JM, Wolf ED and Morel W (2006) The effect of solar irradiance on the mortality of *Phakopsora pachyrhizi urediniospores*. **Plant Disease 90**: 941-945.

Lima WF, Prete CEC, Ribeiro AS and Toledo JFF (2012) Resistência da soja à ferrugem-asiática avaliada pela análise da produtividade de grãos. **Summa Phytopathologica 38**: 73-78.

Mather K and Jinks JL (1984) **Introdução à genética biométrica**. Sociedade Brasileira de Genética, São Paulo, 242p.

Montgomery DC and Peck EA (1981) **Introduction to linear regression analysis**. John Wiley & Sons, New York, 504p.

Parmley KA, Higgins RH, Ganapathysubramanian B, Sarkar S and Singh AK (2019) Machine learning approach for prescriptive plant breeding. **Scientific Report 9**: 17132.

Pérez O, Diers B and Martin N (2024) Maturity prediction in soybean breeding using aerial images and the random forest machine learning algorithm. **Remote Sensing 16:** 4343.

Piekutowska M, Niedbała G, Piskier T, Lenartowicz T, Pilarski K, Wojciechowski T, Pilarska AA and Czechowska-Kosacka A (2021) The application of multiple linear regression and artificial neural network models for yield prediction of very early potato cultivars before harvest. **Agromony 11***:* 885.

Prasad AM, Iverson LR and Liaw A (2006) Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. **Ecosystems 9**: 181-199.

Ribeiro AS, Toledo JFF, Arias CAA, Godoy CV, Soares RM, Moreira JUV, Pierozzi PHB, Vidigal MCG and Oliveira MF (2008) Genetic control of soybean (*Glycine max*) yield in the absence and presence of the Asian rust fungus (*Phakopsora pachyrhizi*). **Genetics and Molecular Biology 31**: 98-105.

Sacon D, Stefanski FS, Gallina A, Milanesi PM and Borges LL (2020) Integration between genetic and chemical control on the progress of Asian soybean rust and yield. **Summa Phytopathologica 46**: 198-204.

Santos JVM, Yamanaka N, Marcelino-Guimarães FC, Toledo JFF, Arias CAA and Abdelnoor RV (2018) Molecular mapping of quantitative trait loci for agronomical traits in soybean under Asian soybean rust infection. **Crop Breeding and Applied Biotechnology 18**: 390-398.

Silva Junior AC, Moura WM, Bhering LL, Siqueira MJS, Costa WG, Nascimento M and Cruz CD (2023) Prediction and importance of predictors in approaches based on computational intelligence and machine learning. **Agronomy Science and Biotechnology 9**: 1-24.

Silveira EB, Michereff SJ and Mariano RLR (2003) Severidade da mancha-aquosa em meloeiro sob diferentes condições de molhamento foliar e concentração de inóculo de *Acidovorax avenae* subsp. *citrulli*. **Fitopatologia Brasileira 28**: 171-175.

Sousa IC, Nascimento M, Silva GN, Nascimento ACC, Cruz CD, Silva FF, Almeida DP, Pestana KN, Azevedo CF, Zambolim L and Caixeta ET (2021) Genomic prediction of leaf rust resistance to Arabica coffee using machine learning. **Scientia Agricola 78**: e20200021.

Toledo JFF (1991) Programa de computador para estimar parâmetros genéticos, componentes de médias e variâncias, pelo método dos quadrados mínimos ponderados. **Pesquisa Agropecuária Brasileira 26**: 1023-1039.

Vuong TD, Walker DR, Nguyen BT, Nguyen TT, Dinh HX, Sleper DA, Lee JD, Shannon JG and Nguyen HT **(**2016) Molecular characterization of resistance to soybean rust (*Phakopsora pachyrhizi* Syd. & Syd.) in soybean cultivar DT 2000 (PI 635999). **PLoS One 11**: e0164493.

Wartha CA and Lorenz AJ (2021) Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. **Crop Breeding and Applied Biotechnology 21:** e394621S15.

Yoosefzadeh-Najafabadi M, Tulpan D and Eskandari M (2021) Application of machine learning and genetic optimization algorithms for modeling and optimizing soybean yield using its component traits. **PLoS ONE 16**: e0250665.

Zambolim L, Reis EM, Guerra WD, Juliatti FC and Menten JOM (2022) Integrated management of Asian soybean rust. **European Journal of Applied Sciences 10**: 602-633.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.