International Journal of *Plant Biology*

MDPI

Article

Genomic Prediction of Root Traits via Aerial Traits in Soybean Using Canonical Variables

Vitor Seiti Sagae [1,2], Noé Mitterhofer Eiterer Ponce de Leon da Costa [2], Matheus Massariol Suela [1,2], Dalton de Oliveira Ferreira [1], Ana Carolina Campana Nascimento [2], Camila Ferreira Azevedo [2], Felipe Lopes da Silva [1] and Moysés Nascimento [2,*]

[1] Departamento de Agronomia, Universidade Federal de Viçosa, Viçosa 36570-260, Brazil; vitor.sagae@ufv.br (V.S.S.); matheus.suela@ufv.br (M.M.S.); daltonferreira.ufv@gmail.com (D.d.O.F.); felipe.silva@ufv.br (F.L.d.S.)

[2] Departamento de Estatística, Universidade Federal de Viçosa, Viçosa 36570-260, Brazil; noe.costa@ufv.br (N.M.E.P.d.L.d.C.); ana.campana@ufv.br (A.C.C.N.); camila.azevedo@ufv.br (C.F.A.)

* Correspondence: moysesnascim@ufv.br

Abstract: The phenotypic evaluation of root traits in soybeans presents challenges in breeding due to its high cost and the requirement for experimental plot destruction. Establishing relationships between aerial and root traits is crucial, given the relative ease of phenotypic evaluations for aerial traits. Therefore, this study aims to utilize the canonical correlation technique to estimate latent variables, subsequently employing GBLUP for the genomic prediction of the root traits (length, volume, surface area, and dry mass) using phenotypic information from aerial part traits (hypocotyl diameter and dry mass). Our results demonstrate the effectiveness of the technique in predicting the root part, even when not directly evaluated. The agreement observed between the top 10% of individuals selected based on the canonical variable and each root trait individually was considered moderate or substantial. This enables the simultaneous selection of genotypes based on both trait groups, providing a valuable approach for soybean breeding programs.

Keywords: genome-wide selection; canonical correlation; *Glycine max* (L.) Merr.; molecular markers

Citation: Sagae, V.S.; Costa, N.M.E.P.d.L.d.; Suela, M.M.; Ferreira, D.d.O.; Nascimento, A.C.C.; Azevedo, C.F.; Silva, F.L.d.; Nascimento, M. Genomic Prediction of Root Traits via Aerial Traits in Soybean Using Canonical Variables. *Int. J. Plant Biol.* **2024**, *15*, 242–252. https://doi.org/10.3390/ijpb15020020

Academic Editor: Adriano Sofo

Received: 15 March 2024
Revised: 29 March 2024
Accepted: 2 April 2024
Published: 5 April 2024

1. Introduction

Soybean (*Glycine max* (L.) Merril) is an herbaceous oleaginous plant with a C3 metabolism and great genetic variability [1,2]. The grain is highly versatile, finding applications in the production of various products and by-products across the agroindustry, chemical, and food sectors [3]. Consequently, soybeans are being increasingly cultivated worldwide [3,4].

Drought stress stands out as a significant limitation to soybean yield globally [5–7]. One strategy employed to mitigate issues related to drought stress involves identifying materials with superior phenotypes in root development, facilitating better plant performance through more efficient water absorption [8]. Specifically, deeper root systems with higher densities enable more efficient soil water extraction. To define the phenotype of the material regarding root superiority, it is essential to measure root traits such as root angle, diameter, length, surface area, and depth [8].

Despite obtaining interesting results [9,10], phenotyping root traits is a time-consuming and challenging technique, complicating the improvement process based on these traits [11,12]. Given the delay in measuring root traits, understanding relationships between these and easily measurable traits, such as aerial traits, becomes crucial. These relationships may allow the selection of better genotypes for root traits based on aerial traits without significant losses in the selection of these genotypes. In practical terms, this possibility would reduce the time and cost of evaluating the roots of genotypes in a breeding program. A statistical method that can be used to study the relationships between groups

of traits, for example, aerial and root, is canonical correlation. Canonical correlation (CC) analysis is a multivariate technique that seeks to determine linear combinations, denoted by canonical variables, that maximize the correlation between the two sets of characteristics of interest [13].

Another approach that can be employed to accelerate the development of new cultivars, can increase genetic gain, and is particularly useful when measuring phenotypes proves difficult or expensive is genome-wide selection (GWS) [14]. GWS uses information from molecular markers, based on the principles of linkage disequilibrium between genes and markers, to predict the genetic merit of genotypes, making it possible later to select genotypes based on their respective genotypic information [15].

Genomic best linear unbiased prediction (GBLUP) is the most applied genomic prediction method in soybeans [16]. However, genomic prediction in soybeans is usually used for individual traits, disregarding any relationship between traits or groups of traits [17,18]. Nevertheless, since the traits could be correlated, multi-trait approaches could improve predictive ability by borrowing information between traits controlled by pleiotropy or linkage genes [19–21]. But generally, these methods require defining a balance in breeding objectives during selection by the breeders [22]. Thus, an approach that simultaneously incorporates information from multiple traits through latent variables that maximize the correlation between those groups of traits in the selection process may be of interest.

In this context, the present study aims: (1) to investigate the relationship between root and aerial traits in soybean through canonical correlation analysis; (2) to use canonical latent variables as pseudo-phenotypes in genomic prediction; (3) to compare selected individuals using canonical latent variables and individual traits with univariate GBLUP.

2. Materials and Methods

2.1. Plant Material

The experiment was conducted in the year 2021 during the period from September to November in the greenhouse of the Federal University of Viçosa, located in Viçosa-MG at latitude S 20°45′14″, longitude W 45°52′54″, and an altitude of 649 m [23]. One hundred commercial soybean cultivars from different seed companies, distinct transgenic events, growth types, and relative maturity groups were sown. The experiment was conducted according to the methodology proposed and described by Nascimento et al. [24] in a randomized block design with three replications, each corresponding to a plot.

Due to destructive nature of root phenotypic evaluations, to avoid root intertwining, each experimental plot comprised a plant that was grown in a three-liter pot containing a substrate composed of soil and sand in a 2:1 ratio. The substrate was moistened to near field capacity, considering a clayey loam soil with tension values of −10 kPa until the V2 stage [25,26]. After the V2 stage, the tension was maintained at −1200 kPa for a period of 20 days. The pots were regularly weighed, and the volume of evapotranspiration water was replenished [24].

Cultural practices followed technical recommendations for soybean cultivation [27]. Phenotypic traits related to aerial and root traits were destructively evaluated for each plot. Hypocotyl diameter in millimeters (HD) was measured with a digital pachymeter for aerial traits. Plant height in centimeters (PH) was determined as the length of plants from the ground to the peak of the main stem.

For root traits, the roots were washed to remove substrate materials and then stored in a 70% alcohol solution at 4 °C. The WinRHIZOPro® software (version 2009) [28] was used to measure total root length in centimeters (TRL), root volume (RV), and projected surface area (PSA). After evaluation in WinRHIZOPro®, the roots were dried in an oven at 65 °C for 48 h and then weighed using a precision balance to obtain the root dry mass (RDM) in grams.

2.2. Genotypic Data Analysis

Individuals were genotyped using the iScan Illumina platform (Illumina, Inc., San Diego, CA, USA) with the "BARC-SoySNP6k" chip at Deoxi Biotecnologia Ltd.a®, in

Araçatuba, SP. The 5403 SNP markers obtained were subjected to quality control through filtering based on the minor allele frequency (MAF) criterion, removing markers with MAF < 0.05. After filtering, 3957 SNP markers were retained for subsequent analyses.

2.3. Phenotypic Data Analysis

The collected data were initially adjusted for effects of the experimental blocks. Correction for experimental effects was performed using mixed linear models (REML/BLUP) [29]. The model considered a randomized block design with three replications and the evaluation of 100 genotypes, expressed as:

$$\mathbf{y} = \mu + \mathbf{Xb} + \mathbf{Zg} + \varepsilon, \tag{1}$$

where the vector $\mathbf{y}$ represents the phenotypes, μ is the general mean, and the vectors $\mathbf{b}$ and $\mathbf{g}$ correspond to the random effects of blocks and genotypes, respectively. Here, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2)$ and $\mathbf{g} \sim N\left(\mathbf{0}, \mathbf{I}\sigma_g^2\right)$, where $\mathbf{I}$ represents an identity matrix, σ_b^2 and σ_g^2 represents the variance components of blocks and genotypes, respectively, and ε represents the vector of residual effects, with $\varepsilon \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$, where σ_ε^2 is the residual variance components. $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices of random effects, associated with block and genotype effects, respectively. Genetic random effects values were predicted and utilized in subsequent analyses.

The adjusted phenotypic data was used as input (pseudo-phenotypes) in the canonical correlation analysis. The two trait vectors of aerial (**A**—group I) and root (**R**—group II) parts are defined as follows:

$$\mathbf{A}^t = [u_{HD} u_{PH}], \tag{2}$$

$$\mathbf{R}^t = [v_{RDM} v_{RV} v_{PSA} v_{TRL}], \tag{3}$$

where u_{HD}, u_{PH}, v_{RDM}, v_{RV}, v_{PSA}, and v_{TRL} represent the corrected traits of hypocotyl diameter, plant height, root dry mass, root volume, projected surface area, and total root length, respectively. $\mathbf{A}^t$ and $\mathbf{R}^t$ represent the concatenate vectors for u_{HD} and u_{PH}, v_{RDM}, v_{RV}, v_{PSA}, and v_{TRL}, respectively.

Given $\mathbf{A}^t$ and $\mathbf{R}^t$ as vectors of observations for the traits constituting groups I and II, respectively. $\mathbf{A}_1$ and $\mathbf{R}_1$ represents the first linear combination (first canonical pair) of the traits in groups I and II, it follows that:

$$A_1 = a_{HD} u_{HD} + a_{PH} u_{PH}, \tag{4}$$

$$R_1 = b_{RDM} v_{RDM} + b_{RV} v_{RV} + b_{PSA} v_{PSA} + b_{TRL} v_{TRL}, \tag{5}$$

where $\mathbf{a}^t = [a_{HD} a_{PH}]$ and $\mathbf{b}^t = [b_{RDM} b_{RV} b_{PSA} b_{TRL}]$ represent two-column row vectors of group I trait weights and a four-column row vector of group II trait weights, respectively.

In this manner, the first canonical correlation was defined as the one that maximized the relationship between A_1 and R_1. Thus, the functions A_1 and R_1 constitute the first associated pair of the canonical correlation, according to the expression below:

$$r = \frac{\widehat{Cov}(A_1, R_1)}{\widehat{V}(A_1)\widehat{V}(R_1)}, \tag{6}$$

$$\widehat{Cov}(A_1, R_1) = \mathbf{a}^t \mathbf{S_{12}} \mathbf{b}, \tag{7}$$

$$\widehat{V}(A_1) = \mathbf{a}^t \mathbf{S_{11}} \mathbf{a}, \tag{8}$$

$$\widehat{V}(R_2) = \mathbf{b}^t \mathbf{S_{22}} \mathbf{b}, \tag{9}$$

where $\mathbf{S_{11}}$ is the 2×2 covariance matrix between the traits of group I, $\mathbf{S_{22}}$ is the 4×4 covariance matrix between the traits of group II, and $\mathbf{S_{12}}$ is the 2×4 covariance matrix between the traits of groups I and II. Therefore, the first canonical correlation (r^1) between

the linear combinations of the traits of groups I and II is given by $r^1 = (\lambda_1)^{1/2}$, where $\lambda 1$ is the largest eigenvalue corresponding to the matrix $\mathbf{R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}}$, which is square and has asymmetric order 2. The first canonical pair is given by $A_1 = \mathbf{a^tA}$ and $R_1 = \mathbf{R^tY}$, where **a** is the associated eigenvector to the first eigenvalue of $\mathbf{R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}^{-1}}$ and **b** is the eigenvector associated with the first eigenvalue of $\mathbf{R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}}$.

The other correlations and canonical pairs are estimated using the eigenvalues and eigenvectors corresponding to the order-specific correlation estimate. The canonical correlations were carried out between group I, formed by the aerial part traits (HD and PH), and group II, composed by the root part traits (RDM, RV, PSA and TRL), according to the method proposed by Hotelling [30,31], through the statistical package "mVar.pt" [32] implemented in the R software (R Version 4.3.1) [33].

2.4. Genomic Prediction Model

The adjusted phenotypes and scores from the latent canonical variables related to aerial traits (A_1) were utilized as pseudo-phenotypes to predict the genetic merit of the individuals. The GBLUP [34] was used to estimate the additive genetic values, as expressed below:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \boldsymbol{\varepsilon}, \tag{10}$$

where **y** is the vector of latent canonical variables or adjusted phenotypes, **b** is the vector of fixed effects, u is the vector of genomic estimated breeding values of the genotypes, with $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}\sigma_u^2)$, where σ_u^2 is the additive genetic variance and G is the genomic relationship matrix given by $\mathbf{G} = \frac{\mathbf{WW^t}}{\sum_{i=1}^{n} 2p_iq_i}$ [35], where p_i and q_i are the allele frequencies of the ith marker and W is the incidence matrix for SNPs; $\boldsymbol{\varepsilon}$ is the residual effects vector, with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$, where σ_ε^2 is the residual variance and **I** represents an identity matrix. **X** and **Z** are the incidence matrices of fixed and random effects, respectively.

Genomic prediction analyzes were conducted using the "sommer" package [36] in the R software [33].

2.5. Evaluation of the Methodology

In order to assess the predictive ability (PA) of all of the fitted models, the Pearson's correlation between the estimated genetic merit and the adjusted phenotypic values on each trait individually and by latent canonical variable related to aerial traits (A_1) were calculated. A five-fold cross-validation (CV) random process was carried out. This process was repeated randomly 50 times. The data set is divided into five populations. At the k-th fold (k = 1, . . ., 5), the k-th population is used as a validation population. The remaining populations were used as a training population. For each of the five folds, the PA was estimated by Pearson's correlation coefficient between the estimated genetic merit from each evaluated model, the adjusted phenotypic values on each trait individually, and by latent canonical variable related to aerial traits (A1). Additionally, based on genetic merit, the top 10% individuals were selected for each approach used. In possession of these individuals, the agreement between the selected individuals was calculated, based on each trait individually and by the A_1, using Cohen's Kappa coefficient (K) [37]:

$$K = \frac{P_o - P_e}{1 - P_e}, \tag{11}$$

where P_o is the proportion of cases correctly classified calculated as $P_o = \frac{tp+tn}{n_T}$, P_e is the probability of agreement calculated as $P_e = \frac{tp+fn}{n_T} \times \frac{tp+fp}{n_T} \times \frac{fp+tn}{n_T} \times \frac{fn+tn}{n_T}$, where n_T is the total of individuals in the testing sets, tp denotes the true positives, tn denotes the true negatives, fp denotes the false positives, and fn denotes the false negatives. The cross-validation approach was repeated 50 times.

3. Results

3.1. Canonical Correlation Analysis

The estimated values of the Pearson correlation coefficient between the evaluated traits were significant according to a *t*-test considering 1 and 5% of significance and ranged from 0.15 to 0.96 (Figure 1).

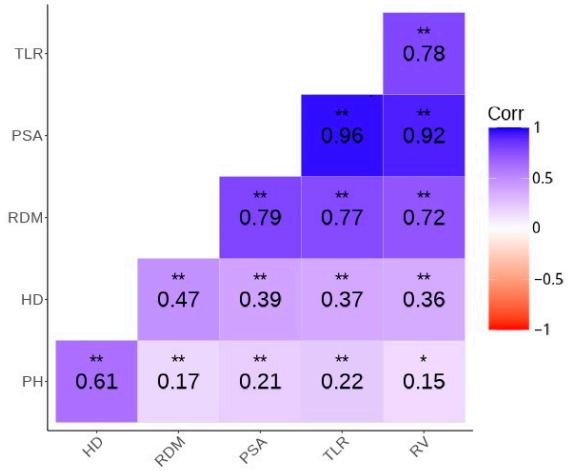

Figure 1. Pearson's correlation (Corr) between phenotypic traits. HD: Hypocotyl diameter; PH: Plant height; RDM: Root dry mass; PSA: Projected surface area; TRL: Total root length; RV: Root volume. * Statistically significant at 5% of significance. ** Statistically significant at 1% of significance. The color scale nearest to blue corresponds to positive and red to negative correlations.

The estimated canonical correlation values were equal to 0.59 ($p < 0.01$) and 0.26 ($p < 0.01$) for, respectively, the first and second canonical pairs (Table 1). A one-unit increase in HD lead to a 13.79 decrease in the first canonical variable (Table 1). The canonical coefficients from the first canonical pair showed that, for the root dry mass (RDM), a one-unit increase resulted in a 49.58 decrease in the first canonical variable when all of other variables are held constant. In a similar way, a one-unit increase in root volume led to a 77.64 increase in the second canonical variable, with the other predictors held constant (Table 1).

Table 1. Canonical correlations (r) and estimated canonical pairs (A_1, R_1 as first pairs and A_2, R_2 as second pairs) between aerial (**A**—Group I) and root (**R**—Group II) traits in commercial soybean cultivars.

Groups	Traits	Canonical Pairs	
		1°	**2°**
Aerial (A_1 and A_2)	HD	−13.79	1.89
	PH	0.39	−0.10
Root (R_1 and R_2)	RDM	−49.58	12.20
	RV	−10.28	77.64
	PSA	1.05	−8.04
	TRL	−0.03	0.29
	r	0.59 *	0.26 *
	p-value	4.80×10^{-10}	1.07×10^{-2}

* Statistically significant at 5% of significance by the Wilks test. HD: hypocotyl diameter; PH: plant height; RDM: root dry mass; RV: root volume; PSA: projected surface area; TRL: total root length.

3.2. Heritability and Prediction Accuracy

The estimates of heritability obtained with the GBLUP model, as the proportion of genetic variance divided by the total phenotypic variance, for the six evaluated traits (HD, PH, RDM, RV, PSA, and TRL) ranged from low (0.09) to high (0.66) (Table 2). The estimate of heritability for the latent variable (A_1) (0.32) was higher than for the univariate traits analysis, except for RDM (0.66) and PH (0.37) (Table 2).

Table 2. Heritability (h^2) obtained using the univariate GBLUP model for each trait and based on the latent variable obtained from canonical variable analysis.

Groups	Traits	h^2
Aerial	HD	0.14
	PH	0.37
Root	RDM	0.66
	RV	0.09
	PSA	0.19
	TRL	0.21
Latent Variable	A_1	0.32

HD: hypocotyl diameter; PH: plant height; RDM: root dry mass; RV: root volume; PSA: projected surface area; TRL: total root length; A_1: Latent Variable obtained for the aerial traits group.

The estimated predictive ability (PA) ranged from 0.47 to 0.73 (Table 3). For these traits, the highest accuracy values were 0.73 and 0.61, observed for PH and HD, respectively (Table 3). The PA obtained for A1 (0.57) was higher than those estimated for the traits related to Root (RV, PSA and TRL), except for RDM (0.59). In relation to aerial traits, the PA of A_1 was lower in comparison with PH and close to the HD trait (0.61) (Table 3). The low standard deviation associated with PA revealed the good precision of the estimates of PA.

Table 3. Predictive ability (PA), standard error of predictive ability SE(PA), mean Cohen's Kappa coefficient (K), and standard error of Kappa coefficient of Cohen SE(K) of genomic prediction using univariate GBLUP in each trait and based on the latent variable obtained by canonical correlation.

Groups	Traits	PA	SE(PA)	K	SE(K)
Aerial	HD	0.61	0.0219	0.29	0.0401
	PH	0.73	0.0153	0.52	0.0291
Root	RDM	0.59	0.0208	0.44	0.0297
	RV	0.55	0.0271	0.18	0.0431
	PSA	0.47	0.0272	0.25	0.0415
	TRL	0.51	0.0231	0.24	0.0397
Latent variable	A_1	0.57	0.0246	0.35	0.0305

HD: hypocotyl diameter; PH: plant height; RDM: root dry mass; RV: root volume; PSA: projected surface area; TRL: total root length. A_1: Latent Variable obtained for the aerial traits group.

The Cohen's Kappa coefficient (K) represents the agreement between the top 10% of individuals selected based on each trait individually, compared with the those selected based on A_1. The estimated Cohen's Kappa coefficient ranged from 0.32 to 0.88 (Table 4). Specifically, the Cohen's Kappa values between A_1 and the root traits were equal to 0.50 (A_1 versus PSA), 0.49 (A_1 versus TLR), 0.69 (A_1 versus RDM), and 0.46 (A_1 versus RV) (Table 4). The associated low standard error of the mean leads to good precision of the calculated values.

Table 4. Estimates of the Cohen's Kappa coefficients (lower triangular) for the evaluated traits and the latent variable obtained for the aerial traits group (upper triangular) The value in parentheses is the standard error of the mean considering 50 repetitions.

Trait	PSA	HD	TLR	PH	RDM	RV	A_1
PSA	1	-	-	-	-	-	-
HD	0.54 (0.03)	1	-	-	-	-	-
TLR	0.88 (0.03)	0.54 (0.03)	1	-	-	-	-
PH	0.40 (0.02)	0.51 (0.02)	0.52 (0.02)	1	-	-	-
RDM	0.63 (0.03)	0.60 (0.02)	0.60 (0.03)	0.41 (0.02)	1	-	-
RV	0.79 (0.03)	0.51 (0.02)	0.72 (0.04)	0.46 (0.02)	0.56 (0.024)	1	-
A_1	0.50 (0.02)	0.70 (0.03)	0.49 (0.02)	0.32 (0.02)	0.69 (0.01)	0.46 (0.02)	1

HD: hypocotyl diameter; PH: plant height; RDM: root dry mass; RV: root volume; PSA: projected surface area; TRL: total root length. A_1: Latent Variable obtained for the aerial traits group.

4. Discussion

This study introduces a novel approach that combines canonical correlation analysis and genomic prediction. The idea is to use a latent variable given by a linear combination of aerial traits to predict the genetic merit of genotypes for root traits, which requires destructive, laborious, and expensive evaluation. The joint phenotyping of aerial and root traits was used to evaluate predictive ability of our proposal through a cross-validation approach.

Some studies employ selection indexes considering several methods to identify promising genotypes for interested traits in soybeans [38,39], including associating these with genomic prediction approaches [40]. However, selection indexes do not allow us to explore the correlations between traits to predict genotypic values for unobserved phenotypes. To circumvent this bottleneck, multi-trait models could be used to improve predictive ability accounting for covariance between traits [21]. Nevertheless, these models do not eliminate the need to gather the trait information into an index for select genotypes [41]. The proposal in this study could be a manner to explore the advantages of both methods in a simpler way.

The high correlations observed between the root traits were similar to those observed by Dayoub et al. [42], except for TLR and PSA, which the study did not observe a significant correlation. Overall, the Pearson correlation between aerial and root traits was low but statistically significant (Figure 1). These results suggest the possibility of using an indirect selection approach, despite the low magnitude of Pearson correlation showing that effective results might not be achieved.

On the hand, the magnitudes of the canonical correlation coefficients were high (0.59) and moderate (0.26) for the first and second pair of canonical variables, respectively. The canonical variables are constructed with the aim to maximize the correlation between the two sets of traits [43]. Thus, our first pair of canonical variables (latent variables) maximizes the association between the aerial and root trait sets. In a practical context, it suggests that the canonical variable referent to aerial traits (A_1) may be used as a pseudo-phenotype to perform an indirect selection to root traits.

The use of latent variables has been used with success in plant breeding. For example, Paixão et al. [44] used factor analysis (FA) to elucidate the structure of relations of the evaluated traits, form new variables, and use them, as pseudo-phenotypes, to predict the genetic merit of canephora coffee individuals. On the other hand, principal component analysis (PCA) can be employed to extract latent variables, effectively reducing dimensionality while enabling the assessment of genetic diversity [45,46]. The crucial distinction between these approaches lies in their interpretation (construction) of latent variables.

While canonical correlation analysis seeks to maximize the shared information between two sets of traits, both factor analysis (FA) and principal component analysis (PCA) focus on explaining internal structure. FA aims to capture the underlying "factors" influencing covariance (trait associations), while PCA prioritizes explaining the bulk of variability (individual trait variances).

The estimated heritability of the latent variable (A_1) was higher than those for RV, PSA and TRL indicating that selection based on A_1 may be a promising strategy. Specifically, the lowest estimated heritability, that is, RV = 0.09, was similar to that observed by Getnet [47]. The estimated heritabilities for RDM (0.66), HD (0.14), PH (0.37), PSA (0.19), TRL (0.21), and PSA (0.19) were similar to those presented in Xavier et al. [48], Yan et al. [49], and Conte et al. [50].

Overall, the PA for A_1 (0.57 ± 0.02) was equal or higher than those obtained for all root traits (RDM = 0.59 ± 0.02; RV = 0.55 ± 0.03; PSA = 0.47 ± 0.03; TRL = 0.51 ± 0.02). Regarding the aerial traits, the PA for A_1 was similar to (HD = 0.61 ± 0.02) or lower (0.73 ± 0.02) than those obtained by individual analysis. These results demonstrate that it is possible to use the latent variable A_1 to predict the genetic merit of soybean genotypes. The PA for some soybean traits was reported by Bandillo et al. [51], corroborating with the results obtained in our study.

To assess the agreement between A_1 and individual root traits, we estimated Cohen's Kappa coefficients using the top 10% of GEBVs. According to Landis and Koch [52], A_1 exhibited moderate agreement with PSA (0.50) and TRL (0.49) and substantial agreement with RDM (0.69). While agreement with RV was slightly lower (0.46), these findings collectively suggest A_1's potential as a promising pseudo-phenotype for root trait evaluation in practical applications.

Overall, this study demonstrates the effectiveness of combining the latent variable A_1 with genomic selection to predict individual genetic merit for root traits in soybean breeding. This is particularly interesting because root trait evaluation is challenging and expensive, often hindering selection for improved genotypes. As Crossa et al. [53] emphasize, genomic prediction shines in such scenarios, enabling the efficient culling of inferior genotypes and reducing phenotyping costs. Our findings pave the way for utilizing A_1 as a powerful tool to accelerate development of superior soybean varieties with enhanced root systems.

5. Conclusions

The approach provides a way to predict and select promising genotypes considering multiple traits simultaneously, differing from the use of selection indexes by exploiting a latent variable that maximizes correlation between groups of traits. Also, it differs from multi-trait approaches by gathering all traits' information into a single variable, making the selection process simpler.

The combination of canonical correlation analysis with genomic selection was efficient to predict individual genetic merit for root traits (length, volume, surface area, and dry mass) in soybean breeding. The agreement observed between the top 10% individuals selected based on the canonical variable and each root traits individually was considered moderate or substantial.

Author Contributions: Conceptualization, V.S.S., M.M.S., A.C.C.N. and M.N.; Data curation, D.d.O.F. and F.L.d.S.; Formal analysis, V.S.S. and M.M.S.; Investigation, N.M.E.P.d.L.d.C., D.d.O.F., A.C.C.N., C.F.A. and F.L.d.S.; Methodology, V.S.S., M.M.S. and M.N.; Software, V.S.S. and N.M.E.P.d.L.d.C.; Supervision, A.C.C.N., Camila Azevedo, F.L.d.S. and M.N.; Validation, V.S.S., M.M.S., A.C.C.N. and M.N.; Visualization, N.M.E.P.d.L.d.C.; Writing—original draft, V.S.S., N.M.E.P.d.L.d.C., M.M.S. and M.N.; Writing—review and editing, A.C.C.N., C.F.A., F.L.d.S. and M.N. All authors have read and agreed to the published version of the manuscript.

Funding: We would like to thank the Foundation for Research Support of the state of Minas Gerais (FAPEMIG, APQ-01638-18), by the National Council of Scientific and Technological Development (CNPq, 408833/2023-8), and by the National Institutes of Science and Technology of Coffee (INCT/Café). MN and CFA are supported by scientific productivity (310755/2023-9 and 306772/2020-5), respectively, from Brazilian Council for Scientific and Technological Development (CNPq).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from one of the authors, Dalton de Oliveira Ferreira, upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Borém, A.; Miranda, G.V.; Fritsche-Neto, R. *Melhoramento de Plantas*, 7th ed.; Editora UFV: Viçosa, Brazil, 2017.
2. Silva, A.F.; Sediyama, T.; Silva, F.C.S.; Bezerra, A.R.G.; Ferreira, L.V. Correlation and Path Analysis of Yield Components in Soybean Varieties. *Turk. J. Field Crops* **2015**, *10*, 177–179. [CrossRef]
3. Shea, Z.; Singer, W.M.; Zhang, B. Soybean Production, Versatility, and Improvement. In *Legume Crops*; Hasanuzzaman, M., Ed.; Intechopen: London, UK, 2020; Volume 1, pp. 1–22. [CrossRef]
4. Ponnusha, B.S.; Subramaniyam, S.; Pasupathi, P.; Subramaniyam, B.; Virumandy, R. Antioxidant and Antimicrobial Properties of Glycine Max-A Review. *Int. J. Curr. Biol. Med. Sci.* **2011**, *1*, 49–62.
5. Daryanto, S.; Wang, L.; Jacinthe, P.A. Global Synthesis of Drought Effects on Food Legume Production. *PLoS ONE* **2015**, *10*, e0127401. [CrossRef] [PubMed]
6. Polania, J.A.; Poschenrieder, C.; Beebe, S.; Rao, I.M. Effective Use of Water and Increased Dry Matter Partitioned to Grain Contribute to Yield of Common Bean Improved for Drought Resistance. *Front. Plant Sci.* **2016**, *7*, 660. [CrossRef] [PubMed]
7. Waraich, E.A.; Ahmad, R.; Ashraf, M.Y. Role of Mineral Nutrition in Alleviation of Drought Stress in Plants. *Aust. J. Crop Sci.* **2011**, *5*, 764–777.
8. Fenta, B.A.; Beebe, S.E.; Kunert, K.J.; Burridge, J.D.; Barlow, K.M.; Lynch, J.P.; Foyer, C.H. Field Phenotyping of Soybean Roots for Drought Stress Tolerance. *Agronomy* **2014**, *4*, 418–435. [CrossRef]
9. Bucksch, A.; Burridge, J.; York, L.M.; Das, A.; Nord, E.; Weitz, J.S.; Lynch, J.P. Image-Based High-Throughput Field Phenotyping of Crop Roots. *Plant Physiol.* **2014**, *166*, 470–486. [CrossRef]
10. Falk, K.G.; Jubery, T.Z.; Mirnezami, S.V.; Parmley, K.A.; Sarkar, S.; Singh, A.; Ganapathysubramanian, B.; Singh, A.K. Computer Vision and Machine Learning Enabled Soybean Root Phenotyping Pipeline. *Plant Methods* **2020**, *16*, 5. [CrossRef]
11. Andrade, L.R.B.d.; Sousa, M.B.; Oliveira, E.J.; Resende, M.D.V.; Azevedo, C.F. Cassava Yield Traits Predicted by Genomic Selection Methods. *PLoS ONE* **2019**, *14*, e0224920. [CrossRef]
12. Heffner, E.L.; Jannink, J.L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Sci.* **2011**, *51*, 2597–2606. [CrossRef]
13. Hotelling, H. Relations Between Two Sets of Variates. *Biometrika* **1936**, *28*, 321–377. [CrossRef]
14. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef] [PubMed]
15. Voss-Fels, K.P.; Cooper, M.; Hayes, B.J. Accelerating Crop Genetic Gains with Genomic Selection. *Theor. Appl. Genet.* **2019**, *132*, 669–686. [CrossRef] [PubMed]
16. Hemingway, J.; Schnebly, S.R.; Rajcan, I. Accuracy of genomic prediction for seed oil concentration in high-oleic soybean populations using a low-density marker panel. *Crop Sci.* **2021**, *61*, 4012–4021. [CrossRef]
17. Jia, Y.; Jannink, J.L. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* **2012**, *192*, 1513–1522. [CrossRef] [PubMed]
18. Persa, R.; Bernardeli, A.; Jarquin, D. Prediction Strategies for Leveraging Information of Associated Traits under Single- and Multi-Trait Approaches in Soybeans. *Agriculture* **2020**, *10*, 308. [CrossRef]
19. Hayashi, T.; Iwata, H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinform.* **2013**, *14*, 34. [CrossRef] [PubMed]
20. Cheng, H.; Kizilkaya, K.; Zeng, J.; Garrick, D.; Fernando, R. Genomic Prediction from Multiple-Trait Bayesian Regression Methods Using Mixture Priors. *Genetics* **2018**, *209*, 89–103. [CrossRef] [PubMed]
21. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Gianola, D.; Hernández-Suárez, C.M.; Martín-Vallejo, J. Multi-Trait, Multi-Environment Deep Learning Modeling for Genomic-Enabled Prediction of Plant Traits. *G3 Genes Genomes Genet.* **2018**, *8*, 3829–3840. [CrossRef]
22. Saba, M.; Aaron, K.; Hu, G.; Wang, L.; Patrick, S.S. Multi-trait Genomic Selection Methods for Crop Improvement. *Genetics* **2020**, *215*, 931–945. [CrossRef]

23. Apresentação do Município de Viçosa. Available online: https://www.vicosa.mg.gov.br/abrir_arquivo.aspx/Anexo_I__Apresentacao_Vicosa?cdLocal=2&arquivo=%7BC1D6CDDA-DDE4-5D26-DEA7-CE57C00D1CB7%7D.pdf (accessed on 2 January 2024).
24. Nascimento, H.R.d.; Oliveira, L.M.; Duarte, A.B.; Dantas, S.A.G.; Ferreira, D.d.O.; Rosmaninho, L.B.d.C.; Cavallin, I.C.; da Cunha, F.F.; da Silva, F.L. A New Methodological Approach for Simulating Water Deficit in Soybean Genotypes. *J. Agron. Crop Sci.* **2021**, *207*, 946–955. [CrossRef]
25. Bernardo, S.; Mantovani, E.C.; Silva, D.D.; Soares, A.A. *Manual de Irrigação*, 9th ed.; Editora UFV: Viçosa, Brazil, 2014.
26. Fehr, W.R.; Caviness, C.E. *Stages of Soybean Development*; Iowa State University: Ames, IA, USA, 1977.
27. Silva, F.; Borém, A.; Sediyama, T.; Câmara, G. *Soja: Do Plantio à Colheita*, 2nd ed.; Oficina de Textos: São Paulo, Brazil, 2022.
28. WinRHIZO 2021. Available online: https://regentinstruments.com/assets/images_winrhizo/WinRHIZO_2021.pdf (accessed on 2 January 2024).
29. Resende, M.D.V. *Genética Biométrica e Estatística No Melhoramento de Plantas Perenes*; Embrapa, Informação Tecnológica: Brasília, Brazil, 2002; 975 p.
30. Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
31. Hotelling, H. Simplified Calculation of Principal Components. *Psychometrika* **1936**, *1*, 27–35. [CrossRef]
32. Ossani, C.; Cirillo, M.A.; Paulo, M.; Ossani, C. Package 'Mvar.Pt'. 2023. Available online: https://cran.r-project.org/web/packages/MVar.pt/index.html (accessed on 2 January 2024).
33. R Core Team. *R: A Language and Environment for Statistical Computing*, R Version 4.3.1; R Foundation for Statistical Computing: Vienna, Austria, 2023. Available online: https://www.R-project.org/ (accessed on 2 January 2024).
34. Xu, S. Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* **2013**, *195*, 1209–1222. [CrossRef] [PubMed]
35. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
36. Covarrubias-Pazaran, G. Genome-Assisted Prediction of Quantitative Traits Using the r Package Sommer. *PLoS ONE* **2016**, *11*, e0156744. [CrossRef] [PubMed]
37. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
38. Leite WD, S.; Unêda-Trevisoli, S.H.; Silva FM, D.; Silva AJ, D.; Mauro, A.O.D. Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Rev. Ciência Agronômica* **2018**, *49*, 491–500. [CrossRef]
39. Woyann, L.G.; Meira, D.; Matei, G.; Zdziarski, A.D.; Dallacorte, L.V.; Madella, L.A.; Benin, G. Selection indexes based on linear-bilinear models applied to soybean breeding. *Agron. J.* **2020**, *112*, 175–182. [CrossRef]
40. Beche, E.; Gillman, J.D.; Song, Q.; Nelson, R.; Beissinger, T.; Decker, J.; Shannon, G.; Scaboo, A.M. Genomic prediction using training population design in interspecific soybean populations. *Mol. Breed.* **2021**, *41*, 15. [CrossRef]
41. Khan, M.A.; Rai, A.; Mishra, D.C.; Budhlakoti, N.; Satpathy, S.; Majumdar, S.G. Comparative study of multi-trait genomic and phenotypic selection indexes for selection of superior genotypes. *Indian J. Genet. Plant Breed.* **2023**, *83*, 88–94.
42. Dayoub, E.; Lamichhane, J.R.; Schoving, C.; Debaeke, P.; Maury, P. Early-Stage Phenotyping of Root Traits Provides Insights into the Drought Tolerance Level of Soybean Cultivars. *Agronomy* **2021**, *11*, 188. [CrossRef]
43. Ferreira, D.F. *Estatística Multivariada*, 1st ed.; Editora UFLA: Lavras, Brazil, 2008.
44. Paixão, P.T.M.; Nascimento, A.C.C.; Nascimento, M.; Azevedo, C.F.; Oliveira, G.F.; da Silva, F.L.; Caixeta, E.T. Factor Analysis Applied in Genomic Selection Studies in the Breeding of Coffea Canephora. *Euphytica* **2022**, *218*, 42. [CrossRef] [PubMed]
45. De Ron, A.M.; Rodiño, A.P. Analysis of the Genetic Diversity of Crops and Associated Microbiota. *Agronomy* **2023**, *13*, 2132. [CrossRef]
46. Karim, K.M.R.; Rafii, M.Y.; Misran, A.; Ismail, M.F.; Harun, A.R.; Ridzuan, R.; Chowdhury, M.F.N.; Hosen, M.; Yusuff, O.; Haque, M.A. Genetic Diversity Analysis among *Capsicum annuum* Mutants Based on Morpho-Physiological and Yield Traits. *Agronomy* **2022**, *12*, 2436. [CrossRef]
47. Getnet, B.A. Genetic variability, heritability and expected genetic advance in soybean [*Glycine max* (L.) Merrill] genotypes. *Agric. For. Fish. J.* **2018**, *7*, 108–112. [CrossRef]
48. Xavier, A.; Muir, W.M.; Rainey, K.M. Assessing Predictive Properties of Genome-Wide Selection in Soybeans. *G3 Genes Genomes Genet.* **2016**, *6*, 2611–2616. [CrossRef] [PubMed]
49. Yan, C.; Song, S.; Wang, W.; Wang, C.; Li, H.; Wang, F.; Li, S.; Sun, X. Screening diverse soybean genotypes for drought tolerance by membership function value based on multiple traits and drought-tolerant coefficient of yield. *BMC Plant Biol.* **2020**, *20*, 321. [CrossRef]
50. Conte, M.V.D.; Carneiro, P.C.S.; Resende, M.D.V.; Silva, F.L.; Peternelli, L.A. Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. *PLoS ONE* **2020**, *15*, e0233290. [CrossRef]
51. Bandillo, N.B.; Jarquin, D.; Posadas, L.G.; Lorenz, A.J.; Graef, G.L. Genomic Selection Performs as Effectively as Phenotypic Selection for Increasing Seed Yield in Soybean. *Plant Genome* **2023**, *16*, e20285. [CrossRef]

52. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]
53. Crossa, J.; Pérez, P.; Hickey, J.; Burgueño, J.; Ornella, L.; Cerón-Rojas, J.; Zhang, X.; Dreisigacker, S.; Babu, R.; Li, L.; et al. Genomic Prediction in CIMMYT Maize and Wheat Breeding Programs. *Heredity* **2014**, *112*, 48–60. [CrossRef] [PubMed]