



RESEARCH ARTICLE

Comparison of traditional selection indices and the Multi-Trait Genotype Ideotype Distance Index (MGIDI) for improving rice (*Oryza sativa* L.) cultivars

Vinoth Kumar Paramasivam¹, Manonmani Swaminathan², Raveendran Muthurajan^{3,4}, Sritharan Natarajan², Sudha Manickam⁴, Puja Mandal¹, Bonipas Antony¹ & Babu Chakrapani^{1,3*}

¹Department of Genetics and Plant Breeding, Centre of Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, India

²Department of Rice, Centre of Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, India

³Directorate of Research, Tamil Nadu Agricultural University, Coimbatore 641 003, India

⁴Department of Plant Biotechnology, Centre for Plant Molecular Biology & Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641 003, India

*Email: babu.c@tnau.ac.in



ARTICLE HISTORY

Received: 03 October 2024

Accepted: 21 October 2024

Available online

Version 1.0 : 11 January 2025

Version 2.0 : 26 January 2025



Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). 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 author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

CITE THIS ARTICLE

Paramasivam VK, Swaminathan M, Muthurajan R, Natarajan S, Manickam S, Mandal P, Antony B, Chakrapani B. Comparison of traditional selection indices and the Multi-Trait Genotype Ideotype Distance Index (MGIDI) for improving rice (*Oryza sativa* L.) cultivars. Plant Science Today.2025;12(sp1):-1-09.
<https://doi.org/10.14719/pst.5475>

Abstract

Rice is a vital cereal crop that is crucial to ensuring global food security. Developing high-yield varieties is essential to feed the growing world population. Selecting the best-performing genotypes based on genetic gain is critical for improving rice production. In this process, various selection indices help identify superior genotypes. Thus, the current study aimed to select superior genotypes through a multi-trait selection index. Among the multivariate indices, the widely used classical Smith-Hazel selection index (SI) and the multi-trait genotype ideotype index (MGIDI) index were used to evaluate the studied genotypes for eleven quantitative traits simultaneously. Three scenarios were formulated namely, retaining multi-collinearity (SI-1), removing multi-collinearity (SI-2) and using a path analysis-based index (SI-3) for estimating the Smith-Hazel selection index. Using these indices, maximum selection gains of 9.2% with the Smith-Hazel index and 18.5% with the MGIDI index for grain yield were predicted. Notably, genotypes K5 and H7 were selected by the Smith (SI-1, SI-2, SI-3) and MGIDI selection index at a selection intensity of 15%. This shows that these genotypes exhibited strong performance across various traits and were classified as elite, highlighting their potential to contribute significantly to future breeding programs aimed at improving grain yield along with multiple traits in rice. The selection of these genotypes makes a valuable resource for developing productive rice varieties.

Keywords

selection index; Smith Hazel; MGIDI; multicollinearity; selection gains

Introduction

Rice is essential for global nutrition, providing a major source of calories for over half of the world's population (1). Beyond its role in food security, rice contributes significantly to economic growth, job security and maintaining social stability in many parts of the world (2). Due to the rapid expansion of the global population and growing urbanization, the need for rice is escalating highlighting the urgent need for sustainable and improved production strategies to ensure future food security. Genetic improvement has played a key role in enhancing the grain yield. However, achieving greater grain yields depends on numerous yield-contributing factors such as the number of productive tillers per plant, the number of filled grains per panicle, thousand-grain weight and so on. Breeders play a critical role in this process by designing the ideotypes as defined by(3),which means optimal

plant types exhibiting the ideal traits for fulfilling specific breeding objectives. By developing rice with ideal characteristics like early maturity, reduced plant height, high number of productive tillers and more number of grains per panicle will ensure that crops meet present and future agricultural demands.

Selection indices play a crucial role in this context. These tools help breeders to develop high-yielding genotypes based on several yield and its component traits. Simultaneous selection of multiple traits improves the efficiency of breeding programs as depending alone on yield may result in the selection of genotypes with other undesirable traits. Additionally, yield is highly influenced by environmental factors. Historically, the Smith index has been used to select genotypes based on multi-trait performance and it relies on the phenotypic and genotypic variance and covariance matrices (4). However, this index has faced many limitations, such as multicollinearity when assessing multiple traits, which affects the accuracy of genetic gain estimation. However, in addition to this, allocating the economic weights for different traits leads to difficulties for the breeders in computing the Smith-Hazel (SH) index (5,6). By removing multi-collinearity and using path coefficients these limitations may be addressed. However, a comparison of these methods with the classical index is needed.

To address these issues, Olivoto and Nardino (7) introduced a novel multivariate selection approach called Multi Trait Genotype - Ideotype Distance Index (MGIDI), which avoids problems such as multi-collinearity and weightage. Further, the MGIDI index is more effective in identifying the best-performing genotypes compared to the traditional Smith Index, particularly in addressing challenges like the inversion of phenotypic variance and multicollinearity issues, which affect genetic gain estimation. Unlike the Smith Index, the MGIDI index not only tackles multicollinearity issues but also overcomes the difficulties faced by breeders in deciding the genetic weightage for traits (8). Apart from this, MGIDI can also assess the genotypes based on the strength and

weaknesses plot and evaluate all the variables under study. Thus, in the current study, both the Smith Index and MGIDI index were applied to compare their efficiency in selecting the genotype.

Materials and Methods

The current experiment involved 43 genotypes, including 21 stabilized lines from the F₆ generation of the cross, CO51 × Rp Bio (Kasalath) and 10 stabilized lines from the F₅ generation of the cross, CO51 × Habataki along with commercial cultivars and landraces were listed in (Table1). These lines were evaluated using a randomized complete block design (RCBD) with three replications at the Department of Rice, Tamil Nadu Agricultural University, Coimbatore.

Biometrical Traits Evaluated

The traits under study included days to fifty percent flowering (DFF), plant height (PHT), flag leaf length (FLL), flag leaf width (FLW), the number of productive tillers per plant (NPTP), panicle length (PL), number of grains per panicle (NGP), number of filled grains per panicle (NFGP), thousand-grain weight (TGW), grain length-width ratio (GLW) and single plant yield (SPY). These traits were observed on three randomly chosen plants per replication.

Statistical Analysis

Smith Indices

The Smith-Hazel index (SH) is used to select the genotypes based on their high genetic value, where genotypes are ranked based on their genetic values. The index is calculated by using the following formula

$$I_i = \sum_k b_k \bar{y}_{ik}$$

where, I_i - index value of progeny i ; b_k - the weight or Regression coefficient of trait k ; and $\bar{y}_{ik}$ - average value of trait k for progeny i .

The values of b_k are calculated using the following formula

$$b_k = A^{-1}C \times D,$$

Table 1. Summary of 43 genotypes along with their parentage and commercial checks used in the research

S. No	Genotypes	Parentage	S. No	Genotypes	Parentage
1	K1	CO51 × Rp Bio	22	H1	CO51 × Habataki
2	K2	CO51 × Rp Bio	23	H2	CO51 × Habataki
3	K3	CO51 × Rp Bio	24	H3	CO51 × Habataki
4	K4	CO51 × Rp Bio	25	H4	CO51 × Habataki
5	K5	CO51 × Rp Bio	26	H5	CO51 × Habataki
6	K6	CO51 × Rp Bio	27	H6	CO51 × Habataki
7	K7	CO51 × Rp Bio	28	H7	CO51 × Habataki
8	K8	CO51 × Rp Bio	29	H8	CO51 × Habataki
9	K9	CO51 × Rp Bio	30	H9	CO51 × Habataki
10	K10	CO51 × Rp Bio	31	H10	CO51 × Habataki
11	K11	CO51 × Rp Bio	32	CO51	ADT43 × RR 272-1745
12	K12	CO51 × Rp Bio	33	RpBio (Kasalath)	Land Race
13	K13	CO51 × Rp Bio	34	Habataki	Land Race
14	K14	CO51 × Rp Bio	35	NootriPathu	Land Race
15	K15	CO51 × Rp Bio	36	Karungan	Land Race
16	K16	CO51 × Rp Bio	37	Thooyamalli	Land Race
17	K17	CO51 × Rp Bio	38	Mapillai Samba	Land Race
18	K18	CO51 × Rp Bio	39	Karthi Samba	Land Race
19	K19	CO51 × Rp Bio	40	Purple Puttu	Land Race
20	K20	CO51 × Rp Bio	41	IR 64	IR 5657-33-2-1 × IR 2061-465-1-5-3
21	K21	CO51 × Rp Bio	42	ADT 46	ADT38 × CO 45
			43	BPT 5204	GEB-24 × T(N)1 × Mahsuri

K, CO 51 × Rp Bio (Kasalath) derivatives; H, CO 51 × Habataki derivatives

where A^{-1} - inverse of the mean phenotypic variance-covariance matrix between traits,

C - matrix of genotypic variances and co-variances averaged across progenies for the traits

D - vector of economic weights for the traits. (4,9)

Three methods were employed in the current study by altering the genetic weights for estimating the Smith-Hazel index.

Method 1: The Smith Index 1 / Selection Index 1 (SI-1) was calculated by allocating equal genetic weights without considering the multicollinearity issues. Multicollinearity may reduce the effectiveness of selection and it can be identified by using the code “cold2 <- colindia(Data [...])”, which tells the Variance Inflation Factor (VIF) and Condition number. If a trait has a VIF value higher than 10, it indicates multicollinearity (10), a Condition number above 10 suggests multicollinearity, with values between 10 and 30 showing moderate multicollinearity and values above 30 indicating severe multicollinearity (11).

Method 2: In Smith Index 2 / Selection Index 2 (SI-2), by using the function “non-collinear vars()” traits that were suspected to show multicollinearity issues are removed and proceed with the remaining traits by allocating equal weightage to them.

Method 3: While, in Smith Index 3 / Selection Index 3 (SI-3), the genetic weights were differentially allotted to traits based on the path analysis. The weights are given as 2 for traits that show high direct effects, 1 for traits that show moderate direct effects, 0.5 for low direct effects and -0.5 for traits that show negative direct effects and 2 is given for single plant yield trait (SPY) (12)

MGIDI Index

The MGIDI index was introduced by the scientists, Olivoto and Nardino, it operates through four key steps: adjusting trait values, followed by factorial analysis to reduce the number of dimensions, creating an ideal plant model and then calculating the Euclidean distance between each genotype to the ideal model. In the first step, the rescaling of the variables is done in such a way that all the variables fall within the range of 0 -100. The rescaling is done according to the following equation (7)

$$rX_{ij} = \frac{N_{nj} - \phi_{nj}}{N_{oj} - \phi_{oj}} \times (\theta_{ij} - N_{oj}) + N_{nj}$$

where N_{oj} and ϕ_{ij} are the original highest and lowest values for a specific trait j , while N_{nj} and ϕ_{nj} are the new highest and lowest values after adjusting the scale. θ_{ij} represents the original value for the j^{th} trait of the i^{th} genotype. For DFF and PHT, where lower values are desirable, $N_{nj} = 0$ and $\phi_{nj} = 100$ are used. For other traits where higher numbers are better, the scale is flipped, with $N_{nj} = 100$ and $\phi_{nj} = 0$. As a result, the ideal genotype would have a value of 100 for all traits after rescaling, except for DFF and PHT.

Following this, exploratory factor analysis was conducted on the rescaled traits (rX_{ij}) to categorize related traits and reduce data dimensionality. This analysis generated factor loadings for each genotype, described by the equation:

$$X = \mu + Lf + \varepsilon$$

In this formula, X represents a $1p \times 1$ vector of rescaled data, μ is a $1p \times 1$ vector of standardized means, L is a $1p \times 1$ vector of common factors and ε is a $1p \times 1$ vector of errors. The variables p

and f refer to the number of traits chosen and the number of shared factors, respectively. To compute the factor scores, eigenvalues and eigenvectors are first derived from the correlation matrix of rX_{ij} . Only eigenvalues greater than 1 are kept and these are then used to calculate the factor scores.

In this context, F is a matrix displaying the factor scores, with dimensions $g \times f$. Z is a $g \times f$ matrix of standardized means, A is a matrix of canonical loadings sized $p \times f$ and R is a correlation matrix of the traits, sized $p \times p$. Here, g represents the number of genotypes, f indicates the number of factors retained and p refers to the number of traits being analyzed. After this, the Euclidean distance between the scores of the genotypes and the ideal genotypes is calculated using the MGIDI index with the following formula

$$MGIDI_i = \left[\sum_{j=1}^f (y_{ij} - y_j)^2 \right]^{0.5}$$

In this equation, y_{ij} represents the score of the i^{th} genotype on the j^{th} factor, where i ranges from 1 to g (the total number of genotypes) and j ranges from 1 to f (the total number of factors). y_j is the score for the j^{th} ideal genotype. The genotype with the lowest MGIDI value is the closest to the ideal type, signifying that it has the most favorable traits overall. The selection differential was determined for all traits with a selection intensity of 15%, resulting in the selection of genotypes with lower MGIDI values, meaning those that are nearer to the ideal genotype.

$$\omega_{ij} = \frac{\sqrt{D_{ij}^2}}{\sum_{i=1}^f \sqrt{D_{ij}^2}}$$

Where, D_{ij} denotes the distance of the i^{th} genotype from the j^{th} factor

Coincidence Index (CI)

The coincidence Index is a term used to measure how often the accessions/traits got selected by different selection indices which overlap or align with each other. In our study also, we calculated the coincidence index calculated by using the formula (13)

$$CI = \frac{A - C}{M - C} \times 100$$

where A - No of genotypes common in several selection indices, M - Total number of genotypes selected and C - Number of genotypes included due to random chance. i.e., selection intensity 15% ($M \times 0.15$).

Software used

All the statistical analysis was performed using the “metan” package (14) in R-Studio software version 2024.04.02.

Results and Discussion

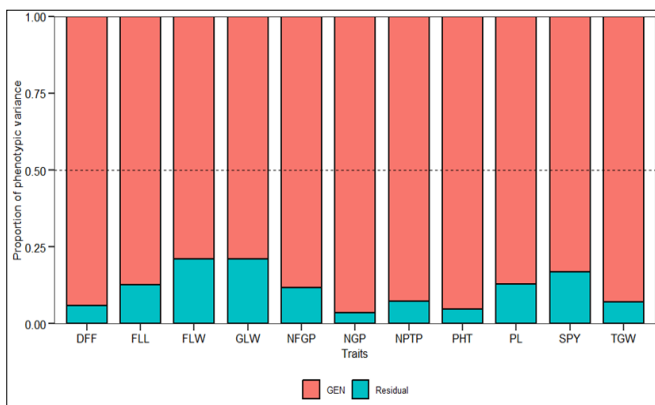
The traits studied for all genotypes were found to be significant, indicating true variation between the lines. This suggests that there is some potentiality for the breeder to make selection to identify the best genotypes. Heritability ranged from 96.4% for NGP to 78.9% for FLW. The selection accuracy was high for all the eleven traits under study, which shows that there are very less, marginal errors in the selection of genotypes based on the traits.

Table 2. Variance components for all the 11 traits

Parameters	DFF	PHT	FLL	FLW	NPTP	PL	NGP	NFGP	TGW	GLW	SPY
GV	51	248	47.5	0.0405	112	17.8	1349	465	41.6	0.253	45.8
G(%)	94.3	95.3	87.5	78.9	92.4	87.4	96.4	88.5	93.2	79.1	84.3
Rv	3.11	12.3	6.79	0.0108	9.17	2.57	50.2	60.5	3.05	0.0665	8.51
R(%)	5.75	4.72	12.5	21.1	7.57	12.6	3.59	11.5	6.83	20.9	15.7
Pv	54.1	261	54.3	0.0513	121	20.4	1400	525	44.7	0.319	54.3
h ² (%)	94	95	88	79	92	87	96	89	93	79	84
Acc (%)	99.0	99.2	97.7	95.8	98.7	97.7	99.4	97.9	98.8	95.9	97.0
Cvg	7.5	16.9	25.3	16.8	59.9	18.6	25	22.7	31.3	16	20.1
Cvr	1.85	3.76	9.55	8.68	17.1	7.06	4.83	8.18	8.49	8.2	8.66
CVR	4.05	4.49	2.64	1.93	3.5	2.63	5.18	2.77	3.69	1.95	2.32

DFF. Days to fifty percent flowering, PHT. Plant height, FLL. Flag leaf length, FLW. Flag leaf width, NPTP. Number of productive tillers per plant, PL. Panicle length, NGP. Number of grains per Panicle, NFGP. Number of filled grains per panicle, TGW. Thousand grain weight, GLW. Grain length width ratio, SPY. Single plant yield. Gv - Genetic Variance, G(%) - Genetic variance in percentage, Rv - Residual Variance, R(%) - Residual Variance (%), Pv. Phenotypic Variance, h²(%) - Heritability in percent, Acc - Accuracy of selection, Cvg.- Genotypic coefficient of variation, Cvr.- Residual coefficient of variation, CVR. The ratio between genotypic and residual coefficient of variation.

As noted by Resende (15), moderate to high accuracy values were observed, indicating reliable identification of superior traits. NGP was reported with a high accuracy of 99.4% and yield showed 97.0 % accuracy. The variance components for the biometric traits are mentioned in (Table 2). The proportion of variation contributed by the genotypes and residuals (both block and error) is shown in (Fig. 1). The high proportion of variance contributed by the genotypes, compared to the residual variance for all traits, indicates that the variability seen is true to type. It is essential to understand the heritability of the traits which helps the breeders for assisting selection from a population. High heritability of the traits implies that the phenotype is a true reflection of the genotype hence selection would be promising and the selections performed are not affected by the environment (16).

**Fig.1.** Variance components explained by the genotypes and residual.

Smith Index

The Smith-Hazel Index is a classic selection method introduced by Smith and Hazel. It is used for selecting plant lines based on Fisher's discriminant function, which creates a linear equation using observable traits to estimate the genetic value of each line. It helps identify top-performing genotypes by assigning weights to each trait, making it more efficient than focusing on a single yield-related trait at a time (4,9). A selection intensity of 15% was applied to choose the genotypes. In Selection Index 1 (SI-1), the top six ranked genotypes are K5, K3, K15, K14, Mapillai Samba and H7, determined by allocating the equal genetic value of the individuals while taking into account traits that exhibit multicollinearity issues.

To address this issue, a new selection model SI-2 was then derived like SI-1 (equal economic weightage system) and the top six genotypes ranked by SI-2 are K5, K3, H7, K14, K15 and Mapillai Samba on removal of the trait that lead to multi-collinearity. Genetic weights of the individual traits and the top-ranked genotypes were mentioned in (Table 3) and (Table 4) respectively. The correlation plot (Fig. 2) shows a general idea about the collinearity issues between the traits. Particularly in the current study, the trait panicle length showed minor multicollinearity issues and it is also confirmed by a variation inflation factor (VIF) and condition number. The trait panicle length had a VIF less than 10, which suggests that minimal multicollinearity issues among the traits and the value of condition number is 14.4, which is closer to the lower end of this

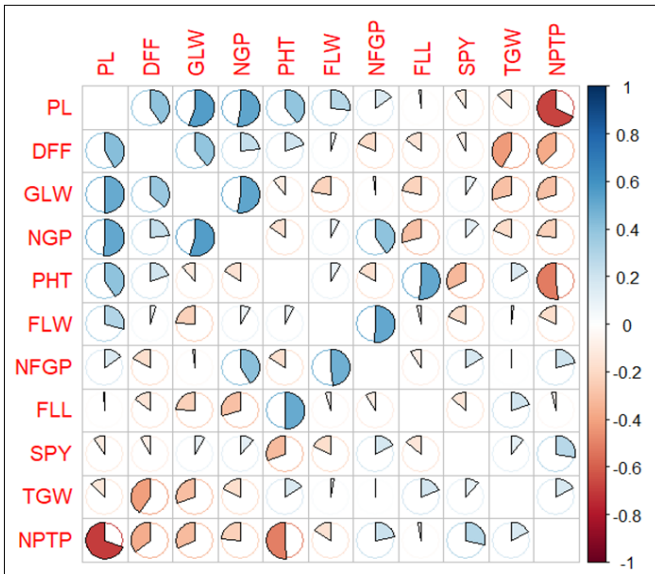
Table 3. Regression coefficients (b) and genetic weights (a) for all the traits under study based on SI-1, SI-2 and SI-3

S.No.	Traits	Equal weights	Selection Index 1	Equal weights	Selection Index 2	Differential weights	Selection Index 3
1	DFF	1	0.86	1	0.87	0.5	0.3583
2	PHT	1	0.94	1	0.95	-0.5	-0.4809
3	FLL	1	0.96	1	0.96	0.5	0.4654
4	FLW	1	4.93	1	4.98	-0.5	5.611
5	NPTP	1	1.00	1	0.98	0.5	0.6165
6	PL	1	1.10	1	-	-	-
7	NGP	1	0.98	1	0.99	-0.5	-0.4802
8	NFGP	1	0.90	1	0.90	2	1.869
9	TGW	1	0.95	1	0.95	1	0.9336
10	GLW	1	1.66	1	1.78	-0.5	0.7485
11	SPY	1	0.86	1	0.86	2	1.831

DFF. Days to fifty percent flowering, PHT. Plant height, FLL. Flag leaf length, FLW. Flag leaf width, NPTP. Number of productive tillers per plant, PL. Panicle length, NGP. Number of grains per Panicle, NFGP. Number of filled grains per panicle, TGW. Thousand grain weight, GLW. Grain length width ratio, SPY. Single plant yield.

Table 4. Top 10 ranked best-performing genotypes and their index score based on SI-1, SI-2 and SI-3

RANK	Selection Index 1		Selection Index 2		Selection Index 3	
	Genotypes	Index score	Genotypes	Index score	Genotypes	Index score
1	K5	660.4	K5	634.9	K5	361.3
2	K3	603.6	K3	576.9	H9	290.6
3	K15	589.8	H7	571.2	H7	284.0
4	K14	589.4	K14	564.9	CO51	281.2
5	Mapillai Samba	588.6	K15	564.1	BPT5204	279.8
6	H7	587.9	Mapillai Samba	560.3	H1	267.2
7	K4	584.1	K4	557.9	H6	261.9
8	H9	574.7	H9	552.4	Mapillai Samba	243.3
9	K18	571.6	K18	544.2	H2	238.0
10	K11	566.9	K11	540.8	K6	231.3

**Fig. 2.** Correlation Plot for 11 biometrical traits.

scale *i.e.*, 10 revealing the presence of very weak multicollinearity. Multicollinearity results in poorly conditioned matrices and biased index coefficients when assessing the multiple traits which result in affecting the estimation of selection gains (10,11). Further, the results show that there was no significant differences in selection gains (%) for grain yield traits in SI-1 and SI-2 but these genotypes had a notable impact on the index score by contributing to a higher or lower overall ranking based on their performance across multiple traits (Table 5). This indicates that excluding the trait with weak multicollinearity or minimal multicollinearity issues (panicle length) did not substantially impact the selection but changed the ranking of the genotypes got selected.

Table 5. Selection gain (%) for all the 11 traits based on SI-1, SI-2 and SI-3

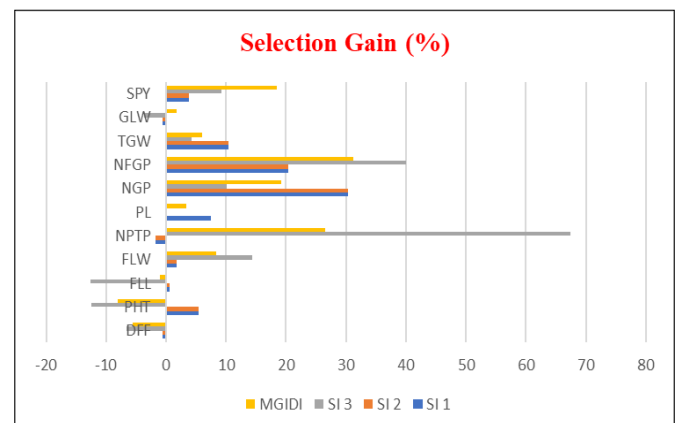
S. No.	Traits	Selection Index 1	Selection Index 2	Selection Index 3
1	DFF	-0.61	-0.61	-6.61
2	PHT	5.35	5.35	-12.39
3	FLL	0.58	0.58	-12.64
4	FLW	1.70	1.70	14.39
5	NPTP	-1.77	-1.77	67.37
6	PL	7.52	-	-
7	NGP	30.27	30.27	10.06
8	NFGP	20.34	20.34	40.06
9	TGW	10.41	10.41	4.23
10	GLW	-0.65	-0.65	-3.63
11	SPY	3.76	3.76	9.23

DFF. Days to fifty percent flowering, PHT. Plant height, FLL. Flag leaf length, FLW. Flag leaf width, NPTP. Number of productive tillers per plant, PL. Panicle length, NGP. Number of grains per Panicle, NFGP. Number of filled grains per panicle, TGW. Thousand grain weight, GLW. Grain length width ratio, SPY. Single plant yield.

While both SI-1 and SI-2 assign equal genetic weights to all traits, SI-3 adopts a differential weightage system based on the direct effects of traits on yield derived from path analysis. For example, traits such as DFF, FLL and NPTP were assigned with weights of 0.5, while traits NFGP and SPY were assigned with genetic weights of 2. Traits that negatively affect yields, such as PHT, FLW, NGP and GLW, were given a genetic weight of -0.5 (Table 3). A similar way differential weightage system allocated to different traits was done by (12). New regression coefficients were then calculated, similar to the approach used for SI-1 and SI-2. The genotypes K5, H9, K7, CO51, BPT5204 and H1 were ranked as the top 6 based on the genetic worth of the individual (Table 4). The genotype K5 was common among all three selection indices (SI-1, SI-2, SI-3). In the SI-1 and SI-2 selection indices, where traits were assigned equal weights, the genetic gain for single plant yield traits was the same for both indices *i.e.*, 3.7%. Meanwhile, in SI-3, where genotypic weights were adjusted based on the positive and negative direct effects on yield from path analysis (differential weightage system), the selection gain of plant yield increased from 3.7% to 9.2%, which demonstrated a significant improvement in yield compared to SI-1 and SI-2 (Table 5) and (Fig. 3). This shows the effectiveness of SI-3 in selecting genotypes as it applies unequal weights to traits and considers the direct effects of different traits on yield through path analysis. This approach leads to better selection gains for traits compared to SI-1 and SI-2.

MGIDI Index

The MGIDI index has numerous practical applications due to its unique and easy-to-interpret selection process. Unlike the SH index, which requires economic weights to predict genetic and economic gains for various combinations of genetic parameters, the MGIDI index does not deal with such procedure and in addition, it effectively addresses multicollinearity issues (17).

**Fig. 3.** Comparison of selection gains for the 11 traits under study for SI-1, SI-2, SI-3 and MGIDI.

In MGIDI, traits are rescaled to a 0-100 scale. For traits with negative gains (DFF, PHT), 100 is the minimum and 0 is the maximum. For others, 0 is the minimum and 100 is the maximum. These values are used to compute the MGIDI score.

The MGIDI selection process relies on factorial analysis and incorporates the correlation of component variables for selection. This factorial analysis effectively differentiates genotypes, particularly concerning the variables included in the factors (18). This study utilized Principal Component Analysis (PCA) to classify the traits into components. This technique helps to reduce the data complexity by transforming correlated traits into uncorrelated components, with each component showing how much each trait contributes. Further, these components are ranked by variance with the first component explaining the most variation. The number of components to retain can be decided using criteria like eigenvalue or scree plot. MGIDI utilizes PCA to group traits into factors which helps in identifying genotypes based on multiple traits and avoids redundancy and multicollinearity (19). PCA was conducted on 43 genotypes to evaluate 11 yield and yield-attributing traits. The PCA results indicated that only four principal components (PCs) had eigenvalues exceeding 1.00, accounting for approximately 72.4% of the total variability. Specifically, PC1, with an eigenvalue of 2.95, explained 26.8% of the variability. This was followed by PC2, which had an eigenvalue of 2.20 and accounted for 20% of the variability; PC3, with an eigenvalue of 1.68, explained 15.3% of the variability; and PC4, with an eigenvalue of 1.14, accounted for 10.3% of the variability (Table 6). Similar reports were also done by (20).

The number of PCs crossing the eigenvalue value of 1 defines the number of factors to be included in the factorial analysis. In the current study, as four PCs have crossed the eigenvalue of 1, hence four factors have been considered and all 11 traits have been grouped among these four factors (Table 6). Factor 1(FA1) comprises traits like NPTP, PL, NGP, GLW and FA2 associated with PHT, FLL, FLW and FA 3 includes traits like FLW, NFGP and finally, FA4 consists of traits like DFF, TGW, SPY. The average communality and uniqueness represented 72.45% and 27.54% of the total genetic variability in the dataset, respectively, as detailed in (Table 7). According to Al-Ashkar (21), genotypes with the lowest MGIDI scores computed by Euclidean distance, show a closer genetic distance between the genotype and the ideotype and such genotypes are demonstrated to be superior in performance. The MGIDI scores identified six genotypes, K5, CO51, H9, H7, K6 and K3 as, as the top performers across all 11 traits, with a selection intensity of 15%. The top-performing

Table 6. MGIDI Selection Index - Eigenvalues, Variance (%), Cumulative Variance (%)

PC	Eigenvalues	Variance (%)	Cumulative Variance(%)
PC1	2.95	26.8	26.8
PC2	2.2	20	46.8
PC3	1.68	15.3	62.1
PC4	1.14	10.3	72.4
PC5	0.78	7.1	79.5
PC6	0.72	6.53	86
PC7	0.49	4.44	90.5
PC8	0.4	3.63	94.1
PC9	0.31	2.8	96.9
PC10	0.19	1.68	98.6
PC11	0.15	1.39	100

PC = Principal component.

Table 7. MGIDI Selection Index - Factorial Loadings, communality and Uniqueness for all the 11 traits under study

VAR	FA1	FA2	FA3	FA4	Communality	Uniquenesses
DFF	-0.5	0.09	-0.15	-0.55	0.59	0.41
PHT	-0.16	-0.86	-0.01	-0.25	0.82	0.18
FLL	-0.19	0.72	0.07	-0.1	0.57	0.43
FLW	-0.03	0.09	-0.87	0.26	0.83	0.17
NPTP	-0.58	-0.43	0	-0.45	0.73	0.27
PL	0.82	0.31	-0.25	0.18	0.87	0.13
NGP	0.76	-0.27	-0.28	-0.16	0.76	0.24
NFGP	0.14	-0.2	-0.81	-0.31	0.81	0.19
TGW	-0.23	0.47	-0.08	-0.6	0.64	0.36
GLW	0.8	-0.23	0.26	0	0.76	0.24
SPY	0.16	-0.28	0.12	-0.69	0.59	0.41

DFF. Days to fifty percent flowering, PHT. Plant height, FLL. Flag leaf length, FLW. Flag leaf width, NPTP. Number of productive tillers per plant, PL. Panicle length, NGP. Number of grains per Panicle, NFGP. Number of filled grains per panicle, TGW. Thousand grain weight, GLW. Grain length width ratio, SPY. Single plant yield.

genotypes identified by the Smith index method were different from those selected using the MGIDI method. This is because the MGIDI method takes into account the distance between each genotype and its ideal type across all traits evaluated in the study. The MGIDI approach has previously been used to select top-performing genotypes in various crops such as strawberries (22), maize (23), wheat (24) and Eggplant (25).

Further, this can also be visualized using a radar graph, (Fig. 4) which illustrates that genotypes located outside the circle have the lowest MGIDI scores. From the radar graph genotype K3, positioned near the cut point on the circumference of the circle marked by the red line, displays interesting characteristics that require further examination (26). It is also recommended that researchers focus on genotypes close to the cut point for further analysis (Fig. 4). In terms of selection gain, the MGIDI index achieved a notable 18.5% improvement in single-plant yield, significantly outperforming the classical SI-1, SI-2 and SI-3 selection indices. Further, the MGIDI index showed a negative selection differential and selection gain for the traits DFF and PHT, which is crucial for plant breeders who aim for the early duration, sturdy dwarf variety. Similar reports of negative selection differential for days from emergence to flowering for the early duration was reported by (27). The percentage selection gains achieved using the MGIDI index are provided in Table 8. Additionally, (Fig. 3) presents a bar graph that offers a comprehensive comparison of the selection gains for the traits assessed using both the Smith and MGIDI indices.

Strength and Weakness view of the genotypes

A distinctive feature of the MGIDI is the strength and weakness plot, which highlights the strengths and weaknesses exhibited by the selected genotypes. It is helpful to understand the various variables that contribute to the selection approach. A factor line approaching the center of the circle for a specific genotype suggests that the factor has a strong influence on that genotype. Consequently, the traits linked to this factor are poorly performed by that genotype and vice versa.

Genotypes associated with Factor 1 (FA1), including K6, K3 and K5, exhibit notable strengths in traits such as NPTP, PL, NGP and GLW. Conversely, genotypes linked to FA2, such as K3, H9 and K6, show strengths in traits like PHT and FLL as these genotype stands out as an optimal genetic source to obtain compact growth (low stature). Additionally, genotypes associated with FA3,

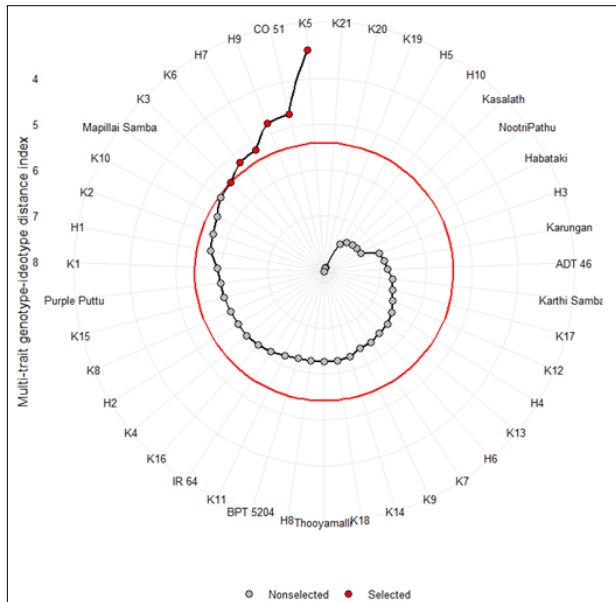


Fig. 4. Ranking of selected genotypes in ascending order based on MGIDI (The red line indicates a selection intensity of 15%, representing to the selection of six genotypes).

including H9 and CO51, perform well in traits such as FLW and NFGP. Finally, Factor 4 (FA4), which includes genotypes such as H7, K5 and K6, demonstrates strengths in traits like DFF, TGW and SPY (Fig. 5). It reveals that these genotypes have shorter durations coupled with higher yields, making them promising selections for future cultivation and breeding programs. Similar reports were also carried out in different crops where the strengths and weaknesses were described by employing MGIDI (22,28,29).

Coincidence Index (CI)

The coincidence index helps in evaluating the reliability, consistency and level of agreement between different selection indices in identifying the most desirable traits or genotypes. The results reveal a complete agreement (100%) between SI-1 and SI-2, indicating that removing panicle length (which shows minor multicollinearity issues) did not impact selection gains. However,

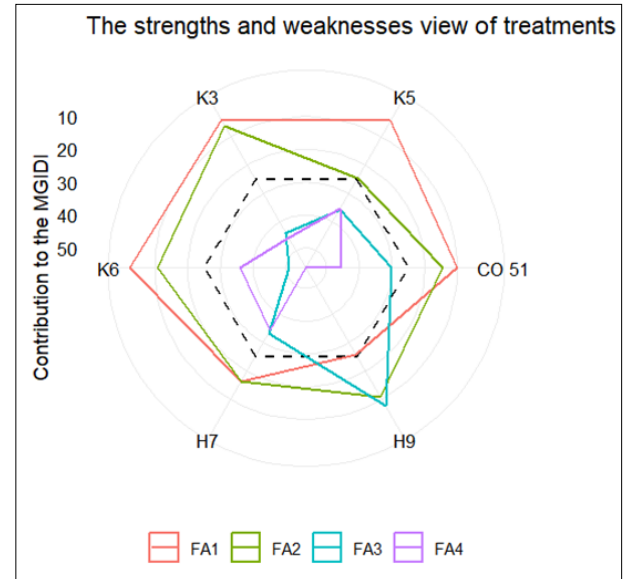


Fig. 5. Strength and Weakness view of the selected genotypes studies for 11 quantitative traits based on MGIDI. (If a factor is positioned closer to the outer edge, it indicates that the traits associated with that factor are more aligned with the ideal plant characteristics. The dashed line represents the theoretical value assuming all factors contribute equally).

the agreement between SI-1 and SI-3, which uses differential genetic weighting, was only 21.56%. Additionally, there was only a 41.17% CI overlap between SI-1 and MGIDI, as well as SI-2 and MGIDI, showing that MGIDI selects genotypes differently from traditional indices. Interestingly, SI-3 and MGIDI exhibited a higher CI overlap of 60.78%, suggesting that SI-3 results were somewhat relatively close to MGIDI selection results. Genotypes K5 and H7 were selected more often in all the selection indices which are outperformed compared to check and commercial varieties. This indicates that different indices are identifying similar traits or genotypes, underscoring the strength and reliability of the selection process (Table 9). Similarly, coincidence indices have been applied to other selection methods to identify superior-performing genotypes (30).

Table 8. Factors, Selection gain (%) for all the traits under study - MGIDI Selection Index

VAR	Factor	SD	h ² (%)	SG	SG(%)	sense
NPTP	FA1	4.83	97.3	4.71	26.6	increase
PL	FA1	0.793	95.4	0.757	3.33	increase
NGP	FA1	28.6	98.8	28.2	19.2	increase
GLW	FA1	0.0593	91.9	0.0545	1.73	increase
PHT	FA2	-7.7	98.4	-7.57	-8.12	decrease
FLL	FA2	-0.297	95.5	-0.284	-1.04	increase
FLW	FA3	0.108	91.8	0.0995	8.3	increase
NFGP	FA3	31	95.8	29.7	31.2	increase
DFF	FA4	-5.43	98.0	-5.33	-5.6	decrease
TGW	FA4	1.27	0.976	1.24	6.03	increase
SPY	FA4	6.62	0.942	6.24	18.5	increase

DFF. Days to fifty percent flowering, PHT. Plant height, FLL. Flag leaf length, FLW. Flag leaf width, NPTP. Number of productive tillers per plant, PL. Panicle length, NGP. Number of grains per panicle, NFGP. Number of filled grains per panicle, TGW. Thousand grain weight, GLW. Grain length width ratio, SPY. Single plant yield. FA. Factorial analysis, SD. Selection differential, h²(%). Heritability in percent, SG. Selection gain.

Table 9. Coincidence index for SI-1, SI-2, SI-3 and MGIDI selection Index

S.NO	Variable 1	Variable 2	Index	Common Genotypes
1	MGIDI	Selection Index 1	41.17	K5,H7,K3 (3)
2	MGIDI	Selection Index 2	41.17	K5,K3,H7(3)
3	MGIDI	Selection Index 3	60.78	K5,CO 51,H9,H7(4)
4	Selection Index 1	Selection Index 2	100	K5,K3,K14,K15,K7 Mappilal Samba(6)
5	Selection Index 1	Selection Index 3	21.56	K5,H7(2)
6	Selection Index 2	Selection Index 3	21.56	H7,K5(2)

Conclusion

This study utilized both the Smith-Hazel selection index and the MGIDI multitrait index to identify top-performing genotypes. While the Smith index addressed multicollinearity, its effectiveness was limited by difficulties in assigning economic values to traits, showing selection gains of 3.76% and 9.23% in plant yield across different indices. In contrast, the MGIDI index provided a more comprehensive assessment, achieving an 18.50% yield improvement and better balancing trait improvements. Genotypes K5 and H7, identified by both methods, performed exceptionally well, showing early growth and higher yields than commercial checks. These lines can be recommended for release upon multi-environment confirmation. Overall, the MGIDI index outperformed the Smith index, particularly concerning grain yield traits, making it a valuable tool for enhancing breeding efficiency and developing superior rice varieties for sustainable agriculture.

Acknowledgements

We would like to extend our sincere gratitude to the Tamil Nadu Agricultural University for their invaluable support throughout this research. Special thanks to the Department of Rice and the Department of Plant Biotechnology for their guidance, technical assistance and expertise, which played a critical role in the successful completion of this study.

Authors' contributions

PVK was responsible for data curation, execution, original draft writing and visualization. CB, SM, MR and NS contributed to the conceptualization, methodology, resource acquisition and supervision. MR, NS and MS supported the visualization, validation and resource management. CB, SM, PM and BA handled the writing, review and editing. Data analysis was conducted by PVK, BA and PM.

All authors read and approved the final manuscript

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

- Zhao M, Lin Y, Chen H. Improving nutritional quality of rice for human health. *Theoretical and Applied Genetics*. 2020;133:1397-413
- Verma V, Vishal B, Kohli A, Kumar PP. Systems-based rice improvement approaches for sustainable food and nutritional security. *Plant Cell Reports*. 2021;40(11):2021-036. <https://doi.org/10.1007/s00299-021-02790-6>
- Donald CT. The breeding of crop ideotypes. *Euphytica*. 1968;17:385-403. <https://doi.org/10.1007/BF00056241>
- Smith HF. A discriminant function for plant selection. *Annals of eugenics*. 1936;7(3):240-50. <https://doi.org/10.1111/j.1469-1809.1936.tb02143.x>
- Graham MH. Confronting multicollinearity in ecological multiple regression. *Ecology*. 2003;84(11):2809-15. <https://doi.org/10.1890/02-3114>
- Olivoto T, de Souza VQ, Nardino M, Carvalho IR, Ferrari M, de Pelegrin AJ, Szareski VJ, Schmidt D. Multicollinearity in path analysis: a simple method to reduce its effects. *Agronomy journal*. 2017;109(1):131-42. <https://doi.org/10.2134/agronj2016.04.0196>
- Olivoto T, Nardino M. MGIDI: Toward an effective multivariate selection in biological experiments. *Bioinformatics*. 2021;37(10):1383-9.
- Bizari EH, Val BH, Pereira ED, Mauro AO, Unêda-Trevisoli SH. Selection indices for agronomic traits in segregating populations of soybean. *Revista Ciência Agrônômica*. 2017;48(1):110-7.
- Hazel LN. The genetic basis for constructing selection indexes. *Genetics*. 1943;28(6):476-90.
- Cheng J, Sun J, Yao K, Xu M, Cao Y. A variable selection method based on mutual information and variance inflation factor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2022;268:120652.
- Kim JH. Multicollinearity and misleading statistical results. *Korean journal of anesthesiology*. 2019;72(6):558-69.
- Palaniyappan S, Arunachalam P, Banumathy S, Muthuramu S. Introspection of discriminate function analysis and MGIDI selection index for selection to improve yield in rice (*Oryza sativa* L.). *Indian journal of genetics and plant breeding*. 2024;84(02):202-8.
- Gepts P, Janick J. Plant breeding reviews. *Crop Domestication as a Long-Term Selection Experiment*. 2010;24(2):1-44.
- Olivoto T, Lúcio AD. metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*. 2020;11(6):783-789. <https://doi.org/10.1111/2041-210X.13384>
- Resende MA, Freitas JA, Lanza MA, Resende MD, Azevedo CF. Genetic divergence and selection index via BLUP in cotton accessions for technological characteristics of the fiber. *Tropical Agricultural Research*. 2014;44:334-640.
- de Souza YP, Daher RF, Vander Pereira A, da Silva VB, Freitas RS, de Amaral Gravina G. Repeatability and minimum number of evaluations for morpho-agronomic characters of elephant-grass for energy purposes. *Revista Brasileira de Ciências Agrárias*. 2017;12(3):391-7.
- Burdon RD, Li Y. Genotype-environment interaction involving site differences in expression of genetic variation along with genotypic rank changes: simulations of economic significance. *Tree Genetics & Genomes*. 2019;15(1):2. <https://doi.org/10.1007/s11295-018-1308-3>
- Costello AB, Osborne J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research and evaluation*. 2019;10(1):7.
- Debnath P, Chakma K, Bhuiyan MS, Thapa R, Pan R, Akhter D. A Novel Multi Trait Genotype Ideotype Distance Index (MGIDI) for Genotype Selection in Plant Breeding: Application, Prospects and Limitations. *Crop Design*. 2024;2:100074.
- Pallavi M, Prasad BM, Shanthi P, Reddy VL, Kumar AN. Multi trait genotype-ideotype distance index (MGIDI) for early seedling vigour and yield related traits to identify elite lines in rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*. 2024;15(1):120-
- Al-Ashkar I, Sallam M, Ibrahim A, Ghazy A, Al-Suhaibani N, Ben Romdhane W, Al-Doss A. Identification of Wheat Ideotype under Multiple Abiotic Stresses and Complex Environmental Interplays by Multivariate Analysis Techniques. *Plants*. 2023;12(20):3540.
- Olivoto T, Diel MI, Schmidt D, Lúcio AD. Multivariate analysis of strawberry experiments: where are we now and where can we go? *BioRxiv*. 2021:2020-12.
- Palaniyappan S, Ganesan KN, Manivannan N, Ravichandran V, Senthil N. Multi trait genotype-ideotype distance index-A tool for identification of elite parental inbreds for developing heterotic hybrids of fodder maize (*Zea mays* L.). *Electronic Journal of Plant Breeding*. 2023;14(3):841-9.
- Meier C, Marchioro VS, Meira D, Olivoto T, Klein LA. Genetic

- parameters and multiple-trait selection in wheat genotypes. *Pesquisa Agropecuária Tropical*. 2021;51:e67996.
25. Uddin MS, Billah M, Afroz R, Rahman S, Jahan N, Hossain MG, Bagum SA, Uddin MS, Khaldun AB, Azam MG, Hossain N. Evaluation of 130 Eggplant (*Solanum melongena* L.) genotypes for future breeding program based on qualitative and quantitative traits and various genetic parameters. *Horticulturae*. 2021;7(10):376.
 26. Olivoto T, Lúcio AD, da Silva JA, Sari BG, Diel MI. Mean performance and stability in multi-environment trials II: Selection based on multiple traits. *Agronomy Journal*. 2019;111(6):2961-969. <https://doi.org/10.2134/agronj2019.03.0221>
 27. Klein LA, Marchioro VS, Toebe M, Olivoto T, Meira D, Meier C, Benin G, Busatto CA, Garafini DC, Alberti JV, Finatto JL. Selection of superior black oat lines using the MGIDI index. *Crop Breeding and Applied Biotechnology*. 2023 ;23(3):e45112332.
 28. Gabriel A, de Resende JT, Zeist AR, Resende LV, Resende NC, Zeist RA. Phenotypic stability of strawberry cultivars based on physicochemical traits of fruits. *Horticultura Brasileira*. 2019;37:75-81.
 29. Mamun AA, Islam MM, Adhikary SK, Sultana MS. Resolution of genetic variability and selection of novel genotypes in EMS induced rice mutants based on quantitative traits through MGIDI.
 30. Ambrósio M, Daher RF, Santos RM, Santana JG, Vidal AK, Nascimento MR, Leite CL, de Souza AG, Freitas RS, Stida WF, Farias JE. Multi-trait index: selection and recommendation of superior black bean genotypes as new improved varieties. *BMC Plant Biology*. 2024;24(1):525. <https://doi.org/10.1186/s12870-024-05248-5>