RESEARCH ARTICLE

SABRAO Journal of Breeding and Genetics
55 (6) 1886-1896, 2023
http://doi.org/10.54910/sabrao2023.55.6.3
http://sabraojournal.org/
pISSN 1029-7073; eISSN 2224-8978

PRINCIPAL COMPONENT ANALYSIS AND ESTIMATED BREEDING VALUES FOR SELECTING SUITABLE PARENTAL GENOTYPES IN RICE (*ORYZA SATIVA* L.)

M. KHATUN[1], M.R. ISLAM[2], M.A.R. KHAN[3], M.K. HOSSAIN[2], and A.K.M.A. ISLAM[1*]

[1]Department of Genetics and Plant Breeding, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh
[2]International Rice Research Institute (IRRI), Dhaka, Bangladesh
[3]Department of Agronomy, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh
*Corresponding author's email: aminulgpb@bsmrau.edu.bd
Email addresses of co-authors: mariumgpb@bsmrau.edu.bd, mr.islam@irri.org, arif@bsmrau.edu.bd, k.hossain@irri.org

SUMMARY

Appropriate parental selection is the breeder's main concern to exploit the highest genetic diversity and generate superior genotypes for subsequent breeding programs. Hence, the presented investigation proceeded to evaluate 353 breeding lines of rice at three breeding zones (Rajshahi, Cumilla, and Gazipur) in Bangladesh in replicated yield trials during the Boro season of 2018–2019 to identify the best genotypes and utilize them as parental materials. Data recorded on 12 yield-related traits helped to determine the best breeding lines with higher predicted breeding values. The first five principal components (PC1, PC2, PC3, PC4, and PC5) represented more than 70% (75.1%) contribution to the variability of the data. Three hundred fifty-three rice genotypes incurred distribution into five clusters over three environments. Clusters I, II, III, IV, and V comprised 66, 51, 83, 79, and 74 genotypes, respectively. Based on estimated breeding values (EBVs), IR107971-B-B RGA-B RGA-202 showed the highest value (0.395), followed by IR 108000-B-B-B-B-13 (0.329), IR 103309-B-B RGA-B RGA-194 (0.321), IR 107982-B-41-1-2-1 (0.291), IRRI 174 (0.264), and IR 107976-B-B RGA-B RGA-254 (0.234). The lowest EBV (0.022) appeared in IR103309-B-B RGA-B RGA-204 among the top 20 genotypes. Both IR 103309-B-B RGA-B RGA-194 (0.321) and IR 107982-B-41-1-2-1 (0.291) could benefit as parents for further breeding programs having higher EBVs and higher genetic diversity.

Keywords: Rice, cluster analysis, estimated breeding values, parental materials

Communicating Editor: Dr. Kamile Ulukapi

Manuscript received: May 17, 2023; Accepted: September 5, 2023.

Citation: Khatun M, Islam MR, Khan MAR, Hossain MK, Islam AKMA (2023). Principal component analysis and estimated breeding values for selecting suitable parental genotypes in rice (*Oryza sativa* L.). *SABRAO J. Breed. Genet.* 55(6): 1886-1896. http://doi.org/10.54910/sabrao2023.55.6.3.

Key findings: EBVs and cluster analysis can assist breeders in selecting the best parents for the next breeding program. The top 20 genotypes based on EBVs with the highest genetic diversity may serve as parents to produce desired plant progenies crucial for the ever-growing populations. More specifically, IR 103309-B-B RGA-B RGA-194 and IR 107982-B-41-1-2-1 will be more effective as parents due to their higher EBVs and more genetic diversity.

INTRODUCTION

Rice, as part of the genus *Oryza*, is a staple food in many nations worldwide. The expansion of rice food production, particularly in Asia and Africa, is one of the fundamental pillars of food safety (Suela *et al*., 2019). According to HIES 2016, the per capita daily rice intake is about 367 g, providing 60% of the total calories and 50% of the overall protein. In rice breeding, there is competition between yield and other agronomic characteristics, including traits linked to biotic and abiotic stress and qualitative features. As a result, the primary goals of rice breeding are to improve yield under favorable conditions and decrease yield losses under adverse conditions.

The land used to grow rice is steadily shrinking because of Bangladesh's industrialization and expansion of the metropolitan region, despite a decline in arable land and a rising population. A significant amount of rice-growing land also incurs conversion for other purposes. According to BRRI (2019), the yield performance also became 4.7 t/ha as the national average. Additionally, artificial events harm rice breeding; therefore, breeders must prepare to combat any challenges that result from uncertainties. The development of new varieties is now an arduous and resource-intensive endeavor. Production has negative impacts from biotic (fungal, bacterial, viral, and insect pest) and abiotic (drought and flood) stressors for tropical wetland rice and non-irrigated rice, respectively (Khush and Virk, 2005). Producing rice must quadruple over the next 30 years to meet the demand of the growing world population. Rice breeders must also consider the attribute (yield) governed by polygenes with minimal consequences for newly generated lines to increase rice productivity.

Breeders must choose acceptable rice genotypes before beginning the breeding program to achieve effective propagation. Several techniques are available for selecting appropriate genotypes. Among them, the principal component analysis (PCA), cluster analysis, and estimated breeding values are more useful for parental materials' selection by observing the diversity and higher breeding values. According to Jolliffe and Cadima (2016), it is challenging to analyze immense datasets; hence, reducing the dimension of such datasets occurs using PCA. Additionally, it improves interpretation skills and minimizes information loss. PCA extracts data from a high-dimensional space and describes the dataset's multivariate normal distribution. It works to keep the relevant portions with the data that differs more and eliminate the unnecessary parts. An orthogonal transformation process known as the principal component analysis (PCA) transforms a collection of characters with potentially correlated values into a set of observations with linearly uncorrelated characters (Mohan *et al*., 2015; Alshugeairy *et al*., 2023; Zayed *et al*., 2023). Genotypes or breeding lines attain grouping using cluster analysis, such that genotypes of the same type are in the same group (referred to as a cluster), and genotypes of different kinds are in distinct groups.

Estimated breeding values (EBVs) are suitable for deciding which parental materials to use in the upcoming breeding program. Based on their EBVs and diversity, all the materials in the population are assessable. The pertinent study aims to find the best genotypes with the highest breeding values for possible use as future parental materials.

MATERIALS AND METHODS

Experimental sites and climate

During the Boro season in 2018–19, the research transpired in three locations, namely, Rajshahi (BRRI regional station), Cumilla (BRRI regional station), and Gazipur (ACI–Mawna). Rajshahi is a city in Western Bangladesh, geographically situated in the Barind Tract at a height of 23 masl (https://en.wikipedia.org). Cumilla is much lower at 17 masl at 23°27′N latitude and 91°12′E longitudes (https://dateandtime.info). With a 34 m height, Gazipur sits inside the Madhupur Tract at grid lines of 23°53′ to 24°20′N and longitudes of 90°09′ to 90°42′E (Simu *et al*., 2018). For Rajshahi, the highest and lowest mean temperatures were 34.4 °C and 9.7 °C, respectively; 34.9 °C and 14.6 °C for Gazipur, and 34.7 °C and 12.1 °C for Cumilla. Rajshahi records a yearly rainfall of 1,228 mm; Gazipur, 1,796 mm; and Cumilla, 1,938 mm (BBS, 2021).

Experimental materials and design

This experiment used a total of 345 breeding lines, an estimated set of genomic selections generated at IRRI-HQ, Philippines, along with eight check types. Five IRRI varieties (IRRI 104, IRRI 154, IRRI 156, IRRI 174, and IRRI 181) and three BRRI varieties (BRRI dhan28, BRRI dhan67, and BRRI dhan81) served as checks in this study. The generation of all these breeding lines utilized the single seed descent (SSD) technique and rapid generation advance (RGA). Seventy breeding lines attained two-time duplication from 353 total breeding lines and became the experimental material. The other genotypes, evaluated under one replication, received a partially replicated trial classification. A replicated design with a row and column orientation was the experiment's setup. The total number of plots for repeated materials was 140, including eight standard check varieties for comparing the performance of the selected breeding lines with them. Plot dimensions, spacing, area, and planting were 5.4 m^2 (20 cm × 20 cm), 5.4 m (27 hills) × 1.0 m (5 rows), and 2–3 seedlings per hill, respectively.

Data recording

Data recording ensued on days to 50% flowering, days to maturity, plant height (cm), panicle length (cm), branches per panicle, filled grains per panicle, unfilled grains per panicle, sterility %, grain length (mm), grain breadth (mm), thousand-grain weight (g), and yield per plant (g).

Statistical analysis

Principal component analysis (PCA) and cluster analysis

Accomplishing the principal component analysis in the experiment was by the R software. The selection of distance measures is a key step in a hierarchical clustering pattern. Manhattan distance is a simple measurement and equals the sum of absolute distances for each character (Madhulatha, 2012). When using the grid-like path, computing the distance between two points can run through the Manhattan distance function. The formula for calculating the distance between two points, such as, X (X_1, X_2, etc.) and Y (Y_1, Y_2, etc.), follows below:

$$d = \sum_{i=1}^{n} | X_i - Y_i |$$

Where:
n = number of characters and
X_i and Y_i = values of ith characters at points X and Y, respectively.

Euclidean distance is a more common measure, and the formula for it is as follows:

$$d = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}$$

Where:
n = number of characters and
x_j and y_j = values of jth characters at points x and y, respectively.

Breeding value estimation

A simplified best linear unbiased prediction (BLUP) mixed model helped estimate the breeding value of parental lines (Popescu, 2014). It is a linear mathematical model that appears below:

$$Y_{ij} = \mu + s_i + e_{ij}$$

Where:
Y_{ij} = record of jth progeny of ith parental lines,
μ= constant parameter,
and, $s_i = ½\ g_i$,

Where:
g_i = breeding value of ith parental lines,
e_{ij} = the residual effect.

The average value of s and e is equal to zero. Both s and e are non-correlated variables among them with the variances $\sum_s^2 s_i$ and $\sum_e^2 e_{ij}$.

According to Rodriguez *et al*. (2019), the formula for calculating the accuracy of predicted estimated breeding value (EBV) for parental breeding lines follows below:

$$r = \sqrt{\{1 - (PEV / \sigma^2_A)\}}$$

Where:
r = prediction accuracy of EBV,
PEV = prediction error variance gained from the elements of the inverse of the coefficient matrix of the mixed model equations and
σ^2_A = additive variance across environments.

RESULTS

Analysis of variance

Creating the Type III ANOVA of three locations for yield used the ImerTest Package (2020) for linear mixed models through Satterthwaite's degrees of freedom method. Type III ANOVA tables were essential for fixed-effect with Satterthwaite for denominator degrees of freedom for F-tests. The analysis of variance for yield was significant in Rajshahi ($P < 0.01$) and Cumilla ($P < 0.001$) but insignificant in Gazipur (Table 1).

Principal component analysis

The first principal component (PC1) was responsible for most of the variation (23.3%) of the data, and the following principal components are accountable for the rest of the variation (Table 2). The first five principal components (PC1, PC2, PC3, PC4, and PC5) represented 75.1% of the data variability. PC5 contributed only 10% divergence, with 100% contributions from 12 principal components. The influence of PC11 and PC12 variation was minimum, 0.4%, and 0.3%, respectively. The deviation from the mean value of PC1 was higher (1.669) than other principal components. Each succeeding principal component accounted for less deviation from the mean as possible.

Principal component analysis' graphical representation employed an elbow graph. From the point, a more or less obtained straight line was the threshold point, indicating that more than 70% distinction has the first five contributing components (Figure 1). The elbow graph illustrates how many clustering contains more than 70% variation.

Cluster analysis

Five clustering resulted by using the k-means clustering method based on the five most significant principal components (Figure 2). On the five important principal components, k here equals five. A cluster of five different groups has five different colors representing them. The distribution pattern of 70 breeding materials between two principal components sustained labeling, and the arrays were variable. Similarity or dissimilarity is distinguishable by the height of the clades. The clades with same height are similar, while the clades with different are dissimilar. The more tallness difference persists, the more the contrast in clades. Five clusters, generated through the hierarchical clustering of 353 genotypes, appear in Figure 3. Dendrogram of 353 breeding lines represented five clusters,

Table 1. Type III Analysis of Variance with Satterthwaite's method of the breeding lines in the trials of Rajshahi, Cumilla, and Gazipur breeding zones of Bangladesh

Sources of variation	Sum sq	Mean sq	NumDF	DenDF	F-value	Pr (>F)
Rajshahi						
Genotypes	207.37	0.589	352	66.90	1.67	0.006 **
Replication	0.124	0.124	1	67.38	0.353	0.555
Field row	0.031	0.031	1	2.44	0.087	0.792
Field column	1.08	1.08	1	67.24	3.05	0.085
Cumilla						
Genotypes	363.91	1.03	352	66.66	2.15	0.0001 ***
Replication	0.94	0.938	1	68.57	1.95	0.167
Field row	0.03	0.029	1	5.10	0.060	0.816
Field column	0.05	0.053	1	68.18	0.111	0.740
Gazipur						
Genotypes	363.88	1.03	352	68	0.895	0.739
Replication	0.930	0.930	1	68	0.806	0.373
Field row	0.240	0.238	1	68	0.206	0.652
Field column	0.160	0.156	1	68	0.135	0.715

Significant codes: *** 0.001, ** 0.01.

Table 2. Standard deviation, proportion of variance and cumulative proportion obtained from principal component analysis for yield and yield-contributing traits of 353 genotypes

Parameters	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Standard deviation	1.67	1.43	1.27	1.18	1.10	0.933	0.787	0.747	0.674	0.638	0.213	0.192
Proportion of variance	0.232	0.169	0.134	0.116	0.100	0.072	0.052	0.047	0.038	0.034	0.004	0.003
Cumulative proportion	0.232	0.401	0.535	0.651	0.751	0.823	0.875	0.921	0.959	0.993	0.997	1.000

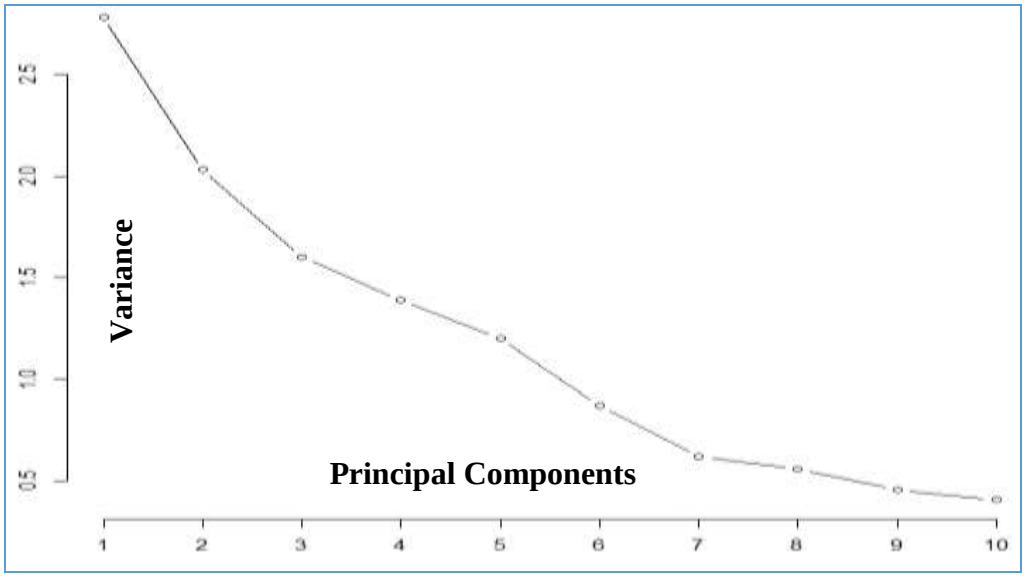

Figure 1. Elbow graph representing principal components versus variances for 353 genotypes.

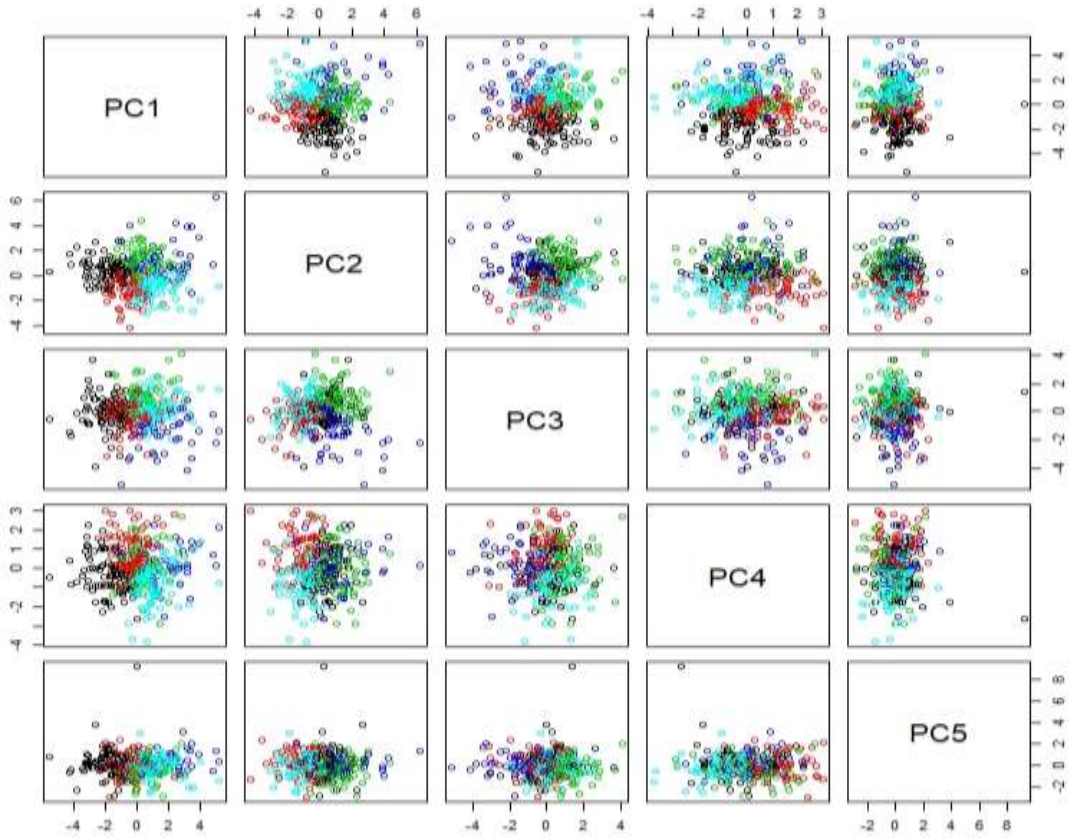

Figure 2. K-means clustering based on the five most significant principal components of 70 replicated breeding materials.

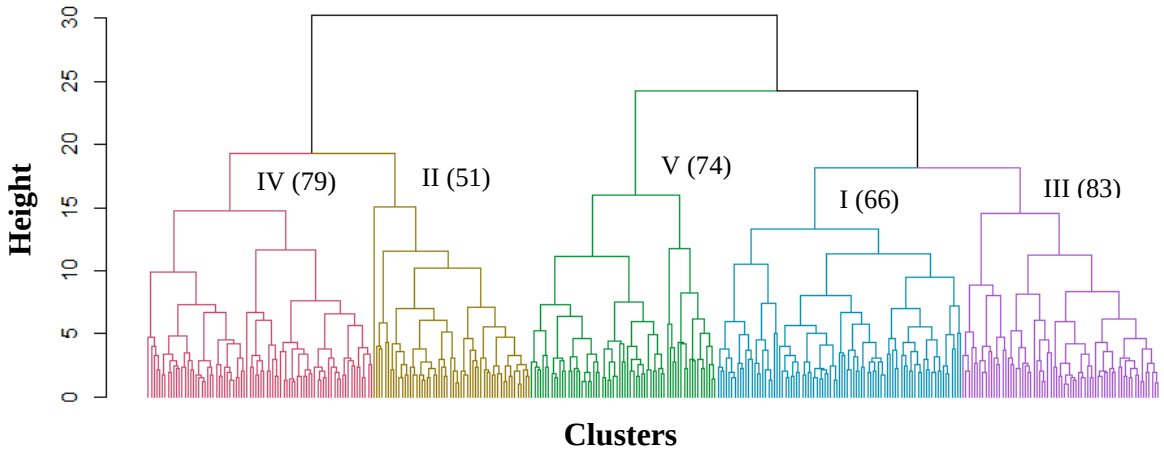

Figure 3. Hierarchical clustering or dendrogram of 353 breeding lines representing five clusters with different colors, each with the total number of genotypes.

i.e., cluster I (66 genotypes), cluster II (51 genotypes), cluster III (83 genotypes), cluster IV (79 genotypes), and cluster V (74 genotypes) (Figure 3). The number of demonstrated genotypes in each cluster was according to their diversity. Based on the elbow graph (Figure 1), the clustering ensued, with 353 genotypes distributed according to their similarity and dissimilarity within the five clusters.

The hierarchical clustering of the top 20 genotypes, constructed based on elbow graph (Figure 4) and estimated breeding values (EBVs), shows that cluster I contained nine genotypes (Figure 5), and they were IRRI 154, IR 106436-B-B RGA-B RGA-B RGA-21, IR 103292-B-B-B-B-B-14, IR 103310-B-B RGA-B RGA-13, IR 100306-B-B RGA-B RGA-B RGA-23, IR 103741-B-B RGA-B RGA-76, IR 107982-B-41-1-2-1, IR 103309-B-B RGA-B RGA-204, and IR 106457-B-B RGA-B RGA-12. Cluster II included three genotypes, i.e., IRRI 174, IR 107971-B-B RGA-B RGA-202, and IR 107982-B-B RGA-B RGA-110. Cluster III, comprising four genotypes, consisted of IR 106449-B-B RGA-B RGA-B RGA-2, IR 108008-B-18-1-3-1, IR 107971-B-B RGA-B RGA-139, and IR 103310-B-B RGA-B RGA-13. IR 106454-B-B-B-B-43, IR 103309-B-B RGA-B RGA-194, and IR 107976-B-B RGA-B RGA-254 were under the cluster IV. Cluster V contained only one genotype, IR 107989-B-B RGA-B RGA-107. Cluster I held nine genotypes, followed by cluster III (4 genotypes), cluster II (3 genotypes), cluster IV (3 genotypes), and cluster V (1 genotype) (Figure 5 and Table 3).

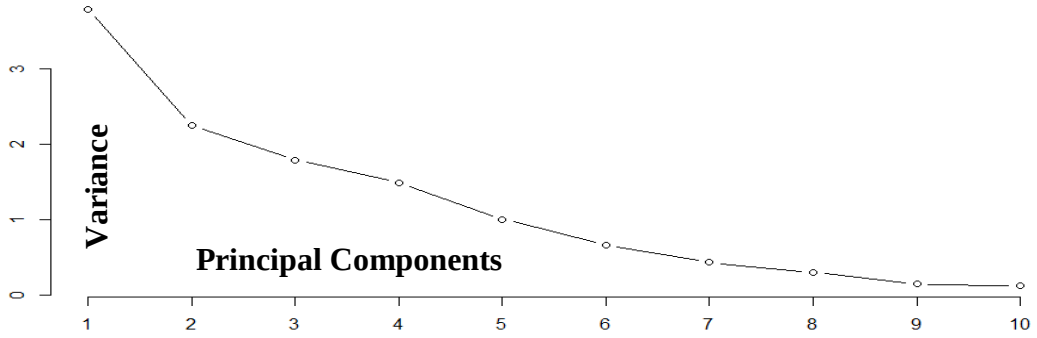

Figure 4. Elbow graph representing principal components versus variances for the top 20 genotypes.

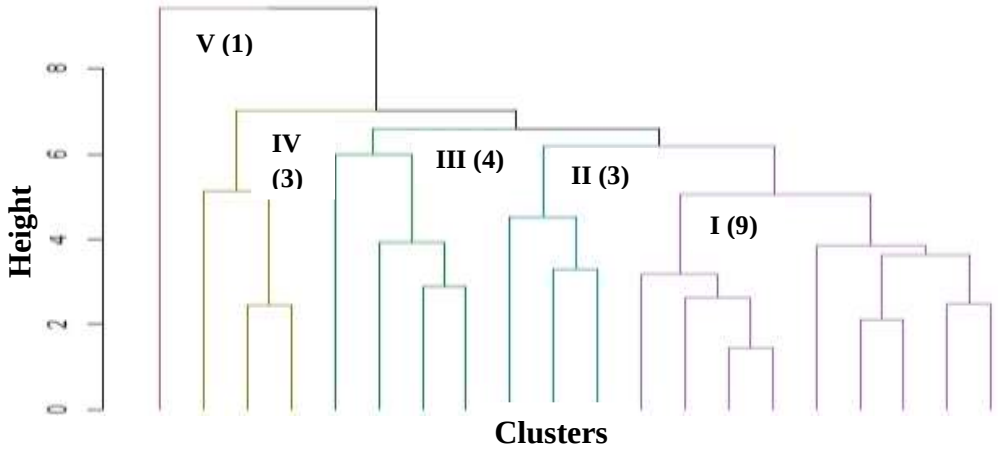

Figure 5. Hierarchical clustering or dendrogram of the top 20 breeding lines representing five clusters with different colors.

Table 3. Distribution of the top 20 genotypes into five clusters over three environments

Clusters	Number of genotypes	Genotypes
I	9	IR 106457-B-B RGA-B RGA-12, IR 103309-B-B RGA-B RGA-204, IR 107982-B-41-1-2-1, IR 103741-B-B RGA-B RGA-76, IR 100306-B-B RGA-B RGA-B RGA-23, IR 103310-B-B RGA-B RGA-13, IR 103292-B-B-B-B-B-14, IR 106436-B-B RGA-B RGA-B RGA-21, and IRRI 154
II	3	IR 107982-B-B RGA-B RGA-110, IR 107971-B-B RGA-B RGA-202, and IRRI 174
III	4	IR 108000-B-B-B-B-13, IR 107971-B-B RGA-B RGA-139, IR 108008-B-18-1-3-1, and IR 106449-B-B RGA-B RGA-B RGA-2
IV	3	IR 107976-B-B RGA-B RGA-254, IR 103309-B-B RGA-B RGA-194, and IR 106454-B-B-B-B-43
V	1	IR 107989-B-B RGA-B RGA-107

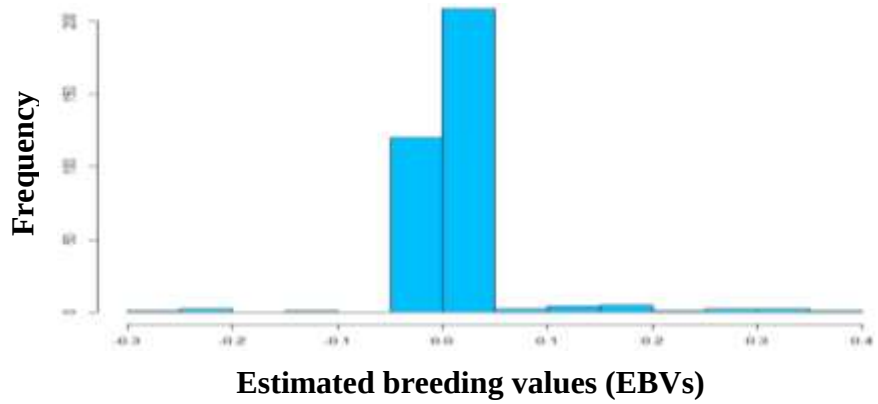

Figure 6. Histogram displaying frequency distribution of 353 breeding lines for estimated breeding values (EBVs).

Estimated breeding values (EBVs)

In public rice breeding programs in Asia, there is a limited use of pedigree information for improving a complex trait like yield (Juma *et al*. 2021). Additive genetic effects, or EBVs, were calculated to enhance genotypic performance in breeding programs. In Figure 6, zero (0.0) represents random effects, and each positive value indicates the higher value of EBVs. For selecting parents in the next breeding cycle, the genotypes containing higher EBVs or above the average value are necessary for selection to increase the genetic gain. The higher EBV value of an individual indicates the better parent used for preference for further breeding programs than others. The highest EBV is 0.395 for IR 107971-B-B RGA-B RGA-202, followed by 0.329, 0.321, 0.291, 0.264, 0.234, 0.196, 0.193, 0.172, 0.163, 0.154, 0.133, 0.129, 0.113, 0.109, 0.092, 0.061, 0.046, 0.039, and 0.022 for IR 108000-B-B-B-B-13, IR 103309-B-B RGA-B RGA-194, IR 107982-B-41-1-2-1, IRRI 174, IR 107976-B-B RGA-B RGA-254, IR 103310-B-B RGA-B RGA-13, IR 106454-B-B-B-B-43, IR 103741-B-B RGA-B RGA-76, IR 106449-B-B RGA-B RGA-B RGA-2, IR 106436-B-B RGA-B RGA-B RGA-21, IR 100306-B-B RGA-B RGA-B RGA-23, IR 107989-B-B RGA-B RGA-107, IRRI 154, IR 107982-B-B RGA-B RGA-110, IR 106457-B-B RGA-B RGA-12, IR 108008-B-18-1-3-1, IR 103292-B-B-B-B-B-14, IR 107971-B-B RGA-B RGA-139, and IR 103309-B-B RGA-B RGA-204, respectively (Table 4). IR 107982-B-41-1-2-1 was the highest yielder (6.06 t/ha), and IR 107971-B-B RGA-B RGA-202 was the lowest yielder (5.44 t/ha) among the top 20 lines based on BLUP, possessing rank one and rank 20, respectively.

Table 4. The top 20 genotypes with estimated breeding values (EBVs) and the ranking based on EBVs

Breeding value ranking	Genotypes	EBV
1	IR 107971-B-B RGA-B RGA-202	0.395
2	IR 108000-B-B-B-B-13	0.329
3	IR 103309-B-B RGA-B RGA-194	0.321
4	IR 107982-B-41-1-2-1	0.291
5	IRRI 174 (G64)	0.264
6	IR 107976-B-B RGA-B RGA-254 (G24)	0.234
7	IR 106454-B-B-B-B-43	0.196
8	IR 103310-B-B RGA-B RGA-13 (G37)	0.193
9	IR 103741-B-B RGA-B RGA-76	0.172
10	IR 106449-B-B RGA-B RGA-B RGA-2	0.163
11	IR 106436-B-B RGA-B RGA-B RGA-21	0.154
12	IR 100306-B-B RGA-B RGA-B RGA-23	0.133
13	IR 107989-B-B RGA-B RGA-107	0.129
14	IRRI 154 (G65)	0.113
15	IR 107982-B-B RGA-B RGA-110	0.109
16	IR 106457-B-B RGA-B RGA-12	0.092
17	IR 108008-B-18-1-3-1	0.061
18	IR 103292-B-B-B-B-B-14 (G20)	0.046
19	IR 107971-B-B RGA-B RGA-139	0.039
20	IR 103309-B-B RGA-B RGA-204	0.022

Note: Genotypes with bold fonts were replicated twice.

DISCUSSION

The p-value of ANOVA represented that the reliability index of the result might be decreasing. When the p-value is higher, the reliability of the relationship between respective variables in the population might be lower (Table 1). Using principal component analysis reduced the complexity of the data set, and the division of observed variance among traits was according to the significance of each feature (Ringnér, 2008). PCA is a suitable method for decreasing the dimensionality of big datasets and improving the interpretation ability. It also lessens the loss of information. It creates new uncorrelated variables and thus increases the variation. PCA is also called the adaptive method of data analysis for solving a problem related to eigenvalue and tailoring different data types and structures (Jolliffe and Cadima, 2016). It possibly converts correlated variables into uncorrelated ones, termed principal components. More than 70% of the variation has contributions from the first five principal components (Table 2).

The elbow graph helped to decide how much clustering contained more than 70% variation during hierarchical clustering (Figures 1 and 4). Cluster analysis is a helpful statistical data analysis tool for grouping similar genotypes in the same cluster and different genotypes in different clumps. There are different kinds of data clustering methods (Madhulatha, 2012). Two popular methods used for data clustering include partitional and hierarchical clustering. Partitioning algorithms work depending on the specification of the primary groups and repetitively transferring items among groups for successful convergence. It usually fixes clusters together. The K-means and k-medoids algorithms are a few of the most popular methods of partitioning algorithms. The k-means algorithm determines the aggregate at each point, and the center is called a centroid. In a cluster, the centroid is the mean value of all the points. Using the K-means algorithm is often on its popularity, having no time and space complexity (Madhulatha, 2012).

PCA has been used to determine the optimum number of clusters to complement cluster analysis and investigate patterns of genetic diversity (Mohammadi and Prasanna, 2003). The genotypes located in one clump differed from the other, whereas similarity prevailed within the cluster (Figures 3 and 5). The genotypes were in different aggregates based on dissimilarity. The genotypes of cluster I were more diverse than the genotypes of cluster V, whereas cluster I and cluster II had less diversity compared with the earlier one. With a smaller inter-cluster distance between cluster I and cluster II, the genotypic diversity was lower.

Similarly, the higher the inter-cluster distance, the higher the genotypic diversity (Figure 5). The most complex and quantitative trait of rice is yield, and it has influences from many genetic and environmental factors. BLUP is a technique for estimating heritable components by increasing the relationship between ideal and predicted breeding values by reducing the prediction error variance (PEV) (Quddus *et al*., 2019). The best breeding lines can gain selection to form new breeding populations combined with pedigree information, sometimes called pedigree BLUPs, to offer EBVs. Parental lines' preference based on EBVs helps improve genetic advancement (Bernardo, 2010). In the conventional phenotypic-based QTL recognition method, the major constraint is in the study design comprising only one generation and frequently involving a single full-sib family where both environmental and genetic elements connect. EBVs can replace the phenotypic values.

Phenotyping of the target population is necessary for estimating EBVs with the ranking of breeding lines in this method (Rodriguez *et al*., 2019). Breeding lines for improvement programs require further selection based on EBVs and diversity. It will be more successful or effective for more genetic gain after choosing when the genotypes having higher EBVs and the most diverse genotypes are options. For example, IR 107971-B-B RGA-B RGA-202 (cluster II) and IR 108000-B-B-B-B-13 (cluster III), having higher EBVs but very close diversity, are less effective as parents among the top 20 genotypes. Both IR 103309-B-B RGA-B RGA-194 and IR 107982-B-41-1-2-1, having higher EBVs and better variety, are more effective as parents because one is in cluster IV and the other in cluster I among the top 20 genotypes (Table 4 and Figure 5). Higher EBVs and variation are typically desirable for selecting breeding materials as parents for succeeding breeding programs. Thus, preferred parents should continue for further breeding programs.

CONCLUSIONS

Any crop breeding program must have a set of superior parental lines to generate high-performing genotypes. Identifying genotypes with potential and desirable agronomic features for further breeding programs has several strategies aiding it. In this study, the most significant methods of parent selection were genetic diversity and estimated breeding values. The first five principal components contributed more than 70% (75.1%) of the variation. PC1 accounted for 23.3% variation, which was higher than other components. Five clusters of 353 rice genotypes over three environments included 66, 51, 83, 79, and 74 genotypes in clusters I, II, III, IV, and V, respectively. Besides, the top 20 genotypes, distributed into five clusters over three environments, included nine, three, four, three, and one genotypes in clusters I, II, III, IV, and V, respectively. The highest EBV was 0.395 for IR 107971-B-B RGA-B RGA-202, followed by 0.329, 0.321, 0.291, 0.264, and 0.234 for IR 108000-B-B-B-B-13, IR 103309-B-B RGA-B RGA-194, IR 107982-B-41-1-2-1, IRRI 174, and IR 107976-B-B RGA-B RGA-254, respectively. These breeding lines can be suitable as future parental materials. The genotypes containing higher EBVs or above the average value are the best options to increase the genetic gain. The genotypes having higher EBVs and more diversity will be more effective as parents for the succeeding breeding program. The IR 103309-B-B RGA-B RGA-194 and IR 107982-B-41-1-2-1, having higher EBVs and more divergence, will be better effective as parents.

ACKNOWLEDGMENTS

The Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh, has provided logistical support, which the authors gratefully recognized. The project named "Establishment of Breeding Zone Trial in Bangladesh for Favorable Breeding Environment" received funding from the International Rice Research Institute, also acknowledged by the authors.

REFERENCES

BBS (2019). Bangladesh Bureau of Statistics, Agriculture Wing. Government of the People's Republic of Bangladesh. *Parishankhyan Bhaban, E-27/A, Agargaon*, Dhaka-1207.

Alshugeairy ZK, Alogaidi FF, Hadi BH, Hassan WA (2023). Path coefficient analyses of introduced rice varieties under different planting distances. *SABRAO J. Breed. Genet*. 55(2): 516-524. http://doi.org/10.54910/sabrao2023.55.2.23.

BBS (2021). General Features and Environment, Statistical Yearbook Bangladesh 2020, 40th ed., Bangladesh Bureau of Statistics, Statistics and Informatics Division, *Ministry of Planning*, Dhaka, Bangladesh, pp. 20-22.

Bernardo R (2010). Genome-wide selection with minimal crossing in self-pollinated crops. *Crop Sci*. *50*(2): 624-627.

BRRI (2019). Modern Rice Cultivation (Adunik Dhaner Chash), 22th ed., Bangladesh Rice Research Institute, Joydebpur, Gazipur-1701, Bangladesh, pp. 1-100.

HIES (Household Income and Expenditure Survey) (2016). Bangladesh Bureau of Statistics. Government of Bangladesh, Dhaka.

Jolliffe IT, Cadima J (2016). Principal component analysis: A review and recent developments. *Philos. Trans. Royal. Soc. A. Math. Phys. Eng. Sci. 374* (2065): 20150202.

Juma RU, Bartholomé J, Prakash PT, Hussain W, Platten JD, Lopena V, Verdeprado H, Murori R, Ndayiragije AN, Katiyar SK, Islam MR, Biswas PS, Rutkoski JE, Arbelaez JD, Mbute FN, Miano DW, Cobb JN (2021). Identification of an elite core panel as a key breeding resource to accelerate the rate of genetic improvement for irrigated rice. *Rice* 14: 92(2021). https://doi.org/10.1186/s12284-021-00533-5.

Khush GS, Virk PS (2005). IR varieties and their impact. *Int. Rice Res. Inst.*, Los Banos, Philippines, pp. 163.

Madhulatha TS (2012). An overview on clustering methods. *J. Eng. Sci*. *2*(4): 719-725.

Mohammadi SA, Prasanna BM (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Sci. 43*(4): 1235-1248.

Mohan YC, Thippeswamy S, Bhoomeshwar K, Madhavilatha B, Samreen J (2015). Diversity analysis for yield and gall midge resistance in rice (*Oryza sativa* L.) in Northern Telangana Zone, India. *SABRAO J. Breed. Genet.* 47: 160-171

Popescu A (2014). Research on a simplified mixed model versus contemporary comparison used in breeding value estimation and bulls classification for milk production characters. *Sci. Papers*. *14*(3): 263.

Quddus MR, Rahman MA, Jahan N, Debsharma SK, Disha RF, Hasan MM, Aditya TL, Iftekharuddaula KM, Collard BC (2019). Estimating pedigree-based breeding values and stability parameters of elite rice breeding lines for yield under salt stress during the boro season in Bangladesh. *Plant. Breed. Biotechnol*. *7*(3): 257-271.

Ringnér M (2008). What is principal component analysis? *Nat. Biotechnol*. *26*(3): 303-304.

Rodriguez CA, Li Z, Hallingback HR, Sillanpaa MJ, Wu HX, Abrahamsson S, Garcia-Gil MR (2019). Analysis of phenotypic and estimated breeding values (EBV) to dissect the genetic architecture of complex traits in a Scots pine three-generation pedigree design. *J. Theor. Biol*. *462*: 283-292.

Simu S, Uddin MJ, Majumder RK, Zaman MN, Rahman MA, Abul M (2018). Speciation analysis of elements of soil samples by XRF in Gazipur industrial area, Bangladesh. *Int. J. Sci. Res. Rci. Eng. Technol*. *3*(3): 2456-5628.

Suela MM, Lima LP, Azevedo CF, Resende MDVD, Nascimento M (2019). Combined index of genomic prediction methods applied to productivity. *Cienc. Rural*. *49*(6): 1-9.

Zayed B, Bassiouni S, Okasha A, Abdelhamed M, Soltan S, Negm M (2023). Path coefficient, eigenvalues, and genetic parameters in Egyptian rice (*Oryza sativa* L.) under aerobic conditions. *SABRAO J. Breed. Genet*. 55(1): 131-145. http://doi.org/10.54910/sabrao2023.55.1.13.