

Original Research Article

Assessment of genetic diversity in aromatic short grain rice (*Oryza sativa* L.) genotypes using PCA and cluster analysis

ABSTRACT

A population panel of 90 aromatic short grain rice accessions were evaluated for 26 agro-morphological and quality traits using principal component analysis (PCA) and cluster analysis for the determination of genetic variation pattern, and identification of the major traits contributing to the diversity. First six principal components (PCs) exhibited Eigenvalue more than one with 74.4 per cent of total variability among the 26 characters. The PC1 showed 24.55% while, PC2, PC3, PC4, PC5 and PC6 exhibited 15.48 %, 11.48 %, 9.96 %, 7.89 % and 5.12 % variability, respectively among the accessions for the traits under study. The results of PCA suggested that characters such as effective tillers per plant, number of spikelets per panicle, number of filled spikelets per panicle, spikelet fertility %, milling %, head rice recovery %, kernel length and kernel length after cooking were the principal discriminatory characteristics of aromatic short grain accessions of rice. **Nine divergent clusters** were formed by UPGMA clustering method. The pattern of group constellation proved the existence of significant amount of variability. The intra cluster distance ranged from 0.00 (cluster VI) to 6.33 (cluster V). The inter cluster distance was maximum between cluster VI and VII (18.854) and minimum between cluster II and cluster IV (7.673). To realize much variability and high heterotic effect, parents should be selected from two clusters having wider inter-cluster distance. Considering the importance of genetic distance and relative contribution of characters towards total divergence, the present study indicated that parental lines selected from cluster VI (IGSR -3-1-5) for number of spikelets per panicle, number of filled spikelets per panicle, grain length, kernel length and length breadth ratio, and from cluster VII (Khasakani, Kolijoha) for effective tillers per plant, 1000 grain weight, grain yield per plant, harvest index, grain breadth, length breadth ratio after cooking and elongation index could be used in crossing programme to achieve desired segregants.

Keywords: Aromatic, Agro-morphological, PCA, Cluster analysis, Quality traits.

1. INTRODUCTION

Rice (*Oryza sativa* L. 2n=24) belongs to the family Graminae and sub family Oryzoidea. Rice is the food for more than half the world's population of 7.8 billion. Besides dominating as an indispensable food component in Asia, rice is rapidly emerging as the chief food in Latin America and Africa. In India, it is a dominating staple food crop of fertile and alluvial soils of North-East India, particularly the Indo-gangetic plains [1]. However, it is grown in all states and in all ecologies ranging from high mountains to lowlands and saline coastal areas.

India is very rich in rice genetic resources in general and aromatic rice particular. Aromatic rice constitutes small and special group of rice and highly priced due to their outstanding cooking quality. In India, aromatic rice is popularly known as Basmati rice. Due to low production and high demand of Basmati rice in International market, about two-third of Basmati produced in the country is exported to different countries and it is increasing every year [2]. Presently, the valuable genetic wealth of aromatic and fine-rice genotypes is being eroded because of their poor yield and the introduction of high-yielding varieties [3]. This has resulted in the shortage of Basmati rice for our internal consumption. Besides Basmati rice, hundreds of aromatic short grain rice is grown in specialized pockets in the states like Bihar, Odisha, Madhya Pradesh, West Bengal, Chhattisgarh, Uttar Pradesh etc. [4-5]. These are short grained and having good aroma retention for longer time when grown in prevailing sub-tropical warm climate of the state [4]. Aroma and taste of some of the short grained aromatic rice is known to be superior to Basmati types. Realizing the demand of aromatic rice in market and to improve the economy of farmer efforts are being started to collect, characterize and evaluate the aromatic short grain rice in the state of India for documentation and to find suitable donors for different traits which are prerequisite for varietal development. Characterization of these landraces is a prerequisite for understanding the extent of diversity, identification of valuable traits required for aromatic rice improvement.

Multivariate statistical tools have found extensive use in summarizing and describing the inherent variation among crop genotypes. One of the tools includes Principal Component Analysis (PCA) is a method to reduce the dimensions of a dataset by condensing a number of variables into a smaller subset that preserves most of the information present in the original set [6]. This technique identifies plant traits that characterize the distinctness among selected genotypes. These are often extended to the classification of a population into groups of distinct orders based on similarities in one or more characters, and thus guide in the choice of parents for hybridization [7]. Cluster analysis is also a multivariate method which aims to classify a sample of subjects (or objects) on the basis of a set of measured variables into a number of different groups such that similar subjects are placed in the same group.

Considering the above points, the present study was carried out to capture the potential genetic diversity among aromatic short grain rice accessions by using principal component analysis and cluster analysis and the results have been used in selection of appropriate parents for breeding program.

2. MATERIALS AND METHODS

Plant materials and field experiment: Plant material for the present investigations consisted of Ninety accessions of rice (*Oryza sativa* L.) including six checks viz., Dubraj, Vishnubhog, Bisni, Jawaphool, Badshahbhog and Indira Sugandhit Dhan1 (**Table 1**). The material was grown in

Augmented Completely Randomized Block Design at Research cum Instructional Farm, Department of Genetics and Plant Breeding, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh (India). The experimental material was planted in four blocks and each block comprised of 27 genotypes. Each entry was transplanted in a plot comprising four rows having twometer length at spacing of 20 cm between rows and 15 cm between plants. Check varieties were randomized within the block. The recommended agronomical practices were adopted to raise good crop in the season. Gap filling was done within a week in order to maintain uniform plant population. The standard agronomic practices were adopted for normal crop growth. Field observations were recorded on five randomly chosen plants of each genotype for various agronomical parameters. The quality analysis with the same experimental material was performed at Indian Institute of Rice Research (IIRR), Rajendranagar, Hyderabad (India).

Table 1. List of ninety aromatic short grain accessions of rice along with six checks

S. No.	Accession Name	S. No.	Accession Name	S. No.	Accession Name
1.	Banspatri	31.	Ganjekalli	61.	Kanakjeer- B
2.	Kankjeer A	32.	Neelabati	62.	Nanu
3.	Dhania – B2	33.	Hankesh	63.	Chittimutyalu
4.	RAU 3061	34.	AS GPC – 14	64.	Tulasikanthi
5.	RAU 3036	35.	AS GPC – 19	65.	Tulasighanti
6.	RAU 3048	36.	AS GPC – 38	66.	Khasakani
7.	RAU 3041	37.	Bayasabhog	67.	Achara mati
8.	RAU 3044	38.	R-1462-243-100-7-1-1	68.	Mayur kranti
9.	Gangaballi	39.	IGSR-2-1-46	69.	Mohan bhog
10.	Dhoiabankoi	40.	Loungchoosi A	70.	Randhunipagal
11.	Nalidhan	41.	Loungchoosi B	71.	Danaguri
12.	Barikunja	42.	Banthaphool A	72.	BR-34
13.	Bansadhan	43.	Boga joha	73.	Kedagauri
14.	Basuabhava	44.	Govind bhog	74.	Gayasu
15.	Parijatak	45.	Tarurboga	75.	Kolijoha
16.	Mahulakuchi	46.	Dhana Prasad	76.	Kalia
17.	Thakurabhog	47.	Jeeraga samba	77.	Dangar chudi
18.	Athmashital	48.	Kadam phool	78.	Rani kajal
19.	Amrit bhog	49.	Tulsimanjari	79.	Bastul
20.	RAU 3056	50.	Ganga baru	80.	Lilavati
21.	Kankjeer	51.	Bhainsapunchhi	81.	Basmati
22.	RAU 3076	52.	Loktimachi	82.	Chatianaki
23.	Kanikabhog	53.	Samudratan	83.	Heerakani
24.	Heerakani	54.	Nawab bhog	84.	Saragadhulli
25.	Basnaparijat	55.	Shrikamal	85.	Dubraj (C)
26.	Jalaka	56.	Tilkasturi	86.	Vishnubhog (C)
27.	Lectimanchi –A	57.	Chittimutyalu (bold grain)	87.	Bisni (C)
28.	Kheersai	58.	Badshaha	88.	Jawaphool (C)
29.	IGSR -3-1-5	59.	Bishnubhoga	89.	Badshahbhog (C)
30.	NDR IRRI 67	60.	RB 2816	90.	Indira Sugandhit Dhan-1(C)

Phenotyping of agro-morphological and phenological traits: Twenty six agro-morphological and quality traits were observed according to methods in Standard Evaluation System for Rice [8]. Agro-morphological variables considered in the multivariate analyses were *viz.*, days to 50% flowering (DTF), flag leaf length (FLL), flag leaf width (FLW), plant height (PH), panicle length (PL), effective tillers per plant (ET), number of spikelets per panicle (NOS), number of filled spikelets per panicle (NOFS), spikelet fertility % (SFP), 1000 grain weight (TGW), grain yield per plant (GYPP), biological yield per plant (BYPP), harvest index (HI) and grain quality traits *viz.*, grain length (GL), grain breadth (GB), hulling % (H %), milling % (M %), head rice recovery % (HRR%), kernel length (KL), kernel breadth (KB), length breadth ratio (LBR), kernel length after cooking (KLAC), kernel breadth after cooking (KBAC), length breadth ratio after cooking (LBRAC), kernel elongation ratio (KER) and elongation index (EI).

Statistical analysis: The observations recorded were statistically analyzed using XLSTAT software. Principal Component Analysis (PCA) was performed to reveal the patterns of data matrix for determination of selection criteria and identification of elite genotypes. The principal component analysis was computed using the following equation:

PCA

p

$$PC1 = \sum_{j=1}^p a_j X_j$$

1

where; PC = Principal component, a_j = Linear coefficient – Eigen vectors

Those PCs with Eigen values greater than one were selected as proposed by Jeffers [9]. Cluster analysis was carried out based on genetic distance matrix applying the UPGMA (Unweighted Pair-Group Method using Arithmetic average) clustering method [10].

3. RESULTS AND DISCUSSION

Principal Component Analysis

PCA is useful for identification of determinants of quantitative trait variability when a large number of accessions are to be accessed for several characters of morphological and agronomic importance. PCA is a statistical analysis method that converts many indicators into a few composite indicators, making complex problems simple and their analysis intuitive through dimensionality reduction [11-12]. In this study, twenty-six agro-morphological and quality contributing traits in aromatic short grain accessions of rice (**Table 2**) were simplified into 6 mutually independent principal components (PCs) using PCA as per the criteria set by Brejda et al. [13]. The PCs with eigen values >1 and which explained at least 5 % of the variation in the data were considered in the present study. The PC with higher eigen values and variables which had high factor loading were considered as best representative of system attributes. Out of 26, only six principal components (PCs) exhibited more than 1.33 eigen value, and showed about 74.4 % cumulative variability among the traits studied. So, these 6 PCs were given due importance for further explanation. The PC1 showed 24.55% while, PC2, PC3, PC4, PC5 and PC6 exhibited 15.48 %, 11.48 %, 9.96 %, 7.89 % and 5.12 % variability, respectively among the accessions for the traits under study. The first PC accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining

variability as possible. The PCA usefulness in measuring of diversity in genotypes also reported by Sharma et al. [14]; Christina et al. [15]; Saha et. al [6].

Table 2. Principal component analysis of 26 agro-morphological and quality traits for 90 aromatic short grain accessions of rice

Traits	Components					
	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Days to 50% flowering	0.599	-0.374	0.066	-0.062	0.151	0.162
Flag leaf length (cm)	-0.441	-0.192	0.320	-0.155	0.161	-0.169
Flag leaf width (cm)	-0.653	-0.042	0.323	-0.121	0.204	-0.044
Plant height (cm)	0.305	-0.467	0.161	-0.429	0.324	0.283
Panicle length (cm)	0.105	-0.411	0.232	-0.335	0.391	0.000
Effective tillers per plant	-0.236	0.116	-0.216	0.604	-0.303	-0.037
Number of spikelets per panicle	-0.084	-0.521	-0.056	0.356	0.613	-0.269
Number of filled spikelets per panicle	-0.007	-0.655	-0.056	0.263	0.600	-0.039
Spikelet fertility %	0.186	-0.362	-0.002	-0.249	-0.047	0.650
1000 grain weight (g)	-0.723	0.382	-0.031	-0.339	-0.067	0.229
Grain yield per plant (g)	-0.697	-0.148	-0.202	0.318	0.121	0.447
Biological yield per plant (g)	-0.397	-0.256	-0.032	0.414	0.252	0.128
Harvest index	-0.599	0.021	-0.279	0.004	-0.136	0.426
Grain length (mm)	-0.528	0.528	-0.165	-0.162	0.456	0.069
Grain breadth (mm)	-0.701	-0.170	0.321	-0.403	-0.173	-0.074
Hulling %	0.556	0.088	0.204	-0.357	0.043	0.278
Milling %	0.815	-0.087	0.183	-0.239	-0.060	-0.025
Head rice recovery %	0.799	-0.124	0.097	-0.199	-0.236	-0.131
Kernel length (mm)	0.009	0.735	-0.164	-0.382	0.403	-0.151
Kernel breadth (mm)	-0.677	-0.241	0.336	-0.405	-0.124	-0.213
Length breadth ratio	0.475	0.675	-0.373	0.006	0.361	0.077
Kernel length after cooking (mm)	-0.054	0.733	0.545	0.033	0.293	0.048
Kernel breadth after cooking (mm)	-0.685	-0.051	-0.099	-0.443	-0.011	-0.133
Length breadth ratio after cooking	0.343	0.649	0.520	0.270	0.238	0.132
Kernel elongation ratio	-0.054	0.153	0.812	0.403	-0.046	0.165
Elongation index	-0.145	-0.060	0.888	0.289	-0.141	0.042
Eigen value	6.384	4.025	2.984	2.590	2.051	1.330
% variation	24.553	15.480	11.476	9.960	7.887	5.116
Cumulative % variation	24.553	40.033	51.508	61.468	69.355	74.472

Values in bold represent highly weighted factors in respective PC

The result of the PCA explained the genetic diversity of the aromatic short grain accessions of rice. Higher the coefficients regardless of the direction (positive or negative) more effective they will be in discriminating between accessions [16]. Within each PC, only highly loaded factors or traits (having absolute values within 10 % of the highest factor loading) were retained for further explanation. Component matrix revealed that the PC1 which accounted for the highest variability (24.55 %) was mostly related with traits such as milling % (0.815) and head rice recovery % (0.799) (**Table 2**). As a result, the first component differentiated those accessions that have high milling recovery. The second principal component accounted for 15.48 % of total variance. Variables highly and positively correlated were kernel length, length breadth ratio and kernel length after cooking. The second component thus identified good cooking quality variables presenting positive contributions and the main characters responsible for quality characterization. The third principal component accounted for 11.48 % of the variability and was highly loaded with two most important cooking quality characters viz. kernel elongation and elongation index.

The PC4 was positively more related with effective tillers per plant, while PC5 was highly loaded with number of spikelets per panicle and number of filled spikelets per panicle. The PC6 which account for only 5.11 % variability was highly related with spikelet fertility percent. Thus, the prominent characters coming together in different principal components and contributing towards explaining the variability have the tendency to remain together which may be kept into consideration during utilization of these characters in breeding program. From the first six PCs, it was cleared that the PC1, PC2 and PC3 mostly related to quality characters while PC4, PC5 and PC6 mostly associated with yield related traits. So, for quality aspect a good breeding programme can be initiated by selecting the accessions from PC1, PC2 and PC3. Similar type of finding also reported by Saibu et al. [17]; Salem et al. [18]; Debsharma et al. [19]. Chakravorty et al. [20] identified six principal components with eigen value greater than 1.0 and that explained 75.9 % of the total cumulative variance within the axes further strengthen current results. Excellent breeding lines were identified for different quality traits based on PCA [21].

Top 10 principal component scores (PC scores) for all the accessions were estimated in six principal components and presented in **Table 3**. These scores can be utilized to propose precise selection indices whose intensity can be decided by variability explained by each of the principal component. High PC score for a particular accession in a particular component denotes high values for the variables in that particular accession.

Table 3: Principal component score of different accessions of aromatic short grain rice

Accessions Name	Score					
	PC1	PC2	PC3	PC4	PC5	PC6
Banspatri	-0.720	3.797	0.398	-0.318	-1.149	-1.901
Kankjeer A	3.999	1.827	-1.189	-1.923	0.731	0.425
Dhania – B2	3.578	-1.205	-0.416	1.389	0.188	0.271
RAU 3061	-0.342	2.635	1.171	0.650	-1.397	-2.028
RAU 3036	-0.035	1.689	-1.438	1.911	-1.663	-2.282
RAU 3048	-5.822	4.302	-0.340	1.642	2.682	1.384
RAU 3041	-1.604	-0.049	0.177	2.781	-1.859	-1.337
RAU 3044	1.525	-1.303	-1.048	-1.180	-0.903	0.154
Gangaballi	2.668	1.219	0.107	0.573	0.579	-0.201

Dhoiabankoi	2.242	0.715	-0.398	-2.159	-1.494	-0.129
Nalidhan	-4.526	-3.750	-1.190	-2.021	-0.296	-1.433
Barikunja	1.308	1.371	1.899	-0.394	-0.407	-1.757
Bansadhan	-0.299	-0.414	0.845	2.282	-1.158	1.401
Basuabhava	0.285	-2.496	-0.092	1.270	1.172	0.721
Parijatak	2.836	-0.653	1.310	-0.532	0.177	0.428
Mahulakuchi	0.290	-0.906	1.205	0.351	-0.858	1.154
Thakurabhog	1.288	-1.626	-0.122	0.449	1.620	0.233
Athmashital	-1.828	-2.518	-0.922	1.508	-0.308	0.513
Amrit bhog	0.362	-1.772	-0.193	-0.152	1.042	-1.596
RAU 3056	-1.417	0.833	1.004	4.104	-2.062	-0.944
Kankjeer	-2.291	0.834	1.258	3.290	-1.687	-0.463
RAU 3076	-1.758	1.205	-1.106	2.875	-0.979	-0.888
Kanikabhog	3.858	-0.491	0.655	1.115	0.859	1.403
Heerakani	1.525	0.845	1.715	-0.105	-1.279	1.461
Basnaparijat	1.525	0.845	1.715	-0.105	-1.279	1.461
Jalaka	-0.254	0.915	-1.057	-0.877	1.112	-2.370
Lectimanchi –A	2.064	0.147	-2.610	-0.786	-0.585	-0.094
Kheersai	0.849	0.201	-1.799	0.209	-0.286	0.445
IGSR -3-1-5	-1.852	-0.872	-3.184	1.713	3.538	-1.371
NDR IRRI 67	2.596	2.348	-2.092	-0.308	1.451	1.265
Ganjekalli	0.955	-0.875	2.210	-0.681	-1.228	-0.350
Neelabati	1.389	-1.217	3.170	-0.597	0.001	1.108
Hankesh	1.059	1.230	2.151	2.597	-0.451	-3.045
AS GPC – 14	1.605	2.045	-1.892	1.595	-1.045	-0.598
AS GPC – 19	-0.604	-1.570	0.438	0.659	0.585	0.893
AS GPC - 38	1.064	0.418	1.656	-0.782	0.236	0.088
Bayasabhog	-2.294	3.945	-0.829	0.815	2.410	0.912
R-1462-243-100-7-1-1	1.784	-0.932	0.800	-1.216	-0.476	0.508
IGSR-2-1-46	0.797	-2.379	0.168	-1.317	-0.236	0.224
Loungchoosi A	-0.936	-1.521	-4.196	-0.593	2.428	-0.042
Loungchoosi B	0.320	1.332	-0.340	1.331	1.826	1.991
Banthaphool A	1.444	-0.659	0.407	-1.287	0.443	-1.537
Boga joha	-2.081	-1.740	2.230	-2.966	0.326	-0.964
Govind bhog	2.632	0.420	-0.417	0.758	-1.328	0.555
Tarurboga	-0.823	0.530	-1.640	-0.704	-0.547	-3.411
Dhana Prasad	0.478	1.857	2.070	-1.560	2.002	0.253
Jeeraga samba	0.498	-3.651	-0.067	1.117	0.613	0.958
Kadam phool	-2.431	-0.218	-0.106	0.328	-0.945	-0.866
Tulsimanjari	2.424	-0.690	-1.095	-1.923	0.911	-0.496
Ganga baru	2.566	-2.476	-1.286	-1.286	-0.313	-0.424

Bhainsapunchhi	-6.169	-2.545	4.073	-1.736	0.677	0.101
Loktimachi	-4.822	-0.452	6.017	1.261	1.941	-0.142
Samudratan	2.561	1.935	-0.375	-2.066	-0.083	-0.628
Nawab bhog	-0.107	-1.955	1.371	1.978	1.968	0.876
Shrikamal	-1.037	-1.913	2.303	-1.037	-0.829	-0.962
Tilkasturi	0.427	-2.581	1.976	3.511	1.639	-0.049
Chittimutyalu (bold grain)	-0.725	1.281	1.681	1.629	0.829	-1.975
Badshaha	3.586	1.440	0.312	-0.544	-1.262	-1.618
Bishnubhoga	-2.371	-1.909	0.012	-1.203	-1.176	0.385
RB 2816	0.507	-1.883	1.823	-2.013	-0.775	-0.521
Kanakjeer- B	2.955	-1.294	-0.556	0.304	-0.531	0.204
Nanu	-0.374	-4.943	0.438	1.588	0.227	-0.777
Chittimutyalu	1.770	1.298	-2.317	-2.194	0.213	-0.323
Tulasikanthi	-1.210	-1.861	-0.718	-1.866	0.967	0.329
Tulasighanti	0.641	-1.333	-0.434	-1.670	-0.229	0.080
Khasakani	-6.706	5.309	1.222	0.469	-1.623	1.533
Achara mati	-5.811	-3.478	-2.761	0.314	-0.632	1.217
Mayur kranti	2.258	0.449	2.115	-0.629	-2.873	1.967
Mohan bhog	2.494	-0.946	-2.017	2.184	-0.161	1.520
Randhunipagal	3.257	1.082	1.033	2.061	-1.193	1.246
Danaguri	2.458	1.567	0.430	1.279	-0.441	0.834
BR-34	0.673	-1.568	-2.033	1.743	0.850	0.464
Kedagauri	0.845	1.115	-1.227	-1.591	3.179	-1.503
Gayasu	1.699	3.867	-0.230	-1.322	1.456	-0.390
Kolijoha	0.303	2.467	-3.590	1.545	-4.374	0.884
Kalia	2.000	0.351	1.596	-2.416	0.651	-0.073
Dangar chudi	-7.017	1.543	-4.062	-2.149	-1.780	0.893
Rani kajal	0.382	-0.207	-0.269	-1.582	-0.706	-0.338
Bastul	-7.903	-1.133	-1.916	-2.808	-3.121	1.011
Lilavati	-0.747	0.092	0.698	-1.923	-0.501	0.522
Basmati	-0.174	-0.423	0.042	-0.363	-0.028	0.342
Chatianaki	-0.882	0.971	1.950	-3.395	-1.214	0.822
Heerakani	2.758	1.627	1.276	-0.900	0.717	1.600
Saragadhulli	0.453	-0.105	-0.812	0.729	-0.866	1.340
Dubraj (C)	-2.487	3.099	-0.209	-0.226	3.385	2.121
Vishnubhog (C)	-1.042	-1.192	-0.274	1.140	1.872	-0.947
Bisni (C)	-2.891	-1.414	-0.119	0.655	0.764	-0.156
Jawaphool (C)	0.943	-2.005	-3.130	1.464	0.786	-0.235
Badshahbhog (C)	1.772	-1.684	-1.467	-0.353	-0.743	0.771
Indira Sugandhit Dhan-1 (C)	-2.159	5.810	0.453	-1.412	2.302	-0.106

Figures in bold represent top 10 scores in each principal component

On the basis of top 10 PC scores in each principal component, accessions are selected and presented as summarized form in **Table 4**. Perusal of results revealed that the “Kanakjeer” A had highest PC score followed by Kanikabhog, Dhania-B2, Badshaha, Randhunipagal, Kanakjeer B, Parijatak, Heerkani, Gangabali and Govindbhogin PC1 indicated that they had high milling % and head rice recovery %. In PC2, Indira Sugandhit Dhan 1 had the highest score followed by Khasakani, RAU 3048, Byasabhog, Gayasu, Banspatri, Dubraj, RAU 3061, NDR IRRI 67 and AS GPC -14 for the highly loaded traits kernel length, kernel length after cooking and length breadth ratio. The highest PC score of Loktimachi followed by BhainsaPunchi, Neelabati, Shrikamal, Bogajoha, Ganjekali, Hankesh, Mayur Kranti, Dhana Prasad and Tilkasturi exhibited high kernel elongation ratio and elongation index. Similarly, decreasing order of PC scores of RAU 3056, Tilkasturi, Kanakjeer, RAU 3076, RAU 3041, Hankesh, Mohanbhog, Ranhunipagal, Nawab bhog and RAU 3036 in PC4 had high number of effective tillers per plant. In PC5, IGSR-3-1-5 had highest PC score followed by Dubraj, Kedaguri, Indira Sugandhit Dhan 1, RAU 3048, Loungchoosi A, Bayashabhog, Dhanaprasad, Loktimachi and Vishnubhog indicated that they have high value of number of spikelets per panicle and number of filled spikelets per panicle. Dubrajrecored highest PC score followed by Loungchoosi B, Mayurkranti, Heerakani, Khasakani, Bansadhan, Bansaparijat, Kanikabhog, RAU 3048 and Saragadhulli in PC6 for high spikelet fertility %. Similar findings were reported by Patel et al [22].

Table 4: List of selected accession in each principal component on the basis of top 10 PC score

PC1	PC2	PC3	PC4	PC5	PC6
Kanakjeer A	I.S. Dhan 1	Loktimachi	RAU 3056	IGSR-3-1-5	Dubraj
Kanikabhog	Khasakni	BhainsaPunchi	Tilkasturi	Dubraj	Loungchoosi B
Dhania B2	RAU 3048	Neelabati	Kanakjeer	Kedaguri	Mayurkranti
Badshaha	Bayasabhog	Shrikamal	RAU 3076	I.S. Dhan 1	Heerakani
Randhunipagal	Gayasu	Bogajoha	RAU 3041	RAU 3048	Khasakani
Kanakjeer B	Banspatri	Ganjekali	Hankesh	Longchoosi A	Bansadhan
Parijatak	Dubraj	Hankesh	Mohanbhog	Bayashabhog	Bansaparijat
Heerakani	RAU 3061	Mayurkranti	Randhunipagal	Dhanaprasad	Kanikabhog
Gangabali	NDR IRRI 67	Dhanaprasad	Nawab bhog	Loktimachi	RAU 3048
Govindbhog	AS GPC -14	Tilkasturi	RAU 3036	Vishnubhog	Saragadhuli

Cluster analysis

Cluster analysis groups large number of accessions into few number of homogenous clusters which in turn facilitates the selection of diverse accessions. Analysis performed by Un-weighted variable Pair Group Method of the Average Linkage Cluster Analysis (UPGMA) using Euclidean distance as dissimilarity measure divided the ninety aromatic short grain accessions of rice into seven clusters (**Table 5 and Fig.1**). The accessions were not evenly distributed among the clusters. The cluster II constituted of 39 accessions, forming the largest cluster followed by

cluster III (18 accessions), cluster I (17 accessions), cluster IV (9 accessions), clusters V (4 accessions), cluster VII (2 accessions) and cluster VI (only one accession). The pattern of group constellation proved the existence of significant amount of variability.

Table 5. Distribution of aromatic short grain accessions among seven clusters

Cluster No.	No. of accessions	Accessions Name
I	17	Banspatri, Dhoiabankoi, Mahulakuchi, Heerakani, Basnaparijat, Govind bhog, Dhana Prasad, Samudratan, Badshaha, Chittimutyalu, Mayur kranti, Randhunipagal, Danaguri, Gayasu, Dangar chudi, Bastul, Chatianaki
II	39	Kankjeer A, RAU 3044, Barikunja, Bansadhan, Basuabhava, Parijatak, Athmashital, Jalaka, Lectimanchi –A, Kheersai, NDR IRRI 67, Ganjekalli, Neelabati, AS GPC – 38, R-1462-243-100-7-1-1, IGSR-2-1-46, Loungchoosi B, Banthaphool A, Boga joha, Tulsimanjari, Ganga baru, Bhainsapunchhi, Loktimachi, Shrikamal, Bishnubhoga, RB 2816, Kanakjeer- B, Tulasikanthi, Tulasighanti, Mohan bhog, Kalia, Rani kajal, Lilavati, Basmati, Heerakani, Saragadhulli, Dubraj (Check 1), Bisni (Check 3), Badshahbhog (Check 5)
III	18	Dhania – B2, Gangaballi, Nalidhan ,hakurabhog, Amrit bhog, Kanikabhog, AS GPC – 19, Loungchoosi A, Jeeraga samba, Nawab bhog, Tilkasturi, Chittimutyalu (bold grain), Nanu , Achara mati, BR-34, Kedagauri, Vishnubhog (Check 2), Jawaphool (Check 4)
IV	9	RAU 3061, RAU 3036, RAU 3041, Kankjeer, RAU 3076, AS GPC – 14, Bayasabhog, Tarurboga, Indira Sugandhit Dhan-1 (Check 6)
V	4	RAU 3048, RAU 3056, Hankesh, Kadam phool
VI	1	IGSR -3-1-5
VII	2	Khasakani, Kolijoha

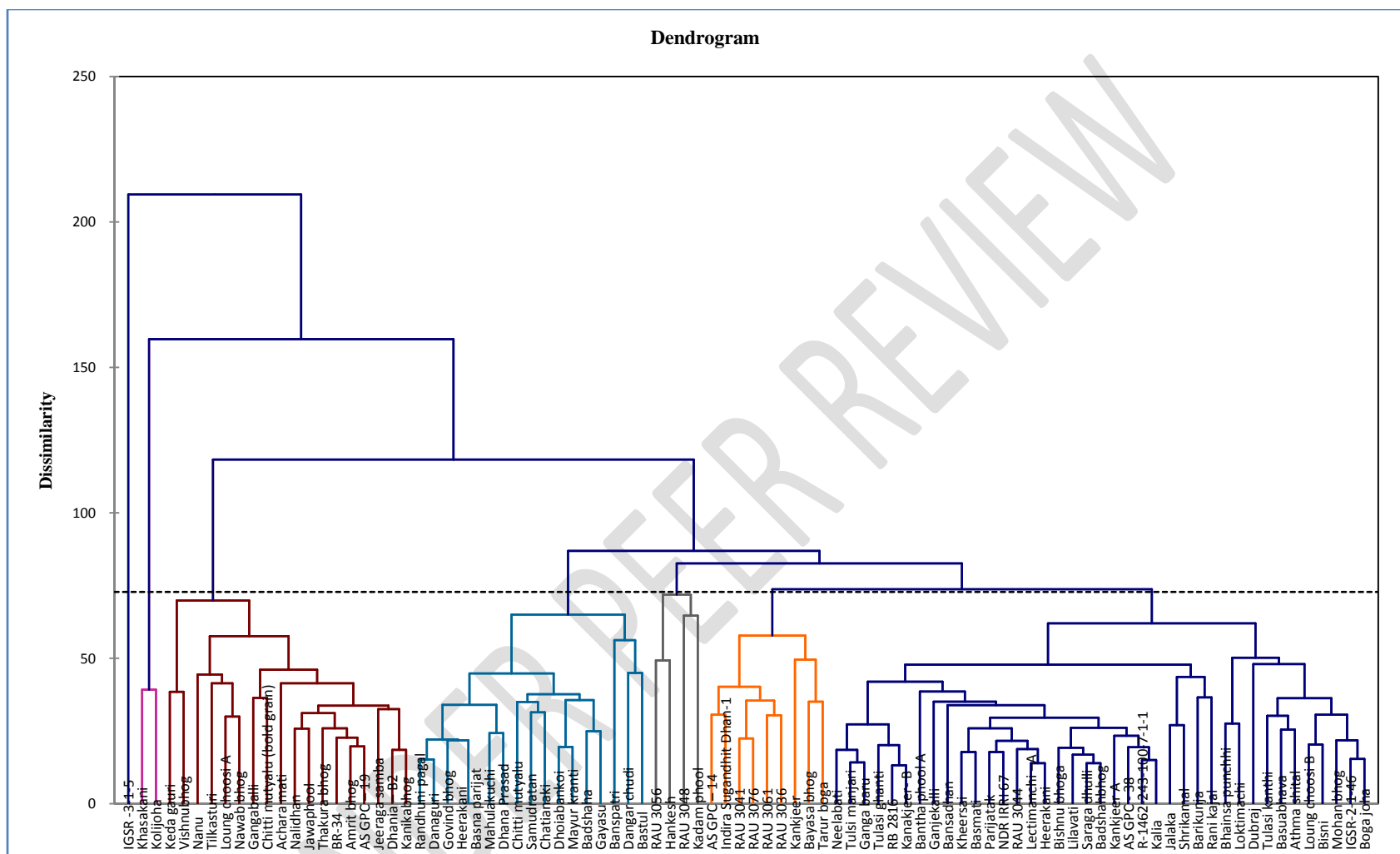


Fig. 1. Dendrogram of 90 aromatic short grain accessions derived by UPGMA from 26 agro-morphological and quality traits.

The inter- and intra-cluster distances among **nine clusters** were computed and have been given in **Table 6**. The intra cluster distance ranged from 0.00 (cluster VI) to 6.33 (cluster V). The inter cluster distance was maximum between cluster VI and VII (18.854) and minimum inter cluster distance was observed between cluster II and cluster IV (7.673). To realize much variability and high heterotic effect, parents should be selected from two clusters having wider inter-cluster distance.

Table 6. Estimates of intra (diagonal and bold) and inter-cluster distances among seven clusters

Class	I	II	III	IV	V	VI	VII
I	5.685	8.718	12.742	7.749	10.814	16.701	9.394
II		5.813	9.342	7.673	7.931	14.353	12.471
III			5.871	11.271	8.486	10.965	15.336
IV				5.668	8.282	15.395	10.660
V					6.330	13.102	13.240
VI						0.000	18.584
VII							4.429

The cluster mean values showed a wide range of variations for all the characters undertaken in the study (**Table 7**). Cluster II exhibited highest mean value for days to 50% flowering (115.562), plant height (157.539), panicle length (27.016), spikelet fertility % (86.393), Hulling % (80.139), Milling % (71.716), head rice recovery % (66.808) and kernel breadth (2.080), while cluster IV contained genotypes with the highest meanvalue for biological yield per plant (74.098). Cluster V recorded the highest value for flag leaf length (35.763), flag leaf width (1.555), kernellength after cooking (8.325), kernel breadth after cooking (2.901) and kernel elongation ratio (1.838), while highest mean value for number of spikelets perpanicle (364.933), number of filled spikelets per panicle (278.283), grain length (7.630), kernel length (4.817) and length breadth ratio (2.451) was recorded by cluster VI. Cluster VII had the highestvalue for effective tillers per plant (18.325), 1000 grain weight (20.365), grain yield per plant (22.708), harvest index (35.306), grain breadth (2.598), length breadth ratio after cooking (2.774) and elongation index (1.212). This result are in confirmation with the findings of Shiva Prasad *et al.* [23]; Kumar *et al.* [24] and Apsath Beevi and Venkatesan [25].

Table 7. Cluster mean values of seven clusters for different quality, yield and its attributing characters in aromatic short grain accessions of rice

Clusters/ Characters	I	II	III	IV	V	VI	VII
DTF	110.086	115.562	114.324	108.574	102.208	107.625	80.958
FLL (cm)	31.702	33.508	31.837	32.365	35.763	33.697	33.203
FLW (cm)	1.381	1.420	1.390	1.430	1.555	1.454	1.296

PH (cm)	152.688	157.539	150.523	118.578	110.102	131.251	108.497
PL (cm)	25.668	27.016	27.185	22.383	22.890	26.928	22.305
ET	12.727	12.189	13.197	14.439	16.233	13.525	18.325
NOS	138.973	194.703	263.400	180.372	229.608	364.933	91.167
NOFS	117.958	168.093	220.819	131.467	175.783	278.283	71.217
SFP (%)	85.343	86.393	84.167	72.194	77.683	76.153	83.435
TGW (g)	14.199	13.359	11.673	13.750	14.489	12.376	20.365
GYPP (g)	14.391	16.182	18.612	16.970	18.650	21.308	22.708
BYPP (g)	57.234	66.778	72.481	74.098	62.842	63.542	65.845
HI	25.320	24.122	25.934	23.674	28.761	34.060	35.306
GL (mm)	6.643	6.462	6.486	7.053	6.880	7.630	6.935
GB (mm)	2.455	2.504	2.421	2.421	2.415	2.262	2.598
H%	79.973	80.139	79.096	77.129	76.496	73.054	77.901
M%	71.263	71.716	70.621	67.351	66.274	68.867	64.668
HRR%	66.781	66.808	63.736	60.529	59.955	50.800	55.067
KL (mm)	4.596	4.451	4.284	4.636	4.550	4.817	4.675
KB (mm)	2.044	2.080	2.037	2.013	2.114	1.966	2.038
LBR	2.278	2.165	2.126	2.317	2.149	2.451	2.320
KLAC(mm)	7.720	7.404	7.071	8.007	8.325	6.958	7.975
KBAC (mm)	2.853	2.869	2.861	2.950	2.901	2.884	2.845
LBRAC	2.749	2.591	2.493	2.715	2.867	2.413	2.774
KER	1.686	1.665	1.652	1.732	1.838	1.436	1.685
EI	1.191	1.213	1.178	1.185	1.343	0.986	1.212

Abbreviations: DTF: Days to 50 % flowering; FLL: Flag leaf length; FLW: Flag leaf width; PH: Plant height; PL: Panicle length; ET: Effective tillers per plant; NOS: Number of spikelets per panicle; NFS: Number of filled spikelets per panicle; SFP: Spikelet fertility %; TGW: 1000 grain weight; GYPP: Grain yield per plant; BYPP: Biological yield per plant; HI: Harvest index; GL: Grain length; GB: Grain breadth; HP: Hulling %; MP: Milling %; HRR: Head rice recovery %; KL: Kernel length; KB: Kernel breadth; LBR: Length breadth ratio; KLAC: Kernel length after cooking; KBAC: Kernel breadth after cooking; LBRAC: Length breadth ratio after cooking; KER: Kernel elongation ratio; EI: Elongation index.

The selection and choice of parents mainly depends upon contribution of characters towards divergence. It is well known that crosses between divergent parents usually produce greater heterotic effect than between closely related ones. Considering the importance of genetic distance and relative contribution of characters towards total divergence, the present study indicated that parental lines selected from cluster VI (IGSR -3-1-5) for number of spikelets perpanicle, number of filled spikelets per panicle, grain length, kernel length and length breadth ratio, and from cluster VII (Khasakani, Kolijoha) for effective tillers per plant, 1000 grain weight, grain yield per plant, harvest index, grain breadth, length breadth ratio after cooking and elongation index could be used in crossing programmesto achieve desired segregants. The accessions possessing extreme phenotype may also be utilized in the development of mapping population for identification of quantitative trait loci [26].

4. CONCLUSION

The results of the present investigation indicate the ample amount of variation present in the rice germplasm studied gesturing towards the usefulness of selection in the improvement of yield and quality traits. From this study it may conclude that Principal component analysis highlights the characters viz., effective tillers per plant, number of spikelets per panicle, number of filled spikelets per panicle, spikelet fertility %, milling %, head rice recovery %, kernel length and kernel length after cooking in first six PCs with maximum variability which were most important in determining variation amongst different genotypes. Thus, intensive selection procedure can be designed to bring about rapid improvement of yield and quality traits. The promising lines exhibited desirable magnitude of specific traits studied can be used as donor in the rice breeding programme. It may further conclude that UPGMA method of clustering grouped the accessions into seven clusters. The pattern of constellation proved the existence of significant amount of variability. Cluster II constituted of 39 accessions, forming the largest cluster. To realize much variability and high heterotic effect, parents should be selected from two clusters having wider inter-cluster distance. Considering the importance of genetic distance and relative contribution of characters towards total divergence, parental lines should be selected from cluster VI (IGSR -3-1-5) and from cluster VII (Khasakani, Kolijoha) to achieve desired segregants.

REFERENCES

1. Walia US, WaliaMK. Scope of direct seeded rice in India, In: Proceedings of *Biennial Conference* of Indian Society of *Weed Science* on News and Emerging Issues in Weed Science. HAU, Hissar, India, November. 2007;2-3: 20.
2. Siddiq EA. Export prospects of Indian Basmati rice. *Indian Farming*, 1990;40(9): 45-47.
3. Islam MZ, Khalequzzaman M, Bashar MK, Ivy NA, Mian MAK, Pittendrigh BR, Haque MM, Ali MP. Variability Assessment of Aromatic Rice Germplasm by Pheno-Genomic traits and Population Structure Analysis. *Scientific Reports*. 2018;8:9911 DOI:10.1038/s41598-018-28001-z
4. Singh RK, Singh US, Khush GS, Rohila R, Singh JP, Singh G, Sekhar KS. Small and medium grained aromatic rice of India. In: *Aromatic rices* by Singh RK, Singh US, Khush GS (Eds.), Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, pp. 2000;154-174.
5. Roy PS, Samal R, Rao GJN, Patnaik SSC, Jambhulkar NN, Patnaik A, Mohapatra T. Differentiation and description of aromatic short grain rice landraces of eastern Indian state of Odisha based on qualitative phenotypic descriptors. *BMC Ecol* 2016;16:36 DOI 10.1186/s12898-016-0086-8
6. Saha SR, Ferdausi A, Hassan L, Haque MA, Begum SN, Yasmin F, Akram W. Rice landraces from haor areas of Bangladesh possess greater genetic diversity as revealed by morpho-molecular approaches along with grain quality traits. *Cogent Food & Agriculture*. 2022;8: 2075130. <https://doi.org/10.1080/23311932.2022.2075130>.
7. Nair NV, Balakrishnan R, Sreenivasan TV. Variability for quantitative traits in exotic hybrid germplasm of sugarcane. *Genetic Resources and Crop Evolution*, 1998;45: 459-464.
8. IRRI. Standard Evaluation System for Rice. International Rice Research Institute, Manila, Philippines. 1996.
9. Jeffers JNR. Two case studies in the application of principal component analysis. *Appl. Stat.*, 1967;16: 225-236

10. Michener CD, Sokal RR. A quantitative approach to a problem of classification. *Evolution*,1957;11: 490–499.
11. Jolliffe I. Principal Component Analysis. In Encyclopedia of Statistics in Behavioral Science. In: Everitt BS, Howell DC, editors. 2005. <https://doi.org/10.1002/0470013192.bsa.501>.
12. Ringnér M. What is principal component analysis? *Nat. Biotechnol.* 2008;26:303–304. <https://doi.org/10.1038/nbt0308-303>.
13. Brejda JJ, Moorman TB, Karlen DL, Dao TH. Identification of regional soil quality factors and indicators. I. Central and Southern High-Plains. *Soil Science Society of America Journal*, 2000;64: 2115–2124.
14. Sharma B, Biswas P, Parikh M. Multivariate analysis for yield and lodging resistant characteristics of different rice genotypes grown under the irrigated transplanted condition. *J. Pharmacognosy and Phytochemistry.* 2020;9 (6):417–425.
15. Christina GR, Thirumurugan T, Jeyaprakash P, Rajanbabu V. Principal component analysis of yield and yield related traits in rice *Oryza sativa* L. landraces. *Electronic Journal of Plant Breeding*, 2021;12(3): 907-911.
16. Sanni KA, Fawole I, Ogunbayo SA, Tia DD, Somado EA, Futakuchi K, Sie M, Nwilene FE, Guei RG. Multivariate analysis of diversity of landrace rice germplasm. *Crop Science*,2012;52: 494–504.
17. Shaibu AA, Uguru MI. Application of Multivariate Analysis to Access Selected Rice Germplasm Phenotypic Diversity. *Int. J Exp. Agri.* 2017;25:1-8.
18. Salem KF, Saleh MM, Aldahak L, Elabd AB. Assessment phenotypic diversity of rice (*Oryza sativa* L.) genotypes by multivariate analysis. *Journal of Arid and Agriculture.* 2021;7:52-59.
19. Debsharma SK, Syed MA, Ali MH, Maniruzzaman S, Roy PR, Brestic M, Gaber A, Hossain A. Harnessing on Genetic Variability and Diversity of Rice (*Oryza sativa* L.) Genotypes Based on Quantitative and Qualitative Traits for Desirable Crossing Materials. *Genes*.2023;14:10. <https://doi.org/10.3390/genes14010010>
20. Chakravorty A, Ghosh PD, Sahu PK. Multivariate analysis of landraces of rice of West Bengal. *American Journal of Experimental Agriculture*, 2013;3(1): 110-123.
21. Wang Q, Li X, Chen H, Wang F, Li Z, Zuo J, Fan M, Luo B, Feng P, Wang J. Mapping combined with principal component analysis identifies excellent lines with increased rice quality. *Scientific Reports.* 2022;12:5969. <https://doi.org/10.1038/s41598-022-09976-2>
22. Patel A, Sao A, Nair S, Mandavi J, Tamrakar N. Principal component analysis for eight quantitative traits in 55 indigenous rice germplasm (*Oryza sativa* L.) *The Pharma Innovation Journal* 2022;11(9): 1201-1206
23. Shiva Prasad G, Radha Krishna KV, Subba Rao LV, Chaithanya U. Quantitative analysis of rice genotypes (*Oryza Sativa* L.). *International Journal of Innovative Research and Development*, 2013;2(9): 14-17.
24. Kumar B, Gupta B, Singh B. Genetic diversity for morphological and quality traits in rice (*Oryza sativa* L.). *The Bioscan.* 2014;9(4): 1759-1762.
25. ApsathBeevi H, Venkatesan M. Genetic divergence studies in rice genotypes under saline condition. *International Journal of Current Advanced Research*, 2015;4(1): 6-8.
26. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. Evaluation of Rice genetic diversity and variability in a population panel by principal component analysis. *Indian journal of Science and technology*, 2014;7(10): 1555-1562.

UNDER PEER REVIEW