Assessment of genetic diversity in aromatic rice (Oryza sativa L.) germplasm using PCA and cluster analysis

D. Shoba, Roshni Vijayan, S. Robin, N. Manivannan, K. Iyanar, P. Arunachalam, N. Nadarajan, M. Arumugam Pillai and S. Geetha
Assessment of genetic diversity in aromatic rice (Oryza sativa L.) germplasm using PCA and cluster analysis

D. Shoba1*, Roshni Vijayan1, S. Robin2, N. Manivannan3, K. Iyanar3, P. Arunachalam3, N. Nadarajan3, M. Arumugam Pillai4 and S. Geetha1

1Department of Plant Breeding and Genetics, AC&RI, Killikulam
2The Regional Agricultural Research Station, Pattambi, Kerala
3Centre for Plant Breeding and Genetics, TNAU, Coimbatore 4. National Pulses Research Centre, Vamban
*E-Mail: shobatnau@gmail.com

(Received: 13 Mar 2019; Revised: 28 Jun 2019; Accepted: 13 Jul 2019)

Abstract
A total of 67 aromatic rice germplasm lines were evaluated for nine traits by PCA and cluster analysis for determining the pattern of variation, relationship among individuals and their characteristics. Out of nine PCs, four exhibited more than 1.0 eigen values and showed 70.14 per cent total variability among the characters. PC1 showed high proportion of total variation 25.9 per cent and PC2, PC3 and PC4 contributed 16.96 per cent, 15.06 per cent and 12.22 per cent of the total variance, respectively. Highest positive eigen vector was observed in grain width (0.46) followed by plant height (0.34) in PC1 indicated the large effect in the overall variation of the genotypes. From the cluster analysis, clusters I, II and III had desirable mean values for days to 50 per cent flowering, plant height, number of productive tillers/plant, grain length, grain width, 100 grain weight, panicle length and single plant yield. The genotypes in these clusters would be utilized for aromatic rice improvement breeding.

Key words
Aromatic rice, Genetic diversity, PCA, Cluster analysis

Introduction
Rice is vital to the lives of billions of people around the world. It is the staple food for 2.5 billion people and covering 9% of the earth’s arable land (International year of rice, 2004). It provides 21% of global human per capita energy and 15% of per capital protein (International Rice Research Institute; 2002). Calories in rice are largely important among the poor, where it accounts for 50-80% of daily caloric intake (International Rice Research Institute, 2001). Rice is grown in all the continents, except Antarctica occupying 165 million hectare area and producing 720 million tonnes (Anonymous, 2012). In India, it accounts for more than 43% of food grain production. It is grown in 44.8 million hectare under 4 major ecosystems: irrigated, rainfall lowland, rainfed upland and flood-prone with average annual production of 100 million tonnes (Sohgaura et al. 2015). Preference for quality rice has expanded from Southeast Asia to Europe, Africa and United States of America.

Aromatic or scented rice have long been highly regarded in Indian society not only because of their excellent quality but also because they are considered auspicious. Aromatic rice is known for its nut-like aroma and taste, which is caused by the chemical compound 2-acetyl-1-pyrroline (Ghareyazie et al. 1997). Many scented varieties of rice have been cultivated in the Indian sub-continent and Basmati distinguishes itself from all other aromatic rice due to its unique aromatic characteristics together with silky texture of its long grain. Aromatic rice have been traditionally cultivated in northern India, confined mainly to Punjab, Haryana, Western Uttar Pradesh and adjoining districts of Rajasthan.

Over the past few years, there has been a noteworthy increase in India’s foreign exchange earnings from export of Basmati rice. The incorporation of basmati quality in the high yielding background has been the most challenging one. It might be due to its complex nature with respect to its constitution and inheritance pattern of component indices (Yumnam et al. 2015). Hence, greater genetic variation with greater diversity in rice gene pool is necessary for utilization in breeding programmes (Sivaranjani et al. 2010). Genetic diversity serves as a way for populations to adapt to changing environments.

With more deviation, it is more likely that some individuals in a population will possess variations of alleles that are suited for the environment. Genetic distance estimates for population grouping can be estimated by different methods as it is critical to realize the usable
variability existing in the population (Nachimuthu et al. 2014) and one of the approaches is multivariate analysis. Multivariate statistical tools include Principal Component Analysis (PCA), cluster analysis and discriminate analysis (Oyelola, 2012) and PCA may be used to reveal patterns and eliminate redundancy in data sets as variations routinely occur in crop species for yield and grain quality (Maji et al. 2012). Principal components (PCs) are generally estimated either from correlation matrix or covariance matrix. Correlation matrix was used to extract the principal components in the present study. The primary benefit of PCA is to quantify the significance of each dimension for relating the variability of a data set.

Cluster analysis is a powerful tool for determining clustering patterns as a way to establish relationships between genetic divergence as well as it is a measure of estimating the cophenetic correlation coefficient. Cophenetic correlation coefficient is a measure of how realistically a dendrogram preserves the pair wise distances between the original unmodeled data points. Hence, the present investigation was carried out on rice aromatic germplasm with an objective of dissecting yield and quality related traits and to identify maximum variability into total variability for grouping the aromatic germplasm accessions by taking into account of several characteristics and relationship between them.

Materials and Methods
A total of sixty seven aromatic rice germplasm accessions collected from different sources were used for study (Table 1). The experiment was carried out at Department of Rice, TNAU during 2013-14. Each germplasm line was sown in plot size of 0.96 m² with spacing of 20 x 20 cm. The experiment was conducted by adopting Randomized Block Design (RBD) replicated twice. Ten randomly selected plants were taken for observations for each replication. Biometrical observations on days to 50% flowering, number of productive tillers per plant, plant height (cm), panicle length (cm), grain length(mm), grain width(mm), hundred grain weight(g) single plant yield(g) and presence of aroma were recorded. Grain length and grain width were measured using vernier caliper. Presence of aroma was also observed by sensory evaluation method suggested by Sood and Siddiq (1978).

In order to classify the patterns of variation, PCA was performed. Those PCs with Eigen values greater than one were selected as proposed by Jeffers (1967). Correlations between the original traits and the respective Principal Components (PCs) were calculated. Statistical analysis was carried out using the software STAR 2.0. The principal component analysis was computed using the following equation: PC1= ΣaXj where, ajXj - Linear coefficient – Eigen vectors. Cluster analysis was performed using agglomerative clustering method adopting Euclidean distance measure.

Results and Discussion
PCA is a powerful tool in modern data analysis because it is a well known multivariate statistical technique which is used to identify the minimum number of components, which can explain maximum variability out of the total variability (Anderson, 1972 and Morrison, 1978) and also to rank genotypes on the basis of PC scores. In the present study, PCA was performed using yield and quality related traits in 67 aromatic rice germplasm accessions. The proportion of variance, cumulative proportion and eigen values are given in Table 2. Out of nine PCs, four exhibited more than 1.0 eigen values and exhibited 70.14 % total variability among the characters studied. The principal components having more than one eigen value showed more variation among the aromatic rice germplasm accessions for the selection of the diverse parents. Among the four principal components, PC1 shared high proportion of total variation 25.9% and the rest of the three principal components viz., PC2, PC3 and PC4 contributed 16.96%, 15.06% and 12.22% of the total variance respectively. Ashfaq et al. (2012) resulted PC1 with 33% and PC2 (14.3%), PC3 (11.4%) and PC4 (9%) variability in forty rice genotypes. Maji and Shaibu (2012) reported that PCA resulted in the first two components with eigen value greater than 1 accounting for 78% of the total variation of sixteen agro-morphological data on twenty three rice germplasm lines. Ahmed et al. (2016) studied that PCA showed that first five components with vector values > 1 contributed 82.90% of the total variations in 31 rice germplasm lines. Gour et al. (2017) reported that out of thirty three traits studied in 83 rice genotypes which include traditional land races, fourteen PCs exhibited more than 0.5 eigen values and showed about 92.75% total variability among the characters studied. Pachauri et al. (2017) studied one hundred twenty four rice germplasm accessions on the basis of nineteen morphological and eleven agronomical traits. From their studies, PC1 expressed 37.12% variability, while PC2, PC3 and PC4 recorded 13.56, 11.04 and 10.76% variability respectively and the the traits such as number of effective tillers and 100 grain weight were the principal discriminatory traits.

Electronic Journal of Plant Breeding, 10 (3): 1095 - 1104 (Sep 2019) DOI: 10.5958/0975-928X.2019.00140.6 ISSN 0975-928X
Sree plot explained the percentage of variance associated with each PC obtained by drawing a graph between eigen values and PC numbers. In the present study, PC1 showed 25.9% variability with eigen value 2.33 which then declined gradually (Fig 1.). Semi curve line is obtained after PC3 tended to become straight with little variance observed in each PC. From the graph, it is clear that the maximum variation was observed in PC1 in comparison to other PCs. Hence, selection of lines from this PC would be useful for further breeding programs. Nachimuthu et al. 2014 reported the highest variability in PC1 with eigen value more than 1.0 in 192 rice genotypes comprising traditional landraces and exotic genotypes.

An eigenvector or characteristic vector of a linear transformation is a non zero vector that changes by only a scalar factor when that linear transformation is applied to it. Eigen value or characteristic value is associated with the eigenvector. To determine the critical limit for the coefficients of the proper vectors, coefficients with greater than 0.3 having large effect to be considered in the overall variation present in the genotypes (Raji, 2002). Eigenvectors of the PCs for different traits are presented in Table 3. The results showed that grain width had highest positive value (0.46) followed by plant height (0.34) in PC1. Plant height recorded highest positive value (0.59) followed by panicle length (0.57), days to 50% flowering (0.43) and presence of aroma (0.33) in PC3.

In the present investigation, the first two PCs (PC1 and PC2) were plotted against each other in biplot to observe the relation among the aromatic rice germplasm lines based on the observed yield and quality attributing characters (Fig.2). Genotypes viz., CANP 496, CANP 295, CANP 510, CANP 265, CANP 90, CANP 91, CANP 168, CANP 325, CANP 318, CANP 355, CANP 108, CANP 51, CANP 307, CANP 132, CANP 302, CANP 58 and CANP 85 formed a group in right top corner of the biplot showed positive values of both the PCs and the character plant height placed in the same quadrant influencing the character on the grain yield. In Fig. 3, PC1 and PC3 were plotted against each other in biplot. Genotypes viz., CANP 58, CANP 307, CANP 422, CANP 17, CANP 510, CANP 427, CANP 108, CANP 404, CANP 302, CANP 51, CANP 403, CANP 355, CANP 91, CANP 325, CANP 121 and CANP 139 formed a group in right top corner of the biplot showed positive values of both the PCs and the characters days to 50% flowering, plant height and grain width positioned in the same quadrant influencing on the grain yield. Ashfaq et al. 2012 reported that highest variability was observed in plant height, tillers per plant and panicle length in 40 rice varieties.

In the present study, 67 aromatic rice germplasms were grouped into five clusters based on average linkage distance method of hierarchical clustering as shown in the diagram (Figure 4 and Table 4). Five major groups were formed as Clusters viz., I, II, III, IV and V consisting of 9, 31, 13, 11 and 3 genotypes respectively. Cluster means of different traits of aromatic rice germplasm accessions are given in Table 5. Cluster I had highest mean value for the traits viz., number of productive tillers/plant (23 No.), grain length (7.7mm), grain width (1.7 mm), 100 grain weight (2.4 g) and single plant yield (3.72g). Cluster III had low mean value for days to 50% flowering (85) and highest mean value for panicle length (25.1 cm). Cluster II had low mean value for plant height (109.1cm). Hence, the genotypes belongs to these clusters can be utilized for the future breeding programs. Same clustering pattern was previously recorded in several reports (Rahman et al. (2011); Suh et al. (1997) and Han-yong et al. (2004)).

A dendrogram is the graphical representation of a cophenetic matrix and hence dendrograms can be compared to one another by comparing their cophenetic matrices (Lapointe et al. 1995). The cophenetic correlation coefficient has been widely used in numerical phenetic studies, both as a measure of degree of fit of a classification to a set of data and as a criterion for evaluating the efficiency of various clustering techniques (Farris et al. 1969). The magnitude of this value should be very close to 1 for a high-quality solution. In the present study, cophenetic correlation coefficient was 0.74 which indicated the more efficiency of the clustering pattern. Huang et al. (2010) reported that the cophenetic correlation coefficient (r=0.918, 0.878 and 0.924 respectively) showed a highly accurate dendrogram represented the genetic distance among 36 rice entries.

A total of sixty seven aromatic rice germplasms were taken for assessing genetic diversity using PCA and cluster analysis for nine yield and quality contributing characters. PCA was utilized to study the variation and to estimate the relative contribution of various traits for total variability. The PC1 showed 25.9%, variability and PC2, PC3 and PC4 exhibited 16.96%, 15.06% and 12.22% variability respectively. It is concluded that PCA highlighted the characters with maximum variability in the present study. Eigen vectors with more than 0.3 value was observed in grain width and plant height in PC1 and also plant height,
panicle length, days to 50% flowering and presence of aroma in PC3. In conclusion, these traits had large effects in the overall variation present in the genotypes and selection of these traits would be useful for further breeding programs.

From the biplot, genotypes viz., CANP 496, CANP 295, CANP 510, CANP 265, CANP 90, CANP 91, CANP 168, CANP 325, CANP 318, CANP 355, CANP 108, CANP 51, CANP 307, CANP 132, CANP 302, CANP 58, CANP 85, CANP 422, CANP 17, CANP 427, CANP 404, CANP 403, CANP 121 and CANP 139 showed positive values for the characters viz., days to 50% flowering, plant height and grain width in biplots for PC1, PC2 and PC3 indicated that these characters are highly influencing on yield. From the cluster analysis, clusters I, II and III had desirable mean values for days to 50% flowering, plant height, number of productive tillers/plant, grain length, grain width, 100 grain weight, panicle length, and single plant yield. The genotypes belongs to these clusters can be utilized for the aromatic rice improvement programs.

References

Ahmed, M.S.U. Mohammad Khalequzzaman, M.D. Khairul Bashar and Abul Kalam Mohammad Shamsuddin. 2016. Agro-morphological, physico-chemical and molecular characterization of rice germplasm with similar names of Bangladesh. *Rice science*, 23(4): 211-218

Anderson, T.W. 1972. An introduction to multivariate analysis. Wiley Eastern Pvt. Ltd. New Delhi.

Anonymous, FAO, 2012. *Rice Market Monitor*, 14(2):1-36.

Ashfaq, M. Khan, A.S. Khan, S.H.U and Ahmad, R. 2012. Association of various morphological traits with yield and genetic divergence in rice (*Oryza Sativa* L.). *International Journal of Agriculture and Biology*, 14:55-62.

Farris, J.S. 1969. On the cophenetic correlation coefficient. *Systematic Zoology*, 18(3): 279-285

Ghareyazie, B. Faramarz Alinia. Corazon A. Menguito. Leila G. Rubia. Justina M. de Palma. Evelyn A. Liwanag. Michael B. Cohen. Gurdev S. Khush and John Bennett. 1997. Enhanced resistance to two stem borers in an aromatic rice containing a synthetic cry1Ab(b) gene. *Molecular breeding*, 3(5): 401-414.

Gour, L. Maurya, S.B. Koutr, G.K. Singh, S.K. Shukla, S.S and Mishra, D.K. 2017. Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot and rotated component matrix. *International Journal of Chemical Studies*, 5(4): 975-983

Han-yong, Y. Xing-hua, W. Yi-ping, W. Xiao-ping, Y and Sheng-xiang, T. 2004. Study on genetic variation of rice varieties derived from Aizzhan by using morphological traits, allozymes and simple sequence repeat (SSR) markers. *Chinese Journal of Rice Science*, 18: 477-482.

Huang, M. Fang-minxie. Li-yunchen. Xiang-qianzhao. Jojee, L and Madonna, D. 2010. Comparative analysis of genetic diversity and structure in rice using ILP and SSR markers. *Rice Science*, 17(4): 257-268.

International Rice Research Institute, 2001. Rice Research and Production in the 21st Century. (Gramene Reference ID 8380)

International Rice Research Institute, 2002. Rice Almanac, 3rd Edition. (Gramene Reference ID 8379)

International Year of Rice, 2004. "Gender and rice" factsheet. [PDF] (Gramene Reference ID 8371)

Jeffers, J.N.R. 1967. Two case studies in the application of principal component analysis. *Applied Statistics*, 16:225-236.

Lapointe, F.J and Legendre, P. 1995. Comparison tests for dendrograms: a comparative evaluation. *Journal of Classification*, 12: 265-282

Maji, A.T and Shaibu, A.A. 2012. Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*, 4(6):87-93

Morrison, D.E. 1978. Multivariate Statistical Methods (2nd ed. 4th Print, McGraw Hill Kogakusta Ltd.

Nachimuthu, V.V. Robin, S. Sudhakar, D. Raveendran, M. Rajeswari, S and Manonmani, S. 2014. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian Journal of Science and Technology*, 7(10):1555-1562.

Oyelola, B.A. 2012. The Nigerian Statistical Association preconference workshop, University of Ibadan, 20-21.

Pachauri, A.K. Sarawgi, A.K. Bhadarkar, S and Ojha, G.C. 2017. Agro-morphological characterization and morphological based genetic diversity analysis of Rice (*Oryza sativa* L.) germplasm. *Journal of Pharmacognosy and Phytochemistry*, 6(6): 75-80
Rahman, M.M. Hussain, A. Syed, M.A. Ansari, A. and Mahmud, M.A.A. 2011. Comparison among clustering in multivariate analysis of rice using morphological traits, physiological traits and Simple Sequence Repeat markers. *American-Eurasian Journal of Agricultural and Environmental Sciences*, **11** (6): 876-882

Raji, A.A. 2002. Assessment of genetic diversity and heterotic relationship in African improved and local cassava (*Manihot esculenta* Crantz) germplasm. Phd Thesis, University of Ibadan, Nigeria.

Sivaranjani, A.K.P. Pandey, M.K. Sudharshan, I. Kumar, G.R. Madhav, M.S and Sundaram, R.M. 2010. Assessment of genetic diversity among basmati and non-basmati aromatic rices of India using SSR markers. *Current Science*, **99**:221-226.

Sohgaura, N. Mishra, D.K. Koutu, G.K. Singh, S.K. Kumar, V and Singh, P. 2015. Evaluation of high yielding and better quality rice varieties using principal component analysis. *Ecology, Environment and Conservation*, (**21**):187-195.

Sood, B.C and Siddiq, E.A. 1978. *Indian Journal of Genetics and Plant Breeding*, **38**: 268-271.

Suh, H.S. Sato, Y.I. and Morishima, H. 1997. Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes and RAPD markers. *Theoretical and Applied Genetics*, **94**: 316-321.

Yumnam, S. Sawarkar, A. Patil, S.G and Senapathi, B.K. 2015. Multivariate analysis of recombinant inbred lines (F6) of basmati derivative. *The Ecoscan*, **9**(1-2):289-294.
Table 1. Details of aromatic rice germplasm accessions used in the study

| Sl. No. | Accession No. | Designation | Source |
|---------|---------------|-------------|--------|
| 1       | CANP232       | Pusa Basmati 1121 | IARI/DRR |
| 2       | CANP199       | Ranbir Basmati | J&K/IARI |
| 3       | CANP551       | IET 22787 (RP 4594-121-148-24-11) | DRR |
| 4       | CANP175       | Basmati 5874 | PU/DRR |
| 5       | CANP140       | Lakhban | PANTNAGAR |
| 6       | CANP90        | Kanak Jee | UP/DRR |
| 7       | CANP210       | Basmati 564 | J&K/IARI |
| 8       | CANP168       | Shyamjeera | RPR |
| 9       | CANP520       | UPR 3565-10-1-1 | PANTNAGAR |
| 10      | CANP413       | Hung-mi-hsiang-ma-tsaan | CHINA/DRR |
| 11      | CANP206       | Domusaali | IRAN/DRR |
| 12      | CANP212       | Chimbilate Basmati | J&K/IARI |
| 13      | CANP255       | Yamini (Basmati CSR 30) | CSSRI/DRR |
| 14      | CANP344       | Basmati Type 3 | PU/DRR |
| 15      | CANP463       | IR 62873-227-1-16 | IRRI/DRR |
| 16      | CANP465       | IR 62873-238-2-3 | IRRI/DRR |
| 17      | CANP444       | IR 841-85-1-1-2 | IRRI/DRR |
| 18      | CANP483       | Basmati 11 | PU/DRR |
| 19      | CANP251       | Pusa Basmati 1 | IARI/DRR |
| 20      | CANP554       | Bas 867 | US PATENTED LINE |
| 21      | CANP250       | Mahasangantha | KOTA/DRR |
| 22      | CANP535       | IR 73428-6-3 | IRRI/DRR |
| 23      | CANP246       | Tarori Basmati | PU/DRR |
| 24      | CANP183       | Basmati 138 | PU/DRR |
| 25      | CANP446       | Kalimosh (Raipur) | RP/RPR |
| 26      | CANP260       | Basmati Bahar | PU/DRR |
| 27      | CANP471       | Hasan Serai | IRAN/DRR |
| 28      | CANP440       | HBC 46 | HA/DRR |
| 29      | CANP129       | Haryana Basmati 1 | HANAGINA |
| 30      | CANP496       | JGL 1798 (Jagithyal Sannalu) | DRR |
| 31      | CANP309       | Improved Pusa Basmati 1 | IARI/DRR |
| 32      | CANP500       | Kasturi | DRR |
| 33      | CANP549       | IET 18033 (RP 3644-9-5-3-2) | DRR |
| 34      | CANP139       | Improved Sarbati | PANTNAGAR |
| 35      | CANP151       | Basmati 386 | LUDHIANA |
| 36      | CANP521       | UPR 2828-7-2-1 | PANTNAGAR |
| 37      | CANP327       | Karnal Local | PU/DRR |
| 38      | CANP355       | Basmati Mahan | DRR |
| 39      | CANP162       | Banspatri | RPR |
| 40      | CANP347       | Basmati 410 | PU/DRR |
| 41      | CANP353       | Basmati 370 | PU/DRR |
| 42      | CANP132       | Bindli | NAGINA |
| 43      | CANP354       | Basmati 6141 | PU/DRR |
| 44      | CANP410       | Daw Leuang | THAILAND/DRR |
| 45      | CANP404       | Binithen | PHILIPPINES/DRR |
| 46      | CANP412       | Guima | PHILIPPINES/DRR |
| 47      | CANP123       | Co ACC 167 (T167) | COIMBATORE |
| 48      | CANP242       | Basmati 217 | PU/DRR |
| 49      | CANP15       | RAU 3043 | BI/DRR |
| 50      | CANP422       | Longku Labat | INDONESIA/DRR |
| 51      | CANP318       | PDKV Shiram | MH/DRR |
| 52      | CANP58        | Kapoor kranti | RPR/DRR |
| 53      | CANP325       | Dhanaprassad | DRR |
| 54      | CANP91        | Bhanta Phool A | UP/DRR |
| 55      | CANP295       | Tarunbhog | RPR/DRR |
| 56      | CANP302       | Neelabati | DRR |
| 57      | CANP307       | Kalikhasa | AS/DRR |
| 58      | CANP510       | Kh.sakani | DRR |
| 59      | CANP17        | Maguraphulla | DRR |
| 60      | CANP427       | Nwaw Hawm Mali | THAILAND/DRR |
| 61      | CANP256       | Vasumati | DRR |
| 62      | CANP85        | Adam Chini B | UP/DRR |
| 63      | CANP51        | Kanika bhog | DRR |
| 64      | CANP121       | KDMIL 105 | THAILAND/CBT |
| 65      | CANP108       | Kamini Johna | AS/DRR |
| 66      | CANP265       | Lectimachi | DRR |
| 67      | CANP403       | Ayepyuang | MYANMAR/DRR |

(Note: IARI-Indian Agricultural Research Institute, New Delhi, DRR- Directorate of Rice Research, Hyderabad, J&K-Jammu and Kashmir, PU-Punjab, RPR-Raipur, CSSRI-Central Soil Salinity Research Institute, Karnal, MH-Maharashtra, UP-Uttar Pradesh, AS-Assam, BI-Bikaner)
Table 2. Proportion of variance, cumulative proportion and eigen values of aromatic rice germplasm accessions

| Statistics          | PC1  | PC2  | PC3  | PC4  | PC5  | PC6  | PC7  | PC8  | PC9  |
|---------------------|------|------|------|------|------|------|------|------|------|
| Proportion of Variance | 25.9 | 16.96| 15.06| 12.22| 10.65| 8.93 | 6.02 | 2.71 | 1.56 |
| Cumulative Proportion | 25.9 | 42.86| 57.92| 70.14| 80.78| 89.72| 95.73| 98.44| 100  |
| EigenValues         | 2.33 | 1.52 | 1.35 | 1.09 | 0.95 | 0.80 | 0.54 | 0.24 | 0.14 |

Table 3. Contribution of different traits towards total variance in aromatic rice germplasm

| Traits                          | Eigen Vectors |
|---------------------------------|---------------|
|                                 | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 |
| Days to 50% flowering           | 0.18 | -0.06 | 0.43 | -0.62 | 0.29 | -0.11 | 0.53 | 0.13 | 0.04 |
| Number of productive tillers/plant | -0.24 | 0.16 | -0.08 | 0.44 | 0.61 | -0.51 | 0.19 | 0.19 | 0.07 |
| Plant height (cm)               | 0.34 | 0.05 | 0.59 | 0.28 | 0.01 | -0.28 | -0.17 | -0.57 | -0.14 |
| Panicle length (cm)             | -0.20 | -0.07 | 0.57 | 0.46 | -0.21 | 0.37 | 0.22 | 0.42 | 0.08 |
| Grain length(mm)                | -0.54 | -0.35 | 0.04 | -0.05 | -0.11 | -0.08 | 0.19 | -0.49 | 0.54 |
| Grain width (mm)                | 0.46 | -0.46 | -0.01 | 0.06 | -0.04 | -0.26 | -0.29 | 0.34 | 0.55 |
| Hundred grain weight(g)         | -0.20 | -0.73 | 0.00 | 0.01 | 0.00 | -0.20 | -0.03 | 0.09 | -0.61 |
| Single plant yield(g)           | 0.25 | -0.30 | -0.13 | 0.19 | 0.59 | 0.62 | 0.08 | -0.25 | 0.04 |
| Presence of aroma               | -0.38 | 0.06 | 0.33 | -0.30 | 0.37 | 0.13 | -0.69 | 0.12 | 0.05 |

Table 4. Hierarchical cluster grouping in aromatic rice germplasm accessions

| Clusters | Frequency | Cluster Memberships |
|----------|-----------|---------------------|
| I        | 9         | CANP252, CANP344, CANP251, CANP246, CANP151, CANP327, CANP353 CANP354, CANP412 |
| II       | 31        | CANP199, CANP551, CANP210, CANP168, CANP520, CANP206, CANP212, CANP255, CANP463, CANP465, CANP383, CANP554, CANP535, CANP385, CANP446, CANP471, CANP440, CANP129, CANP309, CANP500, CANP549, CANP139, CANP521, CANP162, CANP347, CANP242, CANP15, CANP318, CANP325 CANP91, CANP256 |
| III      | 13        | CANP375, CANP140, CANP90, CANP413, CANP444, CANP260, CANP496, CANP132, CANP410, CANP123, CANP422, CANP510, CANP265 |
| IV       | 11        | CANP250, CANP295, CANP302, CANP307, CANP17, CANP427, CANP85, CANP51, CANP121, CANP108, CANP403 |
| V        | 3         | CANP355, CANP404, CANP58 |

Cophenetic correlation coefficient: 0.74
| Traits                      | Cluster | Minimum | Maximum | Mean  | StdDev |
|----------------------------|---------|---------|---------|-------|--------|
| Days to 50% flowering       | 1       | 80      | 94      | 86    | 4.6    |
|                            | 2       | 78      | 104     | 88    | 7.3    |
|                            | 3       | 65      | 98      | 85    | 9.0    |
|                            | 4       | 91      | 99      | 94    | 2.4    |
|                            | 5       | 86      | 103     | 95    | 8.5    |
|                            | 1       | 19      | 26      | 23    | 3.2    |
|                            | 2       | 11      | 21      | 13    | 2.6    |
|                            | 3       | 11      | 15      | 13    | 1.1    |
| Number of productive       | 4       | 12      | 19      | 13    | 2.0    |
| tillers/plant              | 5       | 13      | 23      | 17    | 5.5    |
|                            | 1       | 93.6    | 133.1   | 114.5 | 13.3   |
|                            | 2       | 72.0    | 148.0   | 109.1 | 18.4   |
|                            | 3       | 71.1    | 146.1   | 114.3 | 21.4   |
|                            | 4       | 100.8   | 153.0   | 134.8 | 13.3   |
| Plant height(cm)           | 5       | 126.8   | 167.3   | 146.3 | 20.3   |
|                            | 1       | 21.7    | 25.5    | 23.9  | 1.3    |
|                            | 2       | 17.3    | 26.8    | 23.4  | 2.4    |
|                            | 3       | 19.1    | 24.1    | 21.8  | 1.4    |
|                            | 4       | 19.6    | 25.8    | 22.6  | 2.1    |
| Panicle length (cm)        | 5       | 24.6    | 25.6    | 25.1  | 0.5    |
|                            | 1       | 6.0     | 9.0     | 7.7   | 0.9    |
|                            | 2       | 5.5     | 9.3     | 7.5   | 0.9    |
|                            | 3       | 4.3     | 7.5     | 5.8   | 1.0    |
|                            | 4       | 4.0     | 7.9     | 5.5   | 1.6    |
| Grain length (mm)          | 5       | 4.3     | 7.7     | 6.4   | 1.8    |
|                            | 1       | 1.5     | 2.0     | 1.7   | 0.2    |
|                            | 2       | 1.3     | 2.0     | 1.8   | 0.2    |
|                            | 3       | 1.7     | 2.7     | 2.2   | 0.3    |
|                            | 4       | 1.7     | 3.0     | 2.2   | 0.4    |
| Grain width (mm)           | 5       | 1.6     | 2.2     | 1.8   | 0.3    |
|                            | 1       | 2.1     | 3.1     | 2.4   | 0.3    |
|                            | 2       | 1.0     | 3.0     | 2.2   | 0.4    |
|                            | 3       | 1.3     | 3.4     | 2.2   | 0.6    |
|                            | 4       | 0.9     | 3.8     | 2.0   | 0.9    |
| 100 grain weight(g)        | 5       | 1.1     | 2.3     | 1.6   | 0.6    |
|                            | 1       | 1.92    | 5.87    | 3.72  | 1.4    |
|                            | 2       | 0.95    | 8.87    | 2.61  | 1.7    |
|                            | 3       | 0.90    | 8.46    | 4.10  | 2.5    |
|                            | 4       | 1.76    | 6.49    | 4.46  | 1.4    |
| Single plant yield(g)      | 5       | 0.83    | 5.95    | 2.60  | 2.9    |
Fig. 1. Scree plot for different PCs in aromatic rice germplasm accessions

Fig. 2. The Biplot of aromatic rice germplasm accessions for PC1 and PC2

(Note: A- days to 50% flowering; B- number of productive tillers per plant; C- plant height (cm); D- panicle length (cm); E- grain length (mm); F- grain width (mm); G- hundred grain weight (g); H- single plant yield (g) and I- Presence of aroma. 1 to 67- Aromatic rice germplasm accessions)
Fig. 3. The Biplot of aromatic rice germplasm accessions for PC1 and PC3

(Note: A - days to 50% flowering; B - number of productive tillers per plant; C - plant height (cm); D - panicle length (cm); E - grain length (mm); F - grain width (mm); G - hundred grain weight (g); H - single plant yield (g) and I - Presence of aroma. 1 to 67 - Aromatic rice germplasm accessions)

Fig. 4. Dendrogram using agglomerative clustering method in aromatic rice germplasm accessions
