

# Principal Component Analysis for Evaluation of Rice (*Oryza sativa* L.) Germplasm

Asif B. Shikari\*, G.A. Parray, A.G. Rather and F. A. Sheikh

Sher-e-Kashmir

University of Agricultural Sciences & Technology of Kashmir

Rice Research and Regional Station Khudwani, 192 102, India

## Abstract

**One hundred and fifty rice germplasm lines were evaluated for 31 agro-morphological traits. Principal component analysis was carried out involving 9 and 5 quantitative traits separately which yielded selection differential of 36.18 and 27.07 q/ha respectively. This approach helped in identification of top ranking genotypes with high mean values for primary grain yield components. Preliminary screening was done to identify fine grain rice types.**

Principal component analysis (PCA) is a valuable technique from applied linear algebra, which can be used to classify the relationships among traits in complete multi-trait systems and for identifying the patterns of data by reducing the number of dimensions. An insight into the process contributing to differences in yield among the lines is essential for identification and selection of top ranking genotypes out of diverse germplasm base. The principal component analysis technique has in retrospect been applied in rice by the workers like Rashid *et al.* (2008), Uga (2003), Zamira *et al.* (2003), and Toshinori (1991).

## Materials and Methods

The experimental materials comprised of 150 rice (*Oryza sativa* L.) germplasm accessions having their origin from different geographical regions across the world. The germplasm was grown during *kharif*-2008 as transplanted irrigated rice at Rice Research and Regional Station (SKUAST-K). Thirty five days old seedlings were transplanted in 4 rows of 3 m each at a spacing of 20 x 15 cm in

Augmented Block Design with 3 checks replicated after every 25 test entries. Observations were recorded on ten randomly selected plants from two middle rows for the traits plant height (PH), leaf length (LL), leaf width (LW), panicle length (PL), number of spikelets per panicle (SP), grain density (GD), total tillers per plant (TT), effective tillers per plant (EF), 1000-seed weight (SW), kernel length (KL), kernel breadth (KB), Kernel length breadth ratio (LBR), and on whole plot basis for days to 50% flowering (DF), days to maturity (DM) and grain yield per hectare (GY). Moreover, the characterization data was recorded for the traits leaf blade color, leaf blade pubescence, leaf angle, leaf senescence, flag leaf angle, collar color, ligule color, auricle color, panicle type, secondary panicle branching, panicle threshability, lemma palae color, lemma palae pubescence, awning, awn color, apiculus color and stigma color following SES scale of IRRI (Anonymous, 1996). According to this scale, accessions were grouped on the basis of size i.e. KL (Long: >6.60 mm; Medium: 5.51-6.60 mm; Short: <5.51 mm), and also on the basis of shape i.e. LB (Slender: >3.0; Medium slender: 2.1-3.0; Bold: <2.1). Mean, range and frequency distribution were calculated for the quantitative traits. Simple Anova was carried out following Sharma (1995). Principal Component Analysis was performed to reveal the patterns of data matrix for determination of selection criteria and identification of elite genotypes. Eigen values, eigen vectors and principal components were calculated by R-software-2.7.0 (freely available on <http://CRAN.R-project.org>.) along with 2-D biplot generated between first (PC-1) and second principal component (PC-2).

\* Corresponding author : asifshikari@gmail.com



## Results and Discussion

Significant mean squares were recorded for all the traits which leave a scope for further investigations. High range of variability was recorded for all the characters with GY measuring 3.90-71.71 q/ha (Table 1). Out of 150 genotypes, 106 fell in the range of 80-120 cm and 36 recorded plant height below 80 cm respectively. Majority (90) of the genotypes had 4-8 EF while 117 had SP ranging from 75-150. SW and GY in the range of 20-30 g and 25-50q/ha was recorded for 121 and 60 test entries respectively. GY more than 50 q/ha was exhibited by 12 genotypes. These results indicated that enough variability exhibited for primary and secondary yield contributing traits which provide ample scope for selection and other hybridization based breeding methods for improvement of these important traits. On the basis of kernel size, 16 entries were observed to have long kernels (>6.60 mm); 28 were medium long (5.51-6.60 mm) and 106 had short kernel length (<5.51 mm). With respect to the kernel shape, 9 genotypes were classified as slender (LB > 3.0); 85 medium slender (LB: 2.1-3.0) and 56 as bold (LB < 2.10) kernel types. Considering both size and shape together 56 accessions were short bold, 50 short slender, 4 medium slender (HPR 1070, RP 2421, VL 93-3635 and IR 57301-37-2-2) and 4 had long slender kernels (SAW/GML/286, JOJO, GRALDO and SIJUNG-10). Kernel length and breadth are the components of quality attributes of basmati rice and constitutes besides other traits, the criteria fixed by Directorate of Rice Research (ICAR), Hyderabad, India (Anonymous, 2005) for identification of basmati and basmati type varieties. Though not necessarily be called as basmati types, KL and KB should be greater and lesser than 6.61 mm and 2 mm respectively, before other chemical and cooking properties are worked out. In the present study, JOJO (7.00 mm, 1.60 mm), GRALDO (7.00 mm, 1.90 mm), VL-81 (6.60 mm, 1.80 mm), IR 57301-37-2-2 (6.60 mm, 1.90 mm), CT 6749-21-4-3-M (8.00 mm, 1.90 mm) meet this criteria and need to be tested for chemical and cooking quality, however, as expected they yield meagerly low i.e. 15.09, 30.15, 26.95, 27.32, 17.39 q/ha respectively.

An ideal situation during the perusal of principal component analysis (PCA) in any multi-trait system demands the lack of correlation to exist among different principal component axis so that different dimensions in the data are measured (Everitt and Dunn, 1991). Concurrently this circumstance is met when original xi variables in the data set show significant correlation among themselves or at least with the trait of economic interest. In the present study, non-significant correlation was exhibited by grain yield when paired with mature plant height, leaf length and leaf width (Table 2). To reach out the meaningful conclusions and extract relevant information, these three characters were dropped for PCA. Kernel physical characteristics viz., KL, KB and LB were also not considered for computation of principal components, since they as a rule do not influence GY.

Highly significant positive correlations with GY were exhibited by TT (0.73\*\*), EF (0.71\*\*), SP (0.64\*\*), GD (0.49\*\*) and PL (0.39\*\*). SP showed positive significant association with GD (0.76\*\*), PL (0.57\*\*) and DF (0.25\*\*) and moderate positive correlation with DM (0.20\*). Lack of correlation was shown by EF with all the traits except TT and GY. The two separate PCA's (discussed under Set-1 and Set-2) were carried out, one involving a set of nine out of twelve quantitative traits used in correlation studies and second where only three primary yield components (EF, SP and SW) were involved along DM and GY.

**Set-I: PCA of primary and secondary yield components:** The first five principal components accounted for 99.28% of total variability with the first two explaining the cumulative variability of 93.10%. The individual contributions of PC-1 and PC-2 were 80.17% and 12.93% respectively. Only PC-1 and PC-2 had eigen values ( $\lambda$ ) greater than 1, i.e. 7.215 and 1.164 times the variance of original variables (because variance of standardized variables is 1), therefore, the eigen vector coefficients need to be studied for these two axis (Table 3). PC-1 had high negative loadings for GY (-0.406) and SP (-0.908) (Table 4). PC-2 was positively correlated to GY, TT and EF and

**Table 1: Frequency distribution table for various quantitative traits**

	Mean	Min	Max	Criteria	Fq.	Criteria	Fq.	Criteria	Fq.
<b>PH</b>	93.29	45.60	138.60	<80	36	80-120	106	>120	8
<b>DF</b>	93.01	78.73	108.63	<80	6	80-100	128	>100	16
<b>DM</b>	126.36	112.62	138.53	<120	24	120-130	96	>130	30
<b>LL</b>	32.98	15.00	58.00	<20	3	20-40	125	>40	22
<b>LW</b>	1.16	0.70	2.10	<1	55	1-1.5	84	>1.5	11
<b>PL</b>	18.85	10.00	26.00	<15	19	15-20	81	>20	50
<b>GD</b>	5.67	2.31	12.64	<5	55	5-10	93	>10	2
<b>TT</b>	13.40	4.98	25.91	<10	47	10-20	90	>20	13
<b>PP</b>	5.36	1.99	10.37	<4	47	4-8	90	>8	13
<b>SP</b>	106.63	30.00	224.00	<75	22	75-150	117	>150	11
<b>SW</b>	27.00	11.20	42.25	<20	2	20-30	121	>30	27
<b>GY</b>	26.97	3.90	71.71	<25	78	25-50	60	>50	12
<b>KL</b>	5.31	4.00	8.00	<5.51 (Short)	106	5.51-6.60 (Med long)	28	>6.60 (Long)	16
<b>LB</b>	2.25	1.46	4.37	<2.10 (Bold)	56	2.1-3.0 (Med slender)	85	>3.0 (Slender)	9

Abbreviations used: PH: Plant height (cm); DF: Days to 50% flowering; DM: Days to maturity; LL: Leaf length (cm); LW: Leaf width (cm); PL: Panicle length (cm); SP: No of grains per panicle; GD: Grain density per panicle (per cm); TT: Total tillers per plant; PP: Effective tillers per plant; SW: 1000-seed weight (g); GY: Grain yield per ha (q); KL: Kernel length (mm); LB: Kernel length breadth ratio

**Table 2: Genotypic coefficients of correlation among various quantitative traits**

	PH	DF	DM	LL	LW	PL	GD	TT	EF	SP	SW	GY
<b>PH</b>	1.00	0.23*	0.17	0.49**	0.12	0.49**	-0.13	0.01	0.01	0.16	-0.08	0.08
<b>DF</b>		1.00	0.54**	0.02	-0.14	0.29**	0.10	0.10	0.08	0.25**	-0.21*	0.12
<b>DM</b>			1.00	0.08	-0.14	0.11	0.17	0.16	0.11	0.20*	-0.22*	0.15
<b>LL</b>				1.00	0.34**	0.39**	-0.09	-0.09	-0.08	0.11	0.11	0.05
<b>LW</b>					1.00	0.07	0.10	0.06	0.06	0.08	0.15	0.15
<b>PL</b>						1.00	0.02	0.09	0.09	0.57**	-0.06	0.39**
<b>GD</b>							1.00	0.10	0.09	0.76**	-0.03	0.49**
<b>TT</b>								1.00	0.97**	0.14	-0.16	0.73**
<b>EF</b>									1.00	0.12	-0.16	0.71**
<b>SP</b>										1.00	-0.06	0.64**
<b>SW</b>											1.00	0.18*
<b>GY</b>												1.00

\*\* and\* indicates P(<0.01) and P(<0.05) respectively



negatively correlated to SP. Positive correlation between GY and SP ( $0.63^{**}$ ) stems from their unidirectional negative loadings in PC-1, though it is moderated only to a limited extent by PC-2, because PC-1 explains 6 times more variability than PC-2. The PC-2 axis may be regarded as yield vector having high factor loadings (0.860) for GY. PC-3 is important with regard to selection for early maturity since it possess negative loading for DF (-0.820) and DM (-0.518) and for obtaining high test-weight (0.186) genotypes. However, PC-3 being of little magnitude ( $\bar{e}=0.344$ ), such genotypes would be influenced by PC-1 and PC-2. The PC-1 and PC-2 biplot helped to identify the genotypes with high GY (Figure 1). Highest PC-2 score was observed for 84107-TR 745-12-1 (32.66) followed by HPR 1070 (31.14), AMAROO (29.55), SK/PBG/78 (26.07), IR 62728-2B-5-2-2-2 (25.25), SIJUNG-10 (23.01), DONGHAECHAL (22.27) and VL 93-3548 (20.89). These genotypes yielded a selection differential of 36.18 q/ha for GY. The length of SP vector was greater than the length of GY vector which in turn was longer than other traits under study, where length of the vector is directly proportional to the proportion of variability in principal components and direction tells whether the proportion is positive or negative (Khan and Mir, 2008). GY falls in second quadrant since it has high positive loadings for PC-2 axis and negative loadings for PC-1 axis. PC-3 axis though accounted for very little variability, however, had negative desirable loadings for DM (-0.50) and positive loadings for 1000-seed weight. This axis identified genotypes VL-93-3635 and HPR 1070 both with early maturity of 118 days and 1000-seed weight of 38 and 33 g respectively, HPR 1070 had GY of 57.75 q/ha.

**Set-II: Principal component analysis for the set involving GY, DM and three primary yield traits** (Figure 2) : PC-1 explained 84.31% of the total variability which had eigen value of 1.68. The cumulative proportion of variability accounted for by first two principal components was 96.88% with 12.57% contributed by PC-2. PC-1 had high positive loadings for SP (0.915) and GY (0.412) thus resulted in high significant correlation between

the two ( $r=0.63^{**}$ ). PC-2 showed very high factor loadings for GY (0.902) and low positive loadings for EF (0.124) but was negatively correlated to SP (-0.405). This means that SP contributed to GY in majority of the genotypes followed by smaller group where EF was important. Also, PC-2 depicted of yield component compensation mechanism existing between SP and EF. PC-3 was not of any major significance and use.

Biplot analysis between PC-1 and PC-2 axis revealed that SP vector was the longest and was evenly balanced between the two. Second in magnitude was GY vector which fall in quadrant-I with positive loadings for both PC-1 and PC-2. The genotypes which exhibited high PC-1 scores included YUNLEN-19 (116.46) followed by H 257-2-1-1 (87.68), SKAU-23 (86.63), HPR 1160 (81.66), K-39 (79.69), IR 62443-2B-4-3-2 (75.46), YUNLEN-2 (72.08) and 84107-TR 745-12-1 (71.06) and possessed high mean values in the respective order of 224, 195, 184, 184, 171, 175, 168 and 158 spikelets per panicle. High PC-2 scores of 34.80, 30.13, 28.12, 24.99, 34.74, 22.78, 22.26 and 21.55 were observed for 84107-TR 745-12-1, HPR 1070, AMAROO, SK/PBG/78, IR 62728-2B-2B-5-2-2-2, DONGHAECHAL, WON 124 and SIJUNG-10 respectively so therefore, had high GY of 71.71, 57.79, 59.74, 58.70, 56.60, 44.08, 43.39 and 44.87 q/ha respectively. Here PC-1 and PC-2 seem to be least correlated as they had only 84107-TR 745-12-1 as common genotype. Besides this genotype, cultures namely K-39, SKAU-23 and YUNLEN-2 yielded more than 55 q/ha by virtue of high number of SP. The order of PC-2 scores and GY mean did not go hand to hand among the high yielding genotypes because of the blurring effect of moderate loadings by EF which influenced PC-2. The high yielding genotypes in PC-2 axis had high EF. 84107-TR 745-12-1 however, was identified to possess high average values for the three primary yield traits viz., SP (158), EF (6) and SW (41g). Selection of top 8 genotypes on the basis of PC-1 and PC-2 scores yielded a selection differential of 27.07 and 22.49 q/ha.

As per the Standard Evaluation System of IRRI (Anonymous, 1996), one hundred and fifty

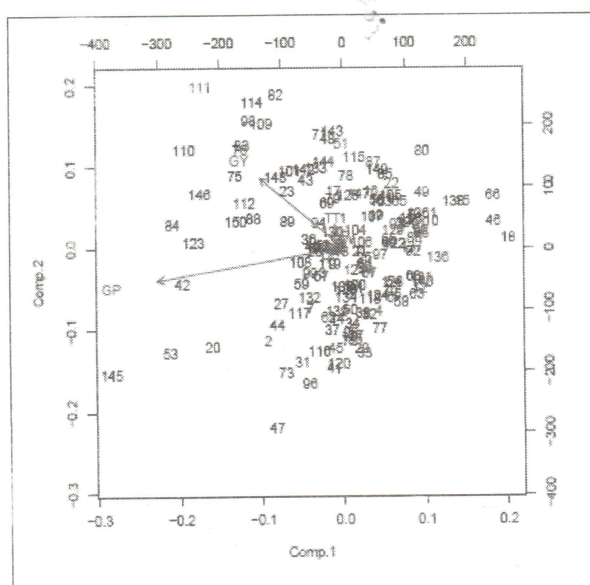


**Table 3: Eigen values of principal components**

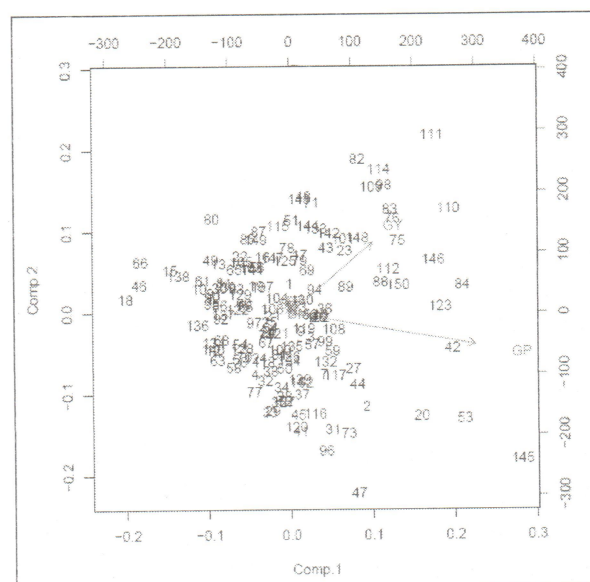
Set-I	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7	PC-8	PC-9
Eigen values	7.215	1.164	0.344	0.119	0.094	0.045	0.014	0.005	0.002
Proportion of variance	80.17	12.93	3.82	1.32	1.04	0.50	0.15	0.05	0.02
Cumulative Proportion	80.17	93.10	96.92	98.24	99.28	99.78	99.93	99.98	100.00
<b>Set-II</b>									
Eigen values	1.685	0.652	0.269	0.191	0.034	-	-	-	-
Proportion of variance	84.31	12.57	2.04	1.03	0.03	-	-	-	-
Cumulative Proportion	84.31	96.88	98.92	99.95	99.98	-	-	-	-

**Table 4: Factor loadings for first three principal components**

Set-I	GY	EF	SP	SW	DM	PL	GD	TT	DF
PC-1	-0.406	0.024	-0.908	0.001	0.001	0.004	0.002	0.012	0.003
PC-2	0.860	0.115	-0.394	0.043	0.058	0.021	0.005	0.286	0.022
PC-3	0.042	0.008	0.010	0.186	-0.518	0.042	0.088	-0.120	-0.820
<b>Set-II</b>									
PC-1	0.412	0.054	0.915	0.020	0.003	-	-	-	-
PC-2	0.902	0.124	-0.405	0.012	0.032	-	-	-	-
PC-3	0.004	0.038	0.011	-0.362	0.931	-	-	-	-



**Figure 1. Set-I Biplot involving nine traits**



**Figure 2. Set-II Biplot involving five traits**

**Name of accessions (1-150):** SAW/GML/214, SAW/GML/224, SAW/GML/234, SAW/GML/294, SAW/GML/252, SAW/GML/270, SAW/GML/280, SAW/GML/286, SAW/GML/292, SAW/GML/295, SAW/GML/310, SAW/GML/317, SAW/GML/327, SAW/GML/188, K-334, K-429, K-312-30-2-2-3-1-3, C-Mochi, Chingshee, HPR 1037, HPR 1064, HPR 1178, K 696-84-1, VL 94-3143, VL 94-3027, K 454-60-1-6-1, K 680-51-1-2-1, VL 93-3635, SKAU-21, HPR 1062, HPR 1068, HPR 1149,



HPR 1171, VL 94-3028, VL 95-3337, K-500, PL11/SAS-2, VL 91-1923, Himdhan, RP 2421, HPR 1155, HPR 1160, K681-33-1, IR 61009-47-3-1-1, JOJO, K-39, CT 6749-36-7-2-M, DONGHAECHAL, ECHUCA, GIGANTE VERCELLI, GILJU 1, GRALDO, H 257-2-1-1, H 270-30-2-1, H 274-16-1-1, Hamnam 15, Hexi-5, HS-379, IR 61728-4B-2-1-1, IR 63347-Ac201-1, L10573 (Acc 68106), MILLIN, MILYANG 80, OLBYE 2, ONPO 1, STEJAREE 45, ONPO 6, RYONSONG 12, RYONSONG 14, RYONSONG 25, SIJUNG 10, YR 5190-16-2-1-2, YUNGEN 135, YUNLEN 15, TATSUMI-MOCHI, K-450-3-2-2, VL 95-3336, K-681-9-1, HPR 2047, VL 93-3635, VL 81, HPR 1070, VL 93-3548, SKAU-23, K-508, K-312-25, K-312-15, Koshihikari, K-332, SAW/GML/08, SK/PBG/50, SK/PBG/51, SK/PBG/56, SK/PBG/58, SK/PBG/62, SK/PBG/67, SK/PBG/68, SK/PBG/78, SK/PBG/79, SK/PBG/81, CNTLR 80076-44-1-1, IR 54465-B-B-3-2-2-2, IR 57301-199-2-3, IR 57301-37-2-2, IR 58614-B-B-2-2, IR 59471-2B-20-2-1, IR 59614-26-2, IR 60010-4B-1-1-1, IR 62728-2B-5-2-2-2, K-39, 84107-TR745-12-1, 84032-TR76-4-1, 84037-TR765-6-2, AMAROO, CHOJANG, CT 6749-21-4-3-M, Hexi 4, HWANGHAE 20, LLABONG, IR 71163-4-1-1-1, IR 61673-Ac201-1-3, IR 61727-4B-1-1-1, IR 62443-2B-4-3-2, IR 62443-2B-7-2-2-1, IR 67087-B-3-2-3-6, IR 68349-131-2-2-3, IR 68352-14-1-1-1, LANGI, NAMAGA, NONG 49, NONG 57, OSMANCIK-97, PADANO, PR 26881-PJ16-4B-78-5-1, PYONGBUK 10, PYONGBUK 21, PYONGBUK 7, SR 12192-T15, SR 13349-59-1, SU 105, SU 98, WON 122, WON 124, WON 125, YUNLEN 19, YUNLEN 2, ZHATONMAXAIGU, BARKAT, STEJAREE, PYONGBUK 5.

**Table 5: Characterization of rice germplasm as per Standard Evaluation System of IRRI-1996**

Leaf blade color	31: Light green; 26: Green; 22: Dark green; 23: Purple tips; 19: Purple margins; 12: Purple blotch; 17: Purple
Leaf blade pubescence	64: Glabrous; 52: Intermediate; 34: Pubescent
Leaf angle	58: Erect; 66: Horizontal; 26: Droopy
Leaf senescence	58: Late and slow (leaves have natural green color); 66: Intermediate (Upper leaves yellowing); 26: Early and fast (all leaves yellow or dead)
Flag leaf angle	90: Erect; 20: Intermediate; 29: Horizontal; 11: Descending
Collar color	61: Light green; 18: Green; 71: Purple
Ligule color	58: White; 53: Purple lines; 39: Purple
Auricle color	49: Light green; 101: Purple
Panicle type	70: Compact; 37: Intermediate; 43: Open
Secondary panicle branching	2: Absent; 55: Light; 48: Heavey; 45: Clustering
Panicle thresh ability	23: Difficult (<1%); 50: Moderately difficult (1-5%); 38: Intermediate (6-25%); 31: Loose (26-50%); 08: Easy (51-100%)
Lemma and Palea color	30: Straw; 32: Gold and gold furrows on straw background; 18: Brown spots on straw; 08: Brown furrows on straw; 04: Brown (tawny); 05: Reddish to light purple; 06: Purple spots on straw; 07: Purple furrows on straw; 08: Purple; 09: Black; 10: White
Lemma and Palea pubescence	08: Glabrous; 11: Hairs on lemma keel; 63: Hairs on upper portion; 15: Short hairs; 53: Long hairs (velvety)
Awning	126: Absent; 15: Short and partly awned; 04: Short and fully awned; 01: Long and partly awned; 04: Long and fully awned
Awn color	128: Awnless; 01: Straw; 06: Gold; 0: Brown; 01: Red; 12: Purple; 02: Black
Apiculus color	25: White; 30: Straw; 20: Brown (tawny); 05: Red; 09: Red Apex; 44: Purple; 17: Purple apex
Stigma color	32: White; 40: Light green; 28: Yellow; 20: Light purple; 30: Purple



different germplasm lines were characterized for 17 qualitative traits. Frequency distribution for various classes of these traits is given in table 5.

Summarizing the above results, SP was established to be the major contributor to the total variability followed by EF which got reflected in differences in GY. At a selection intensity of 5 % greater selection differential (36.18 q/ha) was realized by principal component-2 involving nine characters vis-à-vis principal component-1 (SD: 27.07) and -2 (SD: 22.49) based on five traits. Genotypes identified having high GY were 84107-TR 745-12-1, HPR-1070, YUNLEN-19, H 257-2-1-1, K-39, HPR-1160, AMAROO, SK/PBG/78, etc. The first one had high genotypic values for all the three primary yield traits (SP, SW and EF). HPR 1070 had medium slender kernels, high GY coupled with early maturity of 118 DAS.

### References

- Anonymous. 1996. Standard Evaluation System, INGER Genetic Resources Center, ed. 4<sup>th</sup>, IRRI, Philippines
- Anonymous. 2005. Laboratory Manual on Rice Grain Quality Procedures. In Training for officials of Export Inspection Council of India, New Delhi, held at Directorate of Rice Research, Rajendra Nagar, Hyderabad, 24-30, August, pp:14-17
- Everitt, B.S. and Dunn, G. 1991. Applied multivariate data analysis. Edward Arnold, London.
- Khan, A.A. and Mir, A.H. 2008. Applications of R-Software in Agricultural Data Analysis-II, *SKUAST Journal of Research* 10:260-277
- Rashid, M., Cheema, A. A. and Ashraf, M. 2008. Numerical Analysis of variation among basmati rice mutants. *Pakistan Journal of Botany* 40:2413-2417
- Sharma, J.R. 1995. Experimental Field Designs for Plant Breeding. In Statistical and Biometrical techniques in Plant Breeding. New Age International (P) Limited. pp:15-29.
- Toshinori, A.B.E. 1991. Principal component analysis for rice characters relating with plant stature and yield. *Journal of Yamagata University of Agricultural Sciences* 11:235-247.
- Uga, Y., Fukuta, Y., Ohsawa, R. and Fujimura, T. 2003. Variations of floral traits in Asian cultivated rice (*Oryza sativa* L.) and its wild relatives (*O. rufipogon* Griff.). *Breeding Science* 53: 345-352
- Zamira, A., Barboza, C., Lobo, J. and Espinoza, A. M. 2003. Diversity of native rice (*Oryza* Poaceae:) species of Costa Rica. *Journal of Applied Sciences Research* 50:855-870.