Forage Res., **42** (2) : pp. 90-95 (2016) **http://forageresearch.in**

PRINCIPAL COMPONENT AND CLUSTER ANALYSIS IN SORGHUM (*SORGHUM BICOLOR* (L.) MOENCH)

S. K. JAIN* AND P. R. PATEL

Sorghum Research Station
Sardarkrushinagar Dantiwada Agricultural University, Deesa,
Banaskanta, Gujarat – 385 535 India
**(e-mail: skjain@sdau.edu.in)*

(Received : 06 September 2016; Accepted : 27 September 2016)

SUMMARY

In the current set of an experiment, thirty two sorghum genotypes were assessed for different yield and yield component traits. For evaluation of these traits, basic statistics, correlation, Principal component (PC) and diversity analyses were employed to obtain suitable parents that can be further exploited in future breeding programmes. The estimation of descriptive statistics of nine quantitative traits indicated the existence of variability among the genotypes. Correlation analysis revealed that grain yield was positively correlated with panicle length, leaf length and leaf width whereas fodder yield was positively correlated with number of leaves/plant, leaf width, leaf length, plant height and days to maturity. The positive correlation among these yield contributing traits suggested that these characters are important for direct selection of high yielding genotypes. Principal component (PC) analysis showed first 3 PCs having Eigen value >1 explaining 70.89% of the total variation with different yield and yield component traits. In the biplot analysis between PCs 1 and 2, the genotypes remained scattered in all four quadrants, showing large genetic variability in quantitative traits. The thirty two genotypes were grouped into five clusters on the basis of average linkage and dendogram. The cluster-I having 13 genotypes followed by cluster-III (9), cluster-II (7), cluster-V (2) and cluster-IV (1). Distribution pattern of all the genotypes into five clusters showed the presence of considerable genetics diversity among the genotypes for most of the traits under consideration. Various useful correlations and aforementioned information extracted from cluster and PC analysis will be helpful in designing breeding programmes to obtain high yielding genotypes in sorghum for seed as well as fodder yield.

Key words : Correlation, Genotypes, Principal component analysis, Diversity analysis, Clusters

Sorghum (*Sorghum bicolor* (L.) Moench) is the third most important cereal crop grown in the world. Nigeria is the world's largest producer of sorghum, followed by the United States and India. India contributes about 16% of the world's sorghum production. Sorghum is tolerant to drought and heat and genetically suited to hot and dry agro-ecologies where it is difficult to grow other food grains. It is important nutrition cereals constituting staple diet of the major population in arid and semi arid region of India. It is also used for feed, fodder and the production of alcoholic beverages. Sorghum area declined 74.71 per cent over the past four decades in the country. However, in spite of reduction in the area, all States have recorded a substantial increase in the productivity level but it has higher gap than other countries. To expand the gain in sorghum productivity we will need better breeding strategies that improve the productivity level and boost the total production. House (1985) noted that cultivated sorghums are highly variable and suggested that to enhance the productivity levels of sorghum, prior information on the nature and the magnitude of genetic diversity present in breeding material is a pre-requisite.

Precise information on the nature and degree of genetic diversity help plant breeder in selecting the parents for targeted hybridization. It provides the raw materials from which desirable alleles for improved agronomic traits of interest can be selected and subsequently incorporated into elite lines. The purpose of principal component analysis is to reduce the volume of data. Watson and Eyzaguirre (2002) also reported that PCA of morphological characterization results could identify a few key or minimum descriptors that effectively account for the majority of the diversity observed, saving time and effort for future characterization efforts. Principal components approach

is very helpful in deciding which agronomic traits of crop contributing most to yield, subsequently, these agronomic traits should be emphasized in the breeding program. In order to determine genetic variation genotype classifications and genetic distance among them the cluster analysis is done. Cluster analysis identifies and classifies objects individuals or variables on the basis of the similarity of the characteristics they possess. It seeks to minimize within-group variance and maximize between-group variance. It is also helpful for parental selection in the breeding programme and crop modeling. Therefore, the present study was done to evaluate the genetic diversity among sorghum genotypes specifically for grain and dry fodder yield to select the best genotypes can be exploited in future sorghum breeding programme.

MATERIALS AND METHODS

The plant material comprised of 32 elite entries of sorghum including two checks viz., GJ 39 and CSV 20 pooled under All India Coordinated Sorghum Improvement Project, Sorghum Research Station, Sardarkrushinagar Dantiwada Agricultural University, Deesa. Gujarat. The trial was grown in randomized block design with 3 replications during *Kharif* 2015 at Sorghum Research Station, Deesa (latitude of 24.5° N. longitude of 72° E and elevation of 136 M above the Mean Sea Level). The soil of the field was sandy in texture with pH value of 7.5 to 8.00 having good physical and chemical properties (Organic Carbon= 0.23, EC dsm^{-} = 0.232, K_2O= 259.9 kg/ha and P_2O_5= 46.2 kg/ha). The experimental unit was a four-row plot of 5.0 m long with row to row and plant to plant distance maintained at 0.45 meters and 0.15 cm, respectively. NPK 80:40:00 fertilizers was applied as half basal dose of nitrogen and full dose of phosphorus at the time of sowing and half nitrogen applied after one month of sowing. The all other recommended agronomical practices were followed to raise a good crop during the season. Data were taken on days to 50 % flowering, plant height (cm), number of leaves per plant, leaf length (cm), leaf width (cm), panicle length (cm), grain yield (kg/ha) and dry fodder yield (q/ha). The data was subjected to correlation analysis, principal component analysis (PCA) and cluster analysis using statistical software packages of SAS 9.2. Cluster analysis was performed using average linkage clustering while tree diagram based on euclidian distances was developed by Ward's method. The first two principal components were plotted against each other to find out the patterns of variability among genotypes and characters using SAS 9.2 software.

RESULTS AND DISCUSSIONS

The estimation of descriptive statistics of nine quantitative traits indicated the existence of diversity among the genotypes. Among all the traits investigated, grain yield, dry fodder yield, plant height, leaf length and days to maturity recorded higher variation in mean, range, variance and standard deviation (Table 1). Correlation is measure of strength of linear relationship in between the characters. In the present investigation grain yield was positively correlated with panicle length, leaf length and leaf width. Fodder yield was positively correlated with number of leaves/plant, leaf width, leaf length, plant height and days to maturity (Table 2). Such strong positive correlations recorded among the genotypes, suggest that they are heritable and genetically controlled traits which could be transmitted into desired

TABLE 1
Estimation of basic statistics for nine quantitative traits in 32 genotypes of sorghum

Variable	Mean	Mean deviation	Min-Max	Range	Variance	Standard deviation
Days to 50% flowering	70.56	4018	62.0-82.0	20.00	24.00	4.98
Days to maturity	102.40	4.78	97.0-117.0	20.00	28.75	5.47
Plant height	308.65	30.06	247.0-377.0	130.00	1303.8	36.97
No of leaves/plant	13.66	0.78	12.10-17.10	5.00	0.90	0.98
Leaf length	88.24	5.19	78.70-103.70	25.00	46.06	7.10
Leaf width	8.24	0.81	6.56-11.56	5.00	1.00	1.06
Panicle length	27.25	4.50	20.10-40.10	20.00	32.00	5.65
Grain yield	1996.45	280.12	1217.29-2657.29	1440.00	121500.0	373.85
Fodder yield	146.79	26.42	92.75-187.75	95.00	896.84	31.58

genotypes. The finding of present study was agreed with the Jain *et al*., (2011), Jain and Patel (2012). All the other yield contributing traits were also positively correlated with each other indicated that selection may be done in positive direction based on these traits towards crop improvement program.

The genetic diversity in 32 advance sorghum genotypes including two checks GJ 39 and CSV 20 were observed for the seed and fodder yield and their component traits for the selection of high yielding genotypes for further release and further breeding programmes. The PCA grouped the 9 quantitative traits in to nine components which accounted for the entire (100%) variability among the studied genotypes. According to Chatfied and Collins (1980), components with an eigenvalue of < 1 should be eliminated so that fewer components are dealt with. Furthermore, Hair *et al*. (1998) suggested that, eigenvalues greater than one are considered significant and component loading greater than 0.3 were considered to be meaningful. A scree plot is a simple line segment plot that shows the fraction of total variance in the data and it is a plot, in descending order of magnitude, of the eigen values of a correlation matrix. In the present study the Scree plot of the factors shows that the first three eigenvalues correspond to the whole percentage of the variance in the dataset (Fig. 1). In the present study first three Eigenvectors which has eigenvalues greater than one and cumulatively explained about 70.89 *per cent* of the total variation among the 32 nine quantitative traits in 32 genotypes of sorghum (Table 3). Hence PC-I has eigenvalue 2.91 and accounted for 32.44 % of the variations. This represents an equivalent of five variables viz., leaf length, leaf width, days to maturity, day to 50% flowering and fodder yield and indicated that were important contributing variables for the variation among the genotypes. Genotypes with high PC 1 score therefore would have high level variability of these quantitative traits. Chozin (2007), Mujaju and Chakuya (2008), and Ali et al. (2011) reported important contribution of the first PCs in total variability while studying different traits. The second and third PC explained 2.16 and 1.28 eigenvalues and contributing 24.08% and 14.31% variations, respectively. The second PCA was related to grain yield, panicle length and plant height, days to maturity and day to 50% flowering. In the third PC variation was composed of number of leaves/plant, fodder yield/plant, panicle length and grain yield. Overall, the PCA analysis under this study shows that phenotypic markers are useful in genotypes of sorghum and able to identify few key traits that accounted for the largest variability. The present study supported by earlier workers also (Ali *et al*., 2011 and Akatwijuka et al., 2016). Distribution of biometrical traits in first two components is sown in loading plot (Figure 2). The

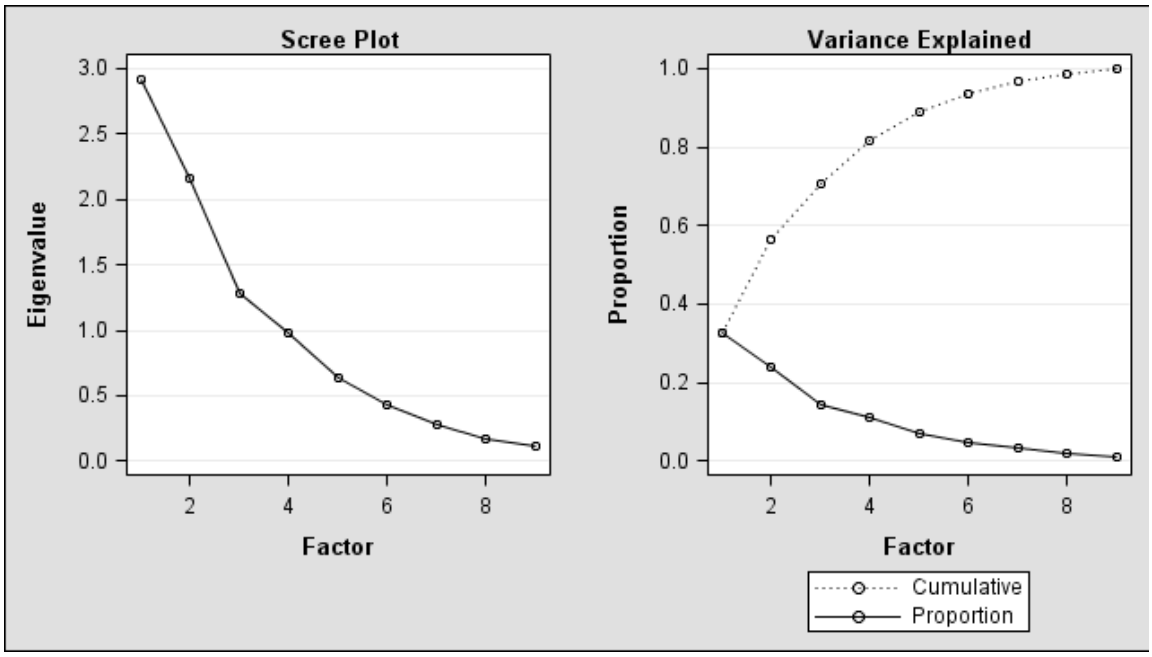

Fig. 1. Scree graph of factor analysis of 32 genotypes of sorghum between Eigen values and the number of factors.

TABLE 2
Pearson phenotypic correlation coefficient between nine quantitative traits in sorghum

	Days to 50% flowering	Days to maturity	Plant height	No of leaves/ plant	Leaf length	Leaf width	Panicle length	Grain yield	Fodder yield
Days to 50% flowering	**1.000**								
Days to maturity	0.855	**1.000**							
Plant height	-0.072	-0.165	**1.000**						
No of leaves/plant	0.091	0.127	0.224	**1.000**					
Leaf length	0.415	0.380	0.440	0.251	**1.000**				
Leaf width	0.302	0.396	0.182	0.175	0.512	**1.000**			
Panicle length	-0.066	-0.004	0.401	0.172	0.482	0.396	**1.000**		
Grain yield	-0.450	-0.363	-0.007	-0.164	0.083	0.133	0.384	**1.000**	
Fodder yield	0.036	0.139	0.296	0.513	0.256	0.483	0.048	0.076	**1.000**

loading plot clearly showed that dry fodder yield, grain yield, leaf width, number of leaves/plant, leaf length, and days to 50% flowering contributed traits towards diversity.

The biplot demarcated the genotypes with quantitative traits explained by the first two dimensions. A breeder in consequence, can simply predict the distance between the genotypes and make a decision for the selection of best genotypes, based on the numerous variables, compressed in the two foremost principal components and examined simultaneously. Genotypes closed to each other in the score plot are similar, genotypes located near the origin are distinctive genotypes and those far from the origin are extreme/ distinct. This is because of the principal component has been constructed with the data centered by the subtracting the average of each variable. The PCA analysis grouped the genotypes in to groups over the four quadrants base on the quantitative traits under study (Figure 2). The genotypes remained scattered in all four quadrants, showing large genetic variability in quantitative traits. The genotypes in the top left quadrants were closely related to grain yield. The right top quadrants consisted of the genotypes with related to number of leaves/plant, leaf length, leaf width, plant height and fodder yield. The right bottom quadrant comprised the genotypes related with the days to 50% flowering and days to maturity. The distance between the locations of any two genotypes on the score plot is

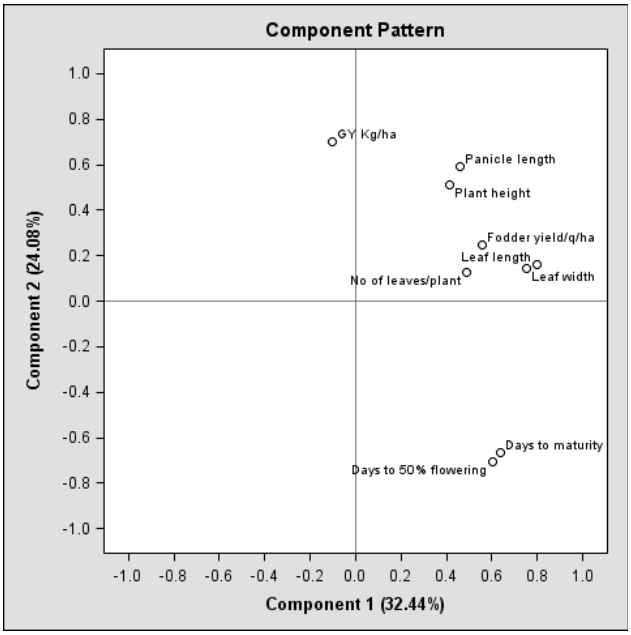

Fig. 2. Plot of the first two PCAs showing relation among various quantitative traits in sorghum.

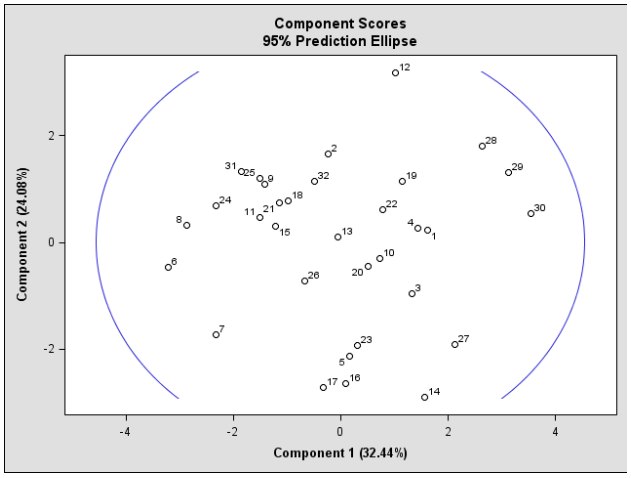

Fig. 3. Distribution of sorghum genotypes for first two principal components based on different quantitative traits.

TABLE 3
Eigenvectors and variance of 9 quantitative traits in genotypes of sorghum

	Eigenvectors		
	PC-1	PC-2	PC-3
Days to 50% flowering	0.352	-0.481	0.166
Days to maturity	0.373	-0.452	0.169
Plant height	0.244	0.346	-0.171
No of leaves/plant	0.287	0.085	-0.586
Leaf length	0.469	0.111	0.227
Leaf width	0.439	0.098	0.160
Panicle length	0.268	0.402	0.355
Grain yield	-0.061	0.474	0.349
Fodder yield	0.325	0.169	-0.497
Eigenvalue	2.919	2.167	1.288
Difference	75.30	87.90	30.50
Proportion %variance	32.40	24.10	14.30
Cumulative %variance	32.40	56.50	70.80

directly proportional to the degree of difference/ similarity between them in terms of the yield and yield components. Figure 3 therefore, revealed that genotype DS 0141 (#12), DS 0157 (#28), DS 0158 (#29), DS 0159 (#30), DS 0156 (#27), DS 0149 (#14), DS 0146 (#17), DS 01135 (#6) and DS 0137 (#8) were the most diverging from the major group which in the principal component axes was concentrated on zero depicting some similarity in term of the quantitative yield traits. The scatter plot from the first two PCs (Fig. 3) generally grouped the 32 genotypes in a similar way to cluster analysis (Fig. 4), using the entire data from all the traits. This showed that PCA is a reliable method in identifying few key traits contributing to the largest variation and could be a reliable method in predicting the important traits influencing clustering of different cultivars observed in Figure 4 under cluster analysis. According to Chahal and Gosal (2002), characters with largest absolute value closer to unity within the first principal component influence the clustering more than those with lower absolute value closer to zero. Therefore, in the present study, differentiation of the genotypes into different clusters was because of relatively high contribution of few characters rather than small contribution from each character.

The thirty two genotypes were grouped into five clusters on the basis of average linkage and dendogram

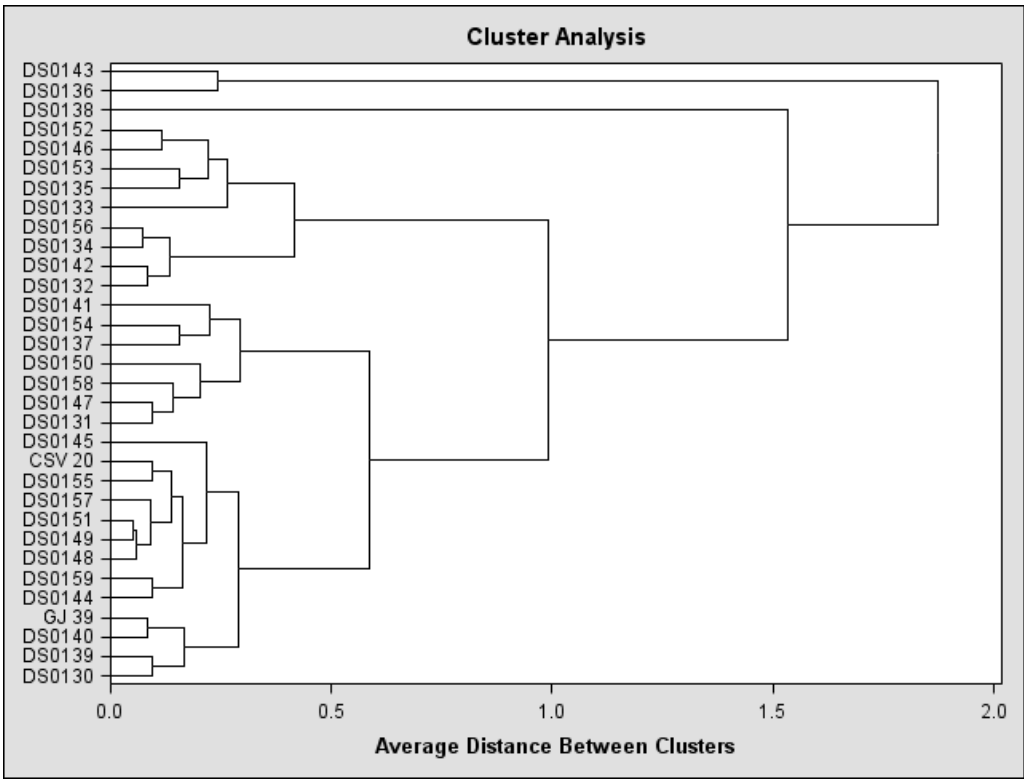

Fig. 4. The dendrogram of sorghum genotypes resulting from cluster analysis using ward method based on standardized data of all the traits.

was cut at a distance of 0.50 are presented in Fig. 4. The cluster analysis sequestrates genotypes into clusters which exhibit high homogeneity within a cluster and high heterogeneity between clusters. The cluster 1 having thirteen genotypes viz., DS130, DS139, DS140, GJ 39, DS144, DS159, DS148, DSF0149, DS151, DS157, DS155, CSV 20 and DSF0145. The cluster 2 includes 7 genotypes viz., DS131, DS147, DS158, DS150, DS137, DS154 and DS 141. Cluster 3 comprised 9 genotypes includes DS132, DS142, DS134, DS 156, DS 133, DS 135, DS 153, DS 146 and DS152. The cluster 4 contained one genotype DS 138 and cluster 5 composed of two genotype DS 143 and DS 136. Distribution pattern of all the genotypes into four clusters showed the presence of considerable genetics diversity among the genotypes for most of the traits under consideration. The clustering pattern showed that there was significant genetic variability among the sorghum genotypes tested that indicated the presence of excellent opportunity to bring about improvement through hybridizing genotypes from different clusters and assemble desirable traits with higher heterotic potential. Thus, the PC analysis, cluster analysis and correlation coefficient in this present set of the experiment provided facilitation in the classification of genotypes and identification of the subset of genotypes having quantitative difference between yield and yield parameters. Various useful correlations and aforementioned information extracted from cluster and PC analysis will be helpful in designing breeding programmes to obtain high yielding genotypes in sorghum for seed as well as fodder yield.

ACKNOWLEDGEMENT

We acknowledge Dr. E. V. D. Sastry, Professor, Plant Breeding and genetics, Rajasthan Agricultural Research Institute, SKN Agricultural University, Jaipur, Rajasthan, for statistical analysis of the data presented here.

REFERENCES

Akatwijuka, R., P. R. Rubaihayo and T. L. Odong, 2016: Genetic diversity among sorghum landraces of southwestern highlands of Uganda. *African Crop Sci. J.*, **24**:179-190.

Ali, M. A., K. Jabran, S. I. Awan, A. Abbas, Z. M. Ehsanullah, T. Acet, J. Farooq and A. Rehman, 2011: Morpho-physiological diversity and its implications for improving drought tolerance in grain sorghum at different growth stages. *Australian J. Crop Sci.*, **5**:311-320.

Chahal, G. S. and S. S. Gosal, 2002: Principles and Procedures of Plant Breeding, Biotechnology and Conventional Approaches. Narosa Publishing House, New Delhi.

Chatfield, C. and A. Collins, 1980: Introduction to multivariate analysis. Chapman and Hall.

Chozin, M., 2007: Characterization of sorghum accessions and choice of parents for hybridization. *J. Akta Agri. EdisiKhusus*, **2**:227-232.

Hair, J. F., R. L. Tatham, R. E. Andrson and W. Black, 1998: Multivariate data analysis. 5th Edn. Prentice-Hall International Inc., London, U.K., ISBN-13: 978-0138948580.

House, L. R., 1985: A Guide to Sorghum Breeding. 2nd edition. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. 206pp.

Jain, S. K. and P. R. Patel, 2012: Genetic variability in land races of forage Sorghum {Sorghum bicolor (L) Moench} collected from different geographical origin of India. *Int. J. Agri. Sci.*, **4**: 182-185.

Jain, S. K., M. Elangovan and P. R. Patel, 2011: Variation and association among fodder yield and other traits in germplasm of forage sorghum (Sorghum bicolor (L.) Moench) *Indian J. Plant Genet. Resou*, **24**: 327-331.

Mujaju, C. and E. Chakuya, 2008: Morphological variation of sorghum landrace accessions on-farm in Semi-arid areas of Zimbabwe. *Int. J. Botany*, **4**:376-382.

Sharma, J. R., 1998: Statistical and biometrical techniques in plant breeding. New Age International (P) Limited Publishers, New Delhi, India. 432 p.

Watson, J. W. and P. P. B. Eyzaguirre, 2002: Home gardens and in situ conservation of plant genetic resources in farming systems. Proceedings of the second international home gardens workshop, 17-19th July 2001, Witzenhausen, Federal Republic of Germany. International Plant Genetic Resources Institute, Rome, Italy.