Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2018; 7(3): 2034-2038
Received: 03-03-2018
Accepted: 07-04-2018

Khushboo Kathayat
Assistant Professor, Lovely Professional University, Phagwara, Punjab, India

ML Kushwaha
Professor, Department of Vegetable Science, G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India

Monisha Rawat
Assistant Professor, Lovely Professional University, Phagwara, Punjab, India

Correspondence
Khushboo Kathayat
Assistant Professor, Lovely Professional University, Phagwara, Punjab, India

Estimation of genetic diversity in taro (*Colocasia esculenta* (L.) Schott. Var. *antiquorum*) germplasm using principal component analysis and cluster analysis

Khushboo Kathayat, ML Kushwaha and Monisha Rawat

Abstract
Taro (*Colocasia esculenta* (L.) is one of the important underground vegetable crops grown vegetatively in various parts of India as well as in Africa. Only few varieties are recognized for cultivation from agricultural point of view, in spite of its large species diversity. Therefore, forty five genotypes of taro were evaluated in Augemented Block Design with three checks accommodated in six blocks during March to October 2016 at the Vegetable Research Centre, GB Pant University of Agriculture and Technology, Pantnagar, Uttarakhand to assess the genetic diversity using principal component analysis and cluster analysis. Approximately 77.052 % of variation was recorded due to first five principal components on various genotypes. The genotypes were further subjected to Hierarchical cluster analysis that resulted into seven non-overlapping clusters. The maximum number of genotypes was included in cluster IV. Cluster VII was recorded to have highest mean for maximum traits under study. Therefore PA-4 is used for developing high yielding taro varieties.

Keywords: taro, genetic diversity, principal component, cluster analysis, cluster mean

Introduction
Taro (*Colocasia esculenta* (L.) Schott var. *antiquorum*) known as eddoe type or *Arvi or ghuia* belongs to the monocotyledonous family Araceae. It is believed that the origin of domesticated taro is from 'wild type' *C. esculenta* var. *aquatilis*, which is from North East India or South East Asia (Matthews 1991). Cultivated ones are mostly diploid (2n=2x=28), although some triploids are found (2n=3x=42). It is a vegetatively propagated crop and cultivation is through the corms and cormels. It is grown widely in Uttar Pradesh, Bihar, Punjab, West Bengal, Assam, Uttarakhand, Orissa, Andhra Pradesh and Tamil Nadu. They are a good source of thiamine, riboflavin, iron, phosphorus, zinc and a very good source of vitamin B6, vitamin C, niacin, potassium, copper, and manganese. Despite its importance as a popular edible tuber crop and great morphological variability in nature, very little research has been done on it in Uttrakhand. At genotypic level and to assess relative contribution of different components to the total divergence both at intra and inter-cluster levels, genetic diversity is a useful tool for measuring the degree of divergence in a biological population (Jatasara and Paroda 1978) [5]. For estimation of diversity within the germplasm, several workers have postulated principal component analysis and clustering of genotypes (Smith *et al*. 1995) [16]. These techniques identify plant traits that help in characterization of the distinctness among selected genotypes.
During phenotypic evaluation of crop plants many traits are simultaneously evaluated. These traits are often highly interrelated, evaluation of all these traits is costly and may not enhance selection response. The principal component analysis (PCA) is one of the powerful statistical methods widely applied to classify phenotypic traits in crop germplasm into groups based on similarities. PCA guides the choice of parents for genetic improvement (Afuape *et al*. 2011) [1]. Principal component analysis or simply PCA is a statistical procedure concerned with elucidating the covariance structure of a set of variables. In particular it allows us to identify the principal directions in which the data varies. For the principal component analysis each genotype was identified on the basis of correlation matrix as a single point in a standardized multidimensional space. The axes of this space were principal components obtained from the original data as orthogonal transformation of the original variety. In this way each principal component becomes a linear combination of the varietal scores corresponding to the original variables. Cluster analysis is also carried out to detect divergent parents for hybridization purposes and to attain meaningful group constellations of a collection of genotype. Hence the

present study was undertaken with 45 germplasm of taro to determine the genetic divergence and to examine relative importance of the characters.

Materials and Methods

The present investigation was conducted in Vegetable Research Center, GB Pant University of Agriculture and Technology during March to October 2016, in Augemented Block Design having 6 blocks. In each block there were 7 genotypes along with three checks and each plot having 1 row of 3 meter long. The basic material for study involves 45 germplasms including three checks of taro germplasm (NDC-1, NDC-2 and Rajendra Arabi) collected from different parts of North India (Uttarakhand, Bihar, Uttar Pradesh and Himanchal Pradesh) and maintained at Vegetable Research Centre. Recommended cultural practices were followed to raise good crops. Observations were recorded for nineteen yield and yield related traits, namely, days to sprouting, plant height (cm), diameter of plants (cm), number of leaves per plant, leaf length (cm), leaf breadth (cm), leaf area index, number of cormels per plant, weight of corms per plant (g), weight of cormels per plant (g), length of corms (cm), width of corms (cm), length of cormels (cm), width of cormels (cm), yield (t/ha), specific gravity (g/cm^3), starch content (%), dry matter (%) and vitamin C content (mg/100g). The mean of the data on different characteristics were used for further analysis. Data were analyzed statistically for principal component and non-hierarchical Euclidean cluster analysis through software developed by Khetan graphics studio, Hyderabad. Principal component analysis and Non-hierarchical Euclidean Cluster Analysis was used for grouping all genotypes into clusters (Pearson 1901[12], Hotelling 1933) [4].

Results and Discussion

Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal component is less than or equal to the number of original variables. The principal component analysis of 45 genotypes based on correlation matrix of the morph-agronomical traits yielded 8 Eigen roots and Eigen vectors (Table 1). The eigen root for seven principal components were calculated as 5.215, 3.206, 2.672, 2.110, 1.523, 1.027, 0.769, respectively. Eigen root of first principal component accounted for 27.448 per cent of variation followed by second to seventh component which accounted for 16.871, 14.061, 11.108, 8.014, 5.407, and 4.048 per cent of total variation present among various genotypes respectively.

The Eigen vectors of six principal components were interpreted as relative weight of variables in each component. The important variables are those which have high positive or negative weights or values. Haydar *et al*. (2007) [3] said that the maximum characters that contribute to the diversity of the genetic material are traits that have the greatest value and positive feature vector.

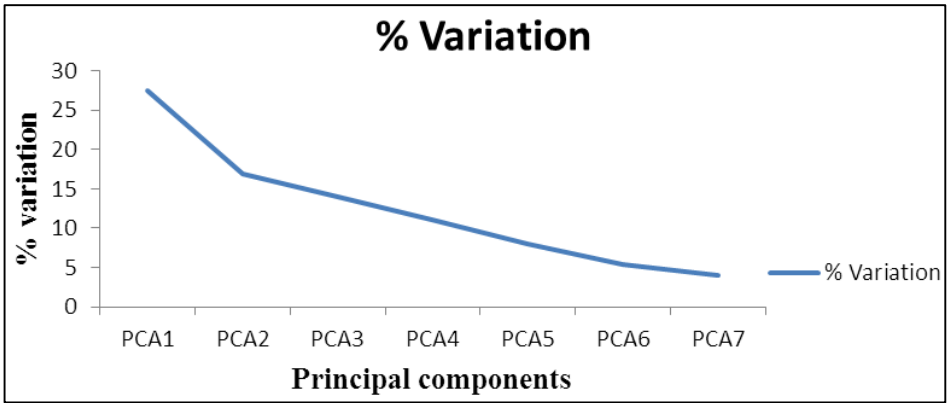
Table 1: Eigen vector, Eigen root and associated variation for principal component in taro based on economic traits

S. No	Traits	1	2	3	4	5	6	7
1.	Days to sprouting	0.281	0.201	0.102	0.323	0.158	0.245	0.051
2.	Plant height (cm)	0.002	0.459	0.142	0.036	-0.014	-0.079	0.019
3.	Diameter of plant (cm)	-0.119	0.182	0.199	0.088	0.231	-0.519	0.529
4.	No. of leaves per plant	0.288	0.072	0.168	0.334	-0.064	-0.072	-0.276
5.	Leaf length (cm)	-0.379	0.173	-0.059	-0.070	-0.107	-0.034	-0.075
6.	Leaf breadth (cm)	-0.290	0.284	-0.089	-0.083	-0.245	-0.045	-0.283
7.	Leaf area index	0.320	-0.315	0.086	0.093	0.203	-0.047	0.178
8.	No. of cormels per plant	-0.218	-0.275	-0.180	-0.157	0.199	0.176	0.340
9.	Weight of corms per plant (g)	0.290	-0.094	0.177	-0.349	0.014	-0.169	-0.130
10.	Weight of cormels per plant (g)	-0.136	0.091	-0.085	0.526	0.168	0.316	0.098
11.	Length of corms (cm)	0.249	0.295	-0.189	-0.319	-0.002	0.096	0.147
12.	Width of corms (cm)	-0.121	-0.029	-0.225	0.223	0.469	-0.299	-0.493
13.	Length of cormels (cm)	-0.116	-0.155	0.171	0.308	-0.378	-0.440	0.115
14.	Width of cormels (cm)	-0.306	0.095	-0.194	-0.060	0.313	-0.041	0.093
15.	Specific gravity(gm/cm^3)	-0.209	-0.252	0.236	0.045	-0.327	0.159	-0.028
16.	Dry matter content (%)	-0.124	0.069	0.469	-0.183	0.245	-0.127	-0.103
17.	Starch content (%)	0.136	0.459	-0.021	0.034	-0.149	0.065	0.205
18.	Vit. C content (mg/100g)	0.275	0.018	-0.389	-0.095	-0.006	-0.329	-0.099
19.	Yield (t/ha)	0.069	-0.087	-0.479	0.185	-0.292	-0.213	0.175
	Eigen root	5.215	3.206	2.672	2.110	1.523	1.027	0.769
	% variation	27.448	16.871	14.061	11.108	8.014	5.407	4.048
	Cumulative variation	27.448	44.319	58.381	69.488	77.502	82.909	86.958

Values in bold indicate the most relevant characters (> 0.42) that contributes to the variation of the components

In the present study first principal component had highest positive weight to leaf area index (0.320) followed by weight of corms per plant (0.290) and number of leaves per plant (0.288) and maximum negative weight due to leaf length (-0.379) followed by width of cormels (-0.306) and leaf breadth (-0.290). while second principal component had highest positive weight to starch content (0.459) followed by plant height (0.459) and length of corms (0.259) and maximum negative weight due to leaf area index (-0.315) followed by number of cormels per plant (-0.275) and specific gravity (-0.252). The third principal component had highest positive weight to dry matter content (0.469) followed by specific gravity (0.236) and diameter of plant (0.199) and maximum negative weight due to yield (-0.479) followed by vitamin C content (-0.389) and width of corms (-0.225). For each factor, a principal component loading of more than 0.42 was considered as being significant (Sheikh and Kumar, 2017) [15]. The maximum variation of 27.448 % was explained by first latent vector followed by 16.871 %, (second vector) and 14.061 % (third vector). Value of Eigen root fell lower than

one after the 6th principal component as has also been supported from previous studies (Mulualem 2013) [9]. First five principal components explained 77.502 % variation. PCA was performed to transform all the variables into new set of independent variables. The principal components that account for more than 90 % of variation were used for non-hierarchical Euclidean cluster analysis (Beale 1969) [2]. Covering 90 % of total variation us useful and should be adopted to explain the variation in the breeding material (Rao 1964) [14]. Similar studies have been done by several scientists in different tuber crops i.e. colocasia, elephant foot yam, aerial yam and sweet potato (Naskar and Sreekumar 2011 [10], Raichal *et al*. 2011[13], Laurie *et al*. 2013 [6], Mulualem and Weldemichael 2013) [10].



Cluster analysis is also carried out to detect divergent parents for hybridization purposes and to attain meaningful group constellations of a collection of genotype. The genotypes were grouped into seven non-overlapping clusters. (Table 2). This analysis can be regarded as efficient tools to categorize germplasm and renders reliable basis in choice of base material to plan future breeding strategies In this analysis, the maximum intra cluster distance was noted in cluster VI (287.23) having ten genotypes followed by cluster III (261.34) having 9 genotype, whereas minimum intra cluster distance recorded in cluster II, V and VII (0.000). It was favored to decide that intra-cluster diversity was the highest in cluster VI *i.e.*, more heterogeneous and intra-cluster diversity was lower in cluster II, V and VII *i.e.* comparatively homogenous.

Table 2: Average distance of inter and intra-cluster centroids

	I	II	III	IV	V	VI	VII
I	137.23	601.84	320.02	424.32	1456.16	763.68	1628.65
II		0.000	805.19	779.69	1989.15	1095.51	2092.33
III			261.34	536.13	1849.28	445.97	1288.15
IV				255.55	797.94	775.88	1027.01
V					0.000	2340.27	1590.79
VI						287.23	966.99
VII							0.000

Table 3: Distributing pattern of 42 germplasms along with checks of taro into 7 clusters

Cluster number	No. of germplasm	Germplasm included
I	11	NDC-6, PA-40, NDC-68, NDC-85, NDC-26, NDC-15, PKS-1, NDC-9, NDC-54, NDC-44 and KCS-2
II	1	PA-23
III	9	Mukta Keshi, PA-38, PA-1, NDC-7, NDC-1, PA-35, Aryan CA-1, NDC-2 and NDC-18
IV	12	PA-41, PA-9, PA-30, KCS-3, Saharasmukhi, PA-8, PA-3, PA-20, PA-42, PA-10, PA-28 and PA-33
V	1	Saatmukhi
VI	10	PA-12, PA-5, PA-22, NDC-71, PA-29, PA-24, Taliya, Rajendra Arabi, PA-16 and PA-39
VII	1	PA-4

The highest inter cluster distance was noticed between cluster V and VI (2340.27) followed by cluster II and VII (2092.33), cluster II and V (1989.15). The highest inter-cluster distance was observed between clusters V and VI indicated the genotypes in these clusters were far diverged than those of other clusters. Similarly, the lowest inter-cluster distance was observed between the cluster I and III (320.02). The inter cluster distance was greater than intra cluster distance implies that considerable amount of genetic diversity is present. These findings are in conformity with the earlier work (Mezhii *et al*. 2015) [8]. Higher inter and intra-cluster distances indicated higher genetic variability among genotypes between and within clusters, respectively. The minimum inter and intra-cluster distance indicated closeness among the genotypes of two clusters and within the cluster also.

An examination of results revealed that genotypes of same series (viz., PA series, NDC series and germplasm from Dholi, Bihar) fell into different clusters indicating that geographical diversity is not related to genetic diversity. The distribution of entries in various clusters showed that there was considerable amount of genetic divergence among the genotypes for all the character studied.

Table 4: Cluster mean for different economic traits in taro germplasms

S. No.	Characters	Clusters						
		I	II	III	IV	V	VI	VII
1.	Days to sprouting	23.75	25.39	29.54	28.64	24.56	30.97	28.22
2.	Plant height (cm)	118.07	139.61	115.23	121.99	109.66	120.33	144.83
3.	Diameter of plant (cm)	10.62	16.07	11.59	12.48	12.70	10.46	13.29
4.	No. of leaves per plant	6.03	7.66	6.51	6.97	5.76	6.76	4.94

5.	Leaf length (cm)	35.21	37.02	32.96	34.27	33.19	32.69	34.30
6.	Leaf breadth (cm)	31.75	33.62	28.89	30.20	26.53	29.01	35.63
7.	Leaf area index	847.25	941.73	728.81	786.33	683.07	716.26	952.76
8.	No. of cormels per plant	15.77	14.64	16.75	14.38	13.58	12.27	13.09
9.	Weight of corms per plant (g)	109.29	130.43	161.86	192.72	185.38	224.91	274.52
10.	Weight of cormels per plant (g)	411.07	326.69	342.11	336.99	253.59	232.78	303.05
11.	Length of corms (cm)	4.44	5.57	6.09	5.13	3.13	3.33	9.97
12.	Width of corms (cm)	4.20	5.15	4.84	3.87	3.31	4.29	6.25
13.	Length of cormels (cm)	5.01	13.03	5.05	5.09	4.06	5.47	4.58
14.	Width of cormels (cm)	1.29	1.15	1.46	0.96	1.59	0.85	0.89
15.	Specific gravity(gm/cm^3)	1.33	1.55	1.43	1.32	1.29	1.40	1.45
16.	Dry matter content (%)	25.41	27.35	24.84	31.91	41.70	25.87	39.37
17.	Starch content (%)	19.61	21.21	19.14	25.01	33.15	20.07	31.21
18.	Vitamin C content (mg/100g)	6.53	7.51	6.59	5.49	5.13	6.76	7.52
19.	Yield (t/ha)	30.82	25.46	29.37	27.42	24.09	24.56	32.15

Cluster I had highest mean value for weight of cormels per plant (411.07) and cluster III had highest mean value for number of cormels per plant (16.75). Highest mean value for diameter of plant (16.07), number of leaves per plant (7.66), leaf length (37.02), length of cormels (13.03) and specific gravity (1.55) were calculated in genotypes that belongs to cluster II. Highest mean value for width of cormels (1.59) and starch content (33.15) and lowest mean value for yield (24.09) were recorded to those genotypes that belong to cluster V. In cluster VI, genotype that has highest mean value for days to sprouting was included. Cluster VII had highest mean value for plant height (144.83), leaf breadth (35.63), leaf area index (952.76), weight of corms per plant (274.52), length of corms (9.97), width of corms (6.25), dry matter content (39.37), vitamin C content (7.52) and yield (32.15).

The investigation results revealed that there was no sharp relationship between the clustering pattern of the genotypes and their geographical sources. Thus, the tendency of cultivars occurring in cluster cutting across geographical boundaries is possible due to genetic makeup of genotypes and subsequently natural selections during their development. So data on inter cluster distances and mean performance of *Colocasia* accession were used to select genetically diverse and agronomical superior genotypes which can be used for crop improvement programme. It can aid taro breeders to make informed decisions in selecting parental accessions for hybridization from diverse clusters. Results of present study revealed that multivariate analysis helps to place the genotypes in different clusters on the basis of PC(s) values. The data on inter cluster distances and mean performance of *Colocasia* accession were used to select genetically diverse and agronomical superior genotypes. Cluster VII were recorded to have highest mean for maximum traits under study.

Conclusion

Important variables in taro germplasm with respect to agronomic traits were number of cormels per plantweight of corms per plant and weight of cormels per plant Considering the magnitude of genetic divergence and magnitude of cluster means for different characters performance, the genotype PA-4 is used for developing high yielding *Colocasia* varieties. These genotypes were found promising for most of characters. Therefore, it is recommended that such genotypes which were found promising for most of traits including yield, can be utilized as commercial cultivar of taro after following the standard release procedure of variety in India.

References

1. Afuape SO, Okocha PI, Njoku, D. Multivariate assessment of the agromorphological variability and yield components among sweet potato (*Ipomoea batatas* (L.) Lam) landraces. African Journal of Plant Science. 2011; 5(2):123-132.
2. Beale EML. Euclidean cluster analysis. 37th session of int. statistical institute, UK, 1969.
3. Haydar A, Ahmed MB, Hannan MM, Razvy MA, Mandal MA, Salihin M. et al. Analysis of genetic diversity in some potato varieties grown in Bangladesh. Middle East Journal of Scientific Research. 2007; 2:143-145.
4. Hotelling H. Analysis of complex of statistical variables into principal components. Journal of Educational pshycology. 1933; 24:417-441.
5. Jatasara DS, Paroda RS. Genetic divergence in Wheat under different environmental conditions, Cereal Research Communications. 1978; 6:307-317.
6. Laurie SM, Calitz FJ, Adebola PO, Lezar A. Characterization and evaluation of South African sweet potato (*Ipomoea batatas* (L.) LAM) land races. South African Journal of Botany. 2013; 85:10-16.
7. Matthews PJ. A possible tropical wild type taro (*Colocasia esculenta* var. *aquatilis*). Indo Pacific Prehistory Association Bulletin. 1991; 11:69-81.
8. Mezhii TL, Changkija S,m Chaturvedi HP. Genetic Diversity Analysis in Indigenous Edible Aroids of Nagaland. Indian Res. J. Genet. & Biotech. 2015; 7(4):442–447.
9. Mulualem T. Morpho- agronomical Characterization of Taro (*Colocasia esculenta)* accessions in Ethiopia. Plant. 2013; 1(1):1-9.
10. Mulualem T, WeldeMichael G. Agronomical Evaluation of Aerial Yam/*Dioscorea bulbifera*/ Accessions collected from South and Southwest Ethiopia. Greener Journal of Agricultural Sciences, 2013; 3(9):693-704.
11. Naskar SK, Sreekumar J. Evaluation of Taro (*Colocasia esculenta* (L.) Schott.) germplasm using Multivariate Analysis. Journal of Root Crops. 2011; 37(2):150-154.
12. Pearson K. On lines and planes of closest fit to systems of point in space. Philospphical Magazine, 1901; 2:559-572.
13. Raichal Anil S, Siril EA, Suhara Beevy S. Morphological variability in 17 wild elephant foot yam (*Amorphophallus paeoniifolius*) collections from Southwest India. Genetic Resources and Crop Evolution. 2011; 58:1263-1274.

14. Rao TS. Advance statistical methods in Biometrical Research. John Willey and sons. New York, 1964.
15. Sheikh N, Kumar Y. Morphological Characterization of Meghalayan *Dioscorea* spp. (yam), North East India. Journal of Agricultural Science and Technology. 2017; 19:487-497.
16. Smith SE, Guarino L, Al-Dos A, Cont, DM. Morphological and agronomical affinities among Middle Eastern alfalfa accessions from Oman and Yemen. Crop Science. 1995; 35:1188-1194.