

Published with Open Access at **Journal BiNET**

Vol. 12, Issue 02: 1029-1035

**Journal of Bioscience and Agriculture Research**Journal Home: [www.journalbinet.com/jbar-journal.html](http://www.journalbinet.com/jbar-journal.html)

## Hierarchical cluster, euclidean distance and principal component analysis based on phenotypic characters of hosta

Hasan Mehraj<sup>1,2\*</sup> and Kazuhiko Shimasaki<sup>2</sup><sup>1</sup>The United Graduate School of Agricultural Sciences, Ehime University, Ehime 790-8556, Japan<sup>2</sup>Lab of Vegetable and Floricultural Science, Faculty of Agriculture, Kochi University, B200 Monobe, Nankoku, Kochi 783-8502, Japan✉ For any information: [ask.author@journalbinet.com](mailto:ask.author@journalbinet.com), [hmehraj02@yahoo.com](mailto:hmehraj02@yahoo.com)\*

Available online: 25 February 2017.

### ABSTRACT

*Hosta* is a widely grown ornamental plants and sometimes used as vegetables. Hybridization of hostas is frequently done by home gardener. We conducted an experiment using twelve hosta taxa (coded from T1-T12) to analyze the morphological traits based hierarchical cluster, euclidean distance and PCA with a view to provide necessary information to the hosta breeders. We used complete randomized design with five replicates for the data collection of morphological traits. *H. sieboldiana* and *H. montana* had similar morphological characters while *H. longipes* var. *gracillima*, *H. kiyosumiensis* were mostly similar morphological characters as well. From the dissimilarity analysis, it was found that the dissimilarity distance between *H. alismifolia* and *H. montana* was maximum (452.2) which was closely followed by *H. longissima* and *H. montana* (447.6) where as minimum dissimilarity distance between *H. sieboldii* and *H. nakaiana* (36.2). PCA indicated that two important components accounted for 97.76% of the total variation among traits in of hosta taxa. The first and second component contributed 81.92% and 15.84% of the total variation, respectively. Leaf area was found as a unique morphological trait. *H. sieboldiana*, *H. montana* and *H. longipes* var. *gracillima* shared the common components; as well as *H. alismifolia* with *H. longissima*, *H. sieboldii* with *H. nakaiana* and *H. tardiva* with *H. longipes* var. *caduca* comprised common components. Overall, results revealed importance morphological components in breeding ability in hosta taxa.

**Key Words:** Plantain lily, Giboshi, Taxa, Dendrogram, Proximity distance matrix and PCA

**Cite Article:** Mehraj, H. and Shimasaki, K. (2017). Hierarchical cluster, euclidean distance and principal component analysis based on phenotypic characters of hosta. Journal of Bioscience and Agriculture Research, 12(02), 1029-1035. **Crossref:** <https://doi.org/10.18801/jbar.120217.125>



Article distributed under terms of a Creative Common Attribution 4.0 International License.

### I. Introduction

*Hosta* is a plant genus, commonly known as hostas, plantain lilies and giboshi (in Japan). It is herbaceous shade tolerant perennial ornamental plant. Sometimes Japanese and Koreans are used

hostas as a vegetable. But hostas are mainly grown for their beautiful foliage that comes with a broad range of leaf shapes, colors, sizes and textures. Delimitation of hosta species has been tricky due to a long history of cultivation, hybridization and selection; particularly in Japan, from the eighth century onwards (Schmid, 1991). Hybridizing and tissue culture propagation results the existence of hundreds of species and thousands of cultivars to the home gardener, recently. Nomenclature of this plant genus is also complicated due to many names i.e., based on types of garden origin or sports originating among wild populations. Their taxonomic classifications were largely based on regionally cultivated materials in Japan (Fujita, 1976), North America (Bailey, 1930), Korea (Chung, 1990; Kim, 1991) and Europe (Hylander, 1960). Hostas have different numbers of species mentioned in previous study, like 23–26 species (Maekawa and Kaneko, 1968; Fujita, 1976), otherwise 40 or more (Maekawa 1940; Schmid, 1991). All hosta species and cultivars showed a numerous morphological variation. Morphological characters variability is the expressive parts of the genetic variability. Genetic variability of breeding materials may lead to the success of the crop improvement program (Prabhu et al. 2009). Selection through morphological characters greatly influence on success of the crop improvement program (Prabakaran, 2010). Phenotypic variability among genotypes gives an indication of potential genotypic variability. Proximate distance analysis helps for the determination of indirect selection of genotypes. Principal component analysis is a suitable multivariate technique in identification and determination of independent principal components; those are effective on plant traits separately. PCA is used in population genetics for the illustration of the object diversity and allocated to group according to similarity through the construction of maps to summarize genetic variation across traits, genotypes or both. A controversial but influential interpretation in these maps indicates continuous spatial variation (Novembre and Stephens, 2008). Application of PCA tool provides a useful means for estimating morphological diversity within and between germplasm (Maji and Shaibu, 2012). Dendrogram and PCA analysis helps to determine effective traits and genotypes for indirect selection of superior genotypes. Wide diversity of hosta gives an idea of its potential for selection work. There is no scientific work for this plant genus to compare of genotypes based on their morphological characters. Therefore this work aimed to analyze the morphological traits based hierarchical cluster, euclidean distance and PCA.

## II. Materials and Methods

A pot experiment was conducted at Kochi University, Kochi 783-8520, Japan from January to September 2016. Single rhizome was planted in each pot at the last week of January 2016. Twelve taxa of hosta were used in the experiment and those were symbolized as T1: *Hosta sieboldiana*, T2: *H. alismifolia*, T3: *H. sieboldii*, T4: *H. longissima*, T5: *H. tardiva*, T6: *H. longipes* var. *gracillima*, T7: *H. nakaiana*, T8: *H. kikutii* var. *caput-avis*, T9: *H. kikutii* var. *polyneuron*, T10: *H. longipes* var. *caduca*, T11: *H. kiyosumiensis*, and T12: *H. montana*. Complete randomized design (CRD) with five replicates (5 pots/taxon, a total of 60 pots) was followed for the data collection. For hosta cultivation, we used Nursery earth® (Takii & Co. Ltd., Kyoto, Japan), a year-round pot soil supplemented with 320 mg/L Nitrogen; 210 mg/L Phosphorus and 300 mg/L Potassium and having mild acidity. We didn't apply any additional fertilizers or other ingredients with this pot soil. Data were collected on different morphological traits [plant height (cm), number of leaves, single leaf area (cm<sup>2</sup>), number of principal vine/leaf, chlorophyll content (%), length of peduncle (cm), length of flower (cm) and number of flower/inflorescence]. We used IBM SPSS (version 16.0) statistics for proximity distance matrix analysis and XLSTAT (Addinsoft, 2016) for principal component analysis. The PCA was assumed by the entire basis in orthogonal matrix (Jolliffe, 2002) which was computed from the eigen values of the covariance matrix.

## III. Results and Discussion

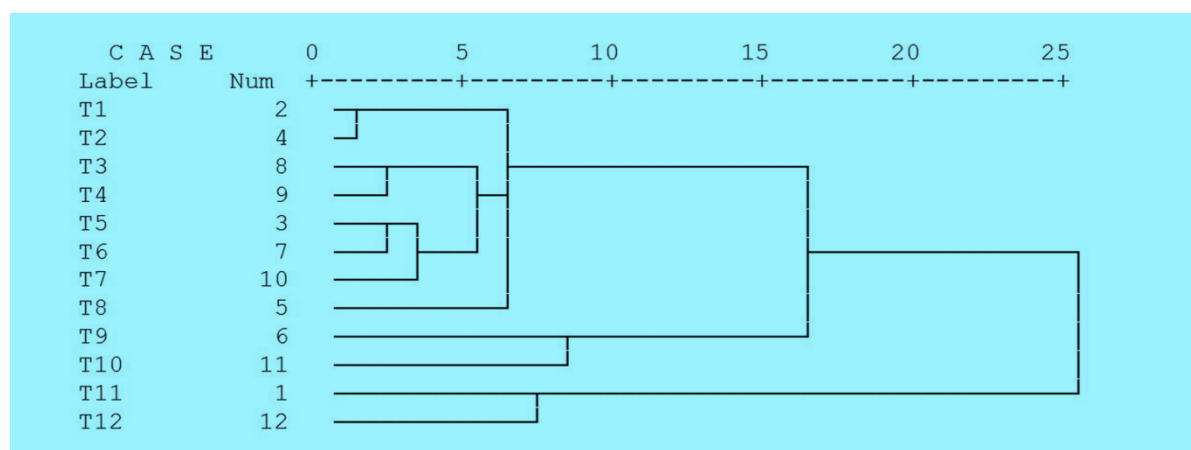
### Cluster analysis and dissimilarity matrix of hosta taxa

We build dendrogram (Ward's method) for 12 taxa of hosta using morphological traits that was represented in Figure 01. Dendrogram represented the fairly and simply interpreted relationship among the taxa. The dissimilarity was represented by the horizontal axis of the dendrogram. Vertical axis represented the objects and clusters. Each joining (fusion) of two clusters on the graph is represented by splitting of a horizontal line into two. The horizontal position of the split, shown by the short vertical bar, gives the distance (dissimilarity) between the two clusters. *H. sieboldiana* and *H. Montana* were in same cluster while *H. longipes* var. *gracillima* and *H. kiyosumiensis* were in the same

cluster (Figure 01). *H. sieboldiana* and *H. montana* had similar morphological characters. *H. longipes* var. *gracillima* and *H. kiyosumiensis* were posing mostly similar morphological characters as well. From the dissimilarity analysis, it was found that the dissimilarity distance between *H. alismifolia* and *H. montana* was maximum (452.2) which was closely followed by *H. longissima* and *H. montana* (447.6). It indicated that *H. alismifolia* and *H. montana* were mostly different on the morphological performance; *H. longissima* and *H. Montana* as well. On the other hand, it was found the minimum dissimilarity distance between *H. sieboldii* and *H. nakaiana* (36.2) (Table 01). *H. sieboldii* and *H. nakaiana* was found as the similarly performing taxa.

**Table 01. The proximity dissimilarity matrix (euclidean distance) between hosta taxa**

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	0	<b>415.2</b>	366.9	<b>410.3</b>	334.6	176.7	363.5	321.4	314.1	344.9	228.1	105.5
T2		0	54.3	13.5	92.2	324.1	69.0	110.9	132.2	88.5	220.2	<b>452.2</b>
T3			0	51.0	60.1	279.3	<b>36.2</b>	70.0	94.0	46.0	175.0	406.9
T4				0	88.1	320.4	65.3	106.9	127.9	85.6	216.3	<b>447.6</b>
T5					0	259.1	81.0	87.4	111.9	77.1	164.0	365.1
T6						0	278.3	232.2	224.0	249.2	111.7	232.4
T7							0	57.0	75.9	43.7	170.7	<b>410.5</b>
T8								0	40.3	41.6	123.9	373.0
T9									0	67.5	115.4	373.5
T10										0	144.0	391.5
T11											0	288.0
T12												0



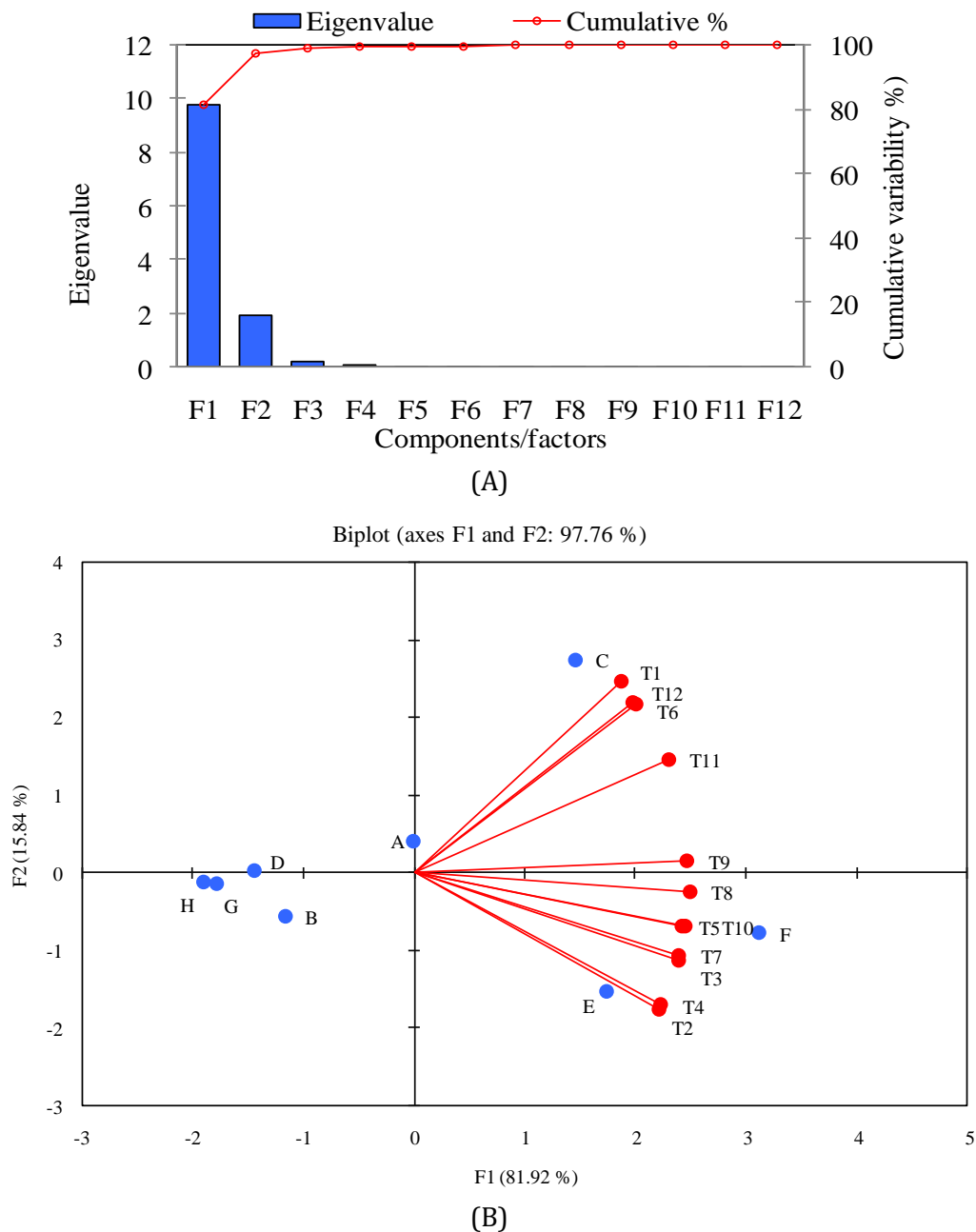
**Figure 01. Dendrogram (hierarchical cluster) of hosta taxa by average linkage (between groups) rescaled distance cluster combine (WARD's method) based on morphological traits.**

T1: *H. sieboldiana*, T2: *H. alismifolia*, T3: *H. sieboldii*, T4: *H. longissima*, T5: *H. tardiva*, T6: *H. longipes* var. *gracillima*, T7: *H. nakaiana*, T8: *H. kikutii* var. *caput-avis*, T9: *H. kikutii* var. *polyneuron*, T10: *H. longipes* var. *caduca*, T11: *H. kiyosumiensis*, and T12: *H. Montana*.

### Principal component analysis for hosta taxa and agro-morphological traits

A high percentage of the variance from the first two eigen values (81.92% and 15.84% of the initial variability) (Figure 02A), ensured us to map using first two factors for quality projection of multi-dimensional PCA. But some of the information might be hidden in the next factors. The definitive goal of PCA was shown in Figure 02B. It's a two dimensional map for the identification of the trends of these hosta taxa. The result of PCA analysis shows that the demographics of twelve hosta taxa and their morphological traits. PCA indicated that two important components accounted for about 97.76% of the total variation among traits in hosta taxa. The first component accounted 81.92% (Figure 02B) whereas second component was 15.84% of the total variation. The first principal component was strongly associated leaf area, chlorophyll content and peduncle length components. For the

morphological traits PCA analysis, we found leaf area as a unique character. Leaf area, chlorophyll content, and peduncle length were found as the highly variable traits. *H. sieboldiana*, *H. longipes* var. *gracillima*, *H. kikutii* var. *polyneuron*, *H. kiyosumiensis* and *H. Montana* belonged to same co-ordinates and rests were in another co-ordinate. A strong association was found between component and all of the hosta taxa. This association confirmed the high level of phenotypic variation among the hosta taxa that is about 81.92% of the total variation. Though the taxa were highly variable with each other but *H. sieboldiana*, *H. montana* and *H. longipes* var. *gracillima* were mostly similar. Similarly very similar traits were showed by *H. alismifolia* with *H. longissima*, *H. sieboldii* with *H. nakaiana* and *H. tardiva* with *H. longipes* var. *caduca* (Figure 02B).



**Figure 02. Principal component analysis for the dataset of twelve hosta taxa (A) Scree plot (B) PCA biplot.**

T1: *H. sieboldiana*, T2: *H. alismifolia*, T3: *H. sieboldii*, T4: *H. longissima*, T5: *H. tardiva*, T6: *H. longipes* var. *gracillima*, T7: *H. nakaiana*, T8: *H. kikutii* var. *caput-avis*, T9: *H. kikutii* var. *polyneuron*, T10: *H. longipes* var. *caduca*, T11: *H. kiyosumiensis* and T12: *H. montana*

A: plant height, B: number of leaves, C: leaf area, D: number of principal vine/leaf, E: chlorophyll content, F: length of peduncle, G: length of flower and H: number of flower/inflorescence  
F1 and F2 belongs to Factor 1 and Factor 2.

## Discussion

From the results of the cluster analysis, proximity dissimilarity matrix and PCA analysis based on morphological characters, we found the similarity and variation among the hosta taxa (Figure 01, Figure 02 and Table 01). Variability for characters was significantly different among the accessions (Shekar et al. 2012; Kumar et al. 2013; Arivalagan et al. 2013). Touhiduzzaman et al. (2016), Sikder et al. (2015) and Hoque and Rahman (2007) were also studied distance analysis. We found maximum distance between *H. alismifolia* and *H. Montana* along with minimum between *H. sieboldii* and *H. nakaiana*. Table 01 showed the proximity dissimilarity by Euclidean Distance of each hosta taxa with other. From this analysis, the result will help to the hosta breeders for the selection of taxa without concerning the gene levels. This proximity distance of genotypes will be helpful for the breeding program. PCA of the phenotypic traits disclosed that the total contribution of the characters to the variation. We found that 1<sup>st</sup> two components accounted for 97.76% of the total variation. The first three principal components are the most significant for reflection of the variation patterns among accessions wherever the traits related with these three components (Guei et al. 2005). Our result showed 97.76% of total variation in first two components that provided an apparent idea of the structure underlying the variables analyzed. It is expected to provide high level of gene transfer during breeding programs due to their highly variable characters (Varthini et al. 2014; Gana et al. 2013; Gana, 2006). On the other hand, the 1<sup>st</sup> and 2<sup>nd</sup> component showed the existence of variation within the components. Leaf area, chlorophyll content and peduncle length showed more variation than the other studied characters. All of the hosta taxa loaded on first principle component that indicated the high level of the variability among them. This high level of variability within the varieties and the phenotypic characters will make room for the improvement hostas through breeding programs. The variation was found among hosta taxa to their respective traits and the results are the agreement with earlier findings reported by many researchers (Hailu, 2016; Mahendran et al. 2015; Gana et al. 2013; Ashfaq et al. 2012; Shahidullah et al. 2009; Bose and Pradhan, 2005; Cheema et al. 2004). These analyses will be helpful to the hosta breeders for genetic improvement traits; that have low heritability for indirect selection (Golparvar et al. 2006).

#### IV. Conclusion

Prominent phenotypic traits and genotypes came together in the first two components of PCA that will help for the selection in hybridization for hosta improvement. It was found a high level of variability among the phenotypic traits and genotypes. First two components were highly related to this variability. Association and proximity euclidean distance among the hosta taxa can identify the diverge genotypes to the breeding program.

#### Acknowledgments

Authors are thankful to laboratory mates who were associated for helping during the experimental period. Authors are also thankful to the personnel who provided immense support for collection of plant materials.

#### Conflicts of Interest

Authors declared no conflict of interest.

#### Authors Contribution

Hasan Mehraj conceived, designed and performed the experiment, analyzed the data and wrote this manuscript as well. Kazuhiko Shimasaki contributed the plant materials and other experimental facilities as a supervisor of the experiment.

#### V. References

- [1]. Addinsoft (2016). XLSTAT v. 2016. Data analysis and statistics software for Microsoft Excel. Addinsoft, New York, USA.
- [2]. Arivalagan, M., Bhardwaj, R., Gangopadhyay, K. K., Prasad, T. V. and Sarkar, S. K. (2013). Mineral composition and their genetic variability analysis in egg plant (*Solanum melongena* L.) germplasm. Journal of Applied Botany and Food Quality, 86, 99-103.

- [3]. 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.
- [4]. Bailey, L. H. (1930). Hosta: The plantain lilies. Gentes Hebarum, 2, 119-142.
- [5]. Bose, L. K. and Pradhan, S. K. (2005). Genetic divergence in deepwater rice genotypes. Journal of Central European Agriculture, 6(4), 635-640.
- [6]. Cheema, A. A., Rashid, M., Ashraf, M. and Qamar, Z. (2004). Genetic divergence in rice collection. Pakistan Journal of Botany, 36(3), 557-565.
- [7]. Chung, M. G. (1990). A Biosystematic study on the genus Hosta (Liliaceae/Funkiaceae) in Korea and Tsushima Island of Japan. Ph. D. Dissertation, University of Georgia, Athens.
- [8]. Chung, M. G. and Kim, J. W. (1991). The genus Hosta Tratt. (Liliaceae) in Korea. Sida, Contributions To Botany, 14(3), 411-420.
- [9]. Fujita, N. (1976). The genus Hosta (Liliaceae) in Japan. Actaphytotaxonomicaet Geobotanica, 27(3-4), 66-96 (in Japanese).
- [10]. Gana, A. S. (2006). Variability studies of the response of rice varieties to biotic and abiotic stresses. Ph.D Thesis, University of Ilorin.
- [11]. Gana, A. S., Shaba, S. Z. and Tsado, E. K. (2013). Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa* L.) grown in a rainfed lowland ecology of Nigeria. Journal of Plant Breeding and Crop Science, 4(6), 120-126.  
<https://doi.org/10.5897/JPBCS12.065>
- [12]. Golparvar, A. R., Ghasemi-Pirbalouti, A. and Madani, H. (2006). Genetic control of some physiological attributes in wheat under drought stress conditions. Pakistan Journal of Biological Sciences, 9(8), 1442-1446. <https://doi.org/10.3923/pjbs.2006.1442.1446>
- [13]. Guei, R. G., Sanni, K. A. and Fawole, A. F. J. (2005). Genetic diversity of rice (*Oryzasativa*). Agron Africaine, 5, 17-28.
- [14]. Hailu, A. F. (2016). Principal component analysis of morphological traits in thirty-six accessions of amaranths (*Amaranthus Spp.*) grown in a rainfed under Mizanand Tepe Conditions, South West Ethiopia. Journal of Biology, Agriculture and Healthcare, 6(1), 113-121.
- [15]. Hoque, M. N. and Rahman, L. (2007). Estimation of euclidean distance for different morpho-physiological characters in some wild and cultivated rice genotypes (*Oryza sativa* L.). Journal of Biological Sciences, 7(1), 86-88. <https://doi.org/10.3923/jbs.2007.86.88>
- [16]. Hylander, N. (1960). The genus Hosta. Journal of the Royal Horticultural Society, 85, 356-369.
- [17]. Jolliffe, I. T. (2002). Principal component analysis, 2<sup>nd</sup> edn. Springer Series in Statistics, Springer – Verlag New York Inc, New York.
- [18]. Kumar, S. R., Arumugam, T. and Anandakumar, C. R. (2013). Genetic diversity in egg plant (*Solanum melongena* L.). Plant Gene Trait, 4(2), 4-8.
- [19]. Maekawa, F. (1940). The genus Hosta. Journal of the Faculty of Science, the University of Tokyo. Sect. 1, Bot. 5, 317-425.
- [20]. Maekawa, F. and Kaneko, K. (1968). Evolution of karyotype in Hosta (Liliaceae). Journal of Japanese Botany, 43, 132-140.
- [21]. Mahendran, R., Veerabathiran, P., Robin, S. and Raveendran, M. (2015). Principal component analysis of rice germplasm accessions under high temperature stress. International Journal of Agricultural Science and Research, 5(3), 355-360.
- [22]. 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. <https://doi.org/10.5897/jpbcs11.093>
- [23]. Novembre, J. and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. Nature Genetics, 40, 646-649.  
<https://doi.org/10.1038/ng.139>  
PMid:18425127 PMCID:PMC3989108
- [24]. Prabakaran, S. (2010). Evaluation of local types of brinjal (*Solanum melongena* L.). M. Sc., (Hort.) Thesis, Agricultural College and Research Institute, TNAU, Madurai.
- [25]. Prabhu, M., Natarajan, S. and Pugalendhi, L. (2009). Variability and heritability studies in F5 and F6 progenies of brinjal. American-Eurasian Journal of Sustainable Agriculture, 3(3), 306-309.
- [26]. Schmid, W.G. 1991. The genus Hosta: Gibōshi Zoku (ギボウシ属). London and Portland: Timber Press. pp. 27-110.

- [27]. Shahidullah, S. M., Hanafi, M. M., Ashrafuzzaman, M., Ismail, M. R. and Salam, M. A. (2009). Phenological characters and genetic divergence in aromatic rices. *African Journal Biotechnology*, 8, 3199–3207.
- [28]. Shekar, C. K., Ashok, P. and Sasikala, K. (2012). Studies on heritability and multivariate analyses in brinjal (*Solanum melongena* L.). *Vegetable Crops Research Bulletin*, 76, 79-88.  
<https://doi.org/10.2478/v10032-012-0005-2>
- [29]. Sikder, R. K., Rahman, M. A., Asif, M. I., Jamal Uddin, A. F. M. and Mehraj, H. (2015). Genetic variability, distance and traits interrelationship analysis of Nerica and Inpari rice varieties. *Scientia Agriculturae*, 10(1), 44-48.
- [30]. Touhiduzzaman, Sikder, R. K., Asif, M. I., Mehraj, H. and Jamal Uddin, A. F. M. (2016). Correlation and genetic distance on sixteen rice varieties grown under SRI. *Advances in Plants & Agriculture Research*, 3(3), 00100.
- [31]. Varthini, V. N., 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.

### HOW TO CITE THIS ARTICLE?

**Crossref:** <https://doi.org/10.18801/jbar.120217.125>  
**Article URL:** <http://www.journalbinet.com/jbar-120117-125.html>

#### APA (American Psychological Association)

Mehraj, H. and Shimasaki, K. (2017). Hierarchical cluster, euclidean distance and principal component analysis based on phenotypic characters of hosta. *Journal of Bioscience and Agriculture Research*, 12(02), 1029-1035.

#### MLA (Modern Language Association)

Mehraj, H. and Shimasaki, K. "Hierarchical cluster, euclidean distance and principal component analysis based on phenotypic characters of hosta". *Journal of Bioscience and Agriculture Research*, 12.02(2017): 1029-1035.

#### Chicago and or Turabian

Mehraj, H. and Shimasaki, K. "Hierarchical cluster, euclidean distance and principal component analysis based on phenotypic characters of hosta". *Journal of Bioscience and Agriculture Research*, 12 no.02(2017): 1029-1035.

#### Journal BiNET | Scientific Publication

- ✓ Faster processing & peer review
- ✓ International editorial board
- ✓ 29 business days publication
- ✓ Greater audience readership
- ✓ Indexing & bibliographic integration
- ✓ Social sharing enabled

Submission or email to [submit@journalbinet.com](mailto:submit@journalbinet.com)

[www.journalbinet.com/article-submission-form.html](http://www.journalbinet.com/article-submission-form.html)