Received: 26 Jul 2021 | Revised: 18 Sep 2021 | Accepted: 22 Sep 2021

Electronic Journal of Plant Breeding

Research Article

Enabling different genetic diversity algorithms to unravel conservation and divergence in moringa germplasm accessions

Allen Eldho Paul[1], M.Williams[1], R. Ajay Prasanth[1], R. Veera Ranjani[1], P. Meenakshisundaram[1], V. Rajasree[2], M. Raveendran[1] and N. Manikanda Boopathi[1*]

[1]Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Coimbatore, India
[2]Department of Vegetable Crops, HC&RI, Tamil Nadu Agricultural University, Coimbatore, India
***E-Mail**: nmboopathi@tnau.ac.in

Abstract
Moringa *(Moringa oleifera* Lam. Moringaceae) leaves have enormously valued for their food, medicinal and industrial uses. To evolve improved leafy biomass producing cultivars, it is essential to systematically analyse the available morphological and genetic diversity. This study investigated the genetic diversity of 55 Moringa germplasm accessions using simple sequence repeat (SSR) markers and leaf morphological traits by employing a wide range of genetic distance measures, clustering procedures and other multivariate methods. Though analysis with DAR win 6.0 and TASSEL 4.0 has generated three clusters, NTSY Spc 2.0 and Power Marker v.3.25 produced only two major clusters. Morphological diversity was also analysed with five leaf traits using both Minitab 20.0 and XL-STAT and they produced two main clusters. Distantly related germplasm accessions identified in this study (such as Karumbu Murungai and Kappalpatti Murungai), would be of great importance in further genetic improvement of Moringa for increased leaf biomass, a valuable trait that can ensure nutritional and economic security.

Key words: Moringa, Molecular markers, SSRs, Genetic diversity algorithms, Leaf biomass

INTRODUCTION

The livelihood of millions of people living in rural India is largely dependent on agriculture, though it encounters several production constraints including unfavourable climatic conditions, labour shortage, increased cost of farming practices and non-availability of water and other inputs. On the other hand, continuous efforts such as harnessing the increased genetic gains through modified plant breeding strategies to ensure food security and fight against the under nourishment of nutrients are being commenced. Farmers, the key stakeholders in agriculture, are always looking for affordable and alternative farming practices to generate at least a minimum income and feed their family members with sufficient and nutrient enriched foods. Moringa (*Moringa oliefera* Lam.,) is valued for its nutritional benefits by several food and healthcare industries. The most commercialized product of Moringa is its leaf powder since it is one of the richest and affordable sources of natural iron, calcium, multivitamins and essential amino acids (Trigo *et al.*, 2021). Though Moringa is recognized as a superfood and low-input responsive crop, the non-availability of leafy Moringa cultivar is the major problem faced by the farmers, as there is a huge demand for Moringa leaves in the market.

Breeding for better Moringa cultivars with improved biomass production requires to sample the available genetic diversity for the leaf characteristics. Traditionally, genetic diversity has been attempted with morphological traits. However, strong environmental influence on the expression of these morphological traits forced the researchers to look into other options to capture the genetic diversity. Molecular markers have successfully shown to be useful in this direction (Boopathi, 2013) and especially the simple sequence repeat (SSR) markers were widely used for sampling the genetic diversity in crop plants (Beşer *et al.*, 2021). Though SSRs were reported in Moringa (Wu *et al.*, 2010), the availability of only twenty SSRs for diversity analysis limits the efficient sampling of genetic diversity that exist in the Moringa. Hence, an additional set of Moringa SSRs was identified, synthesized and validated Boopathi *et al.*, manuscript submitted).

In order to ensure accurate and unbiased estimates of genetic diversity, it is suggested to focus on sampling strategies, employment of various data sets, choice of different genetic distance measures, clustering procedures and other multivariate parameters (Mohammadi and Prasanna, 2003). Unfortunately, there is no simple and effective procedure to include all these parameters with a single algorithm that perform diversity analysis. Hence, it was proposed to utilize different genetic diversity algorithms such as DARwin 6.0, TASSEL 4.0, Power marker v.3.25 and NTSYSpc 2.0 with the SSR data collected from the 55 Moringa germplasm lines and arrive at a consensus to identify genetically diverse Moringa accessions. Diversity analysis based on morphological parameters were also performed using Minitab 20.0 and XL-STAT. It is believed that utilization of such combinations of statistical procedures enables a robust way of identifying intraspecific diversity, which enhances the efficiency of further Moringa genetic improvement program.

Table 1. List of Moringa Germplasm used in this study and their characteristics

S. No.	Sample ID	Description	S. No.	Sample ID	Description
1	PKM MO-01	Long pod, Perennial	29	PKM MO-31	Short poded Pedapally
2	PKM MO-02	Ottakarumbu	30	PKM MO-32	Short poded Armor
3	PKM MO-03	Yalpanamurungai	31	PKM MO-32	Short poded armor
4	PKM MO-04	Kappalpatty murungai	32	PKM MO-33	Short poded Nandipeta
5	PKM MO-05	Karumbu Murungai	33	PKM MO-34	Short poded basara
6	PKM MO-06	Karumpumurunagi	34	PKM MO-35	Short poded Adilabad
7	PKM MO-07	Kutchi Murungai-11	35	PKM MO-36	Medium poded amaravathi
8	PKM MO-08	Kutchi Murungai-12	36	PKM MO-37	Short poded doragiripalli
9	PKM MO-09	Malai Murungai	37	PKM MO-40	Mara murungai
10	PKM MO-10	Kattu Murungai	38	PKM MO-40	Maramurunagi
11	PKM MO-11	Sem Murungai	39	PKM MO-12	Kutchi Murungai
12	PKM MO-13	Sem Murungai	40	PKM MO-41	Kappal patti Murungai
13	PKM MO-14	Karumpu Murungai	41	PKM MO-42	Nool murungai
14	PKM MO-16	PKM -1	42	PKM MO-43	Moolanur Nettai
15	PKM MO-17	PKM -2	43	PKM MO-19	Malai Murungai
16	PKM MO-18	Malai Murungai	44	PKM MO-45	Coimbatore long type
17	PKM MO-20	Nattu murungai	45	PKM MO-46	EMS treated PKM-1
18	PKM MO-54	Nattu Murungai	46	PKM MO-48	Karumbu murungai
19	PKM MO-21	Chedi Murungai	47	PKM MO-49	Yalpanam murungai
20	PKM MO-22	Nattu Murungai	48	PKM MO-50	Medium poded usilampatti
21	PKM MO-23	Nattumurungai	49	PKM MO-52	Kutta Murungai
22	PKM MO-24	Kattumurungai	50	PKM MO-53	Azhagiyavilai
23	PKM MO-25	Kutchi Murungai	51	SAMPLE 47	Short poded pedapalli
24	PKM MO-26	Vayal Murunagi	52	SAMPLE 48	Kattu murunagi
25	PKM MO-27	Medium poded Chozhavanthan	53	SAMPLE 49	Kuchi murungai
26	PKM MO-28	Long poded Warangal	54	SAMPLE 50	Karumpu murunagi
27	PKM MO-29	Long poded Malayal	55	SAMPLE 51	Bushy type
28	PKM MO-30	Medium poded Warangal			

MATERIALS AND METHODS

Totally 55 Moringa germplasm accessions were collected from Moringa Germplasm which was maintained in the College of Horticulture, Periyakulam (**Table 1**). Morphological characterization was performed using the phenotypic data consisting of five leaf traits (fresh weight, dry weight, the number of leaves, leaf length and breadth). Minitab 20.0 (Mathews, 2005) and XL-STAT (XLSTAT, 2013) was used for analysing the variation within the samples by creating a histogram and for investigating the phylogenetic relationship between the accessions. To explain the maximum amount of variance within the morphological leaf traits, Principal Component Analysis was also performed.

Genomic DNA was extracted by method described by Doyle *et al.* (1987) and scoring of allelic data collected from 75 genomic and 41 genic SSR markers was performed with Alpha Ease FC software (Version 4.0.0, Alpha Innotech Corp., USA). DARWin 6 (available at http://darwin.cirad.fr/darwin) was used for cluster analysis by employing the allelic data. The dissimilarity coefficient was estimated by the Jaccard index and Dendrogram was constructed by employing the unweighted pair group with mean average (UPGMA) algorithm. Bootstrapping over loci with 1000 replications was carried out to evaluate the strength of evidence for the branching patterns in the resulting UPGMA dendrogram. Subsequently, Tree distance of the dendrogram was calculated in Darwin 6.0. Similarly, the same genotyping data along with the phenotypical data of leaf traits were used for association mapping using the TASSEL 4.0 software (Bradbury *et al.*, 2007). General Linear Model (GLM) was used to analyse association after transforming genotyping data and subsequently joining the phenotype and genotype data using the intersect join function available in the data menu with default parameters and the phylogenetic tree was constructed. In addition, the genotypic data file was prepared as required by NTSYSpc2.0 and the prepared binary matrix was subjected to statistical analyses using NTSYSpc 2.2(Rohlf, 1998). Dice's similarity coefficient was employed to compute pairwise genetic similarities. The corresponding dendrogram was constructed by applying UPGMA that followed sequential agglomerative hierarchical non-overlapping (SAHN) clustering techniques. Genomic data were also transformed to allelic phase and was processed in text tab delimited format to feed into Power Marker v.3.25. Nei's genetic distance was estimated and the corresponding dendrogram was constructed by the UPGMA method which was viewed using MEGA software (Tamura *et al.*, 2007). PIC value of the SSR markers was estimated using Power marker v3.25(Liu *et al.*, 2005) for understanding the usefulness of SSR used in this study.

RESULTS AND DISCUSSION

Though the importance of Moringa as an essential nutritional and therapeutic plant has long been realized around the world for several decades (Ganesan *et al.*, 2014), there are very scarce varieties developed for high leaf biomass. For identification, conservation, and cultivar development, wild and cultivated plant genetic variability information is vitally important (Hassanein *et al.*, 2018). This study was conducted to analyse the morphological and genetic diversity in fifty-five Moringa germplasm lines that are growing in various agro-climatic regions of South India using phenotypic data and in house generated SSR markers.

Variation among five leaf traits (fresh weight, dry weight, the number of leaves, leaf length and breadth) in 55 ecotypes was calculated (**Fig. 1A**), which clearly shown that there was a normal gaussian kind of distribution for both dry weight and fresh weight of leaves, with a bulge in the middle and a drop in tail length suggesting that data near the mean are more frequent in occurrence than data far from the mean. PCA revealed that 89.3% variance was explained by two components (**Table 2**). Scree plot orders the eigenvalues from largest to smallest, with an ideal pattern of the steep curve, bend and a straight line (**Fig. 1B**). The leaf traits that correlate the most with the second principal component are the number of leaves, fresh and dry weight of leaves (**Table 3**). Therefore, diversity in the samples was attributed to the number

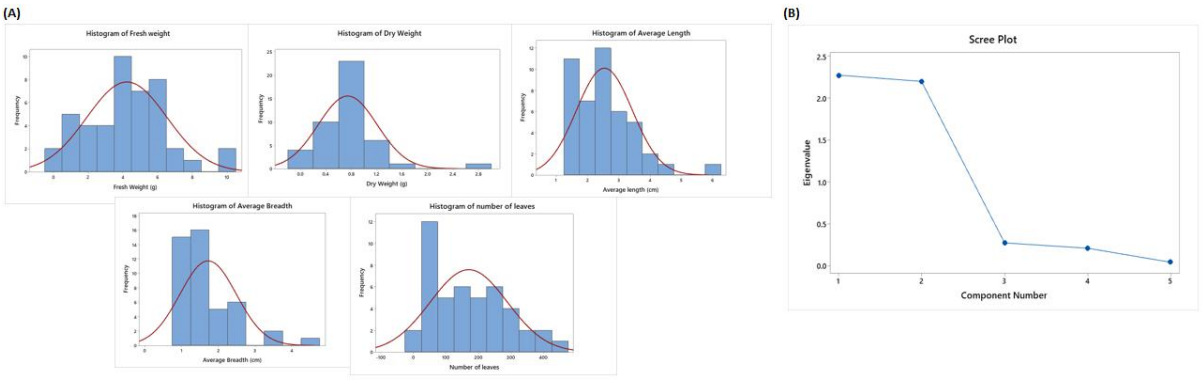


Fig. 1A. Histogram showing variation in leaf traits, 1B. Scree plot of PCA

Table 2. Eigen analysis of phenotypic data

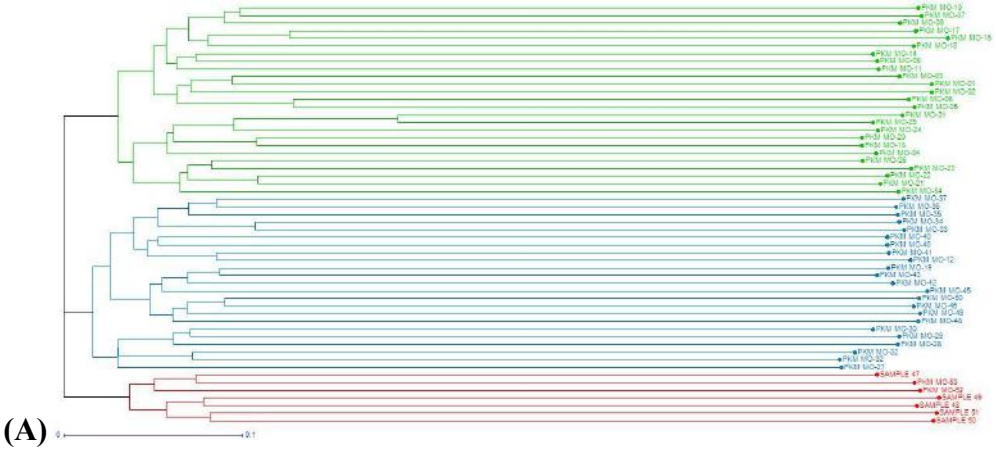
Eigenanalysis of the Correlation Matrix	*PC1*	*PC2*
Eigenvalue	2.2703	2.196
Variation	45.4	43.9
Cumulative variance	45.4	89.3

Table 3. Result of Principal component analysis

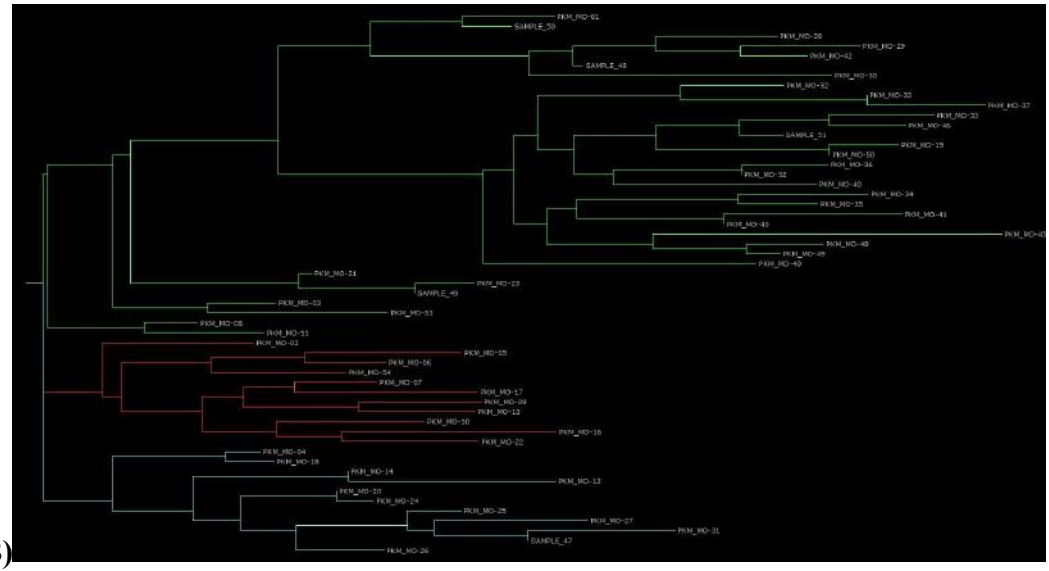
Traits	PC1	PC2
Fresh weight (g)	0.642	0.129
Number of leaves	0.265	0.561
Average length(cm)	0.258	-0.573
Average Breadth(cm)	0.216	-0.58
Dry Weight	0.636	0.065

of leaves, fresh and dry weight of leaves. Clustering observed from Minitab 20.0 and XL-STAT was of a similar kind (**Fig. 2**) irrespective of different genetic distance methods (Multivariate Cluster analysis with Manhattan genetic distance and Multivariate Agglomerative Hierarchical Clustering with Euclidean genetic distance, respectively) were employed.

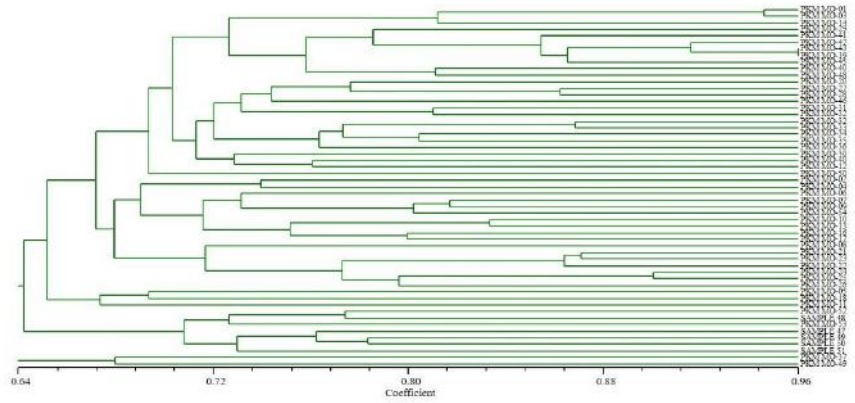
The use of morphological markers in plant breeding has been around for a long time and has been effective (Karaköy *et al.*, 2013). But, these markers aren't reliable since they are limited in quantity and can be influenced by the plant's stage of development and varied environmental circumstances (Eagles *et al.*, 2001). Hence, numerous DNA molecular markers have been discovered and effectively utilised to genetics and breeding operations in numerous agricultural crops. Though the number of methods and software are available for diversity analysis in crop plants, reliant on any one of the methods and



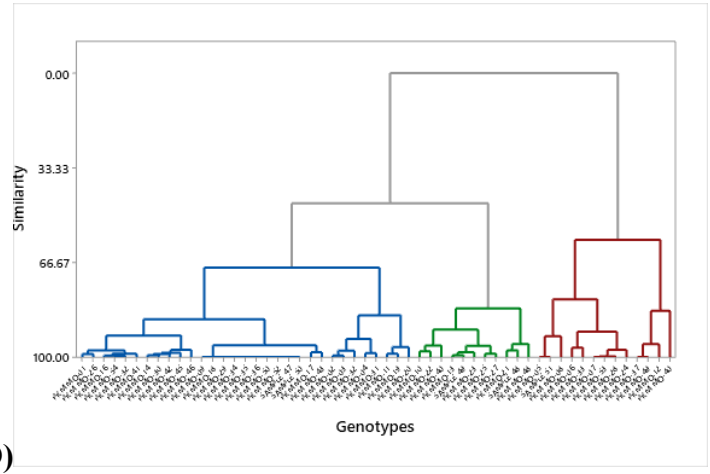
(A)



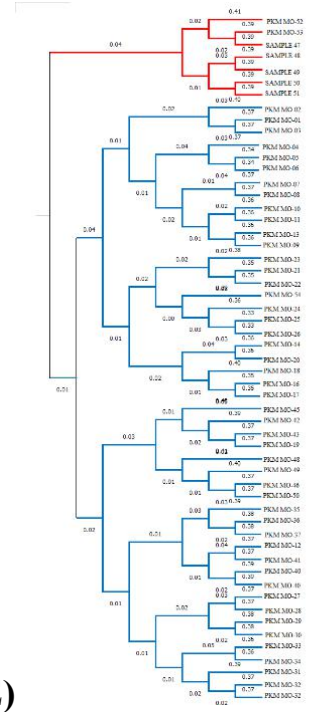
(B)


0.64
0.72
0.80
0.88
0.96
Coefficient

(C)


0.00
33.33
66.67
100.00
Similarity
Genotypes

(D)



(E)

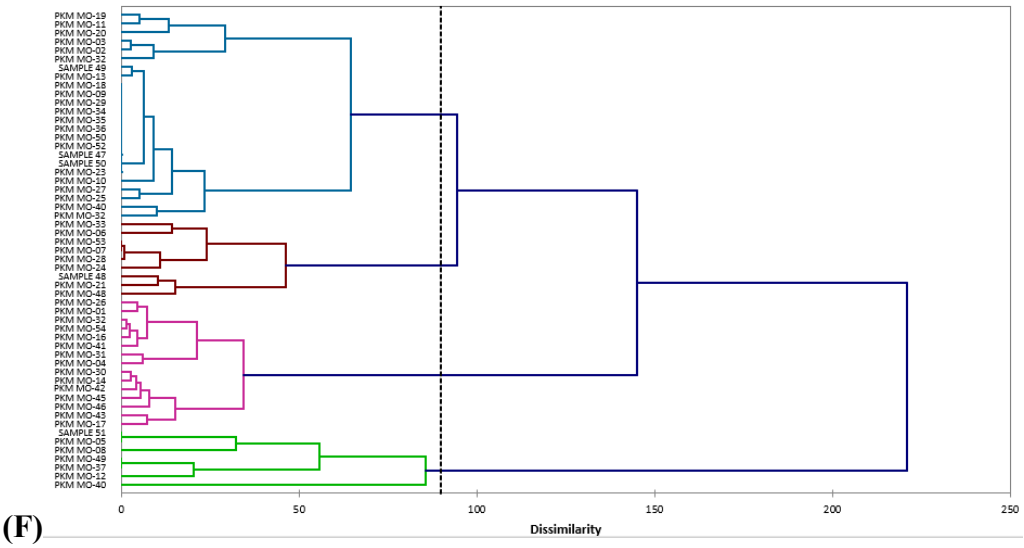


Fig. 2. Dendrograms generated by six different algorithms
(A) Darwin 6.0, (B) TASSEL 4.0 and (C) NTSYSpc 2.0 (D) Power marker v.3.25 (E) Minitab 20.0 and (F) XL-STAT

software may limit the application of the genetic diversity results. Further, the use of a limited number of Moringa specific SSRs (Wu *et al*., 2010) may not provide the complete picture of the genetic diversity that exist in the Moringa germplasm. Hence, in this study, we investigated Moringa genetic diversity using additional SSR markers developed at this laboratory.

Depending on the marker, the number of bands generated varied from two to four (**Fig. 3**). A marker's PIC value is used in genetic studies to determine its informativeness. PIC of SSRs calculated using Power marker v3.25 in this study was ranged from 0 to 0.960. The PIC had a value of zero for monomorphic markers. Nei's genetic distance was found to be in the range of 0.663 to 0.953 indicating that there is a significant genetic difference.

All the four algorithms used in this study has resulted in at least two major clusters (**Table 4; Fig. 2**), even though each one of the clusters was constituted with different germplasm accessions. Though it was not complete, a more similar kind of clustering pattern was observed among the results obtained from Power marker v.3.25 and NTSYSpc 2.2. (**Table 4**). It may be due to the fact that the algorithms used in these software (Dice coefficient and Nei's distance) measures the similarity between two populations as the ratio of the shared bands for a pair of individuals that are randomly drawn (one from each population), and the number of bands exhibited by a randomly sampled individual from the pooled population (Kosman *et al*., 2005).

Invariably, all the six algorithms have placed 13 Moringa accessions (which included serial numbers 30, 31, 33, 34, 35, 38, 40, 41, 42, 43, 44, 45 and 46 of **Table 1**) were placed in the major cluster I. Upon closer examination of their origins, it was revealed that they come from nearby adjacent areas (Theni and Dindigul districts of

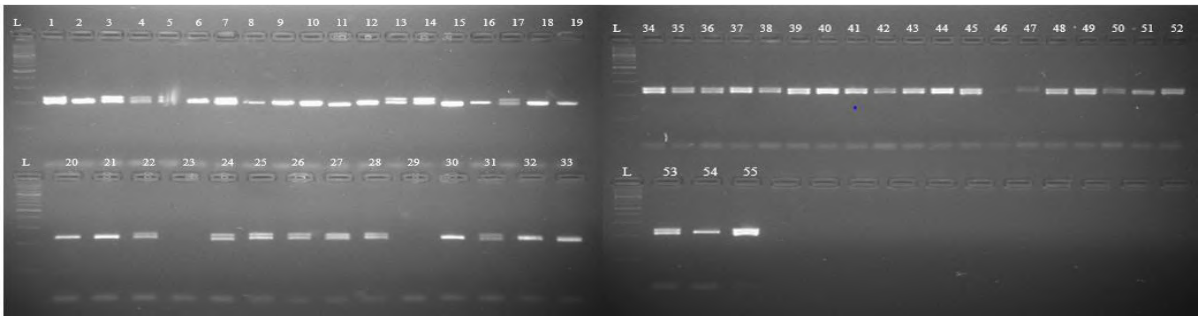


Fig. 3. Agarose gel electrophoresis image of SSR 32565 marker used in the study

Table 4. Genetic diversity analysis of Moringa germplasm accessions by different algorithms generated at least two different clusters.

Cluster	Cluster generated by Darwin 6.0	Cluster generated by TASSEL 4.0	Cluster generated by NTSYSpc 2.2	Cluster generated by Powermaker v.3	Cluster generated by XL-STAT	Cluster generated by Minitab 20.0
I	25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 (30)	1, 3, 8, 11, 19, 21, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55 (33)	(A) 1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 (44)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 (24)	2, 3, 6, 7, 9, 10, 11, 12, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 30, 32, 33, 34, 35, 38, 43, 46, 48, 49, 50, 51, 52, 53, 54, 55 (34)	1, 2, 3, 4, 9, 11, 13, 14, 15, 16, 17, 18, 24, 28, 29, 30, 31, 33, 34, 35, 40, 41, 42, 43, 44, 45, 48, 49, 51, 54 (31)
			(B) 5, 11, 16 (3)	25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 (24)	1, 4, 13, 14, 15, 18, 24, 28, 29, 31, 40, 41, 42, 44, 45 (15)	10, 12, 19, 20, 21, 23, 25, 38, 46, 52, 53 (11)
II	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 24 (19)	2, 5, 6, 7, 9, 10, 12, 14, 15, 18, 20 (11)	36, 49, 50, 51, 52, 53, 54, 55 (8)	49, 50, 51, 52, 53, 54, 55 (7)	5, 8, 36, 37, 39, 47 (6)	5, 6, 7, 8, 22, 26, 32, 36, 37, 39, 47, 50, 55 (13)
III	4, 16, 17, 22, 23, 29 (6)	4, 13, 16, 17, 22, 23, 24, 25, 29, 39, 51 (11)				

Note: Numbers in each column matches with serial number of Table 1 and its corresponding germplasm accession. Numbers in bold and within parenthesis denotes numbers of germplasm lines present in the given cluster.

Tamil Nadu, India). Similarly, almost all the algorithms grouped the Karumbu Murungai (serial number 5 in Table 1) in cluster II, which was originated from Tuticorin.

Interestingly, Moringa accessions were not clustered according to their place of collection in all the investigated algorithms and it was hard to find any geographically based clustering. For example, even though Karumbu Murungai (serial number 05) and Bushy type Murungai (serial number 55) were from the same geographical origin, they were placed in different clusters by the different algorithms used in this study. Some possible explanations for this include comparable agroclimatic conditions of the research region, movement of seeds and the spread of planting materials from cuttings (Mgendi *et al.*, 2010).

An attempt to identify the consensus of Moringa germplasm accessions that have shown reliable genetic diversity in all the analyses that were made in this study had helped to categorize the extremely diverged Moringa germplasm lines. Particularly, Karumbu Murungai (S. No. 54) and Kappalpatti Murungai (S. No. 4 in Table 1) were placed in two different diverged major clusters in all the six-diversity analyses that were conducted in this investigation. DARWIn 6.0analysis revealed a distance between these two accessions of 0.933 units. Hence, these two lines may be used as parental combinations to evolve novel segregating progenies with greater genetic variability. Further, the polymorphic markers found between these proposed parental lines would be used in marker assisted selection to introgress the desirable genomic segments from diverse germplasm into the elite Moringa line upon validating its association with the target traits.

Thus, this study has helped to identify the extremely diverse Moringa germplasm accessions such as Karumbu Murungai and Kappalpatti Murungai, which were invariably identified in all the algorithms used in this study and found to be genetically different at least for leaf morphological traits. Hence, the use of these two lines in a hybridization program will lead to generating the maximum number of recombinants that may have desirable leafy phenotypes.

REFERENCES

Beşer, N., Mutafçılar, Z. Ç. and Hasançebi, S. 2021. Diversity analysis of the rice cultivars (*Oryza sativa L.*) by utilizing SSRs rice diversity by SSRs. *Journal of Food Processing and Preservation,* **45**(2): e15232. [Cross Ref]

Boopathi, N. M. 2013. Marker-assisted selection. In *Genetic mapping and marker assisted selection.* Springer.173-186. [Cross Ref]

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* **23**(19): 2633-2635. [Cross Ref]

Doyle, J. J. and Doyle, J. L. 1987. *A rapid DNA isolation procedure for small quantities of fresh leaf tissue*.

Eagles, H. A., Bariana, H. S., Ogbonnaya, F. C., Rebetzke, G. J., Hollamby, G., Henry, R. J., . . . Carter, M. 2001. Implementation of markers in Australian wheat breeding. *Australian Journal of Agricultural Research,* **52**(12): 1349-1356. [Cross Ref]

Ganesan, S., Singh, R., Choudhury, D. R., Bharadwaj, J., Gupta, V. and Singode, A. 2014. Genetic diversity and population structure study of drumstick (*Moringa oleifera Lam.*) using morphological and SSR markers. *Industrial Crops and Products,* **60:** 316-325. [Cross Ref]

Hassanein, A. M. A. and Al-Soqeer, A. A. 2018. Morphological and genetic diversity of *Moringa oleifera* and *Moringa peregrina* genotypes. *Horticulture, Environment, and Biotechnology,* **59**(2): 251-261. [Cross Ref]

Karaköy, T., Baloch, F. S., Toklu, F. and Özkan, H. 2013. Variation for selected morphological and quality-related traits among 178 faba bean landraces collected from Turkey. *Plant Genetic Resources,* **12**(1): 5-13. [Cross Ref]

Kosman, E. and Leonard, K. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular ecology,* **14**(2): 415-424. [Cross Ref]

Liu, K. and Muse, S. V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics,* **21**(9): 2128-2129. [Cross Ref]

Mathews, P. G. 2005. *Design of Experiments with MINITAB* (Vol. 446): ASQ Quality Press Milwaukee, WI, USA:.

Mgendi, M. G., Manoko, M. K. and Nyomora, A. M. 2010. Genetic diversity between cultivated and non-cultivated *Moringa oleifera Lam*. provenances assessed by RAPD markers

Mohammadi, S. A. and Prasanna, B. 2003. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop science,* **43**(4): 1235-1248. [Cross Ref]

Rohlf, F. J. 1998. NTSYSpc numerical taxonomy and multivariate analysis system version 2.0 user guide.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution,* **24**(8): 1596-1599. [Cross Ref]

Trigo, C., Castello, M. L., Ortola, M. D., Garcia-Mares, F. J. and Desamparados Soriano, M. 2021. Moringa oleifera: an unknown crop in developed countries with great potential for industry and adapted to climate change. *Foods,* **10**(1): 31. [Cross Ref]

Wu, J.-C., Yang, J., Gu, Z.-J. and Zhang, Y.-P. 2010. Isolation and characterization of twenty polymorphic microsatellite loci for *Moringa oleifera* (Moringaceae). *HortScience,* **45**(4): 690-692. [Cross Ref]

XLSTAT, A. 2013. Data analysis and statistics software for Microsoft Excel. *New York, NY: Addinsoft*.