PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.12704

eISSN 2348-1900

RESEARCH ARTICLE

Integrated multivariate analysis for identifying elite female accessions in palmyrah (*Borassus flabellifer* L.)

Pooja Annaji[1], Swarna Priya Ramakrishnan[2*], Rajasree V[1], Manikanda N Boopathi[3], Arumugam M Pillai[4] & Raveendran M[5]

[1]Department of Spices and Plantation Crops, Horticultural College and Research Institute, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India
[2]Palmyrah and Banana Research Station, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Killikulam 628 252, Tamil Nadu, India
[3]Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India
[4]Department of Genetics and Plant Breeding, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Killikulam 628 252, Tamil Nadu, India
[5]Directorate of Research, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Correspondence email - swarnapriya@tnau.ac.in

Received: 13 November 2025; Accepted: 15 December 2025; Available online: Version 1.0: 30 March 2025

Cite this article: Pooja A, Swarna PR, Rajasree V, Manikanda NB, Arumugam MP, Raveendran M. Integrated multivariate analysis for identifying elite female accessions in palmyrah (*Borassus flabellifer* L.). Plant Science Today (Early Access). https:/doi.org/10.14719/pst.12704

Abstract

Borassus flabellifer L., a dioecious palm has significant importance due to its agronomic, ecological and socioeconomic value. Despite its importance, systematic genetic evaluations are limited due to its long juvenile phase and dioeciousness. To bridge the gap a sequential multivariate framework was used in the study to find the best female accessions with balanced yield potential and ideal supporting traits. A randomized complete block design (RCBD) was employed to evaluate thirty accessions for 11 morphological and yield-related traits over a two-year period. Principal component analysis (PCA) used to reduce the dimensionality of the traits and to identify the main contributors to variation as tree height, trunk girth, single fruit weight and fruit yield per palm. The Mahalanobis D^2 analysis identified four phenotypically divergent clusters, with two unique accessions (Acc. 23 and Acc. 58) exhibiting exceptional trait combinations. Finally, the multi-trait genotype-ideotype distance index (MGIDI) consolidated all traits into a singular selection index, identifying Acc.12, Acc.10 and Acc.32 as the accessions most closely resembling the ideotype. This multivariate approach is integrated and goes beyond standard yield-based selection by capturing trait interrelationships and providing a practical framework for the identification of elite palms for targeted breeding and conservation.

Keywords: ideotype breeding; multi-trait genotype-ideotype distance index; palmyrah; phenotypic diversity; principal component analysis

Introduction

Borassus flabellifer L., commonly referred to as palmyrah, is a multipurpose species belonging to the Arecaceae family. It is broadly distributed across tropical regions, extending from the Persian Gulf to the Cambodian-Vietnamese border and is widely cultivated in India, Southeast Asia and Malaysia because of its wide range of products such as padaneer, palm sugar, jaggary, candy, fruit pulp, timber, fiber and leaf products make it a key species for food security, income generation and ecological resilience for millions, particularly in marginalized communities (1-4). The palmyra palm is the state tree of Tamil Nadu (5), accounting for 60 % of India's palmyrah's population and playing a major role in supporting rural livelihoods, cottage industries and an agro-based economy (6). The palmyrah tree is dioecious, which bears male and female on different trees. It takes approximately 15-20 years to reach the reproductive stage and the male produces 10-15 inflorescences annually, each of which bears 4-8 spikelets, while the female inflorescence has two to three branches (7).

Palmyrah palms are unique in that they are remarkably resilient and ecologically adaptable to a wide range of soil and climatic conditions (8). It will thrive in subtropical and tropical regions with 0-45 °C temperature and 250 to 5000 mm rainfall and in harsh environmental conditions viz., drought and water logging (9).

Despite its strong cultivation potential, dioeciousness and extended pre-bearing age limit breeding efforts. Although morphological variability in palmyrah has been identified across various agro-climatic zones, comprehensive multivariate assessments are still restricted, emphasizing the need for more structured evaluations to support breeding and conservation programmes. Breeding of palmyrah operations significantly relies on genetic makeup to lay the groundwork for creating improved varieties with desirable traits. A diversified gene pool enables breeders to improve adaptation, resilience to pests and diseases, yield and quality (10).

Multivariate analysis methods, such as principal component analysis (PCA), Mahalanobis D^2 and multi-trait genotype–ideotype distance index (MGIDI) statistics, are crucial for genetic diversity studies, as they allow for simultaneous research on multiple morphological traits.

Principal component analysis identifies the most significant parameters leading to overall variance (11). Mahalanobis D^2 statistics estimate genetic divergence among accessions, helping genotype grouping (12). They are used in the selection of parents and diverse groups. Multi-trait genotype–ideotype distance index is a new multivariate statistical method that was used to choose genotypes based on multiple characteristics at once. It combines all attributes into single index based on trait correlations using factor analysis. This strategy facilitates direct selection. Previous studies have shown that multivariate approaches are useful for assessing morphological and genetic variations in a wide range of perennial crops. For instance, PCA and cluster analysis have been used successfully to define variations in coconut (13), date palm (14) and arecanut (15) demonstrating their ability to reveal major distinctive traits and genetically distinct groupings.

Multivariate analyses are essential for identifying diverse and superior genotypes required for crop improvement, conservation measures and the establishment of breeding programs for underutilized species such as palmyrah.

Multivariate analysis is essential for identifying diverse and superior genotypes required for crop improvement, conservation measures and establishment of breeding programs for underutilized species like palmyrah. Because this study was primarily concerned with fruit yield, only female accessions were employed in the multivariate analysis. To address this gap, 30 female palmyrah accessions were analyzed using an integrated multivariate technique. Fruit yield is the most important economic characteristic in palmyrah, however it is greatly affected by factors such as the vegetative characters, the weight of the fruit, the number of bunches and the shape of the canopy. A univariate analysis that exclusively concentrates on yield would fail to convey its long-term adaptability and complex, multi-trait nature. Multivariate approaches enable the simultaneous analysis of yield and its associated features, ensuring that selected accessions integrate high productivity with structural stability and longevity.

In this work, we used PCA to find characteristics contributing to variability, Mahalanobis D^2 to categorize and assess divergence among accessions and MGIDI to aggregate traits into a single index expressing similarity to the desired ideotype. These methods work together to create a sequential process for evaluating germplasm.

Materials and Methods

Experimental details

This study included 30 female palmyrah palm accessions 32 years old, previously maintained by V.O. Chidambaranar Agricultural College and Research Institute, Killikulam and currently conserved at Palmyrah Research Station, Killikulam. The experiment was conducted at the Palmyrah Research Station in Killikulam. This Institute is located at 8° 46' N latitude and 77° 42' E longitudes and at an elevation of 40 m above mean sea level. Red sandy soils are predominant in this area with good drainage and moderate water holding capacity. This area receives an annual rainfall of 736.7 mm, which is received during the period from October to December with 40 rainy days. The experiment was conducted using a randomized complete block design (RCBD) and there were two replications. The collection of data took place over the course of two consecutive years (2024 and 2025) from June to August. The homogeneity of variances was initially examined for year-wise values. Since the interactions between year and accession did not show any significant results, the data from all years were combined and the mean values were utilized for further analysis. The traits evaluated were tree height (m), trunk girth (cm) at one meter height from the ground, number of leaves, number of leaflets, leaf length, petiole length (cm) and petiole girth (cm). These measurements were performed in triplicate and number of bunches per palm, number of fruits per bunch, single fruit weight (g) and fruit yield per palm (kg), each measured in triplicate. Measurements were performed using a precision weighing balance and measuring tape.

Statistical analysis

The mean values and replicated data of these traits were used to perform PCA and Mahalanobis D^2 (16, 17), respectively. Principal component analysis identifies key traits contributing to genetic diversity and simplifies complex datasets for effective interpretation and visualization (18). Mahalanobis D^2 helps identify genetically diverse parents for effective hybridization programs. It considers multiple traits simultaneously, offering a more accurate measure of variability than the univariate methods. This facilitates the strategic selection in breeding for maximum heterosis and genetic gain (19). The factoextra package was used for PCA and biplot analyses (20, 21). The Mahalanobis D^2 was analyzed using R software.

The MGIDI was developed to allow for simultaneous direct selection based on multiple traits. This method combines all features into a single index by first doing a factor analysis to account for trait correlations and prevent overweighting redundant traits. The MGIDI score shows each genotype's distance from a predefined "ideotype" an ideal genotype with the optimum value for each characteristic. A lower MGIDI value suggests that a genotype is closer to the ideotype and is therefore more desirable. The method also provides two key tools; a strengths and weaknesses view, which highlights the trait groups in which a genotype performs well or poorly and the contribution plot, which identifies the trait groups that are most influential in differentiating among all genotypes. A weighting coefficient, ω (omega), is determined through factor analysis in the MGIDI approach. It illustrates the impact of each latent factor (FA1–FA4) on the overall index. The MGIDI and ω computation were carried out in R software with the function of 'gamam' and 'MGIDI' of the package 'metan' (22, 23). For MGIDI analysis, the optimal ideotype was defined by assigning desirable directions for each trait: higher values for yield and fruit weight, moderate values for tree height and trunk girth (to balance vigor with ease of management) and optimal ranges for vegetative traits. These desirability criteria are summarized in Table 1.

Results

The mean values of 11 traits of palmyrah accessions were used for multivariate analysis using PCA to identify the key variables and MGIDI was used to identify the selection of ideotype based on multi traits for direct selection. Replicated data was used to analyse Mahalanobis D^2 to identify the unique accessions with desirable traits based on variance-covariance matrix.

Table 1. Criteria for ideotypes

Trait	Desirability	Importance
Tree height (m)	Low	Reduced palm height facilitates harvesting
Trunk girth (cm)	High	Robust trunk girth offers strength and stability
Number of leaves	High	Increases photosynthesis
Leaf length (cm)	High	Increases photosynthetic surface
Number of leaflets per leaf	High	Improves the canopy spread
Petiole length	Moderate	Avoids drooping nature of leaves
Petiole girth	High	Gives good support for heavier bunches
Number of bunches per palm	High	Directly related to yield
Number of fruits per bunch	High	Directly related to yield
Single fruit weight (g)	High	Larger fruits increase the palm yield
Fruit yield per palm (kg)	High	Increases the productivity

Principal component analysis (PCA)

Principal component analysis was conducted on 11 morphological and yield-related traits of 30 female palmyrah palm accessions. The analysis extracted 11 principal components (PCs), of which the first four had eigenvalues greater than one, satisfying the Kaiser criterion. These four PCs cumulatively explained 77.62 % of the total variation, indicating that most of the trait variability was captured by a small number of latent factors (Table 2, Fig. 1).

The PC1 had the highest eigenvalue (3.716), which accounted for 33.78 % of the total variance. It was predominantly influenced by key traits, such as tree height, trunk girth, total leaf length and single fruit weight. The PC2 explained 22.97 % of the variance and was associated with yield traits, number of bunches, number of fruits per bunch and fruit yield per palm. Principal component 3 and PC4 accounted for 11.61 and 9.26 % of the variance, respectively and contributed additional information related to the number of the leaflets per leaf, tree height, petiole length and girth (Table 3).

The biplot of the first two components (Dim1 and Dim2), which together explained 56.75 % of the total variation and revealed that yield-contributing traits were strongly aligned with Dim1, indicated by long vector lengths and high $\cos^2$ values (≥ 0.75). Traits such as petiole length, tree height and trunk girth showed moderate correlations across both dimensions, while vegetative traits such as number of leaflets per leaf were more associated with Dim2. This pattern confirms the central role of fruit yield and associated traits in defining the structure of phenotypic variation among the accessions (Fig. 2).

Principal component analysis effectively reduced 11 traits into four main components that account for 77.62 % of the variation. The PC1 (33.78 %) was characterized by aspects of plant structure and fruit dimensions. The PC2 (22.97 %) was primarily influenced by yield traits such as the number of bunches and fruits. This indicates that plant size and yield are the main contributors to diversity. These components adequately represent the significant phenotypic variation observed in palmyrah accessions.

Mahalanobis D^2 cluster analysis

Mahalanobis D^2 analysis grouped the 30 accessions into four distinct clusters assed on multivariate trait profiles (Table 4). Cluster 1, comprises of 20 accessions, represents genotypes with superior mean performance in yield related traits, especially fruit yield per palm (210.06 kg), number of bunches (16.25) and single fruit weight. Cluster 2 contained eight moderately divergent accessions, which recorded lower yield values but exhibited higher vegetative growth, suggesting potential use in improving stress resilience and canopy traits (Table 5).

Cluster 3 and 4 each consisted of a single highly divergent accession (Acc.23 and Acc.58, respectively). Acc.23 (cluster 3) exhibited the highest mean values for tree height (8.8 m), trunk girth (170 cm), leaf length (300 cm), petiole length (180 cm) and fruit yield per palm (270 kg) as well as highest single fruit weight (1875 g). Cluster 4 also displayed large trunk girth and fruit size, indicating its value as a donor for vigour and yield components.

The inter-cluster distance matrix revealed the highest genetic divergence between cluster 1 and cluster 3 (D^2 = 1487.97), followed by cluster 2 and cluster 4 (D^2 = 867.28). These values suggest

Table 2. Principal component analysis of 11 variables of palmyrah

Principal component	Eigenvalue	Variance (%)	Cumulative variance (%)
PC1	3.716	33.78	33.78
PC2	2.527	22.97	56.75
PC3	1.277	11.61	68.36
PC4	1.019	9.26	77.62

Table 3. Contribution variables for principal components

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
PH	10.25	6.749	11.755	6.71	2.426	0.367	46.299	5.284
TG	14.301	0.014	5.771	17.61	4.714	9.698	4.118	2.592
NL	2.44	9.846	0.047	8.301	71.466	0.24	0.038	1.188
TLL	16.526	2.801	6.715	0.013	0.215	12.521	3.855	37.265
NLLL	0.737	3.104	44.158	26.97	2.215	1.033	2.457	0.117
PG	7.463	3.713	11.163	8.574	3.179	51.785	2.131	0.024
PL	10.722	9.354	1.912	13.286	12.727	3.603	0	2.576
NB	6.752	20.34	5.062	3.838	0.384	5.452	1.409	27.203
NFB	0.121	23.54	4.541	11.115	1.707	13.079	17.486	15.684
SFW	18.149	1.62	8.499	2.326	0.861	0.85	20.727	7.108
FYPY	12.537	18.919	0.378	1.259	0.106	1.371	1.48	0.959

PH: Plant height
TG: Trunk girth
NL: Number of leaves
TLL: Total leaf length
PG: Petiole girth
PL: Petiole length
SFW: Single fruit weight
FYPY: Fruit yield per palm per year
NLLL: Number of leaflets per leaf
NFB: Number of fruits per bunch
NB: Number of bunches per palm
PC: Principle component

Table 4. Clustering pattern of 30 female palmyrah accessions by Mahalanobis D square analysis

Cluster	Total accessions	Accessions
Cluster 1	20	Acc. 10, Acc.11, Acc.16, Acc.20, Acc.24, Acc.30, Acc.38, Acc.39, Acc.40, Acc.42, Acc.44, Acc.47, Acc.52, Acc.55, Acc.57, Acc.59, Acc.60, Acc.62, Acc.63 and Acc.7
Cluster 2	8	Acc.1, Acc.12, Acc.18, Acc.3, Acc.32, Acc.4, Acc.66 and Acc.9
Cluster 3	1	Acc.23
Cluster 4	1	Acc.58

Table 5. Mean values of four clusters estimated by Mahalanobis D square analysis

Trait	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Tree height (m)	6.8275	6.745	8.8	9
Trunk girth (cm)	161.625	147.95	170	190
No. of leaves	27.125	30.85	40	38
Leaf length (cm)	210.125	191.75	300	250
Number of leaflets per leaf	83.5	70.25	59	70
Petiole girth (cm)	27.375	27.45	40	35
Petiole length (cm)	115.5	113.95	180	141
Number of bunches	16.25	10	16	8
Number of fruits per bunch	11.375	9.55	9	9
Single fruit weight (g)	1163.5	833.75	1875	1452
Fruit yield per palm (Kg)	210.0625	82.7835	270	104.54

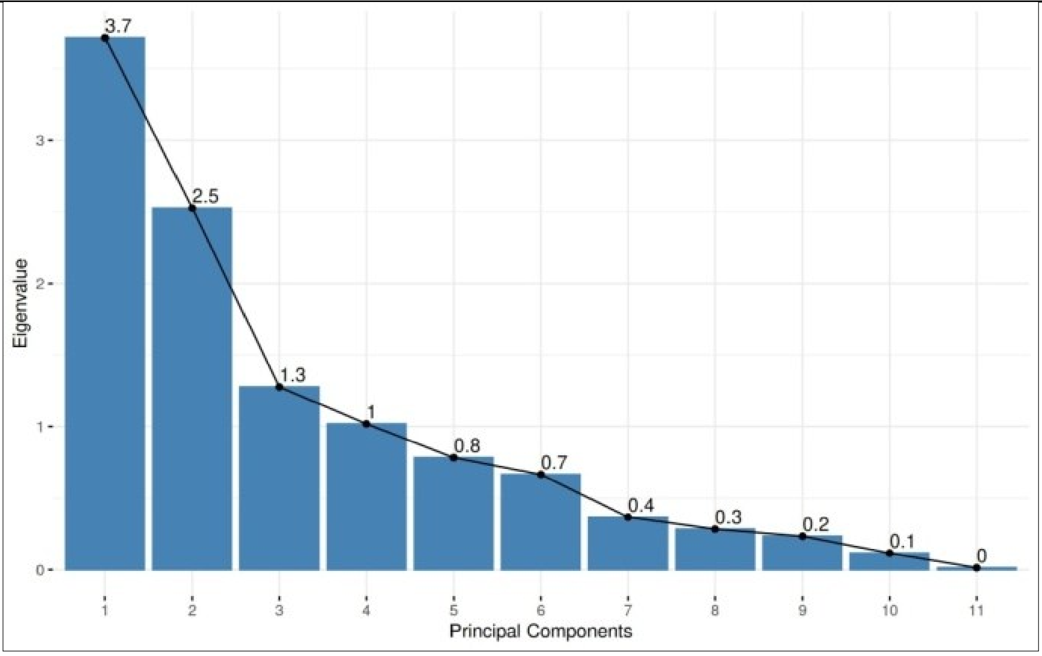

Fig. 1. Scree plot of principal components.

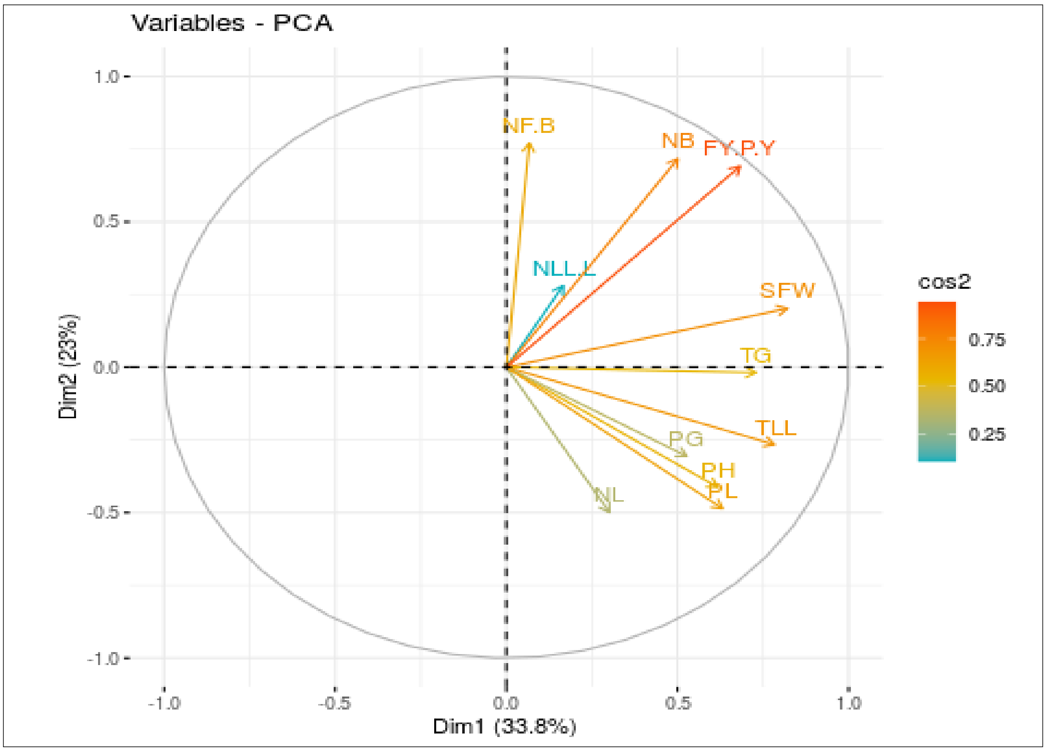

Fig. 2. Biplot of the variables of palmyrah.

that accessions in these clusters are ideal candidates for hybridization to exploit the heterosis. The smallest inter-cluster distance was observed between cluster 3 and cluster 4 ($D^2 = 695.78$), indicating some degree of similarity. Intra-cluster distances ranged from 240.35 (cluster 2) to 295.88 (cluster 1), with cluster 1 showing more internal diversity, which could be useful for recurrent selection (Fig. 3).

The cluster analysis enabled the definitive classification of the thirty palmyrah palms into four groups according to their yield and morphological traits. Cluster 1 had the highest-yielding palms, which produced more fruit. Cluster 2 had a moderate output but higher vegetative growth, such as more leaves. Each palm in clusters 3 and 4 was distinct from the others. Both were outstanding outliers, with Acc.23 performing the best across nearly all characteristics, including yield and size.

Multi-trait genotype-ideotype distance index (MGIDI)

The MGIDI was used to assess the overall performance of accessions based on multi traits into a single index using factorial analysis. The MGIDI scores effectively distinguished between accessions, with lower values indicating a closer resemblance to the ideotype (Table 6). Among the assessed accessions, Acc.12 and Acc.10 had the lowest MGIDI values, implying their exceptional multi-trait performances. Acc.32 had a moderate MGIDI score, while Acc.60 had the highest score, indicating a significant distance from the ideotype.

Fig. 4 reveal the strengths and weaknesses of how each element (FA1–FA4) was added to the MGIDI index. Acc.12 outperformed in FA3 and FA4, whereas Acc.10 excelled in FA3. Acc.32 made a balanced contribution across all categories but showed relative weakness in FA2. In contrast, Acc.60 showed substantial weaknesses in FA1, making it less desirable than the other compositions.

The contribution plot highlights are relative importance of each element in the MGIDI score. Factors FA1 and FA4 accounted for most of the variation, emphasizing their importance in distinguishing accessions. Numerous accessions had contributions that exceeded the threshold (100 divided by the number of factors), indicating that weaknesses in certain categories were disproportionately influencing their MGIDI scores (Fig. 5).

According to the MGIDI index, the most elite, well-balanced genotypes for selection are Acc.12, Acc.10, Acc.32 and Acc.60. The top performances that are closest to the ideal palm are Acc.12 and Acc.10. FA1 and FA4 are the most important trait groups for rising overall performance, according to strength and weaknesses study. Breeders can more successfully identify superior accessions and target particular traits for genetic improvement with this strategy, which effectively goes beyond single-trait selection.

Table 6. Multi-trait genotype–ideotype distance index scores of accessions

S. No.	Genotype	MGIDI	S. No.	Genotype	MGIDI
1	Acc.32	2.471	16	Acc.62	4.601
2	Acc.10	3.002	17	Acc.47	4.738
3	Acc.60	3.660	18	Acc.24	4.771
4	Acc.12	3.677	19	Acc.1	5.018
5	Acc.23	3.707	20	Acc.20	5.046
6	Acc.4	3.733	21	Acc.5	5.179
7	Acc.3	3.987	22	Acc.59	5.206
8	Acc.16	4.019	23	Acc.44	5.296
9	Acc.9	4.069	24	Acc.63	5.301
10	Acc.11	4.188	25	Acc.58	5.324
11	Acc.38	4.224	26	Acc.57	5.335
12	Acc.18	4.386	27	Acc.52	5.443
13	Acc.40	4.470	28	Acc.55	5.492
14	Acc.30	4.473	29	Acc.66	5.574
15	Acc.39	4.537	30	Acc.42	6.444

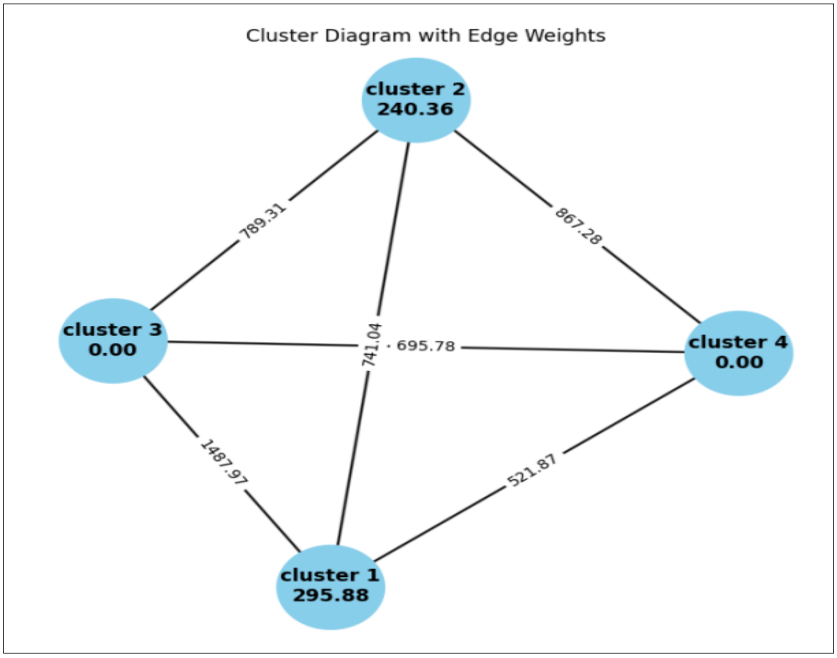

Fig. 3. Cluster diagram with edge weights of 30 palmyrah accessions four clusters.

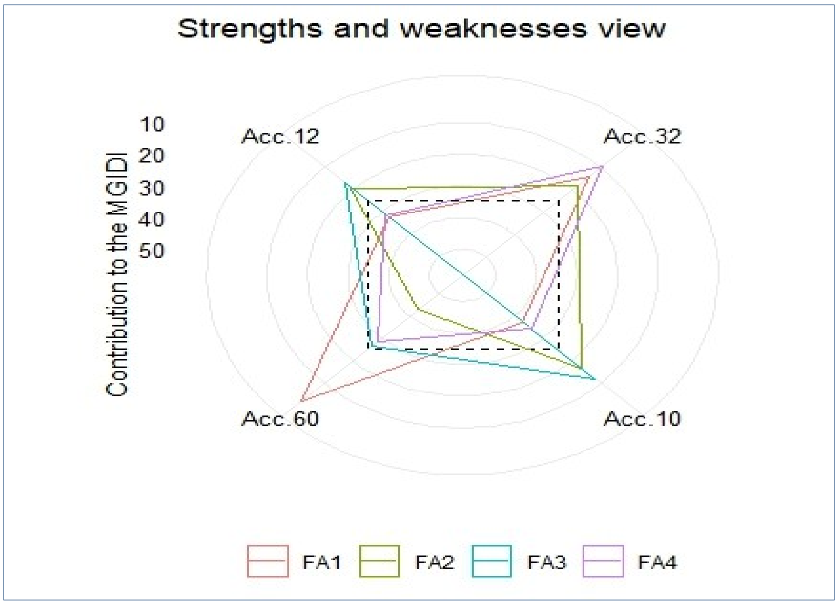

Fig. 4. Strengths and weaknesses view.

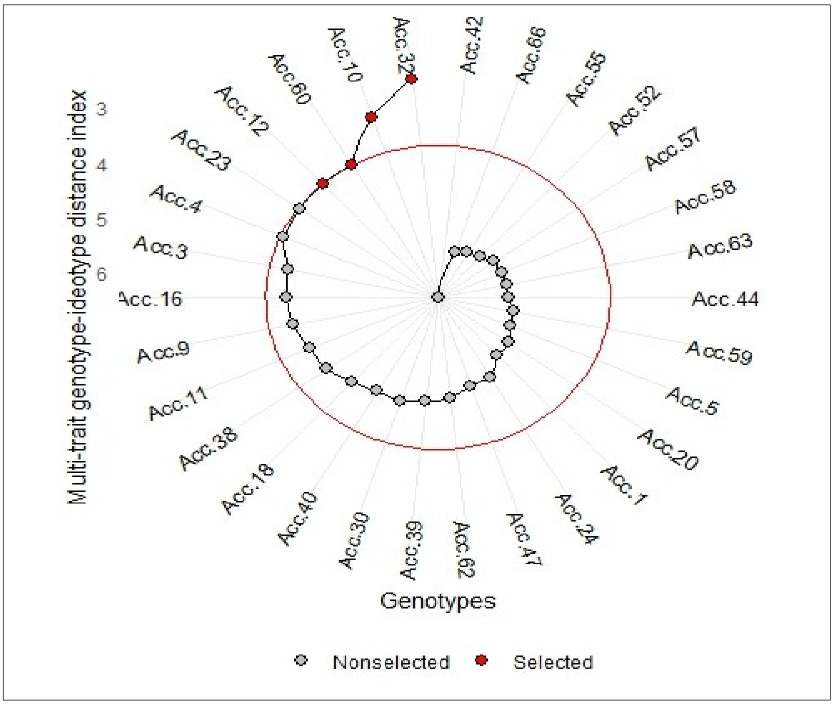

Fig. 5. Contribution plot of the accessions.

Discussion

Our sequential statistical approach successfully provided a comprehensive strategy for selection of elite accessions, enabling progress from understanding diversity to making effective choices. Principal component analysis first diminished trait complexity and emphasized the most significant variables. Mahalanobis D^2 measured divergence based on these parameters and categorized accessions into clusters, identifying unique accessions. Ultimately, MGIDI consolidated these characteristics into a unified index of ideotype similarity, facilitating the direct identification of superior palms that harmonize yield with structural and reproductive attributes. This integration offers a more comprehensive framework than univariate yield selection, guaranteeing that the selected accessions are both productive and agronomically appropriate.

The PCA findings revealed that the first four main components, each with an eigenvalue larger than one, accounted for 77.62 % of the total variability. This high proportion of variance indicates that only a few key traits were primarily responsible for the observed variations between accessions. The PC1, which accounted for 33.78 % of the variation, associated primarily with vegetative traits and single fruit weight. Principal component 2 explained an additional 22.97 % of the variance in essential yield parameters. These findings underscore the importance of reproductive and vegetative characteristics in different accessions and their potential in breeding programs that aim to increase yield. Similar PCA patterns have been reported in other palm species, where yield traits significantly contribute to the genetic variability. For example, studies on coconut (13), date palm (14) and oil palm (24) have consistently identified fruit yield components as key discriminators. These analogies support the use of PCA for trait selection in palms and emphasize the importance of focusing on high-variance traits in improvement initiatives.

Principal component analysis identified these key traits, namely tree height, trunk girth, total leaf length, single fruit weight, number of bunches per palm, number of fruits per bunch and fruit yield per palm, as major factors contributing to variability among female palmyrah accessions. These traits are essential to improve plant stature and yield. Tree height, trunk girth and leaf length are traits that are linked to structural integrity and vegetative vigor, both of which are necessary for long-term productivity. Fruit yield, the most significant economic characteristic, is closely correlated with single fruit weight, number of bunches and number of fruits per bunch. Prioritizing these qualities in breeding and selection efforts can result in increased yield potential, ease of harvesting and adaptability, making them important targets for palmyrah improvement.

Mahalanobis D^2 clustering confirmed the PCA findings and identified four different clusters. The cluster 1 had accessions with balanced performance across characteristics, most notably the yield. Clusters 3 and 4, each consisted of a single accession with high-yielding traits. Notably, Accession 23 in cluster 3 had the highest fruit yield, tree height and single-fruit weight, indicating its potential as a parent for breeding programs aimed at yield improvement. Clusters 1 and 3 had a significant inter cluster distance (D^2 = 1487.97), indicating their genetic differentiation. Cluster 2 had low fruit yield but showed more vegetative growth, with large number of leaves, which could be useful in breeding for canopy structure or photosynthetic efficiency. These findings reflect the genetic richness of the population and provide opportunities for selecting parents with beneficial qualities. Similar findings were reported previously for palmyrah accessions (25).

The use of the MGIDI in the current study allowed for a comprehensive assessment of accessions by combining multi traits into a single selection index. This method was better than single-trait evaluations because it identified genotypes with balanced performance across several trait groups. Acc 12 and Acc 10 were identified as superior accessions, emphasizing their potential application in breeding operations. Their low MGIDI scores and strength in FA3 and FA4 indicate that they process a combination of positive characteristics across desirable traits. Acc.32 had an intermediate score and displayed an overall balanced performance, but it requires targeted improvement in FA2 features to achieve ideotype closeness. In contrast, Acc.60 performed poorly, with weaknesses in FA1 limiting its utilization for direct selection in breeding programs. The significant importance of FA1 and FA4 in MGIDI variation suggests that these trait groups are important for genotypic differentiation. Therefore, breeding methods should focus on improving these parameters to increase total ideotype resemblance. Similar results have been reported in previous studies, where the MGIDI effectively prioritized genotypes by emphasizing trait groups that contribute the most to variability (23). Overall, the study demonstrated that MGIDI is a reliable selection strategy that can identify superior genotypes while identifying particular weaknesses.

The combination of PCA and Mahalanobis D^2 analysis provided a comprehensive view of the trai-based diversity in female palmyrah palms. The identification of genetically distinct and agronomically promising accessions serves as a valuable foundation for breeding programs aimed at improving yield, fruit quality and stress resilience. The incorporation of MGIDI into selection procedures has the potential to expedite genetic advances by ensuring that breeding effort focus on both high performance and trait balance. Future research integrating molecular markers or genomic tools could enhance the resolution of diversity analysis and support marker assisted selection for targeted trait improvement in palmyrah.

Conclusion

This study successfully applied multivariate statistical techniques to reveal the substantial genetic diversity among *B. flabellifer* accessions. Principal component analysis highlighted tree height, trunk girth, single fruit weight and fruit yield as major traits contributing to phenotypic variability among accessions. Mahalanobis D^2 clustering further distinguished the accessions into four genetically diverse groups, with Acc. 23 and Acc. 58 emerging as outliers with superior agronomic traits. These genotypes hold considerable promise as potential parental lines for yield improvement programs. Finally, the MGIDI index served as a practical selection filter, selecting Acc.12, Acc.10 and Acc.32 as elite genotypes with the most balanced and desired multi-trait accessions. Integrating PCA, D^2 and MGIDI sequentially enables an effective strategy for evaluating germplasm. This integrated technique goes beyond basic diversity evaluation to enable precise selection and breeding choices, considerably speeding up genetic improvement. It also supports conservation measures for this underutilized but valuable species.

Acknowledgements

The authors thank to Palmyrah Research Station, Tamil Nadu for providing guidance and support to carry out this research.

Authors' contributions

PA carried out the genetic studies, participated in the data collection, data analysis and drafted the manuscript. SPR participated in the design of the study and performed the statistical analysis and edited the manuscript. RV and AMP participated in the data analysis. MNB and RM participated in design of the study. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

1. Davis TA, Johnson DV. Current utilization and further development of the palmyra palm (*Borassus flabellifer* L., Arecaceae) in Tamil Nadu State, India. Econ Bot. 1987;41(2):247–66. https://doi.org/10.1007/BF02858972 CoLab
2. Morton JF. Notes on distribution, propagation and products of *Borassus* palms (Arecaceae). Econ Bot. 1988;42(3):420–41. https://doi.org/10.1007/BF02860166
3. Upadhyaya A, Sonawane SK. Palmyrah palm and its products (neera, jaggery and candy)-A review on chemistry and technology. Appl Food Res. 2023;3(1):100256. https://doi.org/10.1016/j.afres.2022.100256
4. Mahilrajan S, Thuraisingam S, Prabagar J. Exploring the nutritional, health and economic potential of palmyrah fruit pulp. Food Chem Adv. 2025;1(6):100880. https://doi.org/10.1016/j.focha.2024.100880
5. Davis TA. Palmyra palm, the state tree of Tamil Nadu, is on the verge

of extinction: Protect this very useful tree. Environ Awareness. 1985;8(4):95–106.

6. Subasri K, Senthilnathan S, Elanchezhiyan T, Jeyalakshmi P, Kannan B, Swarnapriya R, et al. Socio-economic determinants of willingness to expand palmyrah palm cultivation: An econometric analysis. Plant Sci Today. 2024;11:5394. https://doi.org/10.14719/pst.5394
7. Sankaralingam A, Hemalatha G, Ali AM. A treatise on palmyrah. All India Coordinated Research Project, Agricultural College & Research Institute, Tamil Nadu Agricultural University & Central Plantation Crops Research Institute, Indian Council of Agricultural Research; 1999.
8. Gnanavelrajah N, Sathasivam T, Jeyavanan K, Selvarajah R. Palmyrah (*Borassus flabellifer*) palm land-use system: A potential key tool for enhancing carbon stock and floristic diversity in a dry zone tropical landscape. Trees For People. 2025;100827. https://doi.org/10.1016/j.tfp.2025.100827
9. Krishnaveni TS, Arunachalam R, Chandrakumar M, Parthasarathi G, Nisha R. Potential review on palmyra (*Borassus flabellifer* L.). Adv Res. 2020;21(9):29–40. https://doi.org/10.9734/AIR/2020/v21i930229
10. Pixley KV, Cairns JE, Lopez-Ridaura S, Ojiewo CO, Dawud MA, Drabo I, et al. Redesigning crop varieties to win the race between climate change and food security. Mol Plant. 2023;16(10):1590–611. https://doi.org/10.1016/j.molp.2023.09.003
11. Haigh J. Applied multivariate statistical analysis, by RA Johnson and DW Wichern. Math Gaz. 1988;72(462):331–32.
12. McLachlan GJ. Mahalanobis distance. Resonance. 1999;4(6):20–26.
13. Sudha R, Niral V, Diwakar Y, Rajesh MK, Samsudeen K. Assessment of genetic diversity in West Coast Tall coconut (*Cocos nucifera*). Indian J Agric Sci. 2019;89(1):88–93. https://doi.org/10.56093/ijas.v89i9.93488
14. Ahmed MV, Bouna ZE, Lemine FM, Djeh TK, Mokhtar T, Salem AO. Use of multivariate analysis to assess phenotypic diversity of date palm (*Phoenix dactylifera* L.) cultivars. Sci Hortic. 2011;127(3):367–71. https://doi.org/10.1016/j.scienta.2010.11.011
15. Wang S, Kang X, Dai J, Dai W, Zhang J, Ji J. Evaluation of areca quality based on principal component and hierarchical cluster analyses in Hainan, China. HortScience. 2023;58(6):699–703. https://doi.org/10.21273/HORTSCI17125-23
16. Mahalanobis PC. A statistical study at Chinese head measurement. J Asiatic Soc Bengal. 1928;25(3):301–77.
17. Rao CR. Advanced statistical methods in biometric research. New York: John Wiley & Sons; 1952. p. 236–72. https://doi.org/10.1002/ajpa.1330120224
18. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans R Soc A. 2016;374:20150202. https://doi.org/10.1098/rsta.2015.0202
19. Mahalanobis PC. On the generalized distance in statistics. Sankhya A. 2018;80(Suppl 1):S1–7. https://doi.org/10.1007/s13171-019-00164-5
20. Kassambara A, Mundt F. Factoextra: Extract and visualize the results of multivariate data analyses. CRAN Contrib Packag; 2016.
21. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686.
22. Olivoto T, Lúcio AD. Metan: An R package for multi-environment trial analysis. Methods Ecol Evol. 2020;11(6):783–89. https://doi.org/10.1111/2041-210X.13384
23. Olivoto T, Diel MI, Schmidt D, Lúcio AD. MGIDI: A powerful tool to analyze plant multivariate data. Plant Methods. 2022;18(1):121. https://doi.org/10.1186/s13007-022-00952-5
24. Balakrishna P, Pinnamaneni R, Pavani KV, Mathur RK. Genetic diversity in oil palm genotypes by multivariate analysis. Int J Curr Microbiol Appl Sci. 2017;6:1180–89. https://doi.org/10.20546/ijcmas.2017.602.146
25. Vasanth S, Premalakshmi V, Ganga M, Manivannan MI, Juliet Hepziba S, Sivaprakash M. Multivariate analysis of morphological variation in palmyrah (*Borassus flabellifer* L.). Biological Forum. 2023;15(5a):353–61.

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.