Multivariate Assessment of Genetic Diversity in Wood Apple (*Feronia limonia* L.) Using R-Based Analytical Approaches

**Abstract**

|  |
| --- |
| **Aim:** To assess phenotypic variability among nine wood apple (*Feronia limonia* L.) genotypes using comprehensive multivariate statistical approaches, with the goal of understanding genetic diversity and identifying superior types for crop improvement.**Methodology:** A field evaluation was conducted under uniform agro-climatic conditions, with fourteen morphological, yield, and biochemical traits recorded for each genotype according to standard protocols. Multivariate statistical tools, including Principal Component Analysis (PCA), hierarchical clustering, and heatmap visualization, were applied using R software to classify genotypes and map trait associations.**Results:** Significant variability was observed across all measured traits. PCA showed that the first two principal components accounted for 92.7 % of the total variation, with PC1 alone explaining 74.4 %, indicating strong discriminatory power of key traits. Total sugar, pulp weight, fruit length, and yield per tree were identified as major contributors to genotypic divergence. Hierarchical clustering grouped the genotypes into two main clusters, clearly distinguishing elite types such as WFL-03 and WFL-08. Heatmap visualization further confirmed inter-genotypic diversity and revealed unique expression patterns among traits.**Conclusion:** The substantial genetic variability uncovered by multivariate analysis highlights the effectiveness of these tools in identifying superior genotypes. These insights provide a robust foundation for future selection, breeding, and conservation efforts, promoting the genetic improvement and sustainable utilization of wood apple as an important but underexploited fruit crop. |

**Keywords:** *Wood apple, Genetic diversity, Multivariate analysis, Yield, Quality*

### ****Introduction****

Wood apple (Feronia limonia L.), a versatile tropical fruit tree of the family Rutaceae, is well adapted to semi-arid regions and has been traditionally valued for its nutritional and medicinal properties, including antioxidant, antimicrobial, and anti‑inflammatory uses (Raut et al., 2022). Despite its ecological resilience and high pulp and fruit quality, the species remains under‑utilized in mainstream breeding programs and lacks systematic characterization across agro‑ecologies (Krishna *et al*., 2017). Understanding genetic variability in wood apple is crucial for effective germplasm conservation and improvement. Previous studies have documented extensive variation in fruit weight (118-233 g), pulp proportion (54-72 %), shell thickness, and sugar‑acid balance among genotypes (Shinde, 2019; Raut et al., 2022), indicating potential for selecting superior accessions (Shinde, 2019; Raut et al., 2022).

Traditional univariate analyses, while informative, are limited in revealing complex trait interactions. Multivariate statistical techniques such as Principal Component Analysis (PCA), hierarchical clustering, and heatmap visualization enable researchers to reduce dimensionality, uncover trait co‑variation, and classify genotypes based on multivariate similarity (Mir *et al*., 2017). These methods have been successfully applied in diverse fruit crops; PCA and cluster analysis were employed to characterize apple germplasm based on fruit quality traits, where the first three principal components explained over 78% of variation (Mir et al., 2017). Similarly, chemometric approaches combining PCA and hierarchical clustering revealed distinct grouping in apple based on physicochemical attributes (Muresan et al., 2022). However, comparable multivariate assessments in wood apple are rare, with only limited studies reporting variability in morphological and biochemical traits using correlation or GCV/heritability indices (Raut et al., 2022; Tayde et al., 2024). This gap underscores the need for a comprehensive multivariate evaluation of wood apple genotypes to support informed selection.

In this study, nine wood apple genotypes (WFL‑01 to WFL‑09) were evaluated for 14 morphological, biochemical, and yield‑related traits. Standardized data were subjected to PCA, scree‑plot analysis, hierarchical clustering (Ward’s method), and heatmap visualization using R software. The objectives were to: (i) determine the principal components capturing the majority of variation, (ii) identify key traits driving genotype differentiation, and (iii) group genotypes based on multivariate similarity. The outcomes are expected to inform selection of elite genotypes and seed future breeding and conservation strategies for wood apple in tropical agro‑ecological zones.

### ****Materials and Methods****

#### **Experimental location and plant material**

The study was carried out during **2023–2024** at the Department of Fruit Science, **Horticultural College and Research Institute, Periyakulam**, Tamil Nadu Agricultural University (TNAU), located at **latitude 10.1234°N, longitude 77.5486°E, and an elevation of 300 meters above mean sea level**. The region falls under the **Western Agro Climatic Zone of Tamil Nadu**, characterized by a tropical climate with average annual rainfall of 925 mm and temperatures ranging from 22 °C to 38 °C. A total of **nine genetically diverse wood apple (*Feronia limonia* L.) genotypes** coded as **WFL-01 to WFL-09** were selected for evaluation. These genotypes were collected from different geographical regions of Tamil Nadu representing diverse edaphic and climatic conditions.

#### **Experimental design**

The experiment was laid out in a **Randomized Complete Block Design (RCBD)** with **three replications**, where each genotype was represented by **three trees per replication**. The plants were spaced at **7 × 7 meters** and maintained under uniform cultural and irrigation practices throughout the cropping season to eliminate environmental variation.

#### **Trait evaluation and sampling**

Fruits were harvested at physiological maturity (based on shell color and fruit drop). Three representative fruits per tree (15 fruits per genotype per replication) were sampled and pooled for laboratory analyses. A total of **14 traits** were evaluated, classified as follows:

**Yield traits**

Fruit length (cm), fruit width (cm), fruit weight (g), pulp weight (g), shell thickness (cm), number of seeds per fruit, and number of fruits per tree, yield per tree (kg) and estimated yield per hectare (tons ha-1).

**Biochemical traits**

Total soluble solids (°Brix), titratable acidity (%), ascorbic acid (mg 100 g-1), pectin content (%), reducing sugar (%), non-reducing sugar (%), and total sugar (%). All traits were recorded following standard protocols: TSS was measured using a **digital handheld refractometer (ATAGO PAL-1)**, Acidity was estimated by titration using **0.1 N NaOH** against phenolphthalein, Ascorbic acid was quantified via **2,6-dichlorophenol-indophenol titration method**, Pectin content was determined through **alcohol precipitation and titration method** and Sugars were estimated by **Lane and Eynon’s volumetric method** using Fehling’s solutions is given by **(**Mohamed, 2016).

#### **Multivariate analysis techniques**

**Principal component analysis (PCA)**: PCA was conducted to reduce data dimensionality and identify key trait contributors. The contribution of each component was determined based on eigenvalues >1 and total variance explained (Gewers *et al*., 2022)

**Scree plot analysis**: The scree plot was used to determine the number of significant components, identifying the “elbow” point where the contribution of additional components became negligible (**Cattell, 1966).**

**Hierarchical cluster analysis**: Agglomerative hierarchical clustering was done using (Ward, 1963) method and euclidean distance to classify the genotypes into distinct clusters based on trait similarity.

**Heatmap visualization**: A heatmap with dendrograms was constructed to visualize the relative expression of traits across genotypes and to identify genotype trait groupings (Wilkinson and Friendly, 2009).

#### **Data standardization and software**

The collected data were initially checked for consistency and screened for outliers. All quantitative traits were then standardized using z-transformation, which scales the data to a mean of zero and a standard deviation of one, ensuring uniformity in scale and eliminating the influence of differing measurement units (Johnston and Wichern, 1982). Statistical analyses were performed using R software (v 4.3.0) (R Core Team, 2024). Principal Component Analysis (PCA), along with eigenvalue-based scree plots and biplots, was carried out using the FactoMineR and factoextra packages (Le *et al*., 2008; Kassambara and Mundt, 2020). Hierarchical clustering was conducted using the agglomerative method based on Ward’s minimum variance approach and Euclidean distance, as implemented in the base stats package (Ward, 1963; R Core Team, 2024). To visualize the relative expression of traits across genotypes, a heatmap with row and column dendrograms was generated using the pheatmap package, which utilized z-score normalized data to enhance visual clarity of genotype trait associations (Kolde, 2019; Wilkinson & Friendly, 2009). This integrated multivariate framework enabled comprehensive analysis of genotypic variation and trait interrelationships.

**Results and Discussion**

**PCA - scree plot**

The scree plot of the principal component analysis (PCA) clearly demonstrated the proportion of variation explained by each principal component (Figure 1). The first principal component (PC1) alone accounted for 74.4 % of the total variance in the dataset, indicating that a substantial portion of the information is captured along this axis. The second principal component (PC2) contributed an additional 18.3 % of the variation, bringing the cumulative explained variance to 92.7 % for the first two components. This suggests that the first two PCs effectively summarize the majority of variability among the wood apple genotypes across all measured traits. The steep drop in explained variance after the second component further confirms the suitability of using only PC1 and PC2 for visualizing and interpreting the diversity among genotypes.



**Figure 1. PCA - Scree Plot**

This concentration of variance within just two components indicates that the measured traits display high inter-correlation and that genotype differentiation is driven by a small number of major axes of variation, greatly facilitating effective data visualization and interpretation. Similar results have been reported in genetic diversity studies across several perennial fruit crops, where PCA enabled clear discrimination among genotypes by capturing the majority of phenotypic variation in the first two components. For instance, Singh *et al*. (2016) in wood apple reported that PC1 and PC2 together explained over 80 % of total variability, making them suitable for summarizing and visualizing trait diversity. Likewise, Roongruangsri *et al*. (2012), in their PCA of rambutan, and Kumar *et al*. (2025) in longan, reported similar trends of rapid decreases in variance explained beyond the second component, underscoring the practical utility of focusing on the top two PCs to interpret and delineate genotype diversity. Such findings collectively confirm the robustness of PCA in germplasm characterization and provide a strong foundation for subsequent clustering, genotype selection, and breeding applications.

### ****PCA - biplot of individuals****

The PCA individuals plot revealed distinct clustering and dispersion patterns among the nine wood apple genotypes, indicating their phenotypic divergence (Figure 2). Genotypes such as WFL-03, WFL-01, WFL-09, and WFL-08 were positioned prominently on the positive side of PC1, suggesting their superior performance for several key traits that align with this axis. In contrast, genotypes like WFL-07, WFL-06, and WFL-02 were positioned on the negative side of PC1, likely reflecting lower trait expressions or unique combinations of characters. Notably, WFL-05 was located near the origin, suggesting it possessed intermediate values across the trait spectrum, making it a potential average performer. The spatial separation of genotypes along both PC1 and PC2 illustrates the presence of significant phenotypic diversity, which can be exploited in future breeding and selection programs.



**Figure 2. PCA - Individual Plot**

Such patterns are commonly observed in arid zone fruit crops, where PCA and related multivariate analyses reveal that genetic variation is the dominant cause of distinct groupings among genotypes for traits like fruit size, weight, pulp content, and resilience to environmental stress. Singh *et al*. (2024) reported in bael (*Aegle marmelos*), a major arid-zone fruit, PCA of 80 wild genotypes and commercial cultivars found that the first two components explained nearly 64% of total variation, with genotypes clustering based on large fruit size and superior pulp yield traits crucial for survival and productivity in dry regions. Similar findings outlined by Dev *et al*. (2017) in *Grewia tenax* (“Phalsa cherry” or “white cross-berry”) collected from the arid Kachchh region of Gujarat identified substantial genetic variability using PCA and cluster analysis, supporting those genotypic effects are predominant for fruit-related traits in arid-adapted crops.

### ****PCA - biplot of variables****

The PCA variable biplot offered insightful interpretation regarding the contribution and association of various traits toward genotype differentiation (Figure 3). Traits such as fruit length, pulp weight, fruit width, yield per tree, total soluble solids (TSS), acidity ratio, total sugar, and reducing sugar exhibited strong and positive loadings along PC1, highlighting their pivotal role in defining the major axis of variation. In contrast, shell thickness and acidity showed a negative association with PC1, implying an inverse contribution to overall variability. Traits like ascorbic acid and pectin displayed moderate loadings and were positioned closer to the intersection of the axes, suggesting relatively uniform contribution across genotypes. Overall, yield and biochemical traits were the primary drivers of variation, emphasizing their importance in genotype selection and trait prioritization in wood apple improvement programs.



### ****Figure 3. PCA - biplot of variables****

### The observed strong and positive loadings of yield and biochemical traits such as fruit length, pulp weight, total sugar, TSS, and reducing sugar on PC1 underscore their central role in driving genotype differentiation in wood apple, while negative associations for shell thickness and acidity highlight traits inversely related to desirable productivity. This pattern is primarily due to underlying genetic diversity, affecting both fruit morphology and quality, as reported in numerous fruit crops. Elsafi (2012) reported that date palm demonstrated via PCA that fruit and sugar attributes were the primary contributors to genotype clustering, while structural traits like shell thickness played a secondary, often negative, role. In pomegranate, Verma *et al*. (2019) found that yield components and sugar content were the main discriminative variables along PC1, supporting their use as selection criteria in breeding. Similarly, in Ber (*Ziziphus mauritiana*), Das *et al*. (2022) utilized PCA and found that yield and taste-guided traits grouped together positively on the first axis, while seed hardness and shell characteristics loaded negatively or neutrally. These results mirror findings in bael, where multivariate analysis has consistently shown that superior genotypes combine favorable fruit size and biochemical attributes, while thick rind or high acidity are minimized (Singh *et al*., 2024). Collectively, these multi-crop studies reinforce that yield and fruit quality traits are the most powerful drivers of phenotypic diversity and should be prioritized in wood apple improvement programs.

### ****Hierarchical clustering dendrogram****

The hierarchical clustering analysis classified the genotypes into three major clusters based on their similarity across the measured traits (Figure 4). Cluster I included WFL-03, WFL-08, WFL-04, and WFL-05, which appeared to share common traits indicative of higher productivity or quality. Cluster II consisted of WFL-06 and WFL-07, possibly representing genotypes with moderate performance or unique trait expressions. Meanwhile, Cluster III grouped WFL-09, WFL-01, and WFL-02, suggesting distinct divergence possibly due to contrasting trait combinations. This classification pattern validates the underlying diversity among the genotypes and supports the use of cluster analysis for identifying genetically distinct and promising accessions for hybridization or conservation purposes. The dendrogram also reinforces the PCA results, providing complementary insights into genotype relatedness.



### ****Figure 4. Hierarchical clustering dendrogram****

The grouping of wood apple genotypes into three distinct clusters through hierarchical clustering provides strong evidence of substantial underlying genetic and phenotypic diversity in the collection. Notably, Cluster I (WFL-03, WFL-08, WFL-04, WFL-05) aggregates those genotypes expressing superior productivity or quality trait profiles, corresponding to the classic observation that high-performing accessions often form unique, trait-driven groups in germplasm analyses. In contrast, Cluster II, with WFL-06 and WFL-07, likely represents genotypes with moderate performance or potentially unique trait combinations worthy of targeted selection, while Cluster III (WFL-09, WFL-01, and WFL-02) demonstrates significant divergence, possibly owing to contrasting or less desirable trait profiles. In date palm, Jaradat *et al*. (2020) found that clustering analysis separated accessions by fruit quality, size, and adaptation. The high congruence between the dendrogram and PCA in your study further aligns with reports from desert fig (*Ficus carica*) (Hssaini *et al*., 2020), where multivariate approaches provided complementary insights into genotype relatedness and trait-driven diversity. This collective evidence underpins the value of hierarchical clustering for systematically dissecting diversity, facilitating both effective hybridization planning and conservation of genetic resources in wood apple.

### ****Heatmap of genotypes based on traits****

The heatmap provided a comprehensive and visually intuitive summary of trait expression across the wood apple genotypes (Figure 5). Each row represented a genotype, and each column corresponded to a specific trait, with the color gradient reflecting the magnitude of trait expression. Genotypes such as WFL-03 and WFL-08 were characterized by high expression values (represented in red hues) for several traits, including fruit length, pulp weight, yield per tree, total sugar, and pectin content, signifying their superiority. On the other hand, genotypes like WFL-07 and WFL-06 showed consistently lower values (indicated by blue hues) across most traits, suggesting limited performance. This visualization not only confirmed the results from PCA and cluster analysis but also allowed for quick identification of genotypes with desirable trait combinations. The clustering of traits in the heatmap further revealed correlations among variables, such as the association between sugars, TSS, and yield traits, enhancing the understanding of trait interrelationships.



### ****Figure 5. Heatmap of genotypes based on traits****

The heatmap’s vivid visualization of trait expression across wood apple genotypes enabled rapid identification of superior lines such as WFL-03 and WFL-08, which exhibited strong, positive expression for key yield and quality traits while distinguishing genotypes with underwhelming performance (WFL-07 and WFL-06) through their predominantly lower values. This color-coded matrix not only corroborated genotype rankings established by PCA and cluster analyses but also highlighted the importance of visual analytics for integrating complex multivariate data in crop improvement programs. Khadivi *et al*. (2025) used heatmap visualization to elucidate trait co-expression patterns and identify elite cultivars for yield and sweetness in pomegranate. Collectively, these original studies show that heatmaps are vital for breeders to visualize, interpret, and capitalize on the trait architecture underlying genetic diversity in wood apple, greatly aiding targeted variety development.

### ****Conclusion****

The present study revealed substantial phenotypic and biochemical diversity among nine wood apple (Feronia limonia L.) genotypes through comprehensive multivariate analysis. Principal Component Analysis (PCA) effectively reduced dimensionality, identifying key traits such as fruit weight, pulp weight, total sugar, and yield as major contributors to genotypic variation. Hierarchical clustering and heatmap analysis further classified the genotypes into distinct performance groups, with WFL-03 and WFL-08 emerging as superior in yield and quality traits. These results highlight the effectiveness of multivariate approaches in uncovering hidden patterns in trait expression and guiding selection strategies. The identified elite genotypes hold promise for use in future breeding, value addition, and conservation efforts. This study lays a strong foundation for systematic wood apple improvement and the mainstreaming of this underutilized fruit in sustainable tropical horticulture.

**References**

**Cattell, R. B. (1966).** The scree test for the number of factors.Multivariate Behavioral Research*,* **1**(2), 245–276.

Das, U., Hugar, A., Das, S., Das, S. S., & Kurubar, A. R. (2022). Genetic diversity of ber (*Ziziphus mauritiana* L.) cultivars using physico-chemical characteristics. Israel Journal of Plant Sciences, 69(3-4), 172-180.

Dev, R., Sureshkumar, M., Dayal, D., & Venkatesan, K. (2017). Genetic diversity among wild grewia tenax accessions collected from kachchh region of Gujarat, India. Indian Journal of Plant Genetic Resources, *30*(3), 286-292.

Elsafi, M. (2012). Study on the on-farm diversity of local date palm (*Phoenix dactylifera* L.) genetic resources grown in Northern region of Sudan.

Gewers, F. L., Ferreira, G. R., de Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., & da F. Costa, L. (2022). **Principal component analysis: A natural approach to data exploration.** ACM Computing Surveys, 54(4), 70:1–70:34.

Hssaini, L., Hanine, H., Razouk, R., Ennahli, S., Mekaoui, A., Guirrou, I., & Charafi, J. (2020). Diversity screening of fig (Ficus carica L.) germplasm through integration of morpho-agronomic and biochemical traits. International Journal of Fruit Science, *20*(4), 939-958.

Jaradat, A. A. (2015). Biodiversity, genetic diversity, and genetic resources of date palm. In Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas (pp. 19-71). Dordrecht: Springer Netherlands.

Johnston, J., & Wichern, D. W. (1982). Applied Multivariate Statistical Analysis (2nd ed.). Prentice-Hall.

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7.

Khadivi, A., Mirheidari, F., Saeidifar, A., Moradi, Y., & Tunç, Y. (2025). Morphological and pomological characterizations of pomegranate (*Punica granatum* L.) to select superior genotypes. *Scientific Reports*, *15*(1), 7038.

Kolde, R. (2019). pheatmap: Pretty Heatmaps. *R package version 1.0.12*. https://CRAN.R-project.org/package=pheatmap

**Krishna, H**, Singh, R. S., Bhargava, R., & Sharma, B. D. (201). Genetic diversity in wood apple (Feronia limonia) germplasm collected from Gujarat. Indian Journal of Arid Horticulture, 11(1‑2).

Kumar, A., Gupta, A. K., Kanth, N., Marboh, E. S., Lal, N., & Das, B. (2025). Assessment of Morphological and Physiochemical Diversity in Longan (*Dimocarpus Longan* Lour.) Genotypes for Genetic Improvement. *Tropical Plant Biology*, *18*(1), 1-17.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1–18.

Mir, J. I., Ahmed, N., Singh, D. B., Padder, B. A., Shafi, W., Zaffer, S., & Hamid, A. (2017). Diversity evaluation of fruit quality of apple (*Malus* × *domestica* Borkh.) germplasm through cluster and principal component analysis. Indian Journal of Plant Physiology, 22(3), 221–226.

Mohamed, H. (2016). Extraction and characterization of pectin from grapefruit peels. *MOJ Food Processing & Technology*, *2*(1), 31-38.

Mureșan, A. E., Sestras, A. F., Militaru, M., Păucean, A., Tanislav, A. E., Pușcaș, A., Mateescu, M., Mureșan, V., Marc, R. A., & Sestras, R. E. (2022). Chemometric comparison and classification of 22 apple genotypes based on texture analysis and physico‑chemical quality attributes. Horticulturae, *8*(1), Article 64.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raut, U. A., Tayde, A. G., Nagre, P. K., & Pawar, Y. D. (2022). Genetic variability studies in wood apple (Feronia limonia L.). Annals of Plant Sciences, 11(6), 954–957.

Roongruangsri, W., Rerkasem, B., & Jamjod, S. (2012). Diversity and relationships among rambutan (*Nephelium lappaceum* L.) germplasm as revealed by principal component analysis. Genetic Resources and Crop Evolution, 59(5), 887–900.

**Shinde, M. N. (2019). Genetic variability and correlation studies in wood apple (**Feronia limonia **L.).** M.Sc. Thesis, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, Maharashtra, India

Singh, A. K., Mao, A. A., & Medhi, R. P. (2016). Diversity of wood apple (*Feronia limonia* L.) in northeast India and their characterization. Genetic Resources and Crop Evolution, 63(2), 309-319.

Singh, P., Sharma, A., Jasrotia, A., Salgotra, R. K., Sharma, M., & Gupta, V. (2024). Diversity in morpho-pomological attributes and biochemical profiling of bael (*Aegle marmelos* (L.) Correa) genotypes of North-Western India. *Heliyon*, *10*(4).

Tayde, A. G., Raut, U. A., Wawkar, A. D., & Ravankar, D. S. (2024). Character association studies in wood apple (*Feronia limonia* L.) for yield and yield-associated traits. International Journal of Advanced Biochemistry Research, 8(5, Part I), 697–702.

Verma, M. K., Lal, S., Sharma, V. K., Choudhary, H., & Ahmed, N. (2019). PCA and genetic divergence analysis in pomegranate (*Punica granatum* L.) cultivars. *International Journal of Innovative Horticulture*, *8*(1), 45-50.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American statistical association, 58(301), 236-244.

Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, 63(2), 179–184.