Journal of Tropical Biodiversity and Biotechnology
Volume 11, Issue 02 (2026): jtbb21013
DOI: 10.22146/jtbb.21013

Research Article

Genetic Relationships and Population Structure of *Momordica cochinchinensis* (Gac) in Peninsular Malaysia using RAPD Markers and BinMat Visualisation

Alsalahi Muaadh[1,2], Ramisah M. Shah[1]*, Fatchurrachman[1], Fauziah T. Ahmad[3], Faridah Yahya[4], Rudiyanto[1]

1)Program of Crop Science, Faculty of Food Science and Agrotechnology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
2)Biology Department, Faculty of Applied Science, Taiz University, 9674 Taiz, Yemen
3)Research Interest Group Postharvest Research and Innovation, Faculty of Food Science and Agrotechnology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
4)Food Science Program, Faculty of Food Science and Agrotechnology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
* Corresponding author, email: ramisah@umt.edu.my

Keywords:
Binary matrix
Cluster analysis
Geographical clustering
Molecular markers
Momordica cochinchinensis
Population structure
Submitted:
24 April 2025
Accepted:
02 December 2025
Published:
08 May 2026
Editors:
Miftahul Ilmi
Sri Nopitasari

ABSTRACT

This study examined the genetic diversity and population structure of *Momordica cochinchinensis* (Gac) accessions from various states across Peninsular Malaysia using Random Amplified Polymorphic DNA (RAPD) markers. Binary RAPD-PCR data were analysed through non-metric multidimensional scaling (nMDS), hierarchical clustering with bootstrap support, and reliability assessment via Shepard plots. Results revealed substantial genetic variation among *M. cochinchinensis* accessions, with no clear geographical clustering, indicating moderate genetic differentiation. The nMDS analysis achieved an acceptable stress value of 14 % in two dimensions, while the Shepard plot demonstrated high reliability (R^2 = 0.84). Hierarchical clustering identified two major genetic lineages with varying bootstrap support, with some accessions—particularly those from Penang—showing high genetic similarity (bootstrap values >90 %). The observed genetic structure suggests that gene flow among populations has prevented complete genetic isolation despite some regional patterns. These findings provided valuable insights for conservation and breeding programs, highlighting the need to preserve populations across multiple states to maintain the full genetic diversity of *M. cochinchinensis*, a species of significant nutritional and medicinal value.

How to cite:
Muaadh, A. et al., 2026. Genetic Relationships and Population Structure of *Momordica cochinchinensis* (Gac) in Peninsular Malaysia using RAPD Markers and BinMat Visualisation. *Journal of Tropical Biodiversity and Biotechnology*, 11(2), jtbb21013. doi: 10.22146/jtbb.21013

INTRODUCTION

Genetic diversity analysis is a cornerstone of modern molecular biology, providing critical insights into species' evolutionary dynamics, population structure, and conservation strategies (Brhane & Hammenhag 2024). Among the various molecular markers utilised for such analyses, Random Amplified Polymorphic DNA (RAPD) markers have proven to be a cost-effective and technically straightforward tool for assessing genetic diversity, particularly in species with limited genomic resources (Sinha et al. 2023). RAPD markers, which generate binary data based on the presence or absence of DNA fragments, have been widely employed in studies of plant species to elucidate genetic relationships and population structures (Ruiz-Chután et al. 2019; Nam et al. 2020).

One such species of interest is *Momordica cochinchinensis*, a dioecious perennial climber belonging to the Cucurbitaceae family. The species is characterised by lobed leaves, pale yellow unisexual flowers, and spiny fruits that turn bright orange-red at maturity (Tran 2016). The arils surrounding the seeds are rich in carotenoids, particularly lycopene and β-carotene, which underlie their high medicinal and nutritional value (Rahman et al. 2024). Native to Southeast Asia, including Peninsular Malaysia, this species exhibits considerable genetic diversity, which is influenced by a combination of ecological heterogeneity, biogeographical history, and human-mediated dispersal (Wimalasiri et al. 2016; Mohd Khairi & Othman 2023).

Understanding the genetic structure of *M. cochinchinensis* is crucial for conservation efforts and breeding programs aimed at improving its agronomic and medicinal properties. This urgency is amplified by the species' restricted natural distribution and increasing pressure from overharvesting for medicinal and nutritional use. Habitat fragmentation has further reduced effective population sizes, raising concerns about genetic erosion (Méndez et al. 2014; Pinto et al. 2024). Moreover, the lack of structured germplasm resources limits the potential for systematic breeding, which is essential for assessing genetic diversity for both conservation and sustainable utilization (Bootprom et al. 2015).

Recent studies have employed RAPD markers to investigate the genetic relationships among *M. cochinchinensis* accessions. These investigations have demonstrated moderate genetic differentiation among accessions, with no distinct clustering based on geographical origin, suggesting substantial gene flow between populations (Bootprom et al. 2015). Random Amplified Polymorphic DNA (RAPD) markers were chosen for this study because they are cost-effective, require no prior sequence information, and have been widely applied in genetic diversity studies of non-model plant species (Marwal et al. 2014). Their ability to generate polymorphic profiles from small amounts of DNA makes them particularly suitable for initial diversity assessments in understudied species such as *M. cochinchinensis* (Dorado et al. 2017). However, analysing binary data from RAPD markers can be challenging due to the complexity of data processing and visualisation. Traditional methods often require multiple software tools and a high level of computational expertise, which can be time-consuming and prone to errors (van Steenderen 2022).

To address these challenges, the development of user-friendly tools for processing and analysing binary data has become increasingly important. One such tool is BinMat, an R package and R Shiny application designed to streamline the analysis of binary data obtained from fragment analysis methods such as RAPD, AFLP (Amplified Fragment Length Polymorphism), and ISSR (Inter-Simple Sequence Repeats) (van Steenderen 2022). BinMat automates the consolidation of replicate samples, generates summary statistics, and produces visualizations such as ordination plots and clustering trees, all within a single platform. This tool not only simplifies the analysis pipeline

but also enhances the reliability and reproducibility of genetic diversity studies.

In this study, we employed RAPD markers to investigate the genetic diversity and population structure of *M. cochinchinensis* accessions collected from various states in Peninsular Malaysia. We utilized BinMat to process and to analyse the binary data, generating non-metric multidimensional scaling (nMDS) plots and hierarchical clustering trees to visualize the genetic relationships among accessions. Our findings provided valuable insights into the genetic structure of *M. cochinchinensis* highlighting the importance of preserving genetic diversity across multiple states for conservation and breeding purposes. Furthermore, we demonstrated the utility of BinMat as a powerful tool for genetic diversity analysis, offering a streamlined and accessible approach for researchers in molecular biology and related fields. By combining traditional RAPD marker analysis with the advanced capabilities of BinMat, this study contributes to a deeper understanding of the genetic diversity and population dynamics of *M. cochinchinensis* while also showcasing the potential of modern computational tools to enhance genetic research.

MATERIALS AND METHODS

Plant Material

Fresh, young leaves of *M. cochinchinensis* was collected from mature accessions across four distinct regions of Peninsular Malaysia to capture the genetic diversity of the accessions across diverse ecological conditions and to facilitate genetic relationship analysis and population structure assessment: North Region including Perlis (Chuping [CHA, CHB], Beseri [BE]), Kedah (Baling [BA], Pokok Sena [PS]), and Penang (Bukit Mertajam [BMA, BMB, BMC]); West Region comprising Perak (Tanjung Rambutan [TR], Batu Gajah [BG1, BG2], Teluk Intan [TI]), Selangor (Sungai Buloh [SB], Telok Panglima Garang [TP], Jenjarom [JE]) and Negeri Sembilan (Port Dickson [PD], Kuala Pilah [KP]); South Region with specimens from Johor (Batu Pahat [BPA, BPB, BPC, BPD]) and Melaka (Merlimau [ME], Alor Gajah [AG]); and East Region encompassing Kelantan (Pasir Putih [PP], Kubang Kerian [KK]), Pahang (Kuala Lipis [KL], Karak Highway [RI]), and Terengganu (Kuala Berang [KB], Bukit Kor [BK]). The young leaves from all regions exhibited no observable morphological variation and were consistent in colour, size, and shape. All leaf samples were immediately placed in labelled tubes containing silica gel to ensure rapid desiccation and preservation of DNA integrity (Sunny et al. 2022), then sealed, catalogued with unique identifiers corresponding to their collection sites, and transported to the UMT laboratory for subsequent molecular analysis. Although the leaf samples collected across regions shared the general palmately lobed structure typical of *M. cochinchinensis*, they exhibited minor natural variability. To avoid redundancy, four representative leaves were selected to illustrate the observable range of variation in lobe number (3–5 lobes), lobe depth, and shape characteristics (Figure 1).

DNA Extraction and RAPD-PCR Analysis

Genomic DNA was extracted from leaf tissues of *M. cochinchinensis* accessions using the Prime Way Plant DNA Extraction Kit (1st Base). RAPD-PCR amplification was performed using 11 universal RAPD primers (Table 1). PCR reactions were carried out according to Wimalasiri et al. (2016), using a T100 Thermal Cycler (Bio-Rad, Singapore). Each 25 µL reaction contained 12.5 µL Green Taq (Promega, USA), 1 µL primer (5 µM), 1 µL template DNA (6 ng μL^{-1}), and 10.5 µL nuclease-free water. The thermal profile consisted of an initial denaturation at 94 °C for 4 min; 45 cycles of 94 °C for 1 min, 32–37 °C for 1 min, and 72 °C for 1 min; and a final extension at 72 °C for 5 min. The

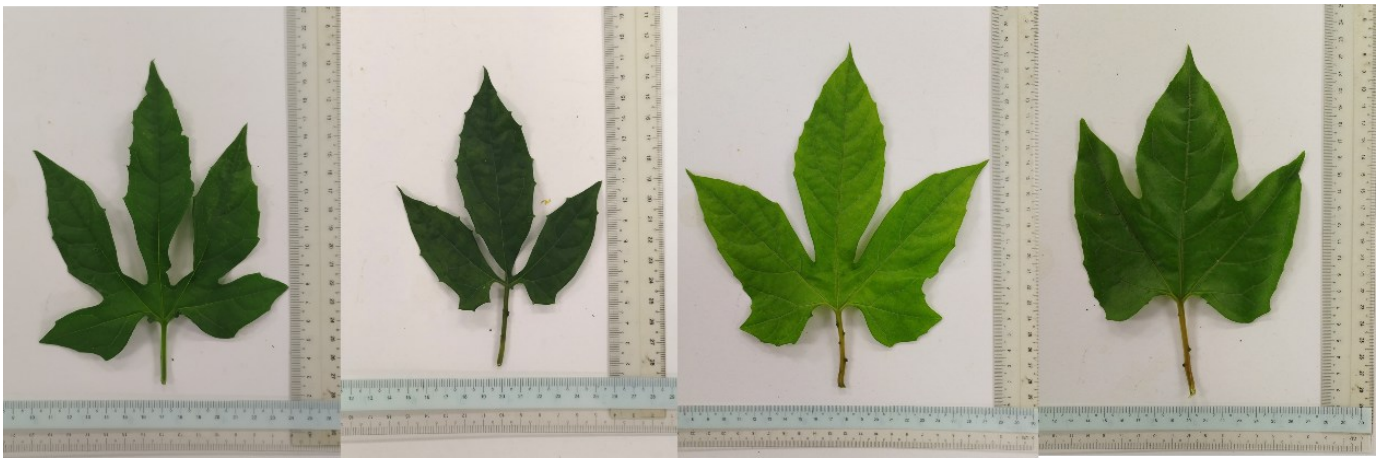

Figure 1. Representative leaf morphology of *M. cochinchinensis* accessions collected from different regions of Peninsular Malaysia. The four samples illustrate the observable range of natural variation, including differences in the number of lobes (3–5), degree of lobing (moderate to deep), and leaf shape.

resulting amplification products were separated by 1.5 % agarose gel electrophoresis. Electrophoresis was performed in 1× TBE buffer at 90 V for approximately 60 minutes. Gels were stained with SYBR Safe (0.1 µL mL^{-1}) and visualized under UV illumination using a gel documentation system (Nu Genius, Syngene, UK). To ensure reproducibility, all RAPD reactions were performed in duplicate, and only bands consistently observed across replicates were scored. Additionally, a subset of samples was rerun on separate gels at different times to confirm the reproducibility of banding patterns across runs.

Table 1. RAPD Primer Sequences for PCR Amplification.

Primer Name	Sequence (5' to 3')
OPC05	GATGACCGCC
OPC15	GACGGATCAG
OPC20	ACTTCGCCAC
OPC08	TGGACCGGTG
OPF1	ACGGATCCTG
OPF7	CCGATATCCC
OPN06	GAGACGCACA
OPO08	CCTCCAGTGT
OPW03	GTAGCGTCAG
D15	CATCCGTGCT
OPA18	AGGTGACCGT

Data Scoring and Analysis

RAPD fragments were scored as binary data with the presence (1) or absence (0) of bands across all *M. cochinchinensis* accessions. Only clear and reproducible bands were considered for analysis. The binary data matrix was utilised to calculate genetic distances between accessions. Each PCR sample was independently replicated to ensure data reliability, following standard protocols for RAPD analysis (Koeleman et al. 1998). The binary matrices were organised with column headings representing loci positions and rows containing sample identifiers followed by binary data. Replicate pairs were arranged consecutively within the dataset, maintaining unique identifiers for each sample. The data was checked for unwanted values utilising BinMat's validation function before processing.

The replicate pairs were consolidated into consensus reads utilising BinMat's consolidation algorithm. According to this algorithm, if both replicates showed presence of a band at a particular locus (1 and 1), a "1" was recorded; if both showed absence (0 and 0), a "0" was recorded; and if one repli-

cate showed presence while the other showed absence (1 and 0), a "?" was recorded to indicate an ambiguous read (van Steenderen 2022).

Software and Data Processing

Binary data obtained from RAPD marker analysis was processed utilising BinMat (van Steenderen 2022), an open-source R package specifically designed for processing and visualising binary genetic data. Raw electropherogram data were initially processed utilising GeneMarker (SoftGenetics) to generate binary matrices representing the presence (1) or absence (0) of amplified DNA fragments at specific loci.

Jaccard (JE) and Euclidean (EE) Error Rate Calculations

Jaccard (JE) and Euclidean (EE) error rates were calculated for the dataset to assess data quality. The Jaccard transformation (dJ) was applied utilising the formula described by Jaccard (1908), where f_{11} represents the total number of times a band occurred at the same locus in both samples, f_{00} represents the shared absence of bands, and f_{10} and f_{01} represent instances where a band was present in only one of the two sample replicates. Samples with Jaccard error rates greater than 0.15 were filtered from the dataset to ensure high-quality data for downstream analyses.

Ordination Analysis

Non-metric multidimensional scaling (nMDS) was employed to visualise relationships between samples based on the RAPD marker data. The optimal number of dimensions was determined utilising a scree plot, with stress values below 0.15 considered to represent a good fit for the data (Clarke 1993). The Jaccard distance metric was applied to create the distance matrix utilised for the nMDS ordination. Goodness-of-fit for the ordination was assessed utilising Shepard plots, which graphically represent how well the ordination fits the original distance data by plotting original Jaccard distances against transformed distances utilized in the nMDS.

Hierarchical Clustering Analysis

Hierarchical clustering analysis was performed utilising the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm with 1,000 bootstrap repetitions to assess the statistical support for the resulting clades. The clustering was based on a distance matrix generated utilising the Jaccard transformation. It was preferred due to its ability to focus on the shared presence of bands while excluding shared absence as biologically meaningful (van Steenderen 2022). This approach is particularly suitable for RAPD data, which typically shows dominance and may not treat the shared absence of bands as biologically informative.

Data Visualisation

The dataset was filtered by setting threshold values for the minimum number of bands present. Summary statistics, including the average number of bands (± standard deviation), the maximum and minimum number of bands, and the total number of loci, were calculated to characterize the dataset. Ordination plots and hierarchical clustering trees were visualised and exported as scalable vector graphics (SVG) files for publication (van Steenderen 2022).

All analyses were conducted within the BinMat platform, which integrates multiple R packages, including pvclust (Suzuki & Shimodaira 2006) for hierarchical clustering and MASS (Venables & Ripley 2002) for nMDS analysis, streamlining the workflow from data consolidation to visualisation on a single platform.

RESULTS AND DISCUSSION

Ordination Analysis

Non-metric multidimensional scaling (nMDS)

A non-metric multidimensional scaling (nMDS) plot generated from binary RAPD marker data (Figure 2) illustrates the genetic relationships among *M. cochinchinensis* accessions collected from Peninsular Malaysian states. This visualisation represents the genetic distances derived from the presence/absence (1/0) scoring of DNA fragments obtained through gel electrophoresis of RAPD-PCR products.

The distribution of *M. cochinchinensis* accessions across the two-dimensional space reveals considerable genetic variation within and between accessions from different Peninsular Malaysia states. The plot demonstrates no distinct, tight clustering by geographical origin, indicating moderate levels of genetic differentiation among the collected accessions. Specifically, Johor accessions (red) appear dispersed across multiple plot regions, suggesting substantial intra-state genetic diversity. Similarly, Kedah samples (blue) display marked variation, particularly along Dimension 2, with points appearing at both extreme positive (approximately 2.7) and negative (-1.9) values. Kelantan accessions (green) are positioned primarily in the upper portion of Dimension 2, while Pahang accessions (black) show considerable spread along both dimensions.

The observed pattern suggests that while some geographical structure exists, gene flow between *M. cochinchinensis* accessions from different Peninsular Malaysia states have been sufficient to prevent complete genetic isolation (Slatkin 1987). The lack of clear state-specific clusters indicates that the genetic variation of *M. cochinchinensis* in Peninsular Malaysia is not strictly compartmentalised by current political boundaries. The genetic diversity patterns displayed in *M. cochinchinensis* through RAPD marker analysis reflects the complex interplay of evolutionary forces, biogeographical history, and human influence across Peninsular Malaysia (Shidfar et al. 2018). The moderate differentiation observed among *M. cochinchinensis* accessions from different states provides valuable insights into the species' population dynamics and distribution patterns (Feder et al. 2013).

The substantial dispersion of accessions from individual states, particularly evident in Johor, Kedah, and Pahang, suggests high levels of intrastate genetic diversity in *M. cochinchinensis*. This pattern may be attributed to several factors. First, these states encompass diverse ecological niches, which may have promoted adaptive differentiation in response to heterogeneous environmental conditions (Tadele 2023). Second, as *M. cochinchinensis* is a dioecious outcrossing species, gene flow via pollen and seed dispersal likely contributes to maintaining genetic diversity within populations (Nora et al. 2011). The extensive genetic variation within states represents an important genetic resource for conservation and breeding programs (Salgotra & Chauhan 2023).

The partial overlap between *M. cochinchinensis* from various states in Peninsular Malaysia in the nMDS plot is suggestive of the presence of limited barriers to historical gene flow throughout the region. This is in good agreement with the biological characteristics of *M. cochinchinensis*, as its fleshy fruits are likely to be dispersed by mammals over long distances. Furthermore, the exploitation of this species as a source of traditional medicine and food might have promoted human-mediated dispersal throughout the region, thereby increasing genetic exchange between populations (Cai et al. 2023). From a conservation perspective, the observed genetic structure highlights the importance of preserving populations across multiple states to capture the full spectrum of genetic diversity in these accessions. The lack of genetic clusters per state implies that conservation efforts must disregard political bound-

aries and focus on conserving genetically diverse populations throughout the species' range (Forgiarini et al. 2025). The considerable genetic diversity uncovered in this study represents a valuable genetic resource for crop improvement projects, especially considering the expanding utilisation of the species as a source of nutrients and bioactive molecules.

For breeding programs of *M. cochinchinensis*, this study highlights the importance of extensive germplasm collection from every state in Peninsular Malaysia. The indicated genetic diversity utilising RAPD markers implies that *M. cochinchinensis* accessions from various regions may possess various adaptations and possibly desirable characteristics. The extensive genetic base manifested in our study provides a promising source for selective breeding programs for the improvement of agronomic performance, nutritional content, or medicinal properties for this plant. While RAPD markers enabled the rapid assessment of genetic diversity in *M. cochinchinensis*, their known limitations in reproducibility and resolution must be recognised. In this study, careful replication and validation were employed to enhance data reliability; however, future studies would benefit from combining RAPD with co-dominant markers, such as SSRs or SNPs, to achieve greater resolution and robustness in population structure analysis. Another limitation of the current study is the lack of integration between molecular and morphological data, as the work primarily focused on establishing a baseline genetic framework. Incorporating traits such as fruit size and shape, aril colour, seed morphology, flowering patterns, and fruit yield would provide deeper insights into how genetic diversity relates to phenotypic diversity. Future research should therefore combine molecular markers with detailed morphological evaluations and more extensive sampling to improve our understanding of the adaptive value of genetic diversity and to guide targeted breeding, conservation, and utilisation strategies for *M. cochinchinensis*.

Determination of the optimal number of dimensions

The genetic relationships among *M. cochinchinensis* accessions were analysed utilising non-metric multidimensional scaling (nMDS) based on binary RAPD marker data. The Figure 3 presents the scree plot for the nMDS anal-

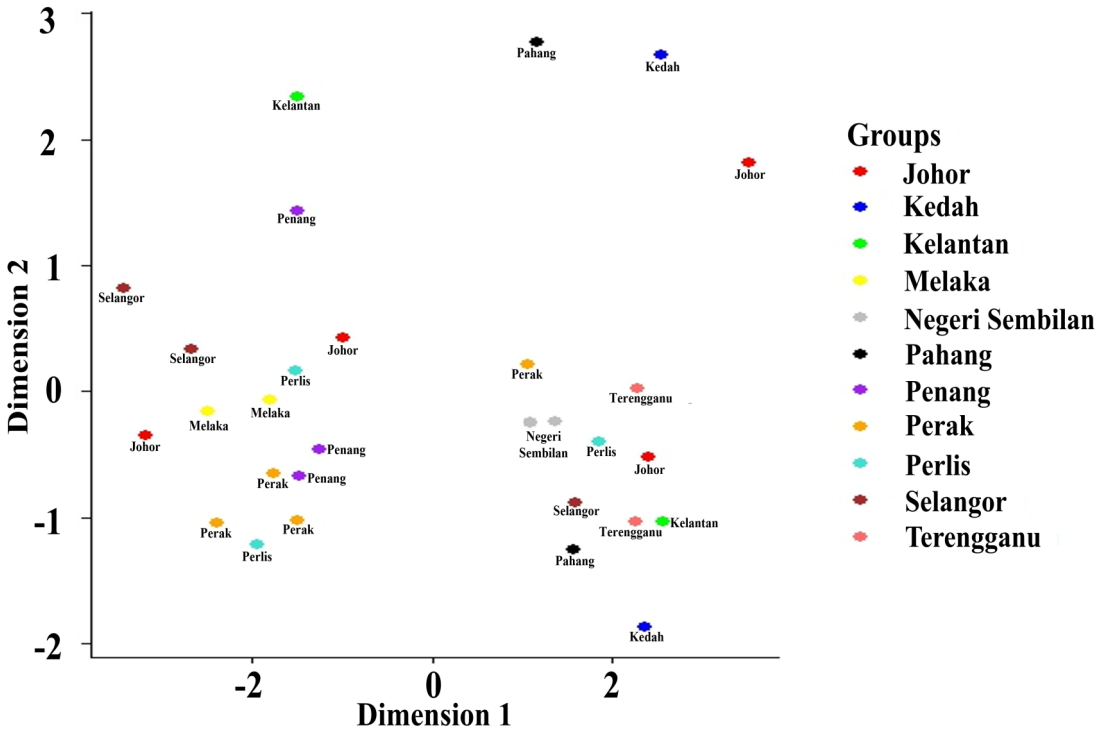

Figure 2. Non-metric multidimensional scaling (nMDS) plot.

ysis, illustrating the relationship between the number of dimensions and the corresponding stress values. The study revealed a considerable reduction in stress value as the number of dimensions increased from one to four. Specifically, the stress values decreased from approximately 25 % at one dimension to 14 % at two dimensions, followed by further reductions to 10 % at three dimensions and 7.5 % at four dimensions. It is interesting to note that the stress value fell below the commonly accepted threshold of 15 % when utilising two or more dimensions, suggesting that a two-dimensional representation provides reasonably accurate depiction of the genetic relationships among the studied accessions (van Steenderen 2022). The scree plot shows an "elbow" at two dimensions, where the rate of stress reduction diminishes substantially, indicating that two dimensions capture most of the meaningful variation in the data (Dexter et al. 2018).

The non-metric multidimensional scaling analysis of RAPD marker data revealed distinct patterns of genetic relationships among the *M. cochinchinensis* accessions analysed. The scree plot indicates that a two-dimensional representation adequately captures the genetic distance relationships, as evidenced by the stress value dropping below 15 %, a threshold widely considered acceptable in nMDS analysis (Kruskal 1964). The pronounced elbow in the scree plot at two dimensions suggests that this is the optimal dimensionality for representing the genetic relationships, balancing complexity with accuracy. This finding is consistent with the principle that the most informative dimensions in ordination techniques are typically identified at the point where the curve begins to flatten (Cattell 1966).

The RAPD markers successfully detected sufficient polymorphism to discern genetic relationships among *M. cochinchinensis* accessions. This finding aligns with previous studies demonstrating the efficacy of RAPD markers in assessing genetic diversity in various plant species (Kindie 2021). Despite criticisms regarding reproducibility, RAPD analysis remains a cost-effective and technically straightforward approach for preliminary genetic diversity assessment, particularly in species with limited genomic resources (Kindie 2021). The relatively low stress value at two dimensions (14 %) indicates that the two-dimensional ordination effectively represents the genetic distances derived from the binary RAPD data (Bruno & Balzarini 2010). This facilitates visualisation and interpretation of genetic relationships and potential clustering patterns that may correspond to geographical origins, adaptive traits, or evolutionary relationships among the accessions (Martín et al. 2003).

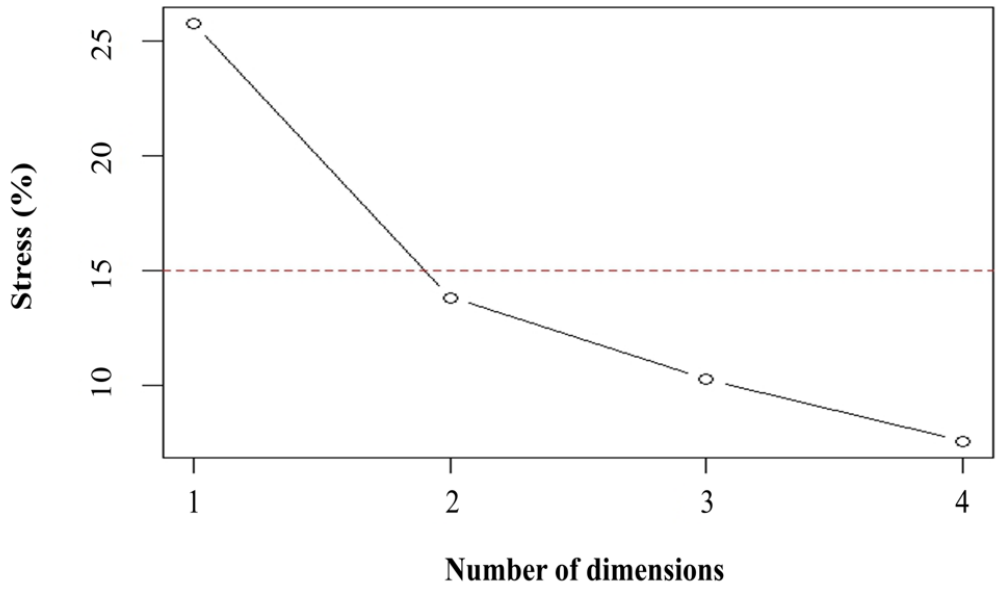

Figure 3. Scree plot shows the relationship between the number of dimensions (x-axis) and stress values as a percentage (y-axis).

Reliability of RAPD Marker Analysis

The RAPD-PCR analysis generated distinct banding patterns which were scored as binary data (presence = 1, absence = 0) across all *M. cochinchinensis* accessions. The resultant distance matrix was subjected to non-metric multi-dimensional scaling (nMDS) to visualise genetic relationships among the studied populations. The Shepard plot (Figure 4) demonstrates the relationship between the original distance matrix derived from the binary RAPD data and the corresponding ordination distances obtained through nMDS analysis (van Steenderen 2022).

The Shepard plot reveals a strong positive linear correlation between the original genetic distances and the ordination distances, with an R-squared value of 0.84. This high coefficient of determination indicates that approximately 84 % of the variance in the original RAPD-based genetic distances is accurately represented in the reduced-dimensional ordination space. The points generally align well along the diagonal reference line, particularly in the mid-range distance values (0.3-0.6), suggesting that the ordination effectively preserves the underlying genetic relationships detected by the RAPD markers. Some deviation is observed at both extremes of the distance spectrum. At the lower end (0.1-0.3), there is a slight overestimation of distances, while at the higher end (0.6-0.7), a minor degree of distance compression is evident. However, these variations are minimal and do not substantially impact the overall representation of genetic relationships, as indicated by the high R-squared value.

The strong goodness-of-fit observed in this analysis validates the reliability of the RAPD markers employed in this study for detecting genetic polymorphisms and accurately representing genetic relationships among *M. cochinchinensis* accessions. This is particularly noteworthy considering the dominant nature of RAPD markers, which can sometimes present challenges in genetic diversity studies (Hossain et al. 2024). These results align with previous findings by Kiran et al. (2010), who reported similar levels of reliability (R-squared values of 0.79-0.86) when using RAPD markers for genetic diversity assessment in different herbal species. The high concordance between original and ordination distances further confirms that the selected RAPD primers effectively captured significant genetic variation within *M. cochinchinensis* accessions.

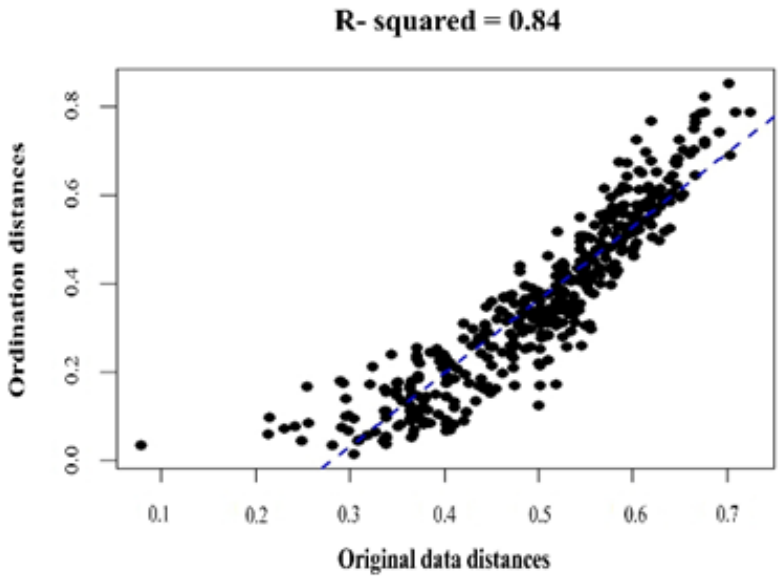

Figure 4. Shepard plot showing the relationship between original data distances (x-axis) and ordination distances (y-axis).

The observed pattern of genetic distances suggests a gradient of genetic differentiation among *M. cochinchinensis* accessions, potentially reflect geographic isolation or adaptive divergence in response to environmental factors (Peña et al. 2022). Further analysis incorporating geographic and environmental data would help elucidate the mechanisms driving this genetic structuring. The high R-squared value of the Shepard plot validates our RAPD-based genetic analysis approach and provides a reliable foundation for interpreting relationships among *M. cochinchinensis* accessions. Despite the limitations of dominant markers like RAPDs, the result of the present study demonstrates their utility for cost-effective genetic diversity assessment in non-model organisms lacking established genomic resources.

Hierarchical cluster analysis

Hierarchical cluster analysis of RAPD marker data revealed distinct genetic relationships among *M. cochinchinensis* accessions. The UPGMA dendrogram (Figure 5), constructed utilising binary data (presence/absence of bands) with the average linkage method, shows clear clustering patterns with p-values (%) indicating the statistical support for each branch. The dendrogram indicates two main clusters that separate at a height of approximately 0.5. The first main cluster predominantly consists of accessions from the east region (KB, BK, PP, RI, KL) and north region (PS, BE, BA), along with several accessions from the west region (TP, PD, KP, TI) and South Region (BPC, BPD). The second main cluster primarily contains accessions from the north region (CHA, CHB, BMA, BMB, BMC), the south region (BPA, BPB, AG, ME), some accessions from the west region (JE, SB, TR, BG2, BG1), and KK from the east region.

Strong correlations with p-values exceeding 90 % are observed among various pairs of accessions, including BMB-BMC from the north region (100 %), JE-SB from the west region (100 %), and BK-KB from the east region (95 %). High p-values are indicative of high genetic similarity among *M. cochinchinensis* accessions. Low height values for certain clusters are reflective of stronger genetic relationships among those accessions. The hierarchical clustering analysis based on RAPD marker polymorphisms provides valuable insights into the genetic relationships among *M. cochinchinensis* accessions (Chikkaswamy 2015). The binary distance measurement effectively translated the gel electrophoresis band patterns into quantifiable genetic distances.

The formation of two distinct major clusters suggests the existence of two primary genetic lineages within the sampled accessions. The high bootstrap value (p-values) at several nodes indicates reliable grouping, particularly for accessions from north region BMB and BMC and west regions JE and SB (100 %), which likely represent very closely related genotypes or possibly duplicate accessions (Singh et al. 2004). The intermediate clustering patterns with moderate p-values (60-80 %) indicate accessions with partial genetic similarity, possibly representing intermediate evolutionary relationships or hybridisation events (Highton 1993). For example, the clustering of BE, BPD, and PP with p-values ranging from 68 % to 87 % suggests a moderate level of genetic relatedness among these accessions.

The average linkage method used in this analysis is appropriate for RAPD data as it considers all pairwise distances between members of different clusters, thus providing a balanced representation of genetic relationships that is less affected by outliers than single or complete linkage methods. The varying heights at which samples join clusters (from 0.1 to 0.5) demonstrate the hierarchical nature of genetic relationships, with samples joining at lower heights being more closely related than those joining at greater heights (Rodrigues et al. 2017). This gradation of genetic similarity provides evidence for continuous rather than discrete variation within the studied accessions.

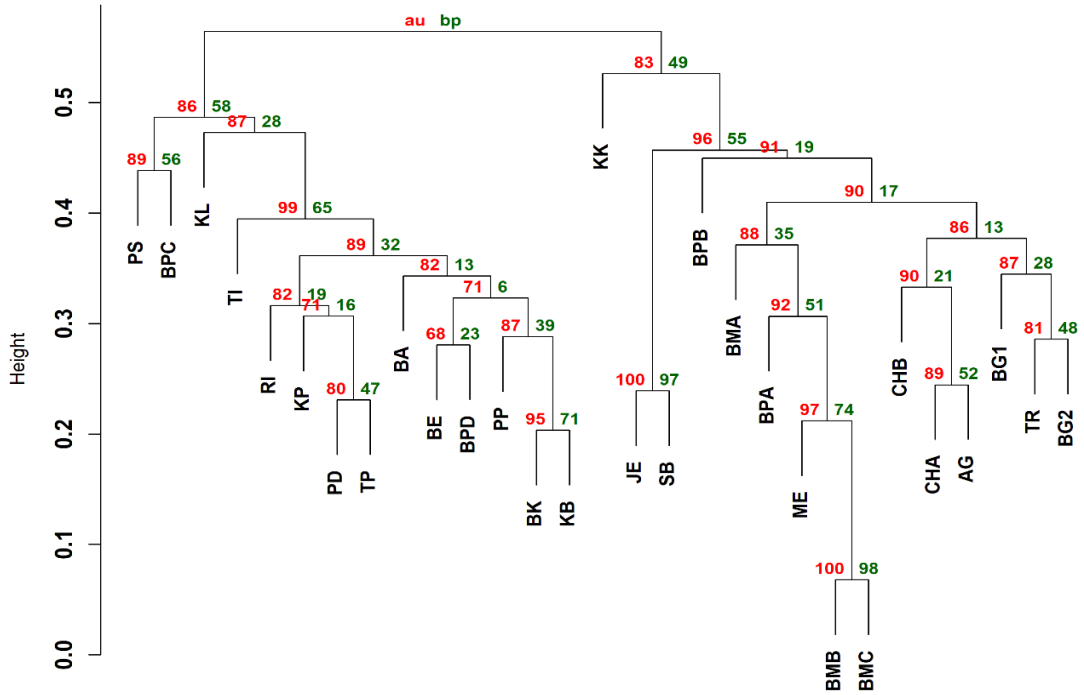

Figure 5. Hierarchical cluster dendrogram with p-values (%) based on binary distance and average linkage method, showing relationship patterns among accessions with statistical support indicated at each branch point.

These findings contribute to understanding the genetic structure and diversity within the studied accessions, which has implications for conservation strategies and breeding programs.

The hierarchical cluster analysis generated a dendrogram with bootstrap support values (expressed as percentages), while the nMDS plot provided a two-dimensional representation of genetic distances derived from binary RAPD marker data. The dendrogram (Figure 5) revealed several distinct clusters with varying levels of bootstrap support. Notably, Penang accessions (BMA, BMB, BMC) formed a tight cluster with exceptionally high bootstrap support (100 % between BMB and BMC), indicating strong genetic similarity among these accessions. This clustering pattern was mirrored in the nMDS plot, where these accessions appeared close. Similarly, certain Perak accessions (TR, TI) clustered together in the dendrogram with moderate bootstrap support (57 %) and showed relative proximity in the nMDS visualisation . Johor accessions (BPC, BPD, BPB, BPA) displayed notable dispersion in both analyses. In the dendrogram, these accessions were distributed across different clusters, while in the nMDS plot, they occupied multiple quadrants, confirming substantial intra-state genetic diversity. Kelantan accessions (KK, PP) formed a distinct group in the dendrogram and were positioned primarily in the upper portion of Dimension 2 in the nMDS plot, suggesting the genetic differentiation of these northern populations.

The dendrogram revealed several instances of cross-state clustering with moderate to strong bootstrap support. For example, certain Perlis accessions (CHA, CHB) showed a genetic affinity with accessions from other states rather than clustering exclusively by geographic origin. This pattern was consistent with the nMDS visualisation , where accessions from different states frequently appeared intermixed rather than forming distinct state-specific clusters. The congruence between hierarchical clustering and nMDS approaches provides robust evidence for a complex population structure of *M. cochinchinensis* in Peninsular Malaysia. The hierarchical clustering quantifies relationship reliability through bootstrap values, while nMDS captures multi-

dimensional similarities without imposing a bifurcating structure. Together, they reveal genetic diversity patterns that transcend administrative boundaries.

High bootstrap support for certain clusters (e.g., Penang accessions BMB/BMC at 100 %) indicates the presence of genetically distinct lineages that may represent isolated populations or those adapted to specific environments. The moderate bootstrap values (25-50 %) for many branches suggest ongoing gene flow or recent divergence among populations (Zaharias et al. 2023). The observed genetic relationships provide evidence for both isolation-by-distance effects and human-mediated dispersal. While some geographically proximate populations show genetic similarity, frequent cross-state clustering indicates that factors beyond geography have shaped diversity in this species, likely including human-mediated transport of this valuable medicinal and food plant.

CONCLUSION

The genetic diversity analysis of *M. cochinchinensis* accessions from Peninsular Malaysia utilising RAPD markers has revealed complex patterns of genetic relationships that transcend administrative boundaries. The congruence between hierarchical clustering and nMDS approaches provides robust evidence for moderate genetic differentiation with substantial intra-state diversity, particularly in Johor, Kedah, and Pahang accessions. The high reliability of our RAPD-based analysis, confirmed by an R-squared value of 0.84 in the Shepard plot, validates the result of the present study's methodological approach despite the limitations inherent to dominant markers.

In breeding programs, such diversity is a valuable resource for crop improvement. Accessions from genetically distinct clusters—such as those from Penang (BMB, BMC) and Johor (BPA, BPC)—should be prioritised as potential parental lines for hybridization. These genotypes may carry unique alleles associated with desirable agronomic traits such as high carotenoid content, large fruit size, or stress tolerance. Moreover, the high intra-state diversity found in Johor and Kedah suggested that multiple local selections could be made within these regions to identify elite lines. Integrating genetic data with future phenotypic evaluations, including fruit yield, nutritional composition, and disease resistance, will help to identify superior genotypes suitable for domestication and cultivar development. Future studies must incorporate co-dominant markers, more extensive sampling strategies, and analyses combining genetic structure with agronomic, morphological, and phytochemical traits to further our understanding of this valuable species.

AUTHOR CONTRIBUTION

A.M. conducts experiments, data analysis and prepares the original manuscript, R.M.S. conceptualisation , data curation, and student supervision, F.T.A runs data curation and student supervision, F.Y. runs the data curation, F. and R. contribute to data analysis and visualisation . All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledged the support and funding provided by Universiti Malaysia Terengganu (UMT) through grant UMT/TAPE-RG/2020/55240. Special gratitude is also extended to the Faculty of Food Science and Agrotechnology for its continuous support throughout this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bootprom, N. et al., 2015. Genetic diversity based on agricultural traits and phytochemical contents in spiny bitter gourd (*Momordica cochinchinensis* (Lour.) Spreng). *SABRAO Journal of Breeding and Genetics*, 47(3), pp.278–290.

Brhane, H. & Hammenhag, C., 2024. Genetic diversity and population structure analysis of a diverse panel of pea (*Pisum sativum). Frontiers in Genetics*, 15, 1396888. doi: 10.3389/fgene.2024.1396888.

Bruno, C.I. & Balzarini, M.G., 2010. Ordinations of genetic data from multidimensional markers. *Revista de la Facultad de Ciencias Agrarias*, 42(2), pp.183-200.

Cai, L. et al., 2023. The complete plastome sequence of *Momordica cochinchinensis* (Cucurbitaceae). *Mitochondrial DNA Part B*, 8(3), pp.329–332. doi: 10.1080/23802359.2023.2181649.

Cattell, R.B., 1966. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), pp.245–276. doi: 10.1207/s15327906mbr0102_10.

Chikkaswamy, B.K., 2015. Assessment of genetic diversity and relationships of medicinal plants using RAPD marker. *International Journal of Advanced Research in IT and Engineering*, 4(2), pp.25-35.

Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1), pp.117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x.

Dexter, E., Rollwagen-Bollens, G. & Bollens, S.M., 2018. The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography: Methods*, 16(7), pp.434–443. doi: 10.1002/lom3.10257.

Dorado, G. et al., 2017. Molecular Markers. In *Reference Module in Biomedical Sciences*. Amsterdam: Elsevier. doi: 10.1016/B978-0-12-801238-3.08996-0.

Feder, J.L. et al., 2013. Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, 44, pp.73–97. doi: 10.1146/annurev-ecolsys-110512-135825.

Forgiarini, C. et al., 2025. Low genetic differentiation despite high habitat fragmentation in an endemic and endangered species of Iridaceae from South America: implications for conservation. *Botanical Journal of the Linnean Society*, 207(1), pp.56–69. doi: 10.1093/botlinnean/boae036.

Jaccard, P., 1908. New research on floral distribution. *Bulletin of the Vaud Society of Natural Sciences*, 44, pp.223–270.

Highton, R., 1993. The relationship between the number of loci and the statistical support for the topology of UPGMA trees obtained from genetic distance data. *Molecular Phylogenetics and Evolution*, 2(4), pp.337–343. doi: 10.1006/mpev.1993.1033.

Hossain, A. et al., 2024. Plant genetic diversity studies: Insights from DNA marker analyses. *International Journal of Plant Biology*, 15(3), pp.607–640. doi: 10.3390/ijpb15030046.

Kindie, B., 2021. Assessing molecular marker applications for genetic variety analysis in biodiversity conservation status. *Molecular Biology*, 10(4), 287.

Kiran, U. et al., 2010. SCAR markers: A potential tool for authentication of herbal drugs. *Fitoterapia*, 81(8), pp.969–976. doi: 10.1016/j.fitote.2010.08.002.

Koeleman, J. G.M. et al., 1998. Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of Acinetobacter genomic species and typing of *Acinetobacter baumannii*. *Journal of Clinical Microbiology*, 36(9), pp.2522–2529. doi: 10.1128/jcm.36.9.2522-2529.1998.

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), pp.1–27. doi: 10.1007/BF02289565.

Martín, A.P.M.S. et al., 2003. Genetic variation within the endangered species *Aldrovanda vesiculosa* (Droseraceae) as revealed by RAPD analysis. *Aquatic Botany*, 75(2), pp.159–172. doi: 10.1016/S0304-3770(02)00173-0.

Marwal, A., Sahu, A.K. & Gaur, R.K., 2014. Molecular Markers: Tool for Genetic Analysis. *Animal Biotechnology: Models in Discovery and Translation*, pp.289–305. doi: 10.1016/B978-0-12-416002-6.00016-X.

Méndez, M. et al., 2014. Joint effects of population size and isolation on genetic erosion in fragmented populations: finding fragmentation thresholds for management. *Evolutionary applications*, 7(4), pp.506–518. doi: 10.1111/eva.12154.

Mohd Khairi, N.N. & Othman, H., 2023. Morphological and genetic analysis of *Momordica cochinchinensis* (Lour.) Spreng. (gac) from different accessions in Malaysia. *Journal of Tropical Life Science*, 13(2), pp.319–334. doi: 10.11594/jtls.13.02.10.

Nam, V.T. et al., 2020. Molecular markers for analysis of plant genetic diversity. *Vietnam Journal of Biotechnology*, 18, pp.589-608. doi: 10.15625/1811-4989/18/4/15326.

Nora, S. et al., 2011. Gene flow (pollen and seed) in fragmented populations of plants. *Ecosistemas*, 20(2-3), pp.35-45.

Peña, J. A., Gómez-Pérez, L. & Alfonso-Sánchez, M. A., 2022. On the trail of spatial patterns of genetic variation. *Evolutionary Biology*, 49(1), pp.84–91. doi: 10.1007/s11692-021-09552-y.

Pinto, A.V. et al., 2024. The impact of habitat loss and population fragmentation on genomic erosion. *Conservation Genetics*, *25*(1), pp.49–57. doi: 10.1007/s10592-023-01548-9.

Rahman, M.N.A. et al., 2024. Physicochemical properties, proximate composition, and carotenoid content of *Momordica cochinchinensis* L. Spreng (Gac) fruit. *Pertanika Journal of Tropical Agricultural Science*, 47(2), pp.461–479. doi: 10.47836/pjtas.47.2.09.

Rodrigues, B. et al., 2017. Dissimilarity measures and hierarchical methods for the study of genetic diversity on soybean. *Bioscience Journal*, 33(6), pp.1544–1555. doi: 10.14393/BJ-v33n6a2017-37909.

Ruiz-Chután, J.A et al., 2019. Assessment of genetic diversity in *Sorghum bicolor* using RAPD markers. *Genetika*, 51, pp.789–803. doi: 10.2298/GENSR1903789R

Salgotra, R.K. & Chauhan, B. S., 2023. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes*, 14(1), 174. doi: 10.3390/genes14010174

Shidfar, M. et al., 2018. RAPD markers reveal genetic variation between *Cichorium spinosum* L. and *Taraxacum* sp.: A substantial medicinal plant of Greece. *Progress in Nutrition*, 20(1-S), pp.153-159. doi: 10.23751/pn.v20i1-s.5993

Singh, A.P. et al., 2004. Phylogenetic relationships as in *Ocimum* revealed by RAPD markers. *Euphytica*, 136(1), pp.11–20. doi: 10.1023/B:EUPH.0000019497.89760.e8.

Sinha, S. et al., 2023. Recent Advancements in Molecular Marker Technologies and Their Applications in Crop Improvement. *Molecular Marker Techniques: A Potential Approach of Crop Improvement*, pp.319–337. doi: 10.1007/978-981-99-1612-2_15.

Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science*, 236(4803), pp.787-792. doi: 10.1126/science.3576198.

Sunny, A.M. et al., 2022. Standardization of extraction of DNA from silica-gel dried leaf sample of *Garcinia Indica* (Thouars) Choisy. *Environmental Sciences Proceedings*, 22(1), 7. doi: 10.3390/IECF2022-13131.

Suzuki, R. & Shimodaira, H., 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), pp.1540–1542. doi: 10.1093/bioinformatics/btl117.

Tadele, Z., 2023. Orphan crops: their role in food and nutrition security. In *Sustainable Food Science-A Comprehensive Approach*. Elsevier, pp.10–22. doi: 10.1016/B978-0-12-823960-5.00049-4.

Tran, X.T., 2016. Effects of maturity on physicochemical properties of Gac fruit (*Momordica cochinchinensis* Spreng.). *Food Science and Nutrition*, 4(2), pp.305–314. doi:10.1002/fsn3.291.

Van Steenderen, C., 2022. BinMat: A molecular genetics tool for processing binary data obtained from fragment analysis in R. *Biodiversity Data Journal*, 10, e77875. doi: 10.3897/BDJ.10.e77875.

Venables, W.N. & Ripley, B.D., 2002. *Modern applied statistics with S. 4th ed.* New York: Springer. doi: 10.1007/978-0-387-21706-2.

Wimalasiri, D. et al., 2016. Morphological and genetic diversity of *Momordica cochinchinensis* (Cucurbitaceae) in Vietnam and Thailand. *Genetic Resources and Crop Evolution*, 63(1), pp.19–33. doi: 10.1007/s10722-015-0232-8.

Zaharias, P., Lemoine, F. & Gascuel, O., 2023. Robustness of Felsenstein's versus transfer bootstrap supports with respect to taxon sampling. *Systematic Biology*, 72(6), pp.1280–1295. doi: 10.1093/sysbio/syad052