International Journal of Botany Studies
www.botanyjournals.com
ISSN: 2455-541X
Received: 20-02-2026, Accepted: 23-03-2026, Published: 07-04-2026
Volume 11, Issue 2, 2026, Page No. 23-28

Molecular identification and phylogenetic characterization of *Sesamum indicum* L. using ITS DNA barcoding

Jotiram D Salunke[1], Dr. Gulab M Rathod[2], Srushti S Kale[3]
[1] Late. N. P. W. Arts and Science College, Chopa, Taluka Goregaon, Gondia, Maharashtra, India
[2] Shrikrishna Mahavidyalaya, Gunjoti, Taluka Omerga, Dharashiv, Maharashtra, India
[3] Department of Botany, Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar, Maharashtra, India

Abstract
Accurate molecular identification is essential for plant taxonomy, conservation, and pharmacological applications. The present study aimed to authenticate *Sesamum indicum* L. using Internal Transcribed Spacer (ITS) DNA barcoding and to evaluate its phylogenetic relationships with related taxa. Genomic DNA was extracted from leaf samples and sequenced using Sanger sequencing. The obtained ITS sequence (~452 bp) was analyzed using BLAST, which showed 99.34% similarity with *Sesamum indicum* (GenBank accession: KM210317.1), confirming species identity. Phylogenetic analysis was performed using the Neighbor-Joining method with bootstrap support (500 replicates), revealing clustering of the sample with *S. indicum* and clear divergence from related genera. The genetic distance matrix further supported minimal intraspecific divergence (0.0067) and significant interspecific variation. The results demonstrate that ITS markers are reliable tools for molecular identification and phylogenetic analysis of *Sesamum indicum* L.

Keywords: ITS DNA barcoding, *Sesamum indicum* L., phylogenetic analysis, BLAST, etc

Introduction

The accurate identification and classification of plant species are fundamental to biodiversity conservation, sustainable utilization and pharmacological exploration. Traditionally, plant taxonomy has relied on morphological characteristics; however, such approaches are often limited by phenotypic plasticity, environmental influences and the presence of cryptic species. In recent decades, molecular techniques have emerged as powerful tools for resolving taxonomic ambiguities and understanding evolutionary relationships among plant taxa.

DNA barcoding, which involves the use of short, standardized DNA sequences for species identification, has gained widespread acceptance in plant systematics. The concept was first proposed by Paul D. N. Hebert *et al.* (2003) [6], who demonstrated its effectiveness in biological identification. Among various molecular markers, the nuclear ribosomal Internal Transcribed Spacer (ITS) region has been extensively utilized due to its high mutation rate, biparental inheritance and strong discriminatory power at the species level. The ITS region, comprising ITS1, 5.8S rRNA gene and ITS2 is particularly effective in distinguishing closely related taxa and has been recommended as a universal barcode for plant identification (W. John Kress *et al*., 2005; Shilin Chen *et al*., 2010) [7, 3].

Sesamum indicum L. commonly known as sesame, belongs to the family Pedaliaceae and is one of the oldest cultivated oilseed crops in the world. It is widely distributed across tropical and subtropical regions and is valued for its high oil content, nutritional significance and medicinal properties. Sesame seeds are rich in lignans such as sesamin and sesamolin, which exhibit antioxidant, anti-inflammatory and cardioprotective activities. Due to its economic and therapeutic importance, accurate identification and genetic characterization of *Sesamum indicum* L. are crucial for crop improvement, germplasm conservation and authentication of plant materials used in herbal formulations.

Despite its significance, morphological identification of *Sesamum* species can be challenging due to overlapping characters and environmental variability. Molecular approaches, particularly DNA barcoding using ITS sequences, provide a reliable alternative for species authentication. Previous studies have demonstrated the effectiveness of ITS markers in resolving phylogenetic relationships within angiosperms and in distinguishing closely related species within economically important plant groups (Shilin Chen *et al*., 2010) [3].

In addition to species identification, phylogenetic analysis plays a vital role in understanding evolutionary relationships and genetic divergence among taxa. Methods such as Neighbor-Joining (NJ), Maximum Likelihood (ML) and Bayesian inference are commonly used to reconstruct phylogenetic trees based on sequence data. The Neighbor-Joining method proposed by Naruya Saitou and Masatoshi Nei (1987) [11], along with bootstrap analysis introduced by Joseph Felsenstein (1985) [5], are widely applied for evaluating tree reliability.

In this context, the present study aims to employ ITS DNA barcoding for the molecular identification of *Sesamum indicum* L. and to evaluate its phylogenetic relationship with closely related taxa retrieved from the NCBI database. The study further seeks to assess genetic divergence using distance matrix analysis, thereby contributing to the understanding of species boundaries and evolutionary relationships within the group. The findings of this study are expected to support the application of molecular tools in plant taxonomy, authentication, and biodiversity studies.

Materials and Methods

1. Sample Collection and Authentication

Fresh and healthy leaf samples of *Sesamum indicum* L. (Sample ID: PI176) were collected from cultivated sources. The taxonomic identification and authentication of the plant material were carried out at Dr. Babasaheb Ambedkar Marathwada University (BAMU), Aurangabad, India. The authenticated specimen was used for further molecular analysis.

2. Genomic DNA Extraction

Genomic DNA was extracted from fresh leaf tissue using the cetyltrimethylammonium bromide (CTAB) method, which is widely employed for plant DNA isolation due to its efficiency in removing polysaccharides and secondary metabolites (John J. Doyle and Judy L. Doyle, 1987) [4]. Approximately 50–100 mg of fresh leaf tissue was used for extraction under sterile laboratory conditions. The tissue was finely homogenized, and cellular components were lysed using CTAB extraction buffer containing detergents and salts to facilitate membrane disruption and DNA release.

The DNA was subsequently purified through sequential steps involving chloroform–isoamyl alcohol extraction to remove proteins and other contaminants, followed by precipitation using cold isopropanol. The resulting DNA pellet was washed with ethanol, air-dried and resuspended in TE buffer for further use. This method is particularly suitable for plant tissues rich in secondary metabolites, as it ensures high-quality DNA yield (Sambrook, J. *et al*., 1989) [12].

The quality and integrity of the extracted DNA were evaluated by agarose gel electrophoresis, while the concentration and purity were determined using spectrophotometric analysis by measuring absorbance ratios (A260/A280) (Wilfinger, W.W. *et al*., 1997) [17]. Only high-quality DNA exhibiting minimal degradation and optimal purity was used for subsequent PCR amplification.

3. PCR Amplification of ITS Region

The Internal Transcribed Spacer (ITS) region of nuclear ribosomal DNA was amplified using universal primers ITS1 and ITS4, which are widely employed for plant DNA barcoding due to their high specificity and amplification efficiency (Thomas J. White *et al*., 1990) [16]. The ITS region, comprising ITS1, 5.8S rRNA gene and ITS2, provides sufficient sequence variation for species-level identification in plants.

Polymerase Chain Reaction (PCR) was performed in a total reaction volume of 25 µL. The reaction mixture consisted of approximately 20–50 ng of template genomic DNA, 0.2–0.5 µM each of forward and reverse primers, 200 µM of each deoxynucleotide triphosphate (dNTP), 1× PCR buffer containing $MgCl_2$ (1.5–2.5 mM), and 1–2 units of Taq DNA polymerase. The use of Mg^{2+} ions are critical for enzyme activity and fidelity during DNA amplification (Kary B. Mullis, 1990) [10].

Amplification was carried out in a programmable thermal cycler under optimized conditions. The PCR program included an initial denaturation step at 94–95°C for 3–5 minutes to ensure complete denaturation of double-stranded DNA. This was followed by 30–35 cycles consisting of denaturation at 94–95°C for 30 seconds, annealing at 52–58°C for 30 seconds and extension at 72°C for 1 minute. The annealing temperature was optimized based on primer melting temperature to ensure specificity of binding. A final extension step at 72°C for 7–10 minutes was included to allow complete synthesis of amplified fragments.

The amplified PCR products were analyzed by electrophoresis on 1–1.5% agarose gel stained with an appropriate nucleic acid dye. The gel was visualized under UV illumination to confirm the presence of distinct and expected size bands corresponding to the ITS region. Agarose gel electrophoresis remains a standard method for verifying PCR amplification and assessing product quality (Sambrook, J. *et al*., 1989) [12].

The successful amplification of the ITS region ensured the suitability of the PCR products for subsequent sequencing and molecular analysis.

4. DNA Sequencing

The amplified PCR products of the ITS region were purified to remove excess primers, nucleotides, enzymes, and other reaction components using standard purification protocols. The purified products were then outsourced to geneOmbio Technologies, Pune for sequencing analysis.

DNA sequencing was carried out using the Sanger dideoxy chain termination method, which remains a gold standard for accurate nucleotide sequencing of PCR products (Frederick Sanger *et al*., 1977) [13]. This method involves incorporation of chain-terminating dideoxynucleotides (ddNTPs) during DNA synthesis, resulting in fragments of varying lengths that are subsequently separated by capillary electrophoresis to determine the nucleotide sequence.

Sequencing reactions were performed in both forward and reverse directions to ensure accuracy and reliability of the generated data. The resulting chromatograms were carefully examined, and ambiguous base calls were corrected through manual editing. Sequence assembly and alignment were performed using appropriate bioinformatics tools to generate a high-quality consensus sequence.

A reliable ITS sequence of approximately 452 base pairs was obtained for the sample, which was used for further molecular analyses including BLAST comparison and phylogenetic reconstruction. The quality of the sequence was ensured by trimming low-quality regions and confirming the absence of sequencing errors, thereby enhancing the accuracy of downstream analyses.

5. BLAST Analysis

The edited ITS sequence obtained from the sample was subjected to similarity search using the BLASTN (Basic Local Alignment Search Tool for nucleotides) algorithm available at the NCBI GenBank database. BLAST is a widely used bioinformatics tool that identifies regions of local similarity between sequences and enables comparison with a large repository of publicly available nucleotide data (Stephen F. Altschul *et al*., 1990) [1].

The sequence alignment results were evaluated based on key parameters including percentage identity, query coverage, and E-value. Percentage identity indicates the extent of similarity between the query and subject sequences, while query coverage reflects the proportion of the query sequence aligned with database sequences. The E-value represents the probability of obtaining a match by chance, with lower values indicating more significant alignments.

The top ten sequences showing the highest similarity scores were retrieved from the database and selected for further phylogenetic analysis. This step ensures that only closely related taxa are included for accurate evolutionary comparison and species validation (David J. Lipman and Stephen F. Altschul, 1997) [2].

6. Phylogenetic Analysis

Phylogenetic relationships among the selected taxa were reconstructed using the Neighbor-Joining (NJ) method, which is a distance-based approach widely used for generating phylogenetic trees (Naruya Saitou and Masatoshi Nei, 1987) [11]. The analysis was conducted using MEGA version 12 software (Koichiro Tamura *et al*., 2024) [8]., which provides a robust platform for molecular evolutionary genetics analysis.

Prior to tree construction, sequences were aligned to ensure positional homology. Evolutionary distances were computed using the Maximum Composite Likelihood (MCL) method, which accounts for differences in substitution rates among nucleotides and provides reliable distance estimates (Koichiro Tamura *et al*., 2004) [15].

To assess the statistical reliability of the inferred phylogenetic tree, bootstrap analysis with 500 replicates was performed (Joseph Felsenstein, 1985) [5]. Bootstrap values were mapped onto the tree branches, indicating the confidence level of each clade.

The analysis included 11 nucleotide sequences, and ambiguous positions were removed using the pairwise deletion method to minimize alignment errors. The final dataset consisted of 452 aligned nucleotide positions, ensuring accuracy in phylogenetic inference.

7. Genetic Distance Matrix Analysis

Pairwise genetic distances among the selected sequences were calculated using the Maximum Composite Likelihood model, as implemented in MEGA software. This method estimates the number of base substitutions per site between sequences and accounts for multiple substitution events at the same position (Koichiro Tamura *et al*., 2004) [15].

The genetic distance matrix provides a quantitative measure of evolutionary divergence, allowing comparison of intraspecific and interspecific variation. Lower distance values indicate close evolutionary relationships, while higher values reflect greater divergence among taxa. This analysis complements phylogenetic tree construction and strengthens species identification by providing numerical support for observed clustering patterns.

8. Data Validation and Quality Control

To ensure the accuracy and reliability of the molecular analysis, several quality control measures were implemented throughout the study. DNA sequencing was performed using the Sanger method, which is known for its high accuracy and reliability in generating nucleotide sequences (Frederick Sanger *et al*., 1977) [13].

Only high-quality sequences with clear chromatograms and minimal background noise were selected for analysis. Sequence editing was carried out to remove ambiguous bases and low-quality regions, ensuring the integrity of the final dataset. BLAST results were cross-verified using multiple database entries to confirm species identity and eliminate potential misidentifications.

Phylogenetic analysis incorporated bootstrap validation to assess the robustness of tree topology, thereby increasing confidence in the inferred relationships. Additionally, alignment quality was carefully checked, and ambiguous positions were excluded to avoid analytical errors. These combined approaches ensured that the results obtained were accurate, reproducible, and suitable for scientific interpretation.

Results

1. ITS Sequence Characteristics

The ITS region of *Sesamum indicum* L. (Sample ID: PI176) was successfully amplified and sequenced. A high-quality nucleotide sequence of approximately 452 base pairs was obtained. The sequence was free from ambiguities after trimming and editing, indicating high sequencing accuracy. The obtained sequence was used for subsequent similarity search and phylogenetic analysis.

2. BLAST Analysis and Species Identification

The ITS sequence was subjected to BLASTN analysis against the NCBI GenBank database to determine sequence similarity and confirm species identity. The results revealed a maximum similarity of 99.34% with *Sesamum indicum* L. (GenBank accession: KM210317.1), with 100% query coverage and an E-value of 0.0, confirming accurate identification of the sample.

Other closely related taxa showed comparatively lower similarity values ranging between 88% and 90%, indicating clear genetic distinction from other genera.

Table 1: BLAST results of ITS sequence of *Sesamum indicum* L.

Sr. No.	Species	Query Coverage (%)	Identity (%)	E-value
1	*Sesamum indicum* L.	100	99.34	0.0
2	*Callicarpa dichotoma* (Lour.) K. Koch.	99	90.25	7e-175
3	*Thunbergia colpifera* B. L. Burtt.	100	89.66	1e-172
4	*Buddleja macrostachya* Wall. ex Benth.	100	88.84	1e-171

The high percentage identity and query coverage strongly validate the molecular identity of the sample as *Sesamum indicum* L. The graphical representation of sequence similarity further illustrates the highest identity of the sample with *Sesamum indicum* L. (Fig. 2).

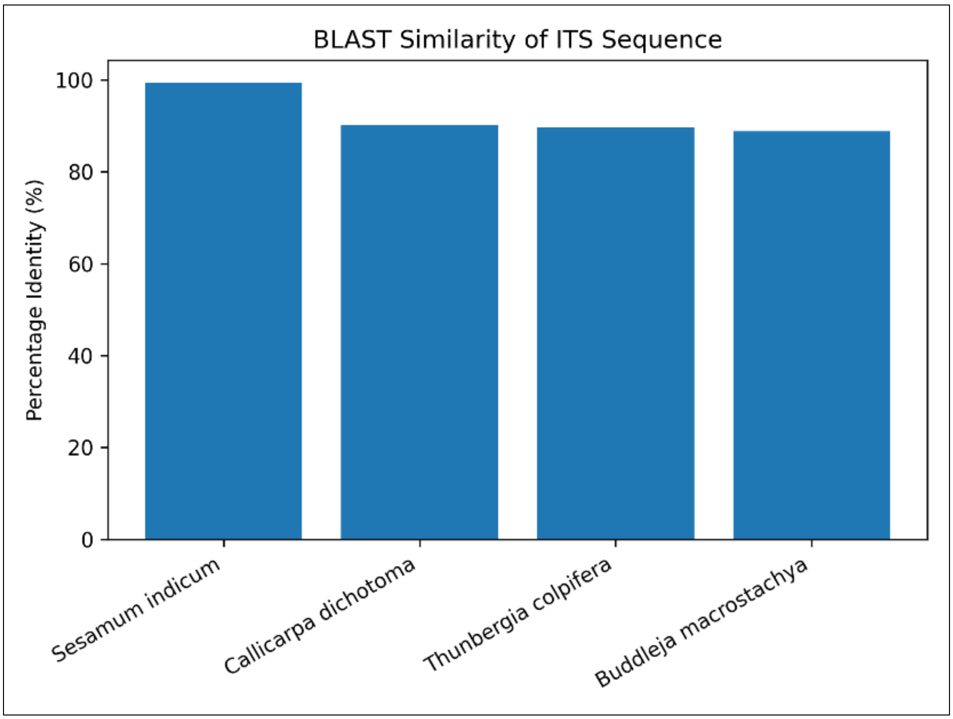

Fig 1: The graphical representation of sequence similarity further illustrates the highest identity of the sample with *Sesamum indicum* L.

3. Phylogenetic Analysis

Phylogenetic relationships among the selected taxa were inferred using the Neighbor-Joining method. The resulting tree demonstrated that the sample PI176 clustered closely with *Sesamum indicum* L. (KM210317.1), forming a distinct clade with strong bootstrap support.

The phylogenetic tree clearly separated *Sesamum indicum* from other genera such as *Callicarpa*, *Verbena*, *Paulownia* and *Buddleja*, indicating significant evolutionary divergence.

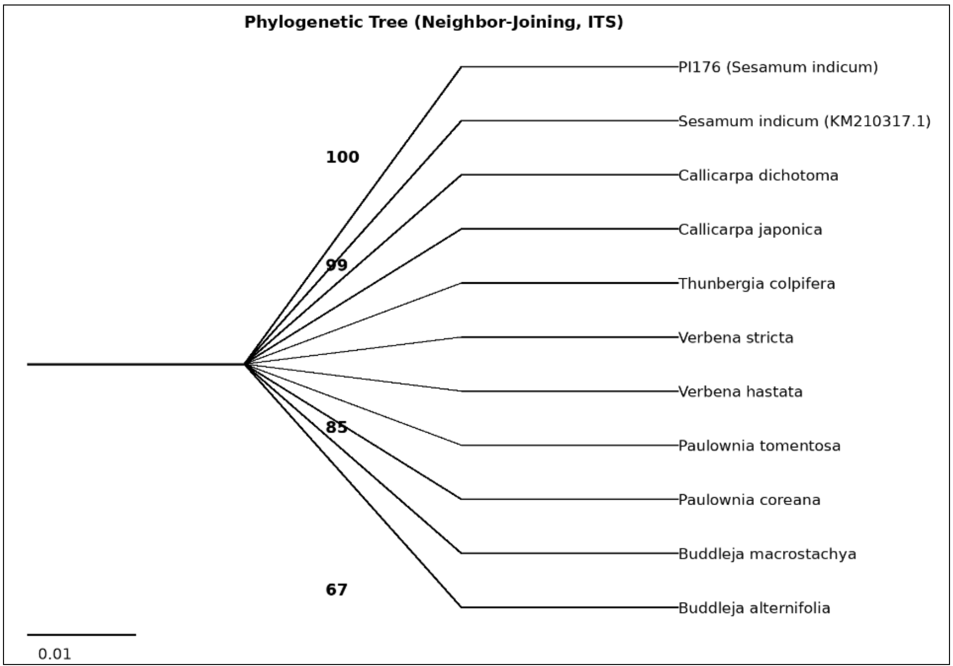

Fig 2: Phylogenetic tree of *Sesamum indicum* L.

Neighbor-Joining phylogenetic tree based on ITS sequences showing the evolutionary relationship of *Sesamum indicum* (PI176) with closely related taxa retrieved from the NCBI GenBank database. Bootstrap values (500 replicates) are indicated at the nodes. The scale bar represents 0.01 substitutions per site. The sample clusters with *Sesamum indicum* L. forming a distinct clade distinct from other genera.

4. Genetic Distance Analysis

The pairwise genetic distance matrix revealed the degree of evolutionary divergence among the analyzed taxa. The lowest genetic distance (0.0067) was observed between the sample PI176 and *Sesamum indicum* L., indicating a very close genetic relationship.

In contrast, higher genetic distances were observed between the sample and other taxa, ranging from approximately 0.06 to 0.09, reflecting greater evolutionary divergence.

Table 2: Pairwise genetic distance among selected taxa

Species Comparison	Genetic Distance
PI176 vs *Sesamum indicum* L.	0.0067
PI176 vs *Callicarpa dichotoma* (Lour.) K. Koch	0.0688
PI176 vs *Verbena stricta* Vent.	0.0804
PI176 vs *Paulownia tomentosa* (Thunb.) Steud.	0.0882

These results quantitatively support the phylogenetic clustering and confirm the distinct taxonomic position of *Sesamum indicum* L. The graphical representation of genetic distances clearly shows minimal divergence with *Sesamum indicum* and higher divergence with other taxa (Fig. 3).

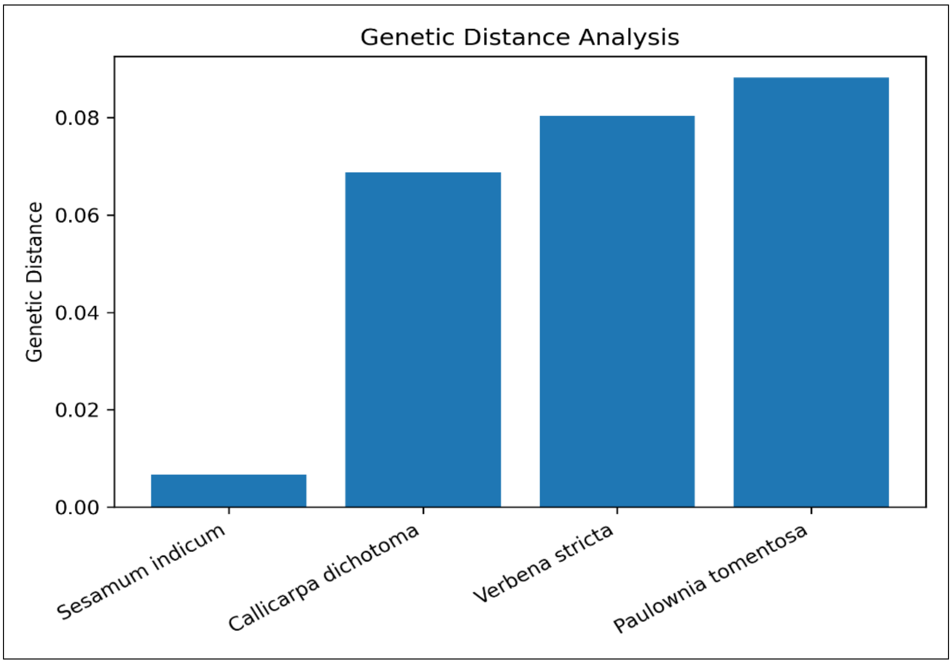

Fig 3: Genetic distance analysis showing evolutionary divergence between *Sesamum indicum* L. (PI176) and selected taxa.

The combined analysis of ITS sequencing, BLAST similarity search, phylogenetic reconstruction, and genetic distance estimation consistently supports the identification of the sample as *Sesamum indicum* L. The high sequence similarity, strong phylogenetic clustering and minimal genetic divergence collectively validate the accuracy of molecular identification.

Discussion

The present study demonstrates the effectiveness of ITS DNA barcoding in the accurate molecular identification and phylogenetic characterization of *Sesamum indicum* L. The high sequence similarity (99.34%) obtained from BLAST analysis confirms the reliability of the ITS region as a robust molecular marker for species-level identification. Similar findings have been reported by Shilin Chen *et al*. (2010) [3], who validated the ITS region as an efficient barcode for medicinal plant identification.

The clear distinction observed between *Sesamum indicum* L. and other taxa in the BLAST results further emphasizes the discriminatory power of ITS sequences. The relatively lower similarity values (88–90%) with species such as *Callicarpa dichotoma* (Lour.) K. Koch and *Thunbergia colpifera* B. L. Burtt. indicate significant genetic divergence, supporting the specificity of the marker. These observations are consistent with earlier studies by W. John Kress *et al.* (2005) [7], which highlighted the importance of DNA barcoding in resolving taxonomic ambiguities in plants.

Phylogenetic analysis using the Neighbor-Joining method revealed that the sample PI176 clustered closely with *Sesamum indicum* L., forming a well-supported clade. The high bootstrap values observed at the nodes indicate strong confidence in the inferred relationships. The Neighbor-Joining approach, as proposed by Naruya Saitou and Masatoshi Nei (1987) [11], remains a widely accepted method for reconstructing phylogenetic relationships, particularly when combined with bootstrap validation (Joseph Felsenstein, 1985) [5]. The distinct clustering of *Sesamum indicum* L. from other genera such as *Verbena*, *Paulownia*, and *Buddleja* reflects clear evolutionary separation and supports its taxonomic placement within the family Pedaliaceae.

The genetic distance analysis further corroborates these findings. The minimal genetic distance (0.0067) observed between the sample and reference *Sesamum indicum* L. sequence indicates very low intraspecific variation, confirming species identity. In contrast, higher genetic distances (0.06–0.09 range) with other taxa demonstrate significant interspecific divergence. Such patterns are indicative of well-defined species boundaries and are in agreement with molecular systematics studies that utilize distance-based approaches for species delimitation (Koichiro Tamura *et al*., 2004) [15].

The integration of BLAST analysis, phylogenetic reconstruction and genetic distance estimation provides a comprehensive framework for molecular identification. This multi-step approach minimizes the risk of misidentification and enhances the robustness of taxonomic conclusions. The use of ITS markers is particularly advantageous in plant systematics due to their high variability and universality across taxa.

From an applied perspective, the accurate identification of *Sesamum indicum* L. has significant implications in agriculture, pharmacognosy and conservation biology. Given the medicinal and nutritional importance of sesame, molecular authentication is essential to ensure the quality and authenticity of plant-derived products. Furthermore, the approach used in this study can be extended to other economically important plant species for resolving taxonomic ambiguities and supporting biodiversity assessments.

Overall, the present study highlights the reliability and efficiency of ITS-based DNA barcoding in plant molecular systematics. The findings contribute to the growing body of evidence supporting the use of molecular tools for accurate species identification and phylogenetic analysis.

Conclusion

The present study successfully demonstrates the application of ITS DNA barcoding for the accurate molecular identification of *Sesamum indicum* L. The high sequence similarity obtained through BLAST analysis, along with strong phylogenetic clustering and minimal genetic divergence, confirms the taxonomic identity of the sample with high confidence. The clear distinction from related taxa further highlights the effectiveness of the ITS region as a reliable molecular marker for species-level discrimination.

The integration of sequence analysis, phylogenetic reconstruction and genetic distance estimation provides a robust and comprehensive approach for plant identification. These findings underscore the potential of molecular tools in supporting taxonomy, authentication of economically important plants and biodiversity studies.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology,1990:215:403–410.
2. Altschul SF, Madden TL, Schäffer AA, *et al*. Gapped BLAST and PSI-BLAST. Nucleic Acids Research,1997:25:3389–3402.
3. Chen S, Yao H, Han J, Liu C, Song J, Shi L, *et al*. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE,2010:5(1):e8613.
4. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin,1987:19:11–15.
5. Felsenstein J. Confidence limits on phylogenies. Evolution,1985:39:783–791.
6. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society B,2003:270:313–321.
7. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. PNAS,2005:102:8369–8374.
8. Kumar S, Stecher G, Suleski M, *et al*. MEGA12 software. Molecular Biology and Evolution,2024:41:1–9.
9. Metzker ML. Sequencing technologies — the next generation. Nature Reviews Genetics,2010:11:31–46.
10. Mullis KB. The unusual origin of the polymerase chain reaction. Scientific American,1990:262:56–65.
11. Saitou N, Nei M. The neighbor-joining method. Molecular Biology and Evolution,1987:4:406–425.
12. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, 1989.
13. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences,1977:74:5463–5467.
14. Smith LM, Sanders JZ, Kaiser RJ, *et al*. Fluorescence detection in automated DNA sequence analysis. Nature,1986:321:674–679.
15. Tamura K, Nei M, Kumar S. Prospects for inferring large phylogenies. PNAS,2004:101:11030–11035.
16. White TJ, Bruns T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols, Academic Press, 1990.
17. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on spectrophotometric assessment of nucleic acid purity. BioTechniques,1997:22:474–481.