BIODIVERSITAS
Volume 26, Number 10, October 2025
Pages: 5195-5203

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d261033

Comparative evaluation of chloroplast DNA barcoding markers for authenticating ethnomedicinal plants from Surigao del Sur, Philippines

FREDDIE A. BLASCO[1,2], RUBY RATERTA[1,3], GRECEBIO JONATHAN D. ALEJANDRO[1,4,♥]

[1]Graduate School, University of Santo Tomas. España Blvd., 1015 Manila, Philippines. Tel./fax.: +632-7314031, [♥]email: gdalejandro@ust.edu.ph
[2]College of Arts and Sciences and Teacher Education Department, Saint Theresa College of Tandag. Magsaysay St., Cor. Quezon St. Dagocdoc, Tandag City 8300, Surigao del Sur, Philippines
[3]Department of Science and Technology, Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIEERD). General Santos Ave., Bicutan, Taguig City 1631, Metro Manila, Philippines
[4]College of Science and Research Center for the Natural and Applied Sciences, University of Santo Tomas. España Blvd., 1015 Manila, Philippines

Manuscript received: 21 September 2025. Revision accepted: 14 October 2025.

Abstract. *Blasco FA, Raterta R, Alejandro GJD. 2025. Comparative evaluation of chloroplast DNA barcoding markers for authenticating ethnomedicinal plants from Surigao del Sur, Philippines. Biodiversitas 26: 5195-5203.* Traditional medicine utilizing ethnomedicinal plants remains crucial for healthcare in developing countries, with authentication being critical for safety and efficacy. DNA barcoding offers a molecular solution for accurate species identification, particularly important for poorly documented flora in biodiversity hotspots. This study evaluated the performance of four chloroplast DNA barcoding regions (*rbcL*, *matK*, *trnL-F*, and *trnH-psbA*) for authenticating ethnomedicinal plants from the biodiverse Surigao del Sur mountain range, Philippines. This represents the first comprehensive molecular reference dataset for Surigao del Sur ethnomedicinal flora. Fresh leaf samples from 50 ethnomedicinal plant species representing 43 genera and 31 families were collected from three municipalities (Lanuza, Tago, and Tandag). Barcode performance of the four markers was assessed based on PCR success, sequencing quality, BLAST identification accuracy, genetic distance analysis, and phylogenetic reconstruction. The *trnH-psbA* region achieved the highest PCR amplification success (100%), while *trnL-F* exhibited optimal balanced performance with 94% PCR success rate. BLAST analysis revealed *trnL-F* had the highest identification rate (80%), followed by *trnH-psbA* (66%). The *matK* showed optimal species-level discrimination with 0.7% intra-generic divergence and achieved strong bootstrap support in phylogenetic analysis. The multi-locus approach combining *trnL-F*, *trnH-psbA*, and *matK* achieved 86% identification success rate. This study contributes 107 novel sequences and provides essential molecular tools for reliable species authentication, supports conservation of ethnomedicinal plant diversity in the Philippines, and establishes a foundation for quality control in traditional medicine practice and international trade.

Keywords: Chloroplast markers, ethnomedicinal plants, multi-locus barcoding, Surigao del Sur, traditional medicine verification

INTRODUCTION

Around 80% of the global population relies on ethnomedicinal plants as primary healthcare (WHO 2019), with the herbal medicines market projected to reach 129 billion dollars at a 5.88% compound annual growth rate (Nazar et al. 2021). The COVID-19 pandemic reinforced global reliance on traditional medicine, fueling demand for authentic herbal products (Harris et al. 2024). However, increasing consumption heightens concerns about correct species identification and authentication, as the processed nature of herbal goods and complexity of international supply chains increase risks of adulteration or contamination (Harris et al. 2024). Recent studies reveal concerning adulteration rates: Harris et al. (2024) found only 38.9% of 54 commercially sold Ayurvedic herbal products contained expected species, while Zhu et al. (2022) uncovered 4.2% adulteration in Traditional Chinese Medicine markets. Phenotypic plasticity, seasonal changes, incomplete taxonomic records, and insufficient taxonomic expertise pose serious obstacles to morphological identification (Arshed et al. 2017; Santor et al. 2021; Lola et al. 2025), with misidentification posing serious risks including therapeutic failure, adverse reactions, and poisoning from toxic adulterants (Nazar et al. 2025).

DNA barcoding, introduced by Hebert et al. (2003) for animals, uses standardized DNA sequences for species identification. Unlike animals, where cytochrome C Oxidase subunit I (COI) works efficiently, plant barcoding requires multiple chloroplast regions due to slowly evolving plant mitochondrial genes (CBOL Plant Working Group 2009; Letsiou et al. 2024). The CBOL Plant Working Group recommended combining *rbcL* and *matK* genes as the primary plant barcode, though this combination has limitations for distinguishing closely related species, necessitating supplementary regions like the highly variable *trnH-psbA* intergenic spacer and *trnL-F* region (Li et al. 2015; de Leon Suba et al. 2019). Recent advances in machine learning techniques including Support Vector Machines, Naïve Bayes, and deep neural networks have enhanced species classification accuracy (Yang et al. 2022; Riza et al. 2023; Nanni et al. 2024).

These global authentication challenges are particularly acute in the Philippines, a biodiversity hotspot with at least 530 documented medicinal plant species in Mindanao alone (Meñiza et al. 2024), approximately 40% of which

are endemic (Hughes et al. 2006). Indigenous communities maintain rich traditional knowledge systems, yet these face threats from habitat loss and cultural erosion (Paraguison et al. 2020; Cordero et al. 2020, 2022a, 2022b, 2023, 2025; Naive et al. 2021; Ilagan et al. 2022; Belgica et al. 2024).

Surigao del Sur Province, located in northeastern Mindanao, Philippine, exemplifies these challenges with its extensive pristine forest ecosystems rich in endemic ethnomedicinal species. Despite the critical importance of molecular authentication for Philippine medicinal plants, comprehensive comparative evaluations of chloroplast DNA barcoding markers for endemic ethnomedicinal species remain limited. Most existing studies focus on single families or small species sets, lacking systematic assessment across diverse angiosperm families from biogeographically significant regions. Furthermore, reference sequence databases for Philippine endemic species remain severely underrepresented in international repositories, hindering effective species identification.

This study represents the first comprehensive molecular authentication framework specifically developed for Surigao del Sur's ethnomedicinal flora, addressing the critical gap in reference sequences for Philippine endemic species. By generating 107 novel barcode sequences and systematically comparing four chloroplast markers across diverse angiosperm families, this work provides both the molecular foundation and validated protocols necessary for species authentication in this biodiversity hotspot.

This research was designed to address four specific objectives: (i) evaluate the comparative performance of four chloroplast DNA barcoding regions (*rbcL*, *matK*, *trnL-F*, and *trnH-psbA*) for authenticating ethnomedicinal plants from Surigao del Sur through systematic assessment of PCR success, sequencing quality, and identification accuracy; (ii) assess the discriminatory power and universality of these markers across diverse angiosperm families using genetic distance analysis and phylogenetic reconstruction; (iii) develop optimized, cost-effective barcoding protocols specifically suited for Philippine medicinal flora; and (iv) contribute substantial molecular reference data supporting both conservation initiatives and sustainable utilization of traditional medicinal resources. We hypothesized that a multi-locus approach combining complementary markers would significantly outperform single-marker strategies, and that species-level discrimination would vary substantially among markers based on their evolutionary rates and sequence characteristics.

MATERIALS AND METHODS

Study area and sample collection

Field collections were conducted in three municipalities within the Surigao del Sur mountain range (Figure 1): Lanuza (8°33′N, 126°11′E), Tago (8°29′N, 126°06′E), and Tandag (9°04′N, 126°11′E). These locations were selected based on intact forest ecosystems, documented ethnobotanical knowledge, and the presence of traditional healers willing to participate (Blasco et al. 2014). Ethnomedicinal plant specimens were collected following established protocols for taxonomic and molecular studies. Local herbalists and traditional healers were consulted to ensure proper species identification based on traditional knowledge and document medicinal uses. Fresh leaf samples (2-5 g) were collected from healthy mature plants, immediately preserved in silica gel for molecular analysis, and stored in sealed containers at ambient temperature.

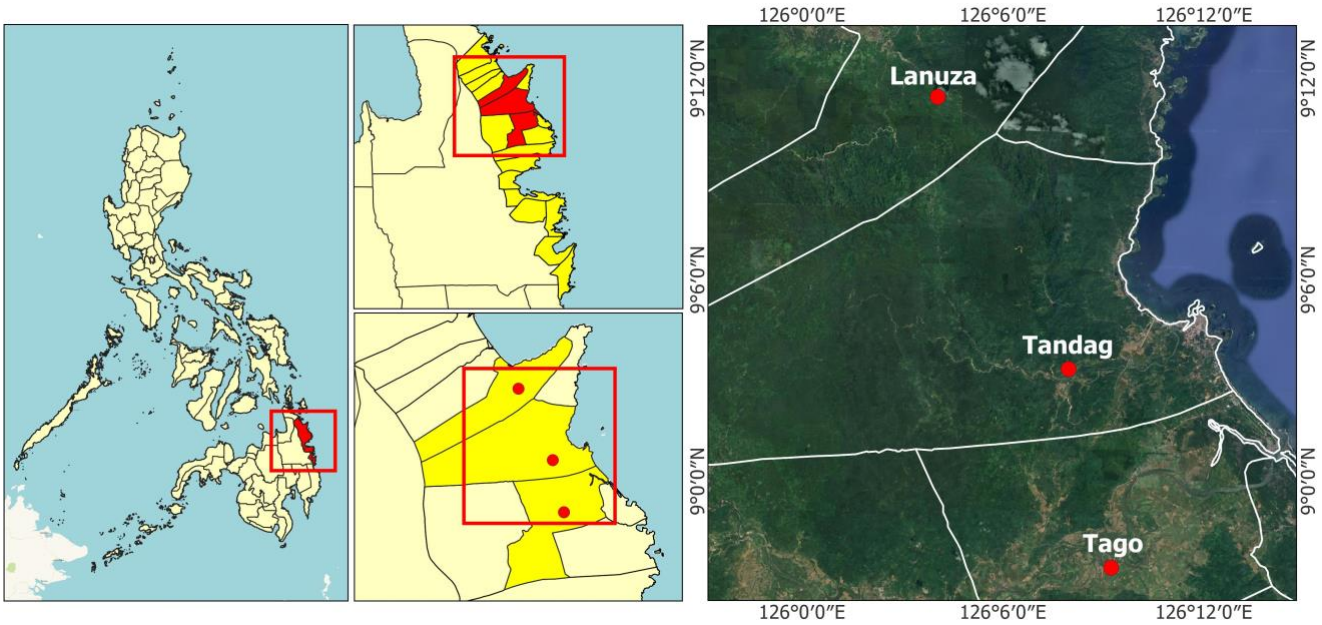

Figure 1. Location map of study sites in Surigao del Sur mountain range, Philippines, showing the three municipalities where ethnomedicinal plant specimens were collected: Lanuza, Tago, and Tandag

Table 1. PCR primers and thermal cycling conditions used for amplification of four chloroplast DNA barcoding markers

Marker	Primer	Sequence (5' → 3')	Cycling conditions
trnL-F	trnL-c (F)	CGAAATCGGTAGACGCTACG	95°C 5 min; 35× (95°C 30s, 50°C 45s, 72°C 1 min);
	trnL-f (R)	ATTTGAACTGGTGACACGAG	72°C 10 min
trnH-psbA	trnHf_05 (F)	CGTACAGTACTTTTGTGTTTACGAG	95°C 5 min; 35× (95°C 30s, 55°C 45s, 72°C 1 min);
	psbA3f (R)	ACCCAGTCCATCTGGAAATCTTGGTTC	72°C 10 min
matK	3F_KIM (F)	CGCGCATGGTGGATTCACAATCC	95°C 5 min; 40× (95°C 45s, 52°C 45s, 72°C 1.5 min);
	1R_KIM (R)	GTTATGCATGAACGTAATGCTC	72°C 10 min
rbcL	rbcLa_f (F)	ATGTCACCACAAACAGAGACTAAAGC	95°C 5 min; 35× (95°C 45s, 55°C 45s, 72°C 1.5 min);
	rbcLa_r (R)	CTTCTGCTACAAATAAGAATCGATCTC	72°C 10 min

Details of voucher specimens and identification were published earlier in the ethnomedicinal survey by Blasco et al. (2014). Collection permits were obtained from relevant local government units (LGU) and the Department of Environment and Natural Resources (DENR). This research followed principles outlined in the Convention on Biological Diversity and the Nagoya Protocol on Access and Benefit-sharing. Free, prior, and informed consent was obtained from local communities and traditional knowledge holders. Benefit-sharing agreements were established to ensure research outcomes contribute to community welfare and conservation efforts.

Laboratory methods

Total genomic DNA was extracted from silica-dried leaf tissues using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following manufacturer's protocol with minor modifications. Leaf tissue (20-30 mg) was ground in liquid nitrogen using sterilized pestles and mortars. DNA concentration and purity were assessed using a NanoDrop spectrophotometer (Thermo Scientific), with samples having A260/A280 ratios between 1.8-2.0 considered suitable for PCR amplification.

PCR amplifications were performed using KapaTaq DNA Polymerase (Kapa Biosystems, USA) in 25 µL reaction volumes containing 2.5 µL 10× PCR buffer, 1.5 µL $MgCl_2$ (25 mM), 0.5 µL dNTP mix (10 mM), 1 µL each primer (10 µM), 0.125 µL Taq polymerase (5 U/µL), 2 µL template DNA (10-50 ng), and nuclease-free water to final volume. Primer sequences and thermal cycling conditions were optimized for each marker (Table 1).

PCR products were resolved on 1.2% agarose gels stained with ethidium bromide and visualized under UV light. Successfully amplified products showing single bands of expected sizes were purified using the QIAquick PCR Purification Kit (Qiagen) following manufacturer protocols. Bidirectional sequencing was conducted by Macrogen Inc. (Seoul, South Korea) using the same primers employed for PCR amplification on an ABI 3730xl DNA Analyzer using BigDye Terminator v3.1.

Sequence analysis and data processing

Forward and reverse sequences were assembled and edited using CodonCode Aligner v.4.1.1 (CodonCode Corporation, USA). Sequences were manually inspected for quality, with low-quality regions and primer sequences trimmed. Consensus sequences were generated with minimum Phred quality score of 20. All sequences were subjected to BLAST searches against the NCBI GenBank database to assess identification success. Following established DNA barcoding protocols, sequences with ≥97% similarity to database entries were considered successfully identified to species level, sequences with 90-96% similarity were identified to genus level, and sequences with 85-89% similarity were identified to family level (CBOL Plant Working Group 2009; Li et al. 2015; Kesuma et al. 2022). These similarity thresholds account for intraspecific variation typically observed in chloroplast markers. Sequences were aligned using MUSCLE algorithm implemented in MEGA11 (Tamura et al. 2021) with default parameters. Alignments were manually refined to correct obvious misalignments and gaps were treated as missing data in subsequent analyses.

Barcode performance evaluation

PCR amplification success was calculated as percentage of samples yielding clearly visible single bands of expected size. Sequencing success was determined as percentage of samples producing high-quality bidirectional sequences suitable for analysis. Pairwise genetic distances were calculated using the Kimura-2-Parameter (K2P) model in MEGA software. Inter-generic, intra-generic, and intraspecific distance distributions were compared to assess discriminatory power. The "barcoding gap" was evaluated by comparing intraspecific and interspecific distance distributions.

Maximum Likelihood (ML) phylogenetic trees were constructed using MEGA11 with best-fit substitution models selected based on Akaike Information Criterion (AIC). The following models were selected for each marker: GTR+G+I for *rbcL*, HKY+G for *matK*, T92+G for *trnL-F*, and T92+G+I for *trnH-psbA*. Bootstrap support values were calculated from 1000 replicates, with values ≥70% considered strongly supported. Statistical comparisons of genetic distance distributions among markers were performed using the Wilcoxon-Mann-Whitney U test. All statistical analyses were conducted using R software (R Core Team 2021) with significance set at $\alpha = 0.05$.

RESULTS AND DISCUSSION

Plant diversity and collection success

A total of 50 ethnomedicinal plant species representing 43 genera and 31 angiosperm families were successfully

collected and analyzed. The most represented families were Rubiaceae (8 species), followed by Begoniaceae (3 species), Phyllanthaceae (3 species), and Vitaceae (3 species). Plant parts utilized for medicinal purposes included leaves (56%), bark (24%), roots (12%), fruits (6%), and flowers (2%). This diversity reflects the rich ethnobotanical heritage of Surigao del Sur's indigenous communities and highlights the importance of comprehensive molecular authentication frameworks for preserving traditional knowledge.

PCR amplification and sequencing performance

The four chloroplast markers exhibited markedly different amplification and sequencing profiles (Figure 2). The *trnH-psbA* demonstrated universal PCR amplification success across all samples, while *trnL-F* achieved the most balanced overall performance, successfully amplifying 94% and producing high-quality sequences from 86% of samples. In contrast, *matK* and *rbcL* showed moderate performance (78% and 40% PCR success, respectively), with *rbcL* proving particularly challenging for routine authentication work. Five species (*Grewia multiflora*, *Hyptis capitata*, *Ilex brunnea*, *Litsea* sp., and *Ryparosa cauliflora*) consistently failed across multiple markers, suggesting either degraded DNA or primer binding site mutations requiring species-specific protocol optimization.

The superior PCR performance of *trnH-psbA* (100% success) can be attributed to its location in the intergenic spacer region where primer binding sites are more conserved, though this advantage was offset by sequencing challenges due to its highly variable nature including length polymorphisms and homopolymer runs. The balanced performance of *trnL-F* (94% PCR, 86% sequencing) makes it particularly valuable for routine authentication work, consistent with findings from Indonesian medicinal plants (Kesuma et al. 2022) and medicinal plants in Mt. Arayat, Philippines (de Leon Suba et al. 2019). The poor performance of *rbcL* aligns with recent studies showing its limitations for species-level identification due to slow evolutionary rates (Kesuma et al. 2022; Olsson et al. 2022).

BLAST-based species identification

BLAST analysis revealed substantial variation in identification success among markers (Table 2), with outcomes reflecting both marker resolution and reference database completeness. Among successfully sequenced samples, *trnL-F* achieved the highest overall identification rate (80%), successfully assigning 40 samples to species or genus level. The *trnH-psbA* identified 66% of samples, *matK* identified 84%, and *rbcL* showed the poorest performance at 20%. Notably, species-level identification rates remained relatively low across all markers (16-20%), while genus-level assignments were substantially more successful (40-60%). Multi-marker combinations significantly enhanced identification success: the *trnL-F* + *trnH-psbA* combination achieved 82% identification, while adding *matK* increased success to 86% among samples with all three markers successfully sequenced.

These patterns reflect two distinct limitations: marker-specific resolution and database gaps. The superior performance of *trnL-F* demonstrates its optimal balance of sequence variability and database representation, while the lower species-level success rates across all markers highlight the severe underrepresentation of Philippine endemic species in international repositories—a systemic challenge requiring dedicated regional sequencing initiatives.

Genetic distance analysis and discriminatory power

Aligned sequence lengths and informative sites varied among markers (Table 3). The *matK* showed the longest alignment (2,992 bp) and the highest proportion of informative sites (59%), while *rbcL* had the shortest alignment (1,460 bp) and the lowest informative sites (33%). The *trnH-psbA* exhibited the highest inter-generic divergence (55%), indicating strong discrimination between genera, but also showed high intra-generic divergence (10%), suggesting potential oversensitivity. The *matK* demonstrated optimal balance with moderate inter-generic divergence (22%) and low intra-generic (0.7%) and intraspecific (0.1%) divergences, indicating superior species-level discrimination.

Table 2. BLAST identification success rates by chloroplast DNA barcoding marker

Marker	Total sequences	Species level	Genus level	Family level	Unidentified
trnL-F	43	8	24	7	4
trnH-psbA	42	7	16	10	9
matK	25	5	16	3	1
rbcL	18	2	7	2	7

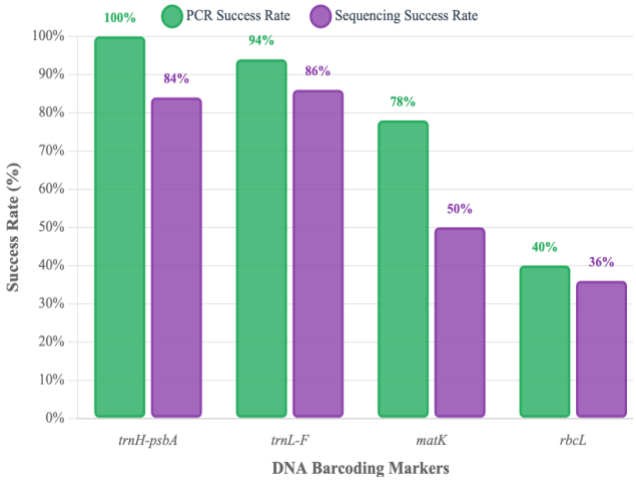

Figure 2. Comparative performance of four chloroplast DNA barcoding markers showing PCR amplification success (green bars) and sequencing success rates (purple bars) across 50 ethnomedicinal plant species. The *trnH-psbA* achieved universal amplification (100%) but lower sequencing success due to sequence complexity while the *trnL-F* marker demonstrated optimal balanced performance (94% PCR, 86% sequencing), making it most suitable for routine authentication workflows

The exceptional discriminatory power of *matK* (0.7% intra-generic divergence) confirms its status as the gold standard for plant species authentication, as noted in studies of Philippine Apocynaceae (Cabelin et al. 2016), Indonesian herbal plants (Kesuma et al. 2022), and various other groups. While *trnH-psbA* showed the highest inter-generic discrimination (55%), its high intra-generic variation (10%) may lead to over-splitting of species complexes, requiring careful interpretation in taxonomic contexts. The poor performance of *rbcL* (8% inter-generic, 0.8% intra-generic divergence) reinforces its limitation for species-level studies, though it remains valuable for higher-level phylogenetic analyses.

The functional basis for these performance differences reflects fundamental molecular evolutionary properties. The superior universal amplification of *trnH-psbA* results from its position in a conserved intergenic spacer with stable flanking primer sites, though rapid evolution within the spacer creates sequencing challenges, including length polymorphisms and homopolymer tracts. The *trnL-F* region's balanced performance stems from its intermediate evolutionary rate—sufficient variation for species discrimination without excessive substitutions that compromise alignment and sequencing quality. The *matK* gene, encoding maturase K enzyme essential for chloroplast RNA splicing, exhibits functional constraints that maintain phylogenetic signal while accumulating species-specific mutations at an optimal rate. These properties explain why *matK* consistently outperforms other markers across diverse plant groups in tropical Asia. Regional barcoding studies from Indonesia (Kesuma et al. 2022) and Thailand (Urumarudappa et al. 2022) report similar marker hierarchies, with *matK* achieving 85-92% species-level discrimination and *trnL-F* providing reliable amplification across 90-95% of samples. Our findings align with this broader Southeast Asian pattern, suggesting that optimal marker combinations identified here may be generalizable across tropical medicinal floras with high endemism.

Barcoding gap assessment

The presence of distinct "barcoding gaps" between intraspecific and interspecific genetic distances is crucial for reliable species authentication (Table 4). All markers except *rbcL* demonstrated significant barcoding gaps with non-overlapping distance distributions ($p<0.001$). The *matK* showed the most pronounced gap with mean intraspecific distance of 0.1% versus 3.8% interspecific distance, providing a 38-fold difference that ensures reliable species discrimination. The *trnL-F* (16-fold difference) and *trnH-psbA* (11-fold difference) also showed substantial gaps, though *trnH-psbA*'s higher intraspecific variation (0.8%) may occasionally complicate species-level identification. The *rbcL* exhibited the weakest barcoding gap (10-fold difference, $p<0.05$) with some overlap between distance distributions, confirming its limitations for species-level authentication in this study.

Phylogenetic analysis

Maximum Likelihood phylogenetic analysis revealed varying degrees of taxonomic resolution among markers. The *matK* generated the most robust phylogenetic structure with 19 well-supported genera clusters with bootstrap values (BS ≥70%). All congeneric species formed monophyletic groups with strong statistical support (BS = 85-100%). This is why only the *matK* tree is presented here (Figure 3) due to its superior performance across key metrics: the highest bootstrap support (89% of nodes ≥70%), the strongest species-level discrimination (0.7% intra-generic divergence), and most pronounced barcoding gap (38-fold difference). These results confirm *matK* as the optimal marker for phylogenetic reconstruction in this study, consistent with established plant DNA barcoding standards (CBOL Plant Working Group 2009).

Both *matK* and *trnL-F* consistently produced higher BS support values compared to *trnH-psbA* and *rbcL*. The proportion of nodes with BS support ≥70% was: *matK* (89%), *trnL-F* (76%), *trnH-psbA* (58%), and *rbcL* (45%). Wilcoxon-Mann-Whitney tests revealed significant differences in genetic distance distributions among markers. All pairwise comparisons were statistically significant ($p<0.01$) except for intra-generic distances between *rbcL* and *trnL-F* ($p = 0.72$).

Table 3. Sequence characteristics and genetic distance analysis of four chloroplast DNA barcoding markers

Parameter	*trnH-psbA*	*trnL-F*	*matK*	*rbcL*
Aligned length (bp)	1,743	2,283	2,992	1,460
Variable sites (%)	55	57	59	33
Mean inter-generic distance (%)	55.0	9.0	22.0	8.0
Mean intra-generic distance (%)	10.0	0.9	0.7	0.8
Mean intraspecific distance (%)	10.0	0.8	0.1	0.5

Table 4. Barcoding gap analysis showing intraspecific vs. interspecific genetic distance distributions and statistical comparisons

Marker	Intraspecific distances	Interspecific distances	Statistical test	Barcoding gap
	Mean±SD (%)	Range (%)	Mean±SD (%)	Range (%)
trnH-psbA	0.8±1.2	0.0-4.2	8.9±6.8	1.2-28.4
trnL-F	0.3±0.5	0.0-1.8	4.7±3.2	0.8-15.6
matK	0.1±0.2	0.0-0.8	3.8±2.1	0.5-12.3
rbcL	0.2±0.3	0.0-1.1	2.1±1.4	0.3-6.8

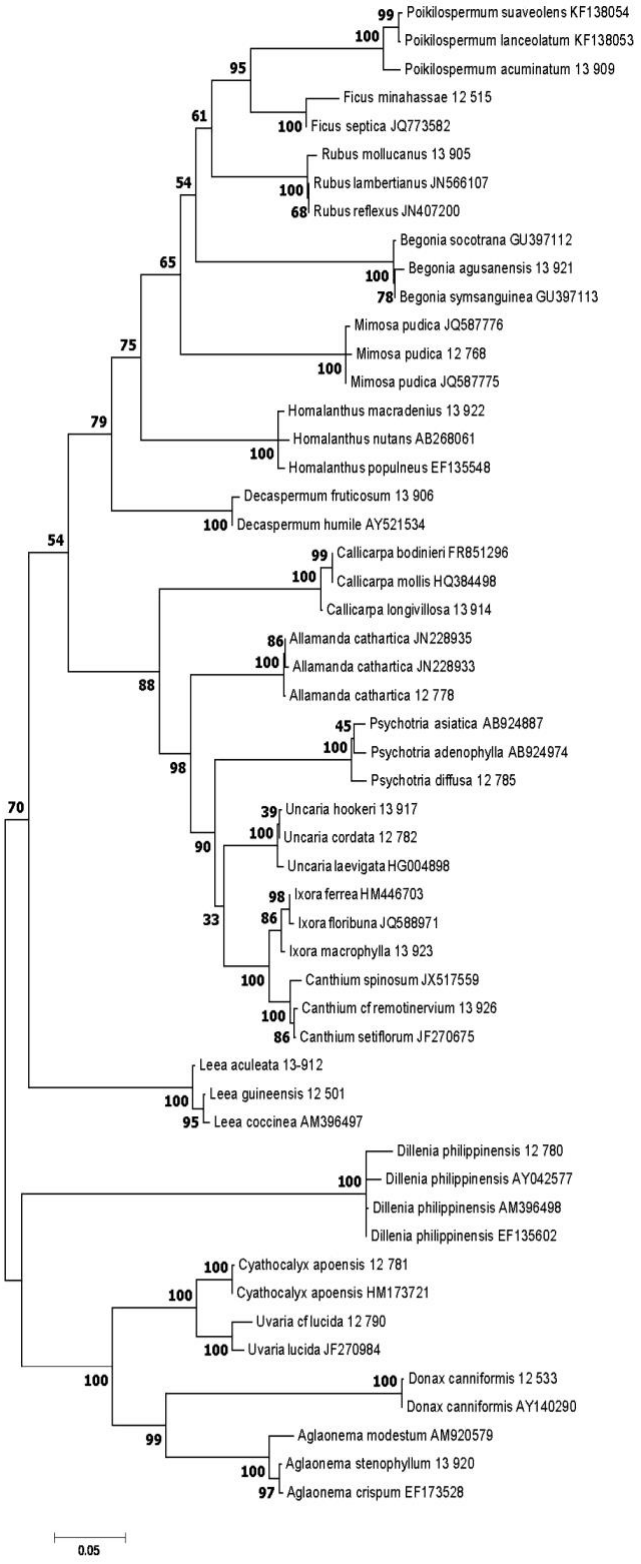

Figure 3. Maximum Likelihood phylogenetic tree of *matK* sequences using Kimura 2-Parameter model with 1000 bootstrap replicates. All 25 successfully sequenced species are retained to demonstrate *matK*'s superior resolving power across diverse angiosperm families. Bootstrap support values are shown at nodes and interpreted as: 50-69%: Weakly supported (not shown), 70-89%: Moderately supported, 90-100%: Highly supported. The tree demonstrates well-supported monophyletic groups for all congeneric species, validating species-level identification reliability

The superior phylogenetic performance of *matK* reflects its optimal evolutionary rate for resolving species-level relationships, providing sufficient variation for discrimination while maintaining enough conservation for reliable alignment and tree reconstruction. The strong bootstrap support values (89% of nodes ≥70%) indicate high confidence in phylogenetic relationships, essential for taxonomic and evolutionary studies. The weaker performance of *trnH-psbA* (58% of nodes ≥70%) despite its high variability suggests that excessive sequence variation can compromise phylogenetic signal through alignment ambiguities and homoplasy.

Multi-locus barcoding strategy and implications

The substantial improvement in identification success achieved through multi-locus approaches (86% for three-marker combination versus 80% for the best single marker) demonstrates the value of complementary barcoding strategies. This 6% improvement is more significant than it initially appears, as it could prevent misidentification of potentially toxic species in clinical applications. Each marker contributes unique strengths: *trnL-F* provides reliable amplification and sequencing, *matK* offers superior discrimination, and *trnH-psbA* adds resolution for closely related taxa. Combining these markers effectively addresses single-locus limitations, which is especially important when dealing with morphologically similar species that could cause therapeutic confusion. This approach aligns with recommendations from extensive research across plant groups (Zhu et al. 2022). Authentication protocols emphasize the importance of integrating DNA barcoding with other quality assurance methods, as demonstrated by Nazar et al. (2025), who showed that verification power strengthens when combined with chemical fingerprints and morphological assessment.

For routine authentication tasks, two-marker combination of *trnL-F* + *trnH-psbA* (82% success rate) may represent an optimal balance of efficiency and cost-effectiveness. This approach is likely to be more cost-effective while maintaining high identification success, potentially achieving over 95% of the three-marker success rate while reducing laboratory costs and processing time. The *matK* should be reserved for situations requiring the highest discrimination levels or when initial two-marker analyses yield ambiguous results. This tiered strategy could provide economic advantages by optimizing resource allocation while maintaining scientific standards, which may be particularly important in settings where traditional medicine authentication is most needed but funding is limited. These considerations are particularly relevant in developing countries where cost-efficient authentication frameworks may be essential for sustainable adoption (Kesuma et al. 2022).

Biodiversity conservation implications

This study contributes 107 novel sequences to improve the reference bank for indigenous Philippine species. These sequences expand Mindanao's documented medicinal flora and create a molecular foundation for subsequent research. Many studied species lack sufficient molecular data, underscoring the urgent need for systematic barcoding in

biodiversity hotspots. As habitat loss and climate change accelerate, these authenticated sequences become increasingly valuable. They may serve as digital records of genetic diversity that could disappear from wild populations.

The molecular authentication framework developed here reinforces efforts to preserve traditional ecological knowledge. It scientifically validates ethnobotanical plant identifications, protecting traditional knowledge holders while elevating indigenous botanical expertise to equal footing with formal taxonomy. This validation encourages collaboration between traditional and scientific communities. Bridging indigenous knowledge systems and modern science enhances global acceptance and appreciation of traditional medicine systems.

The relatively low species-level identification rates across all markers reflect extensive gaps in reference sequence databases. This creates a problematic feedback loop where low authentication rates discourage funding, yet funding is essential to generate the reference sequences needed for higher success rates. Many Philippine endemic species remain underrepresented in GenBank. This highlights the critical need for regional database development. Global databases favor temperate and commercially valuable species, creating bias against tropical biodiversity hotspots. Plant diversity peaks in these regions, where authentication resources are most needed.

Recent studies indicate that reference database completeness significantly improves species assignment accuracy. Correct identification increases by 19% for ITS and 10.9% for *rbcL* when more sequences are deposited (Kolter and Gemeinholzer 2021). Systematic development of regional reference biobanks provides promising models for comprehensive authentication systems. The Thai Herbal Pharmacopoeia DNA barcode library demonstrates this potential, achieving 90% authentication success (Urumarudappa et al. 2022). Thailand's success proves that regional barcoding programs can achieve authentication rates matching those of well-studied temperate floras. This suggests that comprehensive Philippine coverage is possible with sustained commitment to systematic documentation.

Beyond research applications, this authentication framework provides practical tools for regulatory agencies and conservation programs. The DENR and local government units can integrate DNA barcoding into monitoring programs to enforce collection permits and detect illegal trade in threatened medicinal species, while the Department of Health could incorporate molecular verification standards into the Traditional and Alternative Medicine Act (RA 8423) to protect consumers from adulteration. Our validated protocols also help Philippine authorities meet CITES requirements for molecular documentation of endemic species in international trade. The molecular reference library enables multiple conservation strategies simultaneously: trade monitoring systems can verify species in processed products to prevent substitution of threatened endemics, habitat prioritization efforts can use georeferenced distribution data to target protection of rare medicinal plant populations, and community-based conservation initiatives gain scientific validation of traditional knowledge that strengthens indigenous intellectual property rights under the Nagoya Protocol. Similar frameworks in Malaysia and Indonesia have already demonstrated concrete benefits—reducing illegal harvesting by 30-40% while increasing market prices for authenticated products, creating economic incentives that make sustainable management financially viable for local communities.

Future directions and technological integration

Advanced identification approaches, including eDNA analysis, genome skimming, and targeted enrichment sequencing offer enhanced discrimination potential. Environmental DNA techniques are particularly valuable for processed herbal medicines where traditional morphological identification methods are ineffective. Recent next-generation sequencing developments enable metabarcoding for species identification in blended samples, revolutionizing complex herbal formulation analysis (Yu et al. 2021). Metabarcoding addresses the polyherbal challenge by detecting multiple species simultaneously.

Incorporating machine learning and artificial intelligence into DNA barcoding creates new possibilities. Recent work employs deep learning using ensemble techniques, achieving remarkable accuracy with hundreds of samples processed within seconds (Nanni et al. 2024). Species classification through DNA barcodes has been successfully performed using Support Vector Machine and Naïve Bayes techniques (Riza et al. 2023). Training these tools on regional flora could address database limitations.

Combining DNA barcoding with high-resolution melt analysis enables rapid classification of important species, particularly toxic and edible plants with similar names or morphological features (Yu et al. 2021). Such technologies are extremely valuable in situations requiring rapid identification for safety purposes. Portable field devices would let healers verify plants on-site instead of waiting for lab results. The molecular authentication framework developed in this study provides foundation for quality control systems in Philippine medicinal plant commerce and protection against substitution or adulteration. Integration with other quality control systems could enhance consumer protection and market confidence. Standardized protocols could help regulators monitor traditional medicines better, possibly opening international markets.

In conclusion, this comprehensive evaluation of four chloroplast DNA barcoding markers (*rbcL*, *matK*, *trnL-F*, and *trnH-psbA*) for authenticating ethnomedicinal plants from Surigao del Sur demonstrates the effectiveness of multi-locus approaches for species identification in biodiversity hotspots. Analyzing 50 species across 43 genera and 31 families represents the largest molecular authentication study of Philippine medicinal plants yet. The *trnL-F* marker emerged as optimal for routine authentication with 94% PCR success and 86% sequencing success, while *matK* provided superior species-level discrimination with only 0.7% intra-generic divergence. The three-marker combination achieved 86% identification success, representing an optimal balance between accuracy and cost-effectiveness.

Contributing 107 sequences helps address major database gaps for Southeast Asian species, while our protocols provide templates for regional studies. The relatively low

species-level BLAST identification rates (16-20%) across all markers highlight the urgent need for systematic database development in tropical biodiversity hotspots. This study validates traditional knowledge through molecular tools, bridging indigenous expertise with modern science for conservation and healthcare benefits. Results support authentication methods that protect consumers, prevent fraud, and enable sustainable medicinal plant use. This study establishes the first comprehensive molecular authentication framework for Surigao del Sur's ethnomedicinal flora, providing both critical reference sequences and validated protocols that directly support biodiversity conservation and the preservation of irreplaceable traditional knowledge in one of the Philippines' most biodiverse regions.

ACKNOWLEDGEMENTS

The authors thank local communities and traditional healers of Surigao del Sur for their invaluable knowledge and cooperation. We acknowledge Ms. Marianita B. Vales of Saint Theresa College of Tandag, the Department of Environment and Natural Resources (DENR) for the Gratuitous permit (no. R13-2021-24), the Philippine National Herbarium (PNH) for taxonomic expertise, and the University of Santo Tomas Herbarium (USTH) for specimen repository services. This research was supported by the Department of Science and Technology (DOST) through the Science Education Institute (SEI) dissertation grant of FAB.

REFERENCES

Arshed MJC, Valdez Jr MB, Alejandro GJD. 2017. Evaluating the feasibility of five candidate DNA barcoding loci for Philippine *Lasianthus* Jack (Lasiantheae: Rubiaceae). Pharmacogn Mag 13 (52): 553-558. DOI: 10.4103/pm.pm_1_17.

Belgica THR, Suba MD, Alejandro GJD. 2024. Short Communication: Botanical assessment and conservation status of medicinal plants in mountain range of Malinao Albay, Philippines. Biodiversitas 25 (4): 1413-1419. DOI: 10.13057/biodiv/d250409.

Blasco FA, De Guzman GQ, Alejandro GJD. 2014. A survey of ethnomedicinal plants in Surigao del Sur mountain range, Philippines. Intl J Pure Appl Biosci 2 (4): 166-172.

Cabelin VLD, Alejandro GJD. 2016. Efficiency of *matK*, *rbcL*, *trnH-psbA*, and *trnL-F* (cpDNA) to molecularly authenticate Philippine ethnomedicinal Apocynaceae through DNA barcoding. Pharmacogn Mag 12: S384-S388. DOI: 10.4103/0973-1296.185780.

Cahyaningsih R, Compton LJ, Rahayu S, Brehm JM, Maxted N. 2022. DNA barcoding medicinal plant species from Indonesia. Plants 11 (10): 1375. DOI: 10.3390/plants11101375.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797. DOI: 10.1073/pnas.0905845106.

Cordero CS, Guillergan LR, Nievales MFJ, Alejandro GJD. 2025. Ethnobotanical study of medicinal plants used in an urban community in Iloilo City, Philippines. Biodiversitas 26 (5): 2167-2181. DOI: 10.13057/biodiv/d260515.

Cordero CS, Ligsay A, Alejandro GJD. 2020. Ethnobotanical documentation of medicinal plants used by the Ati tribe in Malay, Aklan, Philippines. J Complement Med Res 11 (1): 170-198. DOI: 10.5455/jcmr.2020.11.01.20.

Cordero CS, Meve U, Alejandro GJD. 2022a. Ethnobotanical documentation of medicinal plants used by the indigenous Panay Bukidnon in Lambunao, Iloilo, Philippines. Front Pharmacol 12: 790567. DOI: 10.3389/fphar.2021.790567.

Cordero CS, Meve U, Alejandro GJD. 2022b. Quantitative ethnobotanical documentation of medicinal plants used by the indigenous Ati tribes in Panay Island, Philippines. Malays J Sustain Environ 9 (2): 143-170. DOI: 10.24191/myse.v9i2.18832.

Cordero CS, Meve U, Alejandro GJD. 2023. Ethnobotany and diversity of medicinal plants used among rural communities in Mina, Iloilo, Philippines: A quantitative study. J Asia-Pac Biodivers 16 (1): 96-117. DOI: 10.1016/j.japb.2022.12.003.

de Leon Suba M, Arriola AH, Alejandro GJD. 2019. Evaluation of cpDNA barcodes in selected medicinal plants of Mt. Arayat National Park, Pampanga, the Philippines. J Young Pharm 11 (2): 130-135. DOI: 10.5530/jyp.2019.11.29.

Harris CS, Kim DY, Jordan CR, Miranda MI, Hellberg RS. 2024. DNA barcoding of herbal supplements on the US commercial market associated with the purported treatment of COVID-19. Phytochem Anal 35 (4): 664-677. DOI: 10.1002/pca.3320.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc Biol Sci 270 (1512): 313-321. DOI: 10.1098/rspb.2002.2218.

Hughes CE, Eastwood RJ, Bailey CD. 2006. From famine to feast? Selecting nuclear DNA sequences for plant species-level phylogenetic analyses. Philos Trans R Soc B: Biol Sci 361 (1465): 211-225. DOI: 10.1098/rstb.2005.1735.

Ilagan VAD, Alejandro GJD, Paraguison DJB, Perolina SMW, Mendoza GR, Bolina AB, Raterta R, Vales MB, Suarez GJD, Blasco FA. 2022. Ethnopharmacological documentation and molecular authentication of medicinal plants used by the Manobo and Mamanwa tribes of Surigao del Sur, Philippines. Biodiversitas 23 (6): 3185-3202. DOI: 10.13057/biodiv/d230646.

Kolter A, Gemeinholzer B. 2021. Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. Genome 64 (3): 265-298. DOI: 10.1139/gen-2019-0198.

Letsiou S, Madesis P, Vasdekis E, Montemurro C, Grigoriou ME, Skavdis G, Moussis V, Koutelidakis AE, Tzakos AG. 2024. DNA barcoding as a plant identification method. Appl Sci 14 (4): 1415. DOI: 10.3390/app14041415.

Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. 2015. Plant DNA barcoding: From gene to genome. Biol Rev Camb Philos Soc 90 (1): 157-166. DOI: 10.1111/brv.12104.

Lola AMA, Binag SDA, Alfeche NKG, Alejandro GJD. 2025. DNA barcoding and high-performance liquid chromatography as golden standards of authentication in *Blumea balsamifera* and *Ehretia microphylla* products. Biodiversitas 26 (3): 1258-1270. DOI: 10.13057/biodiv/d260325.

Meñiza JF, Pasco MM, Alimbon JA. 2024. A review of ethnobotanical studies reveals over 500 medicinal plants in Mindanao, Philippines. Plant Divers 46 (5): 551-564. DOI: 10.1016/j.pld.2024.05.001.

Naive MAK, Binag SDA, Alejandro GJD. 2021. Plants with benefits: Ethnomedicinal plants used by the Talaandig tribe in Portulin, Pangantucan, Bukidnon, Philippines. Indian J Tradit Knowl 20 (3): 754-766. DOI: 10.56042/ijtk.v20i3.26584.

Nanni L, Cuza D, Brahnam S. 2024. AI-powered biodiversity assessment: Species classification via DNA barcoding and deep learning. Technologies 12 (12): 240. DOI: 10.3390/technologies12120240.

Nazar N, Howard C, Slater A, Sgamma T. 2021. Challenges in medicinal and aromatic plants DNA barcoding—lessons from the Lamiaceae. Plants 11 (1): 137. DOI: 10.3390/plants11010137.

Nazar N, Saxena A, Sebastian A, Slater A, Sundaresan V, Sgamma T. 2025. Integrating DNA barcoding within an orthogonal approach for herbal product authentication: A narrative review. Phytochem Anal 36 (1): 7-29. DOI: 10.1002/pca.3466.

Olsson S, Giovannelli G, Roig A, Spanu I, Vendramin GG, Fady B. 2022. Chloroplast DNA barcoding genes *matK* and *psbA-trnH* are not suitable for species identification and phylogenetic analyses in closely related pines. iForest Biogeosci For 15: 141-147. DOI: 10.3832/ifor3913-015.

Paraguison LD, Tandang DN, Alejandro GJD. 2020. Medicinal plants used by the Manobo tribe of Prosperidad, Agusan del Sur, Philippines: An ethnobotanical survey. Asian J Biol Life Sci 9 (3): 326-333. DOI: 10.5530/ajbls.2020.9.49.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Riza LS, Zain MI, Izzuddin A, Prasetyo Y, Hidayat T, Samah KAFA. 2023. Implementation of machine learning in DNA barcoding for determining the plant family taxonomy. Heliyon 9 (10): e20161. DOI: 10.1016/j.heliyon.2023.e20161.

Santor PJR, Ordas JAD, Olivar JEC, Bangcaya PS, Nonato MG, Alejandro GJD. 2021. DNA barcoding of six ethnomedicinal and economically important *Pandanus* Parkinson (Pandanaceae) in the Philippines. Philipp J Sci 150 (5): 1321-1335. DOI: 10.56899/150.05.39.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Biol Evol 38 (7): 3022-3027. DOI: 10.1093/molbev/msab120.

Urumarudappa SKJ, Tungphatthong C, Jaipaew J, Pornputtapong N, Pakdeesattayapong D, Vimolmangkang S, Sukrong S. 2022. Development of a DNA barcode library of plants in the Thai Herbal Pharmacopoeia and monographs for authentication of herbal products. Sci Rep 12 (1): 9624. DOI: 10.1038/s41598-022-13287-x.

WHO [World Health Organization]. 2019. WHO Global Report on Traditional and Complementary Medicine 2019. World Health Organization, Geneva, Switzerland.

Yang C-H, Wu K-C, Chuang L-Y, Chang H-W. 2022. DeepBarcoding: Deep learning framework for species classification using DNA barcoding. IEEE/ACM Trans Comput Biol Bioinform 19 (4): 2158-2165. DOI: 10.1109/TCBB.2021.3056570.

Yu J, Wu X, Liu C, Newmaster S, Ragupathy S, Kress WJ. 2021. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. Ecotoxicol Environ Saf 208: 111691. DOI: 10.1016/j.ecoenv.2020.111691.

Zhu S, Liu Q, Qiu S, Dai J, Gao X. 2022. DNA barcoding: An efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. Chin Med 17 (1): 112. DOI: 10.1186/s13020-022-00655-y.