Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Molecular identification of protected Magnoliaceae species: a multi-marker evaluation based on extensive sampling and genome-skimming data

Kai Mu
Beihua University
Xiaoyi Cheng
Northeast Forestry University
Xiaoman Wang
Hebei Normal University
Zhang Jin
Ministry of Public Security of the People's Republic of China
Zhengge Zhu
Hebei Normal University
Jie Zhang
Northeast Forestry University
Shiliang Zhou
Institute of Botany
Ruijian Wang
Beihua University
Chao Xu
xuchao@ibcas.ac.cn
Institute of Botany
Xueying Yang
Ministry of Public Security of the People's Republic of China

Research Article

Keywords: DNA barcoding, Magnoliaceae, multi-marker evaluation, protected species, species identification

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10747250/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Magnoliaceae are among the most threatened plant families, with 24 species listed as nationally protected in China. The reliability of DNA barcoding for their identification has not been systematically evaluated under large-scale individual sampling. Here we employed genome skimming to generate plastid genome and nuclear ribosomal internal transcribed spacer sequences from 102 individuals representing 23 protected Magnoliaceae species. Supplemented with publicly available data, we assessed DNA barcoding performance at both species and individual levels across 196 chloroplast fragments using three analytical methods. At the species level, the BLAST-based method resolved 18 of the 24 species using ten selected candidate fragments, considerably more than standard barcodes or even complete plastid genomes could achieve under the same method. Under tree-based and barcode gap-based methods, however, complete plastid genomes yielded the highest resolution. When comparing the same fragments individually, species-level success rates were generally lower than individual-level success rates owing to the stringent criterion requiring all individuals of a species to be correctly identified. We term this systematic divergence the Species-Individual Discrepancy, which demonstrates that species-level resolution does not guarantee individual-level identification reliability. Nine species could not be correctly identified by any single fragment, likely due to incomplete lineage sorting or chloroplast capture. Our findings show that intraspecific polymorphism and reticulate evolution complicate plastid-based barcoding. We therefore recommend integrating population-level reference data and multi-marker strategies in conservation applications for threatened Magnoliaceae species.

1. Introduction

Magnoliaceae are among the most primitive lineages of angiosperms (Li et al., 2019; Sun et al., 2023), comprising approximately 300 species distributed globally and representing a classic example of the eastern Asian-North American disjunct distribution pattern (Azuma et al., 2001). Members of this family possess significant ornamental, economic, and medicinal values, with many species being important sources of bioactive compounds (Lee et al., 2011; Huang et al., 2025). However, due to habitat destruction and over-collection resulting from human activities, Magnoliaceae have become one of the most threatened plant families (Xie et al., 2022a). According to the IUCN Red List assessments, nearly half (48%) of Magnoliaceae species are facing various levels of extinction risk. In China, 24 Magnoliaceae species are listed in the National Key Protected Wild Plants of China (2021 edition), highlighting the priority of genetic resource conservation and management, which depends on reliable species identification. The widespread hybridization and morphological convergence in Magnoliaceae have long frustrated traditional morphology-based identification (Dong et al., 2022), making molecular identification a viable option (Agarwal et al., 2008).

DNA barcoding provides a standardized molecular toolkit for rapid species identification (Hebert et al., 2003; Perveen et al., 2026). The core plant barcode combination (*rbcL* + *matK*), together with the supplementary barcode ITS, has been widely adopted for species-level discrimination (Hollingsworth et al., 2009; China Plant BOL Group, 2011). In recent years, advances in high-throughput sequencing have enabled chloroplast genome sequencing and plastome-scale “super-barcoding,” considerably improving discriminatory power and facilitating the derivation of lineage-specific barcodes (Li et al., 2015; Coissac et al., 2016; Shah et al., 2026). Unlike standard barcodes that rely on a few short fragments, super-barcodes utilize the complete plastid genome (approximately 150 kb) and the large number of variable sites it contains, theoretically offering significantly enhanced resolution for closely related species (Parks et al., 2009; Wu et al., 2021). Several studies have demonstrated the superiority of super-barcodes over standard barcodes in various plant groups, but have also revealed that even complete plastid genomes may fail to achieve full species-level resolution in recently radiated lineages (Wang et al., 2020; Dong et al., 2022; Yu et al., 2022). Thus, the optimal choice among standard barcodes, highly variable candidate fragments, and plastid super-barcodes depends on the specific application scenario. Practical considerations such as cost-effectiveness, time-efficiency, and operational feasibility are equally decisive alongside resolution power.

Individual-level identification reliability is equally critical as species-level resolution in practical applications (Huang et al., 2020). Applications of DNA barcoding often require individual-level identification. In forensic authentication of timber, quality control of medicinal materials, or genetic resource management of endangered species, the question to be answered is typically which species an unknown sample belongs to rather than whether that species forms a monophyletic clade (Chen et al., 2014; Dormontt et al., 2015). However, most existing barcoding evaluations have focused on species-level resolution, with far less attention paid to systematically assessing the accuracy of molecular data in correctly assigning each individual to its species (Meyer & Paulay, 2005; Wohlmuth et al., 2016). Whether species-level resolution reliably predicts individual-level identification accuracy remains to be tested explicitly across taxonomic groups and marker systems.

Individual-level identification reliability is strongly constrained by the representativeness of reference sequences in databases. Currently, reference sequences in GenBank and BOLD for Magnoliaceae and most taxa are predominantly derived from a small number of individuals per species and the true extent of intraspecific genetic variation has never been systematically evaluated (Yu et al., 2014; Wang et al., 2020). Previous studies have shown that overlap between intraspecific and interspecific genetic distances is a major cause of DNA barcoding failure (Meyer and Paulay, 2005; Wiemers and Fiedler, 2007). In lineages that have undergone recent rapid radiation or frequent hybridization, incomplete lineage sorting and chloroplast capture can cause some individuals to carry plastid haplotypes that are phylogenetically closer to other species than to conspecific accessions (Wyler et al., 2016). Under such circumstances, reference databases built from a few individuals will severely underestimate intraspecific variation, thereby compromising the practical reliability of barcode-based identification (Luo et al., 2015). Recent studies have shown that incomplete reference databases produce not only unclassified queries but also false positive detections and ecologically meaningful misclassifications (McCartin et al., 2026). Therefore, barcode evaluations based on large-scale individual sampling are methodologically essential for testing the impact of intraspecific polymorphism on identification performance (Luo et al., 2015).

In the selection of highly variable fragments for developing lineage-specific barcodes, existing studies often rely on oversimplified approaches. Most studies have relied exclusively on nucleotide diversity (π) to assess fragment polymorphism and screen candidate barcodes (Kuang et al, 2011; Li et al., 2024). However, π reflects the average degree of sequence divergence and does not comprehensively measure a fragment's identification potential. In contrast, the

number of segregating sites (S) reflects the richness of variable sites, and the number of parsimony-informative sites (P) reflects the proportion of variable sites that carry phylogenetic signal. Moreover, many studies have conducted screening only at the family or genus level, overlooking the uneven distribution of variation hotspots across different lineages. A fragment that is highly variable in certain clade may be highly conserved in another (Wyler et al., 2016; Orsholm et al., 2026). Together, these observations point to a key principle for DNA barcoding marker development. The selection of candidate fragments should be guided by multiple complementary criteria rather than a single summary statistic. It should also account for both lineage-wide utility and group-specific informativeness. The relative weight of these criteria depends on the taxonomic scope and practical objectives of the intended application.

Furthermore, the method used to calculate species identification success rates directly influences the evaluation conclusions. Three mainstream approaches are currently employed. The tree-based method, which requires all individuals of a target species to form a monophyletic clade (Hebert et al., 2003). The BLAST-based method, which returns the best matching species through similarity searches (Altschul et al., 1990). And the barcode gap-based method, which requires the maximum intraspecific distance to be smaller than the minimum interspecific distance (Meyer & Paulay, 2005). These methods differ in stringency and data requirements, and the identification success rates they yield can vary considerably. Yet few studies have systematically compared the performance of all three methods in the same dataset within Magnoliaceae. More importantly, all these methods face a common challenge when multiple individuals per species are included: whether to require all individuals to meet the criterion simultaneously (Collins & Cruickshank, 2013). If a lenient criterion is adopted (i.e., a species is scored as successful as long as at least one individual is correctly identified), the evaluation may overestimate the practical utility of barcodes. If a strict criterion is adopted (all individuals must meet the criterion), the evaluation becomes more conservative but may underestimate identifiability due to a few outlier individuals (Little et al., 2013). The choice between these two criteria has direct implications for the reliability and practical relevance of the evaluation conclusions.

To address these gaps, the present study employed genome skimming to generate plastid genome and nrITS sequences from 102 individuals representing 23 of the 24 protected Magnoliaceae species, supplemented with publicly available sequences from NCBI. Specifically, we aimed to address four questions: (1) Can plastid super-barcodes, lineage-specific candidate fragments, and standard barcodes effectively identify the 24 protected species at the species level? (2) Can a multi-indicator screening framework identify reliable candidate barcodes for Magnoliaceae? (3) What are the differences in species identification success rates among BLAST-based, tree-based, and barcode gap-based methods under the species-level stringency criterion? (4) What is the individual-level identification reliability of single fragments under large-scale individual sampling, and what is the relationship between species-level and individual-level identification performance? By addressing these questions, this study provides systematic data support and methodological reference for DNA barcoding practice and genetic resource management of protected Magnoliaceae species. We also offer a technical framework for the precise identification and digital management of rare Magnoliaceae genetic resources in gene banks.

2. Materials and Methods

2.1 Plant materials

This study focused on 24 protected Magnoliaceae species in the list of National Key Protected Wild Plants in China. Of these, 23 species were newly sequenced, with sampling sizes ranging from 1 to 15 individuals per species, totaling 102 accessions. Plant materials were obtained from herbarium specimens deposited at the Herbarium of the Institute of Botany, Chinese Academy of Sciences (PE), and silica-dired leaves stored in the institute's DNA bank. Voucher information and detailed sample data are provided in Supplementary Table S1.

To establish a comprehensive framework for phylogenetic and barcoding analyses, we downloaded all available Magnoliaceae complete plastid genome sequences and nuclear ribosomal internal transcribed spacer (nrITS) sequences from the NCBI GenBank public database. The search was performed using the keyword "Magnoliaceae"[Organism], with sequence length filters set to 140,000-180,000 bp for plastid genomes and 100-8,000 bp for ITS sequences, and the sequence format set to GenBank (full). Details of the downloaded data are listed in Supplementary Table S2.

2.2 DNA extraction and genome skimming sequencing

Total genomic DNA was extracted from plant tissue using a modified CTAB protocol (Li et al., 2013). To mitigate interference from polyphenols and polysaccharides commonly present in herbarium specimens, 2% PVP (polyvinylpyrrolidone) and 0.2% β-mercaptoethanol were added to the extraction buffer. The water-bath incubation was extended to 1.5 h to improve DNA yield from aged materials.

All DNA samples were sent to a commercial sequencing facility for library construction and sequenced on the DNBSEQ-T7 platform with paired-end reads (150 bp). Each sample was planned to sequence approximately 50 Gb of raw data in order to provide sufficient coverage depth for both plastid genomes and the high-copy nrDNA regions. Raw reads were quality-filtered using fastp v0.23 (Chen et al., 2018) to remove adapter sequences, reads with average quality below Q20, and reads shorter than 50 bp. The resulting clean reads were used for subsequent assembly analyses.

2.3 Assembly of plastid genomes and nrITS sequences

2.3.1 Plastid genome assembly and annotation

Plastid genomes were assembled de novo from clean reads of each sample using GetOrganelle v1.7.7 (Jin et al., 2020). Assembly was conducted with 10 iterative rounds, default k-mer parameters, and the embplant_pt mode for land plant plastomes. For samples with unsatisfactory assembly results, we increased the iteration rounds or supplied a reference sequence from a closely related species to obtain complete circular plastid genomes.

Complete plastid genomes were annotated using PLANN v1.1 (Huang and Cronk, 2015), with reference to published Magnoliaceae plastomes, to identify protein-coding genes (PCGs), tRNA genes, and rRNA genes, including their boundaries and functional predictions. Annotation results were manually curated

in Geneious Prime v2023.2 (Kearse et al., 2012), with special attention to the completeness of start and stop codons of protein-coding genes, and the precise positioning of the inverted repeat (IR) boundaries.

2.3.2 Extraction and assembly of nrITS sequences

Nuclear ribosomal internal transcribed spacer (ITS) sequences were extracted from the genome skimming data. Reads were mapped to the ITS sequence of *Magnolia figo* (Lour.) DC. (NCBI accession: KP092905) using BWA v0.7.19. The mapped reads were assembled in Sequencher 5.4.5 (Gibbs and Cockerill, 1995) with a sequence similarity threshold of 99% to exclude fungal contaminants and non-specific sequences. Successful assemblies were verified by BLAST v2.17.0 (Altschul et al., 1990) searches against ITS sequences of closely related Magnoliaceae species in NCBI. The complete ITS1-5.8S-ITS2 region was extracted from the assemblies and used for subsequent DNA barcoding evaluation.

2.4 Phylogenetic analysis and data quality assessment

2.4.1 Dataset construction

Based on the assembled and downloaded sequences, we constructed two datasets: a plastid genome dataset and an nrITS dataset. Each dataset was aligned using MAFFT v7.505 (Katoh and Standley, 2013) with the auto setting. Alignment gaps were trimmed with trimAl v1.4 (Capella-Gutiérrez et al., 2009) using the automated1 option, generating high-quality matrices for subsequent analyses.

2.4.2 Phylogenetic tree reconstruction

To evaluate species identification accuracy and assess monophyly, we reconstructed phylogenetic trees from both the plastid genome dataset and the nrITS dataset. Maximum likelihood (ML) trees were constructed using IQ-TREE v2.2.0 (Minh et al., 2020). The best-fitting nucleotide substitution model was automatically selected by the built-in ModelFinder module based on the Bayesian Information Criterion (BIC). Branch support was assessed with 1000 ultrafast bootstrap replicates, with the convergent bootstrap option enabled to ensure stable bootstrap values. Multi-threading was enabled for computational efficiency. All trees were rooted using *Liriodendron* species as outgroup. Trees were visualized and annotated using iTOL v7 (https://itol.embl.de/) (Letunic and Bork, 2021), and further refined for presentation using Adobe Illustrator v25.0.

2.4.3 Species monophyly testing and data quality assessment

We tested the monophyly of each of the 24 protected species based on the reconstructed ML trees. All individuals of a species formed a monophyletic clade with moderate to high support (ML bootstrap ≥ 70%), indicating that the DNA data can be used for reliable species identification. Individuals of a species might be scattered across multiple, non-contiguous clades, or some individuals grouped with individuals of other species. In such cases, we further examined the pattern. If a particular individual consistently grouped with another species with high support, we considered it as a potential misidentification of the voucher specimen and excluded it from downstream barcoding analyses. If individuals of the same species occupied different well-supported clades but morphological identification was confirmed correct, this was interpreted as possible chloroplast capture or hybridization events, and such individuals were retained. After this quality assessment, we excluded putatively misidentified samples and retained a filtered dataset for subsequent nucleotide polymorphism analyses and barcoding evaluations.

2.5 Plastome fragment partitioning and nucleotide polymorphism analysis

2.5.1 Fragment classification and partitioning

To systematically assess variation patterns across different taxonomic groups within Magnoliaceae and to identify group-specific DNA barcodes, we assigned all plastid genome sequences to sections following the classification system of Wang et al. (2020). Based on the annotation files, we used a custom Python script (https://github.com/puppyborn/A-Python-Toolkit-for-Magnoliaceae-Classification/blob/main/split_chloroplast_genome.py) to partition each plastid genome into three fragment categories: genes, intergenic spacers (IGSs), and introns. Fragments over 100 bp in length and annotated in most of the plastomes were extracted for analysis. We used fragments in each category separately for evaluation.

2.5.2 Nucleotide polymorphism calculation

For each fragment within each taxonomic group, we calculated nucleotide diversity (π) and the number of segregating sites (S) using DnaSP v6.12.03 (Rozas et al., 2017). π is defined as the average number of nucleotide differences per site between two sequences and serves as a core measure of genetic variation within a population or lineage. S is the total number of sites in the alignment where at least two individuals differ, reflecting the overall richness of variation. We also calculated the number of parsimony-informative sites (P) from the plastid genome alignment using MEGA 12 (Stecher et al., 2026). P is defined as the number of variable sites where at least two alternative nucleotides are each present in at least two individuals, reflecting the phylogenetic information content of the variation. It should be noted that DnaSP excludes sites containing gaps when computing S, counting only unambiguous nucleotide substitutions, whereas MEGA treats gaps as valid character states when computing P. Consequently, the absolute values of S and P are not directly comparable across the two indices, as each captures a different facet of sequence variation under different computational conventions. In this study, S and P were used independently for ranking fragments by their respective measures of variation richness and phylogenetic information content, rather than for direct cross-index comparison

These three indices (π, S, and P) were calculated for all of the separated chloroplast fragments across 18 groups: the 15 *Magnolia* sections, the genus *Magnolia*, the genus *Liriodendron*, and Magnoliaceae as a whole. Polymorphism variation was summarized by group and fragment category.

2.5.3 Polymorphism heatmap construction and screening of group-specific barcodes

We compiled the π, S, and P values for all fragments across groups and generated polymorphism heatmaps using a custom Python script (https://github.com/puppyborn/A-Python-Toolkit-for-Magnoliaceae-Classification/blob/main/heatmap_pi_s_p_V1.py). To improve the statistical rigor of candidate barcode selection, we applied two outlier detection strategies to each of the three diversity indices: (1) Global outlier detection. For all fragment-group combinations, we calculated the upper quartile (Q_3) and interquartile range (IQR). Fragments with diversity values exceeding $Q_3 + 1.5 \times IQR$ were defined as global high-variation hotspots (Tukey, 1977). (2) Group-wise outlier detection. For each gene fragment (i.e., each row in the heatmap), we calculated the mean and standard deviation (SD) of its diversity values across the 15 *Magnolia* sections. Z-scores were then calculated for each cell as $Z = (diversity_i - mean_diversity_i) / SD_diversity_i$. Cells with $Z > 1.96$ (corresponding to two-tailed $P < 0.05$) were considered as group-specific high-variation hotspots (Akey et al., 2002). Based on statistical significance, along with consideration of π, S, P values and fragment length, we selected 10 high-variation fragments as candidate group-specific DNA barcodes. Among the three diversity indices, P was given the highest weight in fragment prioritization, as it directly reflects the proportion of variable sites that carry phylogenetic signal and is therefore most informative for species discrimination.

2.6 Evaluation of DNA barcoding identification success

To systematically evaluate the performance of different DNA barcodes for identifying 24 protected Magnoliaceae species, we conducted analyses at two levels: species-level resolution and individual-level accuracy. The former compares the theoretical discrimination power of different marker strategies, while the latter simulates the reliability of single-sample identification in practical applications.

2.6.1 Species-level resolution assessment

This section addresses the question: which type of barcode most effectively distinguishes each protected Magnoliaceae species as an intact genetic entity? Three categories of barcodes were compared: standard DNA barcodes (*rbcL*, *matK*, ITS), group-specific DNA barcodes (the 10 high-variation fragments selected, single and combined), and plastid super-barcodes (complete plastid genomes).

To ensure comparability across methods, we applied a species-level stringency criterion to three approaches (tree-based, BLAST-based, and barcode gap-based). Under this criterion, a species was considered successfully identified only when all of its individuals met the corresponding criterion. The success rate was then calculated as the percentage of successfully identified species out of the total 24 species. The three methods are: (1) Tree-based method. Neighbor-Joining (NJ) trees were constructed for each barcode alignment using IQ-TREE v2.2.0 with the BIONJ algorithm, Jukes-Cantor (JC) model, 1000 bootstrap replicates, and the convergent bootstrap option enabled. A species was scored as successfully identified if all its individuals formed a monophyletic clade with bootstrap support ≥ 70%. (2) BLAST-based method. BLASTn searches were performed with a leave-one-out cross-validation strategy. Each individual was used as query against a reference database consisting of all other individuals with default parameters (E-value ≤ 1e-5). The best-hit species assignment was recorded for each query. Under the species-level stringency criterion, a species was considered successfully identified only when all of its individuals returned themselves as the best hit. (3) Barcode gap-based method. Pairwise genetic distances were calculated using a custom Python script under the Kimura 2-parameter (K2P) model for each barcode dataset (https://github.com/puppyborn/A-Python-Toolkit-for-Magnoliaceae-Classification/blob/main/k2p_barcode_gap_V1.py). For each species, we calculated the maximum intraspecific distance and the minimum interspecific distance. A species was considered to have a clear barcode gap if the difference (minimum interspecific distance − maximum intraspecific distance) was greater than zero, and was thus scored as successfully identified.

2.6.2 Individual-level accuracy assessment

This section addresses the question: which plastid regions are useful for individual-level species identification and with what accuracy? The aim is to provide a reference identification potential dictionary for future studies. The analysis included all partitioned fragments, namely 76 protein-coding genes, ten RNA genes, 93 intergenic spacers, and 17 introns.

In contrast to Section 2.6.1, this section focuses on individual-level identification accuracy. Given the large number of fragments involved (196 fragments × 193 individuals), constructing individual trees and assessing monophyly for each fragment would be impractical and difficult to standardize. Therefore, this section employed only the BLAST-based and genetic distance-based nearest-neighbor methods, both using leave-one-out cross-validation. For BLAST method, each query sequence was searched against the reference database using BLASTn. The best-hit species assignment was recorded and compared with the true species identification. For genetic distance nearest-neighbor method, pairwise genetic distances were calculated using a custom Python script under the Kimura 2-parameter (K2P) model for each barcode dataset (https://github.com/puppyborn/A-Python-Toolkit-for-Magnoliaceae-Classification/blob/main/k2p_barcode_gap_V1.py). The species of the sequence with the smallest genetic distance to the query was taken as the identification result and compared with the true species identification.

3. Results

3.1 Sequencing and assembly statistics

A total of 102 accessions representing 23 Magnoliaceae species were successfully sequenced, generating raw data ranging from 35.12 Gb to 219.13 Gb per accession, with a mean of 61.05 Gb (Table S1). After quality filtering, clean reads were obtained for all samples. Plastid genomes were successfully assembled for all 102 accessions, yielding complete circular genomes with the typical quadripartite structure. The assembled plastid genomes ranged from 159,554 bp to 160,201 bp in length, with an average of 160,045 bp. The plastomes were highly conserved in gene content, each comprising 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes, with only slight variation (85–87 protein-coding genes and 7–8 rRNA genes) among individual accessions.

nrITS sequences (ITS1-5.8S-ITS2) were successfully recovered from all of the 102 accessions, with a success rate of 100%. The assembled nrITS fragments ranged from 558 bp to 609 bp in length. All newly assembled plastid genomes and nrITS sequences have been deposited in NCBI GenBank under accession

numbers provided in Supplementary Table S1. Raw sequencing reads have been deposited in the NCBI SRA database (Supplementary Table S1).

3.2 Phylogenetic analysis and data quality assessment

3.2.1 Phylogenetic tree reconstruction

Maximum likelihood (ML) trees were reconstructed from both the plastid genome dataset (692 accessions, including 102 newly sequenced + 590 downloaded) and the nrITS dataset (207 accessions, including 102 newly sequenced + 105 downloaded). Among the ITS sequences downloaded from NCBI, only 105 passed quality filtering (excluding fungal contamination, non-Magnoliaceae sequences, sequences with obvious sequencing errors, and incomplete ITS regions). Overall, the availability and quality of Magnoliaceae ITS data in GenBank were substantially lower than those of plastid genomes. The alignments of both datasets were deposited in ScienceDB (https://doi.org/10.57760/sciencedb.45292).

The plastid ML tree (Fig. 1 and Figure S1) recovered a well-resolved and highly supported backbone topology largely congruent with the classification system of Wang et al. (2020). The 15 sections of *Magnolia* were consistently recognized as distinct clades, with most receiving strong bootstrap support (≥ 95%). *Liriodendron* was placed as the sister to the rest of Magnoliaceae with maximal support (100%). Within sections, the majority of species formed well-supported monophyletic clades, with most receiving bootstrap support ≥ 85%. Based on the nrITS ML tree (Figure S2), the topology showed substantially lower resolution compared to the plastome tree. At the sectional level, several clades recovered in the plastid tree were either poorly supported or not recovered in the ITS tree. The backbone relationships among major lineages of *Magnolia* were largely unresolved, with most deep nodes receiving low bootstrap support (< 70%). At the species level, the ITS tree failed to resolve several species as monophyletic, and individuals of some species were scattered across different clades, indicating limited phylogenetic signal in the ITS region for distinguishing closely related taxa within Magnoliaceae.

3.2.2 Species monophyly and data quality assessment

Based on the plastid ML tree, we tested the monophyly of each of the 24 protected species. Of these, 14 species formed well-supported monophyletic clades (ML bootstrap ≥ 70%), indicating clear species boundaries and confirming their suitability for DNA-based species identification (Fig. 1; Figure S2). These 14 species included *Liriodendron chinense* (Hemsl.) Sarg., *Magnolia baillonii* Pierre, *Magnolia cathcartii* (Hook.f. & Thomson) Noot., *Magnolia changhungtana* Noot., *Magnolia decidua* (Q.Y.Zheng) V.S.Kumar, *Magnolia ernestii* Figlar, *Magnolia guangdongensis* (Y.H.Yan, Q.W.Zeng & F.W.Xing) Noot., *Magnolia henryi* Dunn, *Magnolia hypolampra* (Dandy) Figlar, *Magnolia kwangsiensis* Figlar & Noot., *Magnolia rostrata* W.W.Sm., *Magnolia shiluensis* (Chun & Y.F.Wu) Figlar, *Magnolia sinica* (Y.W.Law) Noot. and *Magnolia zenii* W.C.Cheng.

For the remaining ten species, non-monophyletic patterns (polyphyly or paraphyly) were observed, with 62 individuals distributed across different well-supported clades. After careful examination, 17 of the 102 newly generated accessions were identified as potential misidentifications of voucher specimens, as they consistently grouped with other species with high support. These 17 individuals were re-identified based on both plastid and ITS phylogenetic trees and subsequently excluded from downstream barcoding analyses of the protected species, as were ten NCBI-derived sequences showing obvious misidentifications (marked in Supplementary Tables S1 and S2). However, all excluded individuals and sequences were retained for nucleotide polymorphism analyses at the group level.

For the remaining 90 individuals from the ten non-monophyletic species, morphological identification could not confidently resolve their taxonomic status, and their conflicting placements across different well-supported clades were interpreted as possible chloroplast capture or hybridization events. Of these, five individuals showing persistently ambiguous placements were also excluded from downstream barcoding analyses (marked in Supplementary Table S1). Collectively, these ten species, together with their closely related congeners, may represent promising species complexes warranting further taxonomic investigation using integrative approaches.

3.3 Plastome fragment partitioning and nucleotide polymorphism analysis

3.3.1 Overview of partitioned fragments

Based on the annotation files, we partitioned the plastid genomes into three fragment categories: 86 genes (76 protein-coding genes and ten RNA genes), 93 intergenic spacers, and 17 introns. Each fragment was labeled according to its position on the plastid genome (Fig. 2). All fragments were assigned to 18 groups: the 15 *Magnolia* sections, the genus *Magnolia*, the genus *Liriodendron*, and Magnoliaceae as a whole.

3.3.2 Nucleotide polymorphism statistics

For each fragment within each group, we calculated three diversity indices: nucleotide diversity (π), segregating sites (S), and parsimony-informative sites (P). Summary statistics are presented in Table S3.

In genus *Magnolia*, IGSs showed the highest levels of polymorphism, with π values ranging from 0.0000 to 0.0119, followed by genes (0.0003–0.0067), while introns exhibited the lowest polymorphism (0.0000-0.0075) (Table S3). This pattern was consistent across all 15 *Magnolia* sections and *Liriodendron* (Figure S3). Among the 15 *Magnolia* sections, sect. *Yulania* displayed the highest overall nucleotide diversity (mean ± SD = 0.0018 ± 0.0016), followed by sect. *Maingola* (0.0016 ± 0.0017) and sect. *Rytidospermum* (0.0014 ± 0.0014). Sect. *Tuliparia* and sect. *Tulipastrum* showed the lowest overall diversity (Table 1). The genus *Liriodendron* exhibited notably lower polymorphism (0.002 ± 0.0022) compared to genus *Magnolia*.

Table 1
Summary statistics of three diversity parameters for Magnoliaceae taxonomic groups

Taxonomic level	Taxonomic group	π (mean ± SD)	π (range)	S (mean ± SD)	S (range)	P (mean ± SD)	P (range)
Section	Sect. Splendentes	0.0011 ± 0.0014	0–0.0088	3.3316 ± 5.1607	0–51	2.2704 ± 3.5233	0–32
	Sect. Talauma	0.0011 ± 0.0012	0–0.0062	6.5663 ± 8.7636	0–70	3.5561 ± 5.0273	0–37
	Sect. Gwillimia	0.0011 ± 0.001	0–0.0058	5.301 ± 7.4659	0–63	3.3776 ± 4.5787	0–35
	Sect. Tuliparia	0.0002 ± 0.0006	0–0.0042	0.3112 ± 1.0328	0–10	0.1531 ± 0.4495	0–3
	Sect. Macrophylla	0.0003 ± 0.0005	0–0.0027	0.9898 ± 1.8453	0–17	0.5969 ± 1.2261	0–11
	Sect. Magnolia	0.0013 ± 0.0012	0–0.0093	5.9031 ± 7.4163	0–66	3.5204 ± 4.5103	0–36
	Sect. Oyama	0.0015 ± 0.0017	0–0.0104	3.5714 ± 5.0779	0–41	3.0153 ± 4.2505	0–32
	Sect. Rytidospermum	0.0014 ± 0.0014	0–0.0089	5.6888 ± 7.8112	0–71	4.9031 ± 6.9648	0–58
	Sect. Manglietia	0.001 ± 0.0012	0–0.0068	6.7194 ± 8.8403	0–61	4.051 ± 5.8841	0–51
	Sect. Kmeria	0.0004 ± 0.0006	0–0.0035	1.7398 ± 2.9454	0–27	0.1173 ± 0.4315	0–3
	Sect. Gynopodium	0.0007 ± 0.0008	0–0.004	2.8673 ± 3.8301	0–31	1.8061 ± 2.7124	0–22
	Sect. Tulipastrum	0.0002 ± 0.0005	0–0.0031	0.2908 ± 0.6178	0–3	0.1684 ± 0.4263	0–2
	Sect. Yulania	0.0018 ± 0.0016	0–0.0108	7.0459 ± 9.7372	0–91	5.4694 ± 8.1527	0–75
	Sect. Maingola	0.0016 ± 0.0017	0–0.0109	3.3878 ± 5.3063	0–48	2.3673 ± 3.8813	0–33
	Sect. Michelia	0.0014 ± 0.0013	0–0.0108	11.7296 ± 15.2451	0–142	8.7551 ± 11.8136	0–109
Genus	*Magnolia*	0.0038 ± 0.0027	0–0.0119	51.5306 ± 66.1107	0–605	47.9643 ± 62.2231	0–564
	Liriodendron	0.002 ± 0.0022	0–0.0128	3.2245 ± 5.2507	0–52	3.0204 ± 4.9871	0–49
Family	Magnoliaceae	0.0042 ± 0.0029	0–0.0138	56.9439 ± 73.9902	0–684	55.8469 ± 72.3223	0–658

For segregating sites (S), the number of variable sites per fragment ranged from 0 to 684 across all fragment-group combinations, with genes contributing the highest average number of S sites (Figure S4). Within genes, the highest three S values in the family were observed in *ycf1* (S = 684), *rpoC2* (S = 338), and *trnK*-UUU (S = 305), indicating that these regions harbor substantial variation. IGSs showed intermediate levels of S, with *matK-rps16* (S = 186) *petA-psbJ* (S = 181) and *rpl32-trnL*-UAG (S = 180) among the most variable. Introns exhibited the lowest S values, with *ndhA* (*intron)* (S = 129) being the most variable, while most other introns had S values ranging from 6 to 80. *TrnK*-UUU (*intron)* was not considered having the highest S because it consists of *matK* in current dataset.

For parsimony-informative sites (P), the average number of P in IGSs is slightly lower than that in genes while the total number of P in all IGSs is higher than that in all genes of the family. The P values ranged from 0 to 658 across all fragment-group combinations (Figure S5). The highest P values were consistently found in genes such as *ycf1* (P = 658), *rpoC2* (P = 264), and *trnK*-UUU (P = 261). IGSs showed moderate P values, with *rps16-trnQ*-UUG (P = 303), *rpl32-trnL*-UAG (P = 217), and *ndhC-trnV*-UAC (P = 211) among the most informative, while introns generally had lower P values, with *ndhA* (P = 118) standing out as the most informative gene. We also noted that for certain fragments, particularly those rich in indels, the P values slightly exceeded the corresponding S values. This reflects the differing gap-treatment conventions between MEGA and DnaSP as described in Methods, and does not affect our use of each index as an independent ranking criterion for fragment evaluation.

3.3.3 Polymorphism heatmaps and outlier detection

Polymorphism heatmaps were generated for π, S, and P values across all fragment-group combinations, including the 15 *Magnolia* sections, *Magnolia*, *Liriodendron*, and Magnoliaceae (Fig. 3; Table S3 for individual indices). To improve the statistical rigor of candidate barcode selection, we applied two outlier detection strategies to each of the three diversity indices specifically for the 15 *Magnolia* sections (Table S4-S5). The heatmaps revealed that polymorphism hotspots were unevenly distributed across the plastid genome. High-variation regions were primarily concentrated in specific IGSs (e.g., *petL-psbE*, *petA-psbJ*, *rps16-trnQ*-UUG, *rpl32-trnL*-UAG, *ndhF-rpl32*, and *trnK*-UUU), with several protein-coding genes (e.g., *ycf1*, *ndhF*, *matK*, *rpoC2*) also showing elevated diversity. Notably, IGSs consistently exhibited the highest levels of variation across all three indices.

The three diversity indices revealed broadly similar hotspot patterns (Fig. 3), but with notable differences in the number of statistically significant outliers (Fig. 3). S and P identified substantially more significant sites than π across the 15 *Magnolia* sections (Fig. 3), while the trends between S and P were highly similar (Figure S4-S5). Global outliers were identified in multiple fragments, predominantly IGSs. The highest number of global outliers was observed in sect. *Yulania* and sect. *Michelia* (S and P), particularly for fragments such as *rps16-trnQ*-UUG, *rpl32-trnL*-UAG, and *trnK*-UUU. In contrast, sect. *Tuliparia*, sect. *Macrophylla*, and sect. *Tulipastrum* showed few or no global outliers, consistent with their overall low polymorphism levels. Group-wise outliers were also detected across various section-fragment combinations, highlighting fragments with significantly elevated diversity in specific lineages. For example, sect. *Yulania* exhibited group-specific high variation in *clpP* and *ndhC-trnV*-UAC ($Z > 1.96$), while sect. Michelia showed elevated values in *ndhA* and *matK-rps16* relative to their respective fragment means across sections.

Based on the combined evidence from the three diversity indices (π, S, and P) and the two outlier detection criteria, we selected 10 high-variation fragments as candidate group-specific DNA barcodes (Table 2). These included seven IGSs (*ndhF-rpl32*, *rpl32-trnL*-UAG, *petL-psbE*, *rps16-trnQ*-UUG, *matK-rps16*, *ndhC-trnV*-UAC, *petA-psbJ*) and three genes (*ndhF*, *clpP*, *ndhA*). The selection of these 10 fragments was supported by multiple lines of evidence. First, all candidate fragments fell within the preferred length range of 1,000–2,500 bp for DNA barcoding applications. Second, all ten fragments ranked among the top performers in at least two of the three diversity indices when assessed at the Magnoliaceae-wide level. Third, all candidates were identified as global outliers (exceeding $Q_3 + 1.5 \times IQR$) for S and P values in six to nine *Magnolia* sections, indicating consistently elevated polymorphism across multiple lineages within the family. Collectively, these criteria ensured that the selected fragments combine adequate length, high overall variability, and broad utility across diverse taxonomic groups within Magnoliaceae.

Table 2
Basic information of the DNA barcode markers developed in this study

Locus	Type	Length(bp)	Region	GC%	π (nucleotide diversity)	S (segregating sites)	P (parsimony-informative sites)	Sequence
ndhF-rpl32	IGS	1451	SSC	28.49	0.0107	159	200	692
ndhF	PCG	2265	SSC	34.64	0.0061	288	233	692
rpl32-trnL-UAG	IGS	1671	SSC	31.04	0.0089	180	220	692
petL-psbE	IGS	2109	LSC	33.79	0.0053	48	194	692
rps16-trnQ-UUG	IGS	2294	LSC	34.79	0.0073	98	303	692
clpP	PCG	2101	LSC	38.46	0.0034	155	137	692
matK-rps16	IGS	1702	LSC	35.42	0.0069	186	168	692
ndhA	PCG	2265	SSC	36.26	0.0046	212	179	692
ndhC-trnV-UAC	IGS	2068	LSC	34.84	0.0067	96	211	692
petA-psbJ	IGS	1348	LSC	35.06	0.0098	181	197	692
matK-CBOL	PCG	793	LSC	36.34	0.0063	99	32	692
rbcL-CBOL	PCG	553	LSC	44.44	0.0030	44	80	692
ITS	ITS	686	NUC	73.34	0.0217	49	159	206

3.4 Species-level resolution assessment

The species-level resolution of three barcode categories (standard barcodes, group-specific barcodes, and plastid super-barcodes) was evaluated using three methods (tree-based, BLAST-based, and barcode gap-based), all under the species-level stringency criterion. Results are summarized in Fig. 4 and Table S6-S9. Overall, the BLAST-based method exhibited the best performance, identifying the greatest number of species, followed by the barcode gap-based method, whereas the tree-based method showed the lowest resolution of the three.

Among the three standard barcodes, ITS exhibited the highest species-level resolution across all three evaluation methods, with the BLAST-based method identifying up to 12 species (Fig. 4A; Table S9). *matK* ranked second, identifying at most four species, while *rbcL* showed the lowest resolution, with only one species successfully identified. When full-length *matK* and *rbcL* sequences were used instead of the CBOL core regions, the number of resolved species increased to seven and six out of 24, respectively. The 10 candidate fragments selected from polymorphism analysis showed substantially higher resolution than standard barcodes when used individually. Among them, *petL-psbE*, *clpP*, and *matK-rps16* performed best, each successfully identifying 11 species using the BLAST-based method. Complete plastid genomes yielded the highest species-level resolution among all marker types, with 14 species successfully identified under the NJ tree-based method. Notably, the barcode gap-based method resolved one additional species compared to the NJ tree-based method, while the BLAST-based method performed relatively poorly for plastome sequences, identifying only 10 species (Fig. 4A). When combining group-specific fragments, the BLAST-based method consistently outperformed the other two approaches (Fig. 4B). However, the gain from adding more fragments was marginal. Combining six fragments increased the number of resolved species by only one compared to the five-fragment combination. Interestingly, when all ten fragments were combined, the number of resolved species reached 18 under the BLAST-based method.

We present detailed species-level resolution results for the 24 protected Magnoliaceae species based on all extracted chloroplast fragments, evaluated using the three methods described above (Fig. 5; Table S9). When the goal is to identify specific species, a broader range of chloroplast fragments is available beyond the ten candidate fragments selected in this study. Nevertheless, nine species, namely, *M. aromatica*, *M. grandis*, *M. lawiana*, *M. megaphylla*, *M. officinalis*, *M. omeiensis*, *M. sieboldii*, *M. ventii*, and *M. yunnanensis*, could not be accurately identified using any single fragment with the current reference database. This suggests that combining multiple variable fragments and selecting appropriate analytical methods may be necessary to improve identification success for these problematic taxa.

We further examined the relationship between species-level resolution and four fragment characteristics (alignment length, π, S, and P) across the three evaluation methods. Longer sequences, which accumulate more variation, tended to resolve more species regardless of the method used (Fig. 6A). Similar trends were observed for S and P values (Fig. 6C, 6D). In contrast, π did not serve as a reliable predictor of identification success, as no clear relationship between π values and resolution was detected (Fig. 6B).

3.5 Individual-level accuracy assessment

The individual-level identification accuracy of all plastid fragments was evaluated using BLAST and genetic distance nearest-neighbor methods with leave-one-out cross-validation. Results are compiled as matrix tables in Supplementary Tables S10 (BLAST method) and S11 (genetic distance nearest-neighbor method).

To examine the relationship between species-level and individual-level identification performance, we plotted the success rates of all 196 chloroplast fragments using two independent methods, BLAST-based assignment (Fig. 7A) and genetic distance-based nearest-neighbor methods (Fig. 7B). For the distance-based approach, species-level success was determined by the presence of a clear barcode gap (i.e., minimum interspecific distance > maximum intraspecific distance), while individual-level success was assessed using a leave-one-out nearest-neighbor criterion, where an individual was scored as correctly identified if its closest neighbor (s) all belonged to the same species.

Under the BLAST framework, species-level success rates for single fragments were generally low, with the vast majority of fragments falling below 0.5 (Fig. 7A). Individual-level success rates showed a broader distribution, ranging from 0 to nearly 0.78. A moderately positive correlation was evident between the two metrics. Fragments with higher species-level resolution tended to achieve relatively higher individual-level success. However, the scatter was considerable, particularly in the mid-range of species-level success (0.2–0.4), where individual-level rates varied widely from 0.2 to above 0.78. Notably, the best-performing fragments at the species level (e.g., *petL-psbE*, *clpP*, *matK-rps16*) did yield the higher individual-level success. Several fragments with moderate species-level resolution outperformed them at the individual level, suggesting that the factors governing individual-level accuracy are not identical to those driving species-level discrimination. The distance-based approach exhibited a broadly similar but more constrained pattern (Fig. 7B). Species-level success rates were generally lower than those observed with BLAST, with most fragments falling below 0.35 and the only one maximum reaching up to 0.5 (*ycf1*). Individual-level success rates ranged from 0 to approximately 0.70, comparable to the BLAST results. The correlation between the two metrics appeared weaker than in the BLAST-based analysis, as many fragments with moderate species-level success (0.2–0.3) showed highly variable individual-level performance. This suggests that the presence of a barcode gap at the species level is a necessary but insufficient condition for reliable individual-level identification using the nearest-neighbor criterion. Overall, the ten selected group-specific fragments performed well at both species and individual levels. This result validates our multi-indicator screening strategy, confirming its capacity to identify markers that are effective across hierarchical identification tasks.

4. Discussion

The present study enriched the genetic resources of protected species and provided the most comprehensive assessment of DNA barcoding performance in Magnoliaceae to date. By systematically evaluating 196 chloroplast fragments across three analytical frameworks at both species and individual levels, we reveal a critical gap between theoretical species-level discrimination and practical individual-level identification reliability. This gap has direct implications for the application of DNA barcoding in conservation genetics and genetic resource management of threatened Magnoliaceae species.

4.1 Phylogenetic Framework and the Challenge of Cytonuclear Discordance

The plastid genome phylogeny recovered in this study is largely congruent with the classification system of Wang et al. (2020). The 15 sections were consistently recovered as distinct clades with strong bootstrap support, and *Liriodendron* was placed as sister to the rest of Magnoliaceae with maximal support. However, 10 of the 24 protected species exhibited non-monophyletic patterns in the plastid tree, with individuals scattered across different well-supported clades. This proportion of non-monophyly is notable and aligns with previous observations of cytonuclear discordance in Magnoliaceae. Dong et al. (2022) contrasted plastid and nuclear phylogenies of 48 *Magnolia* s.l. species and revealed extensive cytonuclear conflicts in both shallow and deep relationships, with hybridization detected within sect. *Yulania* and sect. *Magnolia*. The non-monophyletic species observed in our study, including *M. officinalis*, *M. sieboldii*, and *M. omeiensis*, are likely products of similar evolutionary processes, such as chloroplast capture, incomplete lineage sorting, or historical hybridization (Rieseberg & Soltis, 1991; Wyler et al., 2016; Muranishi et al., 2013). These processes have left widespread imprints on the chloroplast genomes of Magnoliaceae, complicating the use of maternally inherited plastid markers for species identification.

Our study found that species which are non-monophyletic in the plastid tree are particularly challenging for species-level identification under strict criteria, as the presence of divergent haplotypes within a species often prevents all individuals from meeting the identification threshold simultaneously. At the individual level, introgressed or captured haplotypes can cause some individuals of a species to be phylogenetically closer to heterospecific congeners than to conspecific accessions, thereby reducing individual-level identification success. These findings suggest that in lineages with extensive reticulate evolution, the concept of a "species" as a single genetic entity may not be adequately captured by plastid data alone.

4.2 Polymorphism Hotspots and the Selection of Group-Specific Barcodes

Systematic scanning of 196 chloroplast fragments across 15 *Magnolia* sections, the genus *Magnolia*, *Liriodendron*, and Magnoliaceae revealed that intergenic spacers exhibited the highest nucleotide diversity, followed by protein-coding genes, with introns showing the lowest polymorphism. The 10 candidate fragments selected through combined outlier detection strategies represent regions of consistently elevated polymorphism across multiple lineages. Several of these candidates overlap with previously identified high-variation regions in Magnoliaceae. Kuang et al. (2011) identified 16 potential barcode markers from the plastome of *M. kwangsiensis*, including several IGSs that correspond to those identified in our study (i.e., *ndhF-rpl32, rpl32-trnL, petA-psbJ, psbE-petL, trnK-rps16*). Xie et al. (2022b) identified five highly variable regions in six threatened species of subgenus *Gynopodium*, including *psbA-trnH*, *petA-psbJ*, *rpl32-trnL*-UAG, and *ccsA-ndhD*. Li et al. (2024) identified 13 mutational hotspots in four *Manglietia* species. Among these independently identified hotspots, only *petA-psbJ* and *psbE-petL* overlap with our candidate list. This limited overlap likely reflects differences in screening strategies: previous studies relied solely on π and small plastome datasets, whereas our multi-indicator framework incorporated S, P, fragment length, and extensive sampling across 15 sections.

More broadly, the limited overlap illustrates that different screening criteria yield different candidate lists, and a comprehensive framework integrating multiple indices is better suited to identify markers with genuine utility across identification scales. We also noted that for certain fragments, particularly those rich in indels, the P values slightly exceeded the corresponding S values. This reflects the differing gap-treatment conventions between MEGA and DnaSP as described in Methods, and does not affect our use of each index as an independent ranking criterion for fragment evaluation.

4.3 Species-Level Resolution: From Standard Barcodes to Super-Barcodes

At the species level, ITS outperformed the other standard barcodes (*matK* and *rbcL*) across all three evaluation methods. This result is consistent with the widely recognized utility of ITS as a complementary barcode for discriminating closely related taxa in plants (Hollingsworth et al., 2009; China Plant BOL Group, 2011). However, ITS also showed substantially lower phylogenetic resolution at the sectional and backbone levels compared to the plastome tree, reflecting its limited information content for deeper relationships. The modest performance of *matK* and *rbcL* in Magnoliaceae is consistent with earlier observations. Yu et al. (2014) evaluated seven candidate barcode loci across 68 species representing 10 genera of Magnoliaceae and found that *psbA-trnH* and *matK* exhibited the highest amplification and sequencing success rates, with the combination of *psbA-trnH* and *matK* accurately distinguishing multiple closely related species. Although matK ranked among the top ten fragments based on P value and length in our study, we excluded it from our group-specific candidate list because it has already been established as a standard DNA barcode. However, even full-length *matK* resolved only seven protected species under the BLAST-based method, and full-length *rbcL* resolved only six. The use of full-length sequences rather than the CBOL core regions improved resolution for both markers, indicating that fragment length is a critical determinant of BLAST-based identification success.

The 10 candidate fragments, when used individually, showed considerably higher resolution than standard barcodes. When all 10 fragments were combined, the number of resolved species reached 18 (75%) under the BLAST-based method, notably higher than the 10 species resolved by complete plastid genomes using the same analytical approach. Under tree-based and barcode gap-based methods, however, complete plastid genomes outperformed all other marker types, resolving 14 and 15 species respectively. This methodological dependency highlights an important consideration. The relative performance of different barcode types is not intrinsic but varies with the analytical framework. BLAST-based searches, which rely on sequence similarity and are sensitive to fragment length and diagnostic sites, appear to favor concatenated multi-fragment datasets that provide more query coverage. In contrast, tree-based and gap-based methods benefit from the complete evolutionary information embedded in full plastid genomes, making them more suitable for phylogenetic and distance-based assessments. Overall, these observations support the view that plastid super-barcodes offer superior discrimination power for species identification in plants (Shah et al., 2026). We could also infer that when BLAST is the method of choice, as is often the case in practical identification scenarios, carefully selected multi-fragment combinations can serve as a cost-effective and equally powerful alternative.

4.4 The Species-Individual Discrepancy: A Core Methodological Finding

The most significant finding of this study is the systematic discrepancy between species-level and individual-level identification success, a phenomenon we term the Species-Individual Discrepancy (SID). When comparing species-level and individual-level success rates across the 196 fragments, we observed that individual-level success rates were generally higher than their species-level counterparts for the same fragments (Fig. 7). This seemingly counterintuitive pattern is, in fact, a direct consequence of the species-level stringency criterion adopted in this study. Under this criterion, a species is scored as successfully identified only when all of its constituent individuals meet the identification threshold. Consequently, the presence of a single "outlier" individual, whether resulting from introgression, chloroplast capture, cryptic diversity, or potential misidentification, is sufficient to cause the entire species to be classified as unresolved at the species level. At the individual level, however, each accession is evaluated independently. A typical individual may be correctly identified even if its conspecific outlier is not. The general observation that individual-level success rates exceeded species-level rates therefore carries a clear implication. The majority of individuals within most species are intrinsically identifiable using chloroplast fragments. And species-level failures are disproportionately driven by a small number of problematic individuals per species, rather than by a lack of overall interspecific differentiation (Little, 2013).

This pattern reinforces the SID concept by demonstrating that intraspecific heterogeneity, whether biological (hybridization, incomplete lineage sorting) or operational (misidentification), poses a more substantial threat to species-level barcode performance than does insufficient marker informativeness per se (Collins & Cruickshank, 2013). From a practical perspective, these findings suggest that excluding questionable accessions during reference database construction could artificially inflate species-level resolution, whereas including population-level variation, while more realistic, will necessarily lower the apparent success of strict species-level metrics.

The SID finding has important methodological implications. Future DNA barcoding studies should incorporate individual-level assessment metrics alongside species-level metrics; otherwise, the practical utility of barcodes will be systematically overestimated. For applications that require individual-level identification, such as forensic authentication of protected timber, medicinal material quality control, or enforcement of wildlife trade regulations, single-fragment approaches carry an inherent error risk of 25–50% at the individual level for the protected Magnoliaceae species. Multi-fragment combinations or complete plastid genomes are indispensable when unambiguous identification is required.

4.5 Nine Problematic Species: Taxonomic Complexes and Conservation Implications

Nine of the 24 protected species (*M. aromatica*, *M. grandis*, *M. lawiana*, *M. megaphylla*, *M. officinalis*, *M. omeiensis*, *M. sieboldii*, *M. ventii*, and *M. yunnanensis*) could not be accurately identified using any single chloroplast fragment with the current reference database. The biological causes of this poor identifiability are likely multiple. Recent rapid radiation may have resulted in incomplete lineage sorting of plastid genomes, such that species-specific haplotypes have not yet emerged (Wyler et al., 2016; Wang et al., 2020). Alternatively, ongoing or historical hybridization may have facilitated chloroplast capture, leading to shared plastid haplotypes across species boundaries (Rieseberg & Soltis, 1991; Dong et al., 2022). In some cases, the species themselves may represent poorly defined taxonomic entities requiring integrative revision using morphological and nuclear genomic data (Testé et al., 2022).

From a conservation perspective, these findings are concerning. Several of these nine species are listed as threatened on the IUCN Red List. *Magnolia officinalis* Rehder & E. H. Wilson, for example, is a medicinally important species whose wild populations have been severely depleted by over-collection.

Magnolia sieboldii K.Koch is a widely cultivated ornamental with declining wild populations. If DNA barcoding based on single plastid fragments cannot reliably distinguish these species from their congeners, then genetic identification of confiscated materials, verification of cultivated sources, and monitoring of wild populations may all be compromised. For these species, conservation genetic studies should prioritize nuclear genome markers, such as single nucleotide polymorphisms or microsatellites, which offer biparental inheritance and higher discriminatory power (Allendorf et al., 2010; Andrews et al., 2016).

Conclusions

Plastid super-barcodes and multi-fragment combinations provide valuable tools for species-level identification in Magnoliaceae, with the potential to resolve up to 75% of protected species when fragments are combined. However, the Species-Individual Discrepancy revealed in this study demonstrates that species-level success does not guarantee individual-level reliability. Intraspecific polymorphism, hybridization, and chloroplast capture collectively erode the practical utility of single-fragment barcodes for individual specimen identification. For conservation management of threatened Magnoliaceae species, a multi-marker strategy incorporating population-level reference data remains the most prudent approach.

Declarations

Acknowledgments The molecular experiments were carried out at the Plant DNA and Molecular Identification Platform (PDMIF), Institute of Botany, Chinese Academy of Sciences (IBCAS). The authors gratefully acknowledge Mr. Min Li (IBCAS) for assisting contacting photo authors of Magnoliaceae protected species. We thank Mr. Yechun Xu, Dr. Meng Wei and other authors for kindly sharing their beautiful photos.

Author contributions KM performed the molecular experiments, conducted the phylogenetic analyses, DNA barcoding analysis and drafted the manuscript. XYC and XMW contributed to data analysis and figure preparation. JinZ assisted with sequence alignment and tree construction. ZGZ, JieZ, and SLZ participated in data validation and manuscript revision. RJW, CX, and XYY supervised the research, interpreted the results, and critically revised the manuscript. All authors read and approved the final manuscript.

Funding This research was funded by the Institute of Forensic Science, Ministry of Public Security of China (2024JBGS003, 2025JB006) and National Plant Specimen Resource Center, Institute of Botany, Chinese Academy of Sciences (E0117G1001).

Data availability The data generated and analysed during the research are included in the main manuscript and supplementary materials. Additional data are available from the corresponding author upon reasonable request.

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval and informed consent The research followed the ethical guidelines of the International Society of Ethnobiology. Prior informed consent was obtained from all respondents before participation.

Consent for publication All authors approved the final manuscript and consented to its publication.

References

1. Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27 (4):617–631. https://doi.org/10.1007/s00299-008-0507-z
2. Akey J, Zhang G, Zhang K, Jin L, Shriver M (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12 (12):1805–1814. https://doi.org/10.1101/gr.631202
3. Allendorf, F, Hohenlohe, P, & Luikart, G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11, 697–709. https://doi.org/10.1038/nrg2844
4. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. J Mol Biol, 215 (3), 403–410. https://doi.org/10.1016/S0022-2836 (05)80360-2
5. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet, 17, 81–92. https://doi.org/10.1038/nrg.2015.28
6. Azuma H, García-Franco JG, Rico-Gray V. Thien LB (2001) Molecular phylogeny of the Magnoliaceae: the biogeography of tropical and temperate disjunctions. Am J Bot 88: 2275–2285. https://doi.org/10.2307/3558389
7. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348
8. Chen SL, Pang XH, Song JY, Shi LC, Yao H, Han JP, Leon C (2014) A renaissance in herbal medicine identification: from morphology to DNA. Biotechnol Adv 32 (7):1237–1244. https://doi.org/10.1016/j.biotechadv.2014.07.004
9. Chen SF, Zhou YQ, Chen YR, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34 (17):884–890. https://doi.org/10.1093/bioinformatics/bty560
10. China Plant BOL Group, Li DZ, Gao LM et al (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences 108 (49):19641–19646. https://doi.org/10.1073/pnas.1104551108

11. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol 25 (7):1423–1428. https://doi.org/10.1111/mec.13549Digital Object Identifier (DOI)
12. Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. Mol Ecol Resour 13 (6):969–975. https://doi.org/10.1111/1755-0998.12046
13. Dong SS, Wang YL, Xia NH, Liu Y, Liu M, Lian L, Li N, Li LF, Lang XA, Gong YQ, Chen L, Wu E, Zhang SZ (2022) Plastid and nuclear phylogenomic incongruences and biogeographic implications of *Magnolia* s.l. (Magnoliaceae). J Syst Evol 60 (1):1–15. https://doi.org/10.1111/jse.12727
14. Dormontt EE, Boner M, Braun B et al (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging Biol Conserv, 191:790–798. https://doi.org/10.1016/j.biocon.2015.06.038
15. Gibbs RA, Cockerill M (1995) Working on the assembly line - Sequencher for the Macintosh (Version 3.0, March, 1995) Trends Biochem Sci 20 (4):162–163. https://doi.org/10.1016/S0968-0004 (00)88995-1
16. Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond Ser B-Biol Sci 270:313–321. https://doi.org/10.1098/rsbl.2003.0025
17. Hollingsworth PM, Forrest LL, Spouge JL et al (2009) A DNA barcode for land plants. Proc Natl Acad Sci U S A 106 (31):12794–12797. https://doi.org/10.1073/pnas.0905845106
18. Huang CJ, Chu FH, Huang YS et al (2020) Development and technical application of SSR-based individual identification system for Chamaecyparis taiwanensis against illegal logging convictions. Sci Rep 10 (1). https://doi.org/10.1038/s41598-020-79061-z
19. Huang DI, Cronk QCB (2015) Plann: A command-line application for annotating plastome sequences. Appl Plant Sci 3 (8):1500026. https://doi.org/10.3732/apps.1500026
20. Huang YP, Han YJ, Mao RW et al (2025) Water Conservation Capacity of Soil and Litter Layers of Five Magnoliaceae Plants in Hainan Island, China. Forests 16 (3):514. https://doi.org/10.3390/f16030514
21. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21 (1):241. https://doi.org/10.1186/s13059-020-02154-5
22. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30 (4):772–780. https://doi.org/10.1093/molbev/mst010
23. Kearse M, Moir R, Wilson A et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12):1647–1649. https://doi.org/10.1093/bioinformatics/bts199
24. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome 54 (8):663–673. https://doi.org/10.1139/g11-026
25. Lee YJ, Lee YM, Lee CK, Jung JK, Han SB, Hong JT (2011) Therapeutic applications of compounds in the *Magnolia* family. Pharmacol Ther 130 (2):157–76. https://doi.org/10.1016/j.pharmthera.2011.01.010
26. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49 (W1):W293-W296. https://doi.org/10.1093/nar/gkab301
27. Li HT, Yi TS, Gao LM et al (2019) Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants, 5 (5):461–470. https://doi.org/10.1038/s41477-019-0421-0
28. Li JL, Wang S, Yu J, Wang L, Zhou SL (2013) A modified CTAB protocolfor plant DNA extraction. Chinese Bulletin of Botany 48:72–78.
29. Li TZ, Zhang SY, Deng YF, Li YL (2024) Comparative Analysis of Chloroplast Genomes for the Genus *Manglietia* Blume (Magnoliaceae): Molecular Structure and Phylogenetic Evolution. Genes 15 (4):406. https://doi.org/10.3390/genes15040406
30. Li XW, Yang Y, Henry RJ, Rossetto M, Wang YY, Chen SL (2015) Plant DNA barcoding: from gene to genome. Biol Rev 90 (1):157–166. https://doi.org/10.1111/brv.12104Digital Object Identifier (DOI)
31. Little DP, Stevenson DW (2013) A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 29 (2):217–227. https://doi.org/10.1111/j.1096-0031.2006.00126.x
32. Luo AR, Lan HQ, Ling C, Zhang AB, Shi L, Ho SYW, Zhu CD (2015) A simulation study of sample size for DNA barcoding. Ecol Evol 5 (24), 5869–5879. https://doi.org/10.1002/ece3.1846
33. McCartin LJ, Vohsen SA, Wood AL et al (2026) Accounting for Intra- and Intergenomic Sequence Variation in Reference Barcodes Improves eDNA Metabarcoding Biodiversity Assessment. Mol Ecol Resour 26 (3). https://doi.org/10.1111/1755-0998.70130
34. Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3 (12):e422. https://doi.org/10.1371/journal.pbio.0030422
35. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37 (5):1530–1534. https://doi.org/10.1093/molbev/msaa015
36. Muranishi S, Tamaki I, Setsuko S, Tomaru N (2013) Asymmetric introgression between *Magnolia stellata* and *M. salicifolia* at a site where the two species grow sympatrically. Tree Genet Genomes 9:1003–1015. https://doi.org/10.1007/s11295-013-0612-1
37. Orsholm J, Zito A, Somervuo P, Harrison J, Koskela M, Ovaskainen O, Braga MP, Chazot N, Roslin T, Furneaux B (2026) Discovering the unseen: A performance comparison of taxonomic classification methods for unknown DNA barcodes. Methods Ecol Evol 00:1–20. https://doi.org/10.1111/2041-210x.70358
38. Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. Bmc Biol 7 (84). https://doi.org/10.1186/1741-7007-7-84

39. Perveen M, Ashraf S, Masoodi KZ (2026) Genetic fingerprints of flora: revolutionizing plant identification with DNA barcoding. Genet Resour Crop Evol 73 (4):162. https://doi.org/10.1007/s10722-026-02774-z
40. Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5 (1):65–84.
41. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins S, Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34 (12):3299–3302. https://doi.org/10.1093/molbev/msx248
42. Shah P, Jain N, Gawande N, Sharma T, Devanathan K, Sankaranarayanan S, Balaji R (2026) Advancing plant DNA barcoding: integrating chloroplast genome sequencing, cryptic diversity discovery and machine learning. Mol Biol Rep 53 (1):550. https://doi.org/10.1007/s11033-026-11736-8
43. Stecher G, Suleski M, Tao Q, Tamura K, Kumar S (2026) MEGA 12.1: Cross-Platform Release for macOS and Linux Operating Systems. J Mol Evol 94 (1):14–18. https://doi.org/10.1007/s00239-025-10287-z
44. Sun LY, Nie TJ, Chen Y, Li J, Yang AX, Yin ZF (2023) Gene identification and tissue expression analysis inform the floral organization and color in the basal angiosperm *Magnolia polytepala* (Magnoliaceae). Planta 257 (4). https://doi.org/10.1007/s00425-022-04037-4
45. Testé E, Hernández-Rodríguez M, Veltjen E, Bécquer ER, Rodríguez-Meno A, Palmarola A, Samain MS, González-Torres LR, Robert T (2022) Integrating morphological and genetic limits in the taxonomic delimitation of the Cuban taxa of *Magnolia* subsect. *Talauma* (Magnoliaceae). PhytoKeys 213:35–66. https://doi.org/10.3897/phytokeys.213.82627
46. Tukey JW (1977) Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts
47. Wang YB, Liu BB, Nie ZL, Chen HF, Chen FJ, Figlar RB, Wen J (2020) Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences via genome skimming. J Syst Evol 58 (5):673–695. https://doi.org/10.1111/jse.12588
48. Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). Front Zool 4 (8). https://doi.org/10.1186/1742-9994-4-8
49. Wohlmuth H, Leach D, McGrath K, Gordon L, Mouatt P, De Voss J (2016) An assessment of the utility of DNA barcoding in the quality control of herbal raw materials, extracts and finished products. Planta Med 82 (S01): S1-S381. https://doi.org/10.1055/s-0036-1597032
50. Wu L, Wu ML, Cui N, Xiang L, Li Y, Li XW, Chen SL (2021) Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. Chin Med 16 (52). https://doi.org/10.1186/s13020-021-00460-z
51. Wyler SC, Naciri Y (2016) Evolutionary histories determine DNA barcoding success in vascular plants: seven case studies using intraspecific broad sampling of closely related species. BMC Evol Biol 16 (103):1–15. https://doi.org/10.1186/s12862-016-0678-0
52. Xie HH, Tang YG, Fu J, Chi XL, Du WH, Dimitrov D, Liu JQ, Xi ZX, Wu JY, Xu XT (2022a) Diversity patterns and conservation gaps of Magnoliaceae species in China. Sci Total Environ 813. https://doi.org/10.1016/j.scitotenv.2021.152665
53. Xie HH, Zhang L, Zhang C, Chang H, Xi ZX, Xu XT (2022b) Comparative analysis of the complete chloroplast genomes of six threatened subgenus *Gynopodium* (*Magnolia*) species. Bmc Genomics 23:716. https://doi.org/10.1186/s12864-022-08934-6
54. Yu H, Wu KY, Song JY, Zhu YJ, Yao H, Luo K, Dai Y, Xu SJ, Lin YL (2014) Expedient identification of Magnoliaceae species by DNA barcoding. Plant Omics 7 (1):47–53.
55. Yu XQ, Jiang YZ, Folk RA, Zhao JL, Fu CN, Fang L, Peng H, Yang JB, Yang SX (2022) Species discrimination in Schima (Theaceae): Next-generation super-barcodes meet evolutionary complexity. Mol Ecol Resour 22 (8):2995–3011. https://doi.org/10.1111/1755-0998.13683

Figures

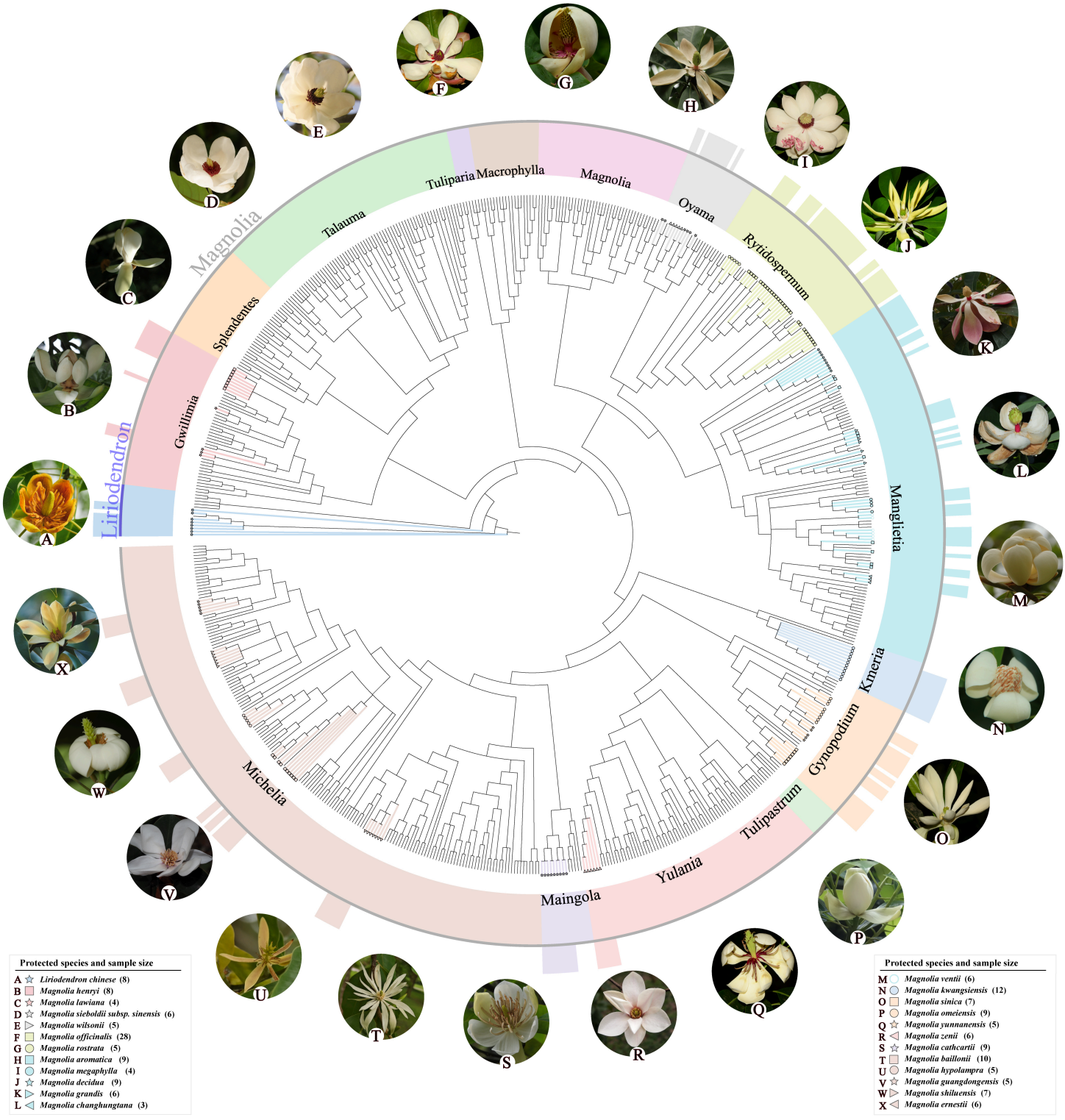


Figure 1

Maximum likelihood (ML) phylogeny of Magnoliaceae reconstructed from 692 complete plastome sequences. *Liriodendron* is used as the outgroup; branch colors denote the sections of *Magnolia* s.l. following Wang et al. (2020), and 24 protected species are highlighted with different symbols. Photos: Jun Liu (A, L, R, X); Yechun Xu (B, C, K, M, V, W); Cehong Li (D, N, S); Qin Tian (E); Xinxin Zhu (F, I, O, Q); Pin Zhang (G); Xirong Zheng (H); Ang Liu (J); Buyun Zhang (P); Xinyang Chen (T); Meng Wei (U)

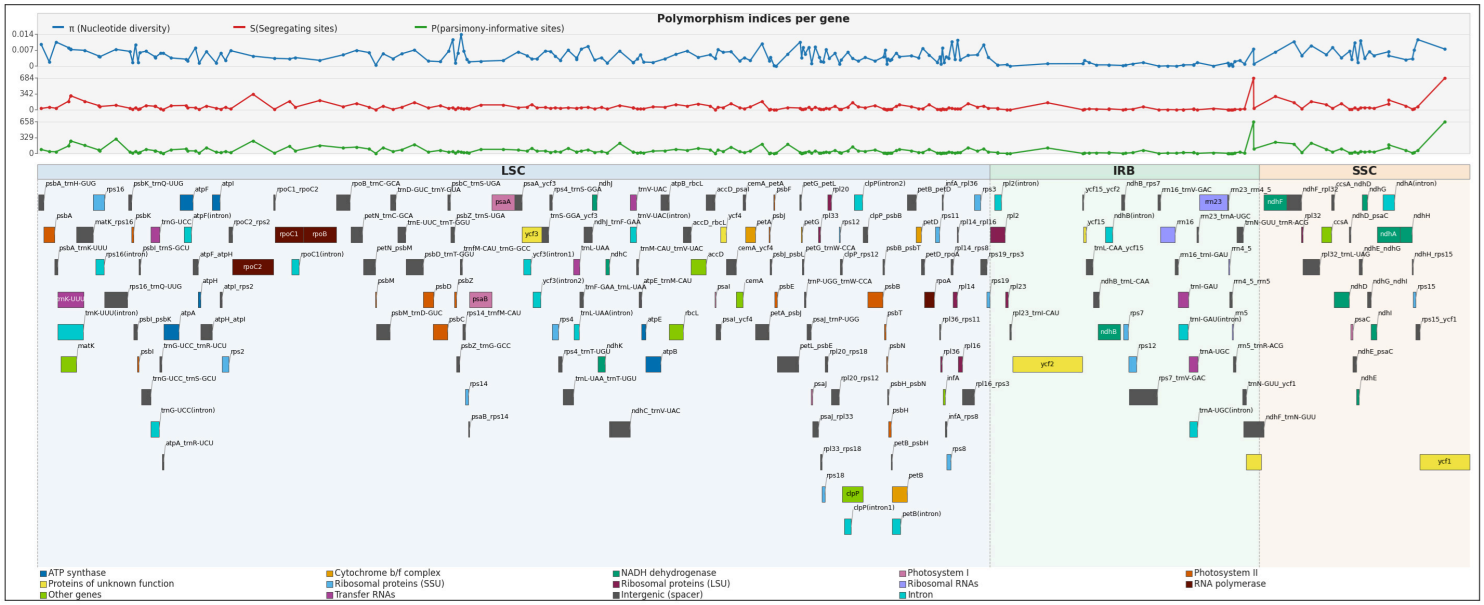


Figure 2

Genomic positional distribution of the partitioned chloroplast fragments and the variation trends of three diversity parameters (nucleotide diversity π, number of segregating sites S, and number of parsimony-informative sites P) along the plastome. In the chloroplast genome map partitioned into LSC, IRB, and SSC regions, the following features are shown in distinct colors: ATP synthase, Cytochrome b/f complex, NADH dehydrogenase, Photosystem I, Photosystem II, proteins of unknown function, ribosomal proteins (SSU and LSU), ribosomal RNAs, RNA polymerase, transfer RNAs, other genes, intergenic spacers, and introns

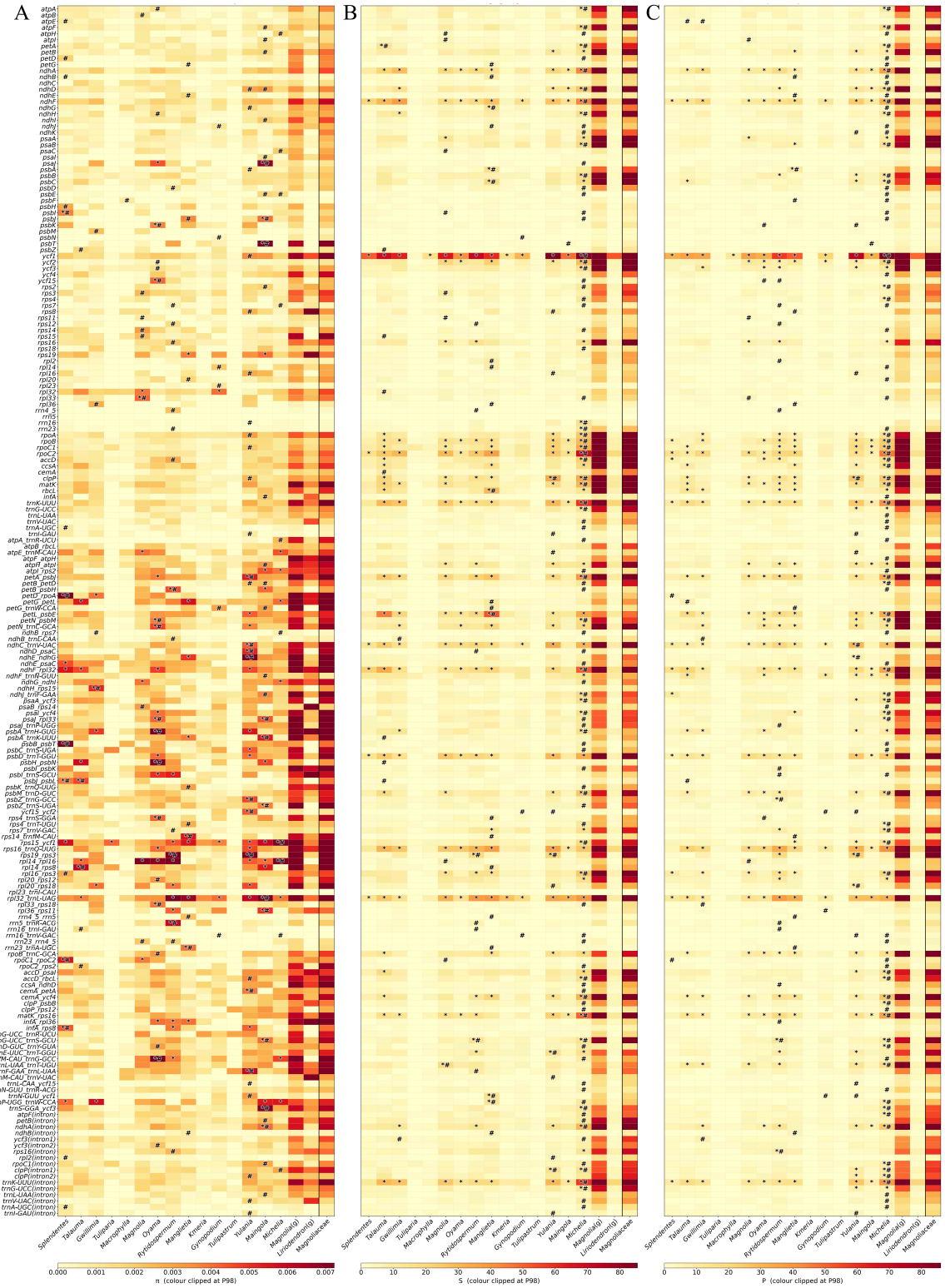


Figure 3

Polymorphism heatmap of the three diversity parameters (π, S, P) for the three fragment types (Genes, IGSs, introns) across 18 taxonomic groups (15 sections, genus *Magnolia* (g), genus *Liriodendron* (g), and Magnoliaceae as a whole). Panel A shows nucleotide diversity (π), panel B shows number of segregating sites (S), panel C shows number of parsimony-informative sites (P). Color intensity increases with the parameter value, the uppermost 2% of the values are shaded with an identical dark color. Asterisks (*) and number signs (#) mark loci meeting the global Tukey outlier criterion and the group-wise Z-score criterion

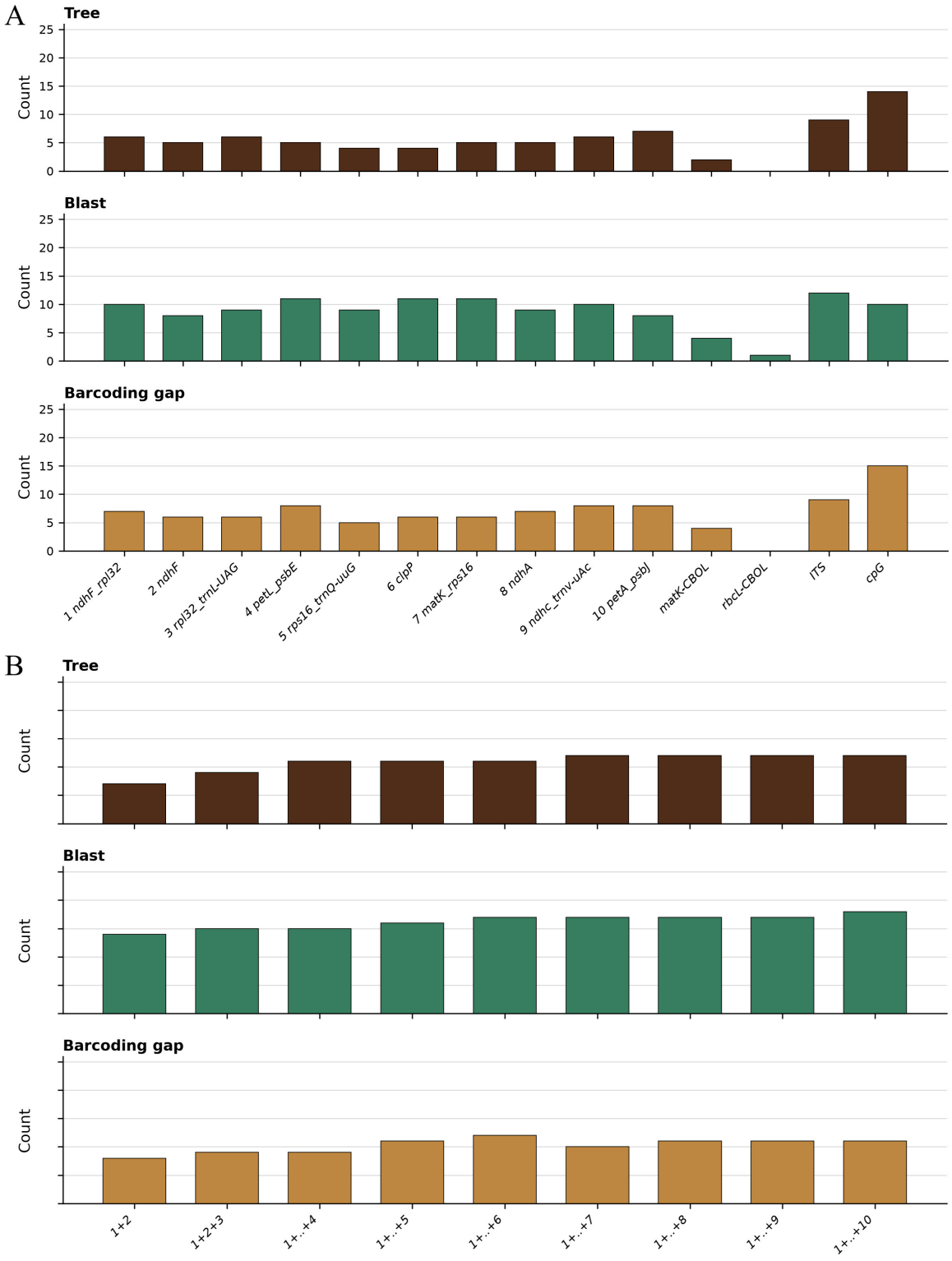


Figure 4

Comparison of species-level identification results across different fragments for 24 protected Magnoliaceae species. The y-axis shows the number of successfully identified species (out of 24), and the x-axis represents different barcode fragments. A) on the left shows identification results of the single fragments, comprising ten manually selected hypervariable regions, the standard barcodes (*rbcL*, *matK*, ITS), and the plastid genome super-barcode (cpG). B) on the right shows identification results of the ten selected fragments concatenated sequentially. Number 1 to 10 represent the fragments of *ndhF-rpl32*, *ndhF*, *rpl32-trnL*-UAG, *petL-psbE*, *rps16-trnQ*-UUG, *clpP*, *matK-rps16*, *ndhA*, *ndhC-trnV*-UAC, *petA-psbJ*, respectively. The results from the tree-based, BLAST-based, and barcode gap-based methods are shown in different colors

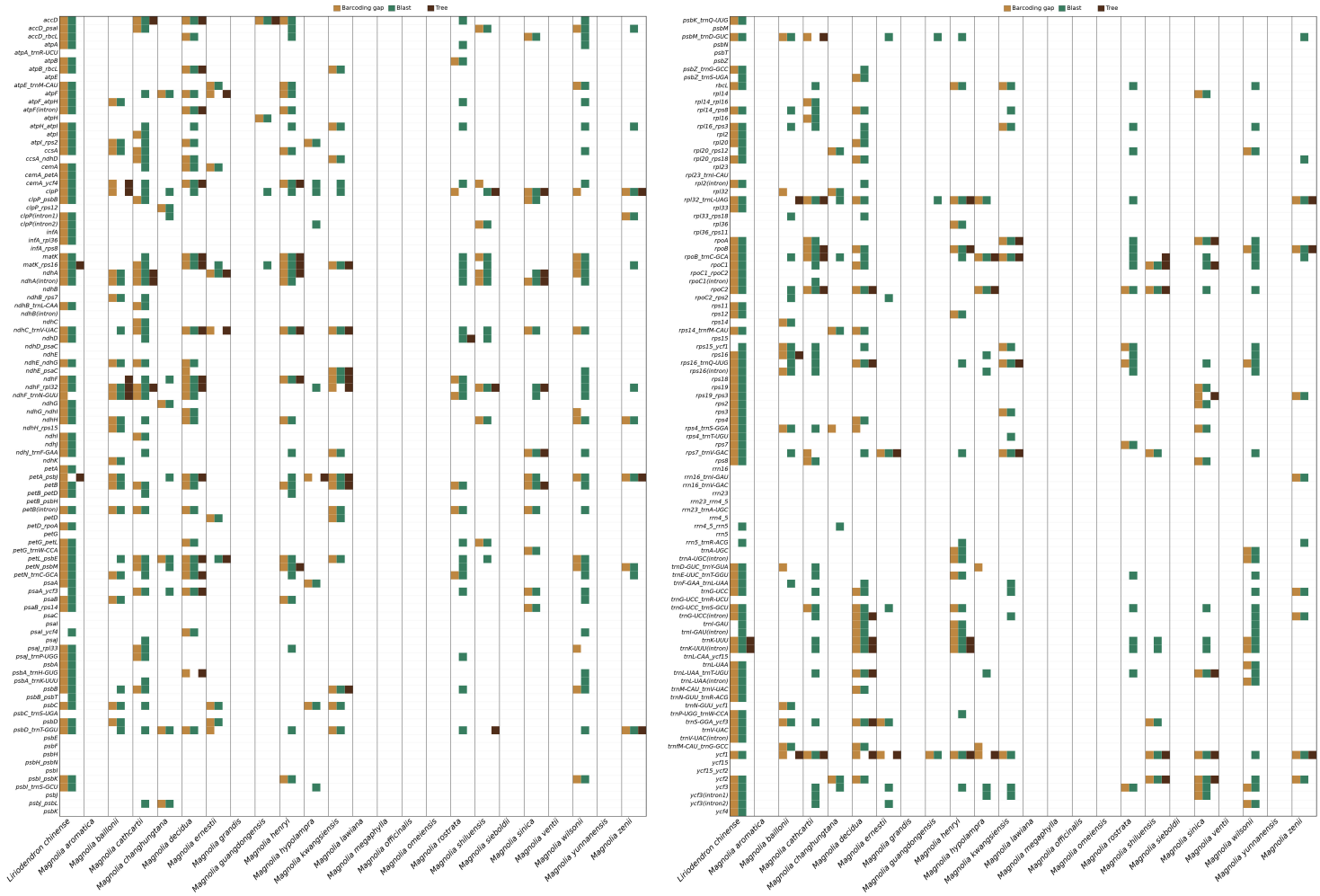


Figure 5

Grid diagram of chloroplast fragments available for successfully identifying the 24 protected Magnoliaceae species. The x-axis lists the 24 species, and the y-axis lists the 196 candidate chloroplast fragments retained from the plastome. Cells are shaded if the fragment successfully discriminated the species by any of the three methods (Tree, BLAST, or Barcode gap-based method). White cells denote unsuccessful identifications

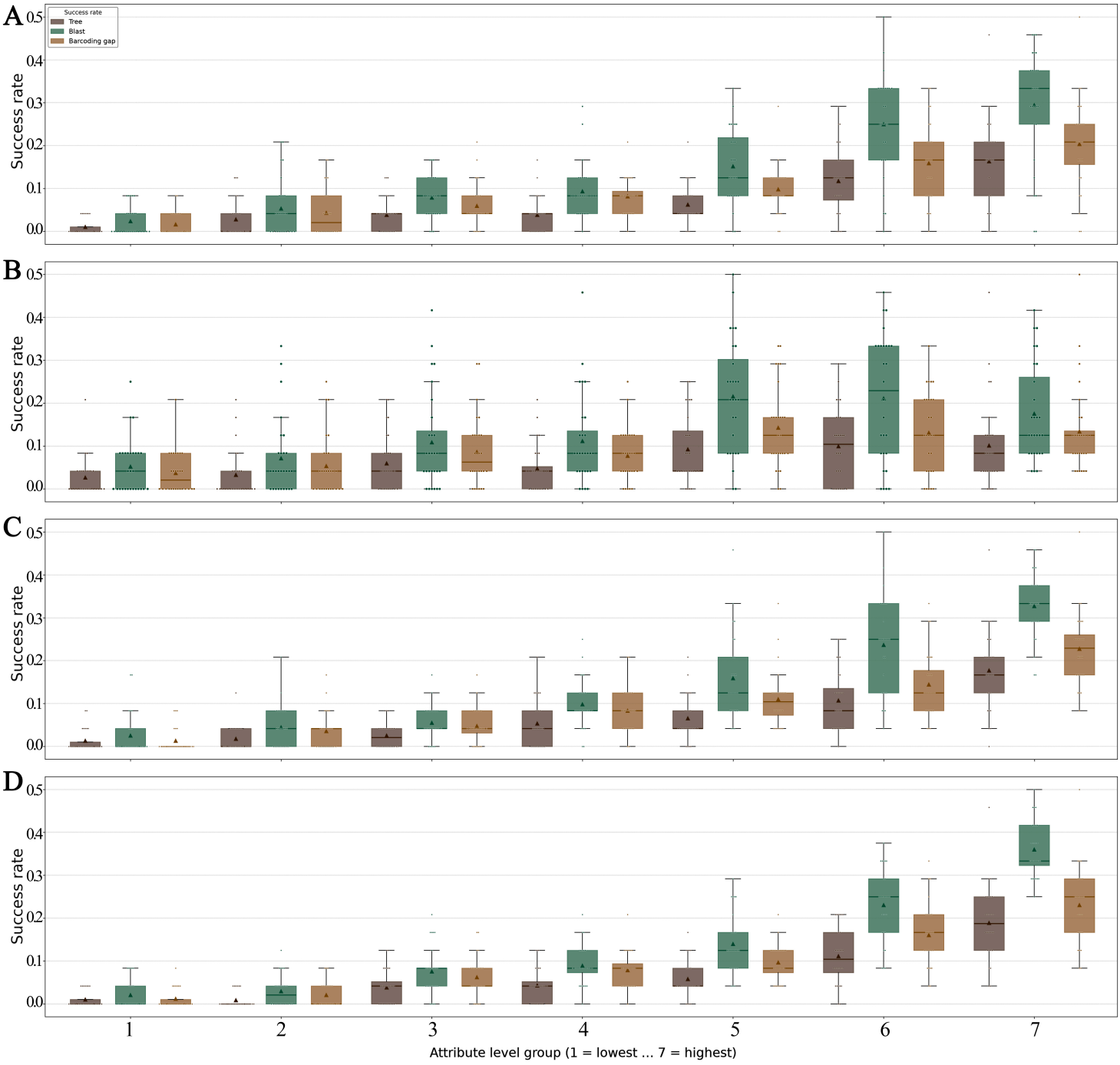


Figure 6

Correlation analysis between species-level identification success rates of 196 chloroplast fragments and four indices: A) sequence length, B) nucleotide diversity (π), C) number of segregating sites (S), and D) number of parsimony-informative sites (P). The 196 fragments are evenly divided into seven groups (n = 28 per group) in ascending order of each index. Identification success rates within each group are shown as boxplots. For each index, results are plotted separately for the tree-based, BLAST-based, and barcode gap-based methods. Dots represent individual fragment values and triangles indicate mean values per group

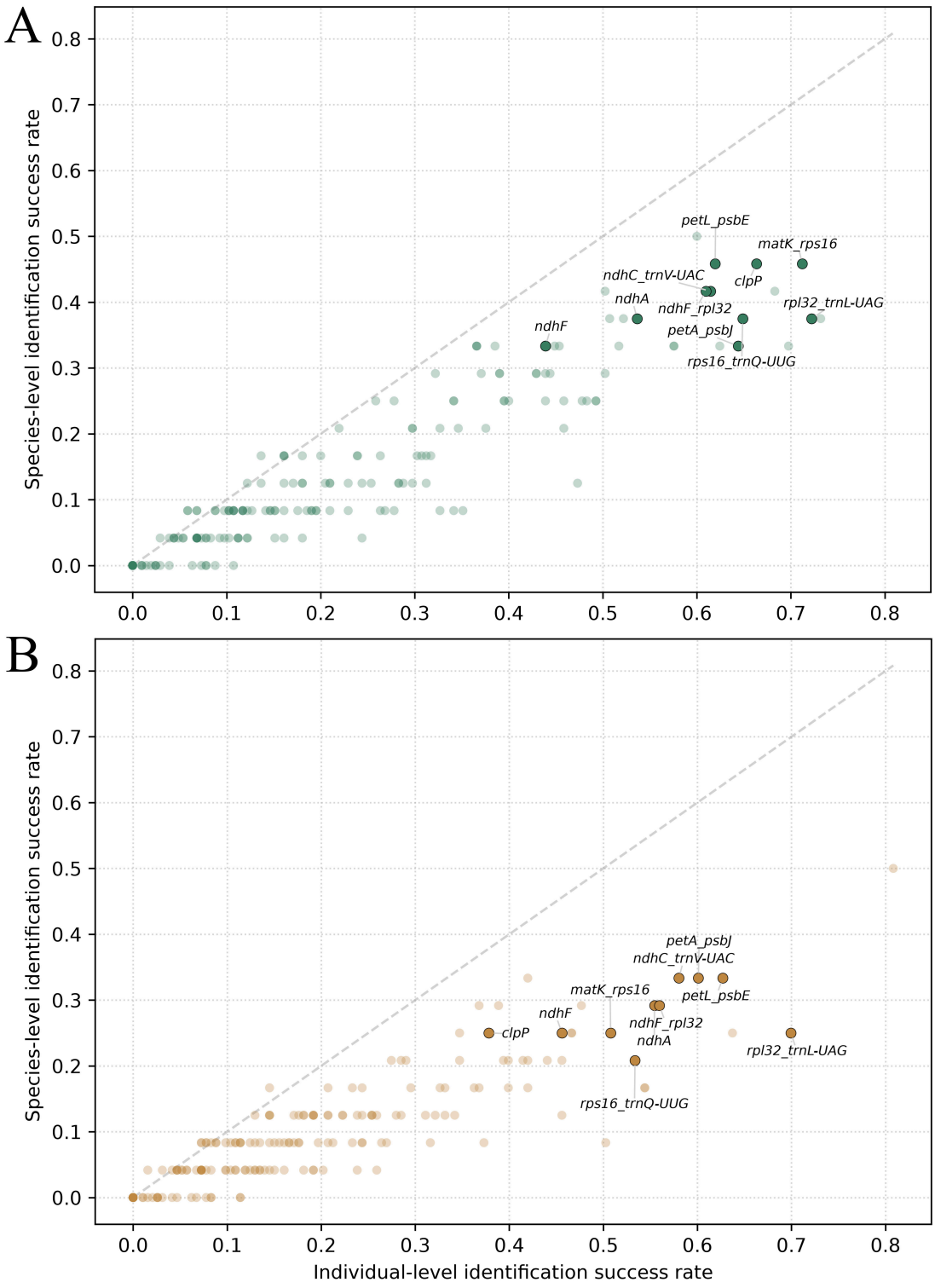


Figure 7

Scatter plot of the correlation between species-level and individual-level identification success rates for 196 fragments. The x-axis represents the individual-level identification success rate of each fragment, and the y-axis represents the species-level identification success rate. All 196 fragments are plotted in the coordinate system, with A based on the BLAST method and B based on the barcode gap method. The names of ten selected fragments are noted specifically in the plot. The dashed line represents the y = x reference line

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS3Variationtrendsofnucleotidediversityalongthesequentialorderofchloroplastgenomefragmentsacrossalltaxonomicsections.jpg
- SupplementaryFigureS4Variationtrendsofthenumberofvariablesites Salongthesequentialorderofchloroplastgenomefragmentsacrossalltaxonomicsections
- SupplementaryTableS1.Summaryofsamplesandnewlyobtainedsequencingdatasetsgeneratedinthepresentstudy.xlsx
- SupplementaryFigureS5Variationtrendsofthenumberofparsimonyinformativesites Palongthesequentialorderofchloroplastgenomefragmentsacrossalltaxo
- SupplementaryTableS2.PublicMagnoliaceaesequencedatadownloadedfromNCBI.xlsx
- SupplementaryTableS3.Valuesofthreediversityparameters.xlsx
- SupplementaryTableS4.Tukeysmultiplecomparisontestfornucleotidediversitysegregatingsites Sandpolymorphicsites Pacrosstaxonomicgroupsin Magnolia
- SupplementaryTableS5.Zvaluestatisticsacross15groupsforthreediversityparameters.xlsx
- SupplementaryTableS6.Specieslevelmonophylybranchsupportforthe24protectedMagnoliaceaespeciesbasedonthetreemethod.xlsx
- SupplementaryTableS7.Specieslevelidentificationsuccessratesofthe24protectedMagnoliaceaespeciesbasedontheBLASTmethod.xlsx
- SupplementaryTableS8.Differencesbetweenminimuminterspecificandmaximumintraspecificdistancesforthe24protectedMagnoliaceaespeciesunderthe
- SupplementaryLegends.docx
- SupplementaryTableS9.Specieslevelidentificationsuccessratesofallfragmentsundertreebasedmethod BLASTmethodandbarcodegapbasedmethod.xlsx
- SupplementaryTableS10.Individuallevelidentificationresultmatrixunder theBLASTmethod.xlsx
- SupplementaryTableS11.Individuallevelidentificationresultmatrixbasedonthegeneticdistancenearestneighbormethod.xlsx