

Complete chloroplast genome sequence of *Pachystachys lutea* Nees: genome structure, adaptive evolution, and phylogenetic relationships

Changmei Du

Xinyang Agriculture and Forestry University

Yan Dong

Xinyang Agriculture and Forestry University

Haishuo Gao

Xinyang Camellia oleifera Industry Development Center

Tingting Fang

Xinyang Agriculture and Forestry University

Jianhua Yue

jhyues@163.com

Xinyang Agriculture and Forestry University

Yan Zhang

Shanghai Vocational College of Agriculture and Forestry

Research Article

Keywords: *Pachystachys lutea*, Acanthaceae, chloroplast genome, comparative analysis, phylogenetic analysis

Posted Date: April 17th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-5848411/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomic Data on November 19th, 2025. See the published version at <https://doi.org/10.1186/s12863-025-01380-9>.

Abstract

Background *Pachystachys lutea* Nees is a typical species of the family Acanthaceae, native to tropical South America. As an evergreen shrub, it has found extensive application in landscape greening due to its unique ornamental value. However, there are few available phylogenetic and genetic studies about the chloroplast (cp) genome of *P. lutea*.

Results This study characterized the cp genome of *P. lutea* by using Illumina sequencing technology and inferred the phylogenetic position of the species. The results indicated that the cp genome had a high degree of conservation in gene structure and gene content, with a typical quadripartite structure. Its total length is 151,574 bp and the total GC content is 38.18%. A total of 132 genes were annotated, including 87 protein-coding genes (PCGs), 37 tRNAs and eight rRNA genes. Through the comparative analysis, the diversity and variation of large single-copy (LSC) and small single-copy (SSC) regions were significantly higher than those of inverted repeat (IR) regions. Genes with high nucleotide polymorphism, such as *rps19*, *ycf1*, and *ndhF* provided potential reference loci for molecular identification within the *P. lutea*. The phylogenetic analysis showed that the *P. lutea* and *Clinacanthus nutans* forms a sister group with 100% bootstrap value, which proves that *P. lutea* develops conservatively in the course of evolution.

Conclusion This paper for the first time reports the phylogenetic study of the complete cp genome within the genus *Pachystachys*. The study provides a theoretical basis for the research on genetic diversity, molecular markers, and species identification of plants in the Acanthaceae family. It enriches the genetic information and supports the evolutionary relationships among plants in this family.

Introduction

The Acanthaceae family contains approximately 250 genera and more than 4000 species [1, 2], mainly distributed in tropical and subtropical areas [3, 4]. Approximately 50 genera and more than 400 species are distributed in China [5]. Many species of the Acanthaceae family are remarkable flowering ornamental plants, characterized by diverse forms, vivid colors, and extended ornamental periods. *Pachystachys lutea* Nees is one of them, is a perennial evergreen flowering shrub that prefers high temperatures, high humidity and sunny environments and is mainly distributed in the southern areas of the Yangtze River in China. The flower has a unique shape and blooms continuously throughout the year in warm environments (the greenhouse of Shanghai Chenshan Botanical Garden), so it is often planted as an ornamental plant. *P. lutea* has very large spikelike inflorescences at the top of each new branch, and the golden bracts are stacked and spread out with small white flowers like shrimp bodies. Furthermore, its roots, leaves, and flowers are also used for medicinal purposes, treating pneumonia [6], diarrhea, worms [7], etc., and are the important medicinal plant resources to be developed. In addition, it has economic value. The leaves serve as the source of endophytic fungi [8, 9] and blue dye [10], and nectar has the potential to serve as a substitute for sugar production in the future [11]. Currently, there are relatively few species within the genus *Pachystachys* that have been named, yet its outstanding ornamental characteristics have made *P. lutea* highly recommended in China, where it has become an

extremely valuable ornamental plant resource, widely loved by the public, and effectively enhancing the diversity of landscape species applications.

The chloroplast (cp), an important organelle in green plants [12], houses the cp genome which is of great significance in plant evolutionary biology research. It presents remarkable benefits due to its genetic stability, well-preserved genome structure, and faster rate of evolutionary change compared to mitochondria. Consequently, the cp genome is widely utilized in the study of genetic relationships and species evolution between different species [13, 14]. Despite its significant ornamental and medicinal value, the complete cp genome of *P. lutea* has not yet been sequenced, limiting our understanding of its evolutionary history and genetic traits.

P. lutea is a species within the genus *Pachystachys* of the family Acanthaceae. Research on *P. lutea* has mainly focused on its horticultural applications, cultivation techniques, and medicinal properties, leaving its genetic and phylogenetic aspects largely unexplored. Nevertheless, scant knowledge exists regarding its phylogenetic relationship and evolutionary history, primarily due to the dearth of research on the modern taxonomic system, especially in the scarcely explored realm of molecular evolution. Scotland and Vollesen (2000) [15], initially applied molecular systematics in conjunction with floral organ development to establish a classification system for the family Acanthaceae. Through a comprehensive study of *trnL-trnF* and internal transcribed spacers (ITS), it was demonstrated that *Pachystachys* shares a common lineage with *Henrya*, *Carlowrightia*, *Anisacanthus*, and *Tetramurium*, and this branch is further grouped with the one encompassing *Dicliptera* and *Justicia*. However, the position of *P. lutea* within the Acanthaceae requires further validation.

This study aims to: (1) analyze the structural characteristics of the cp genome of *P. lutea*, (2) investigate its adaptive evolutionary traits, and (3) clarify its phylogenetic relationships with closely related species.

To achieve these aims, we employed molecular biology and high-throughput sequencing techniques to characterize the cp genome of *P. lutea*. This involved using the Illumina NovaSeq 6000 platform for sequencing and a suite of bioinformatics tools for genome assembly, annotation, and analysis. We also compared the results with those obtained from traditional morphological taxonomic methods. Traditional morphological taxonomic methods have limitations, often leading to ambiguous or inconsistent classifications. Studies at molecular level can provide more precise genetic information. By filling this research gap, we aim to lay a solid foundation for more accurately determining its taxonomic position and facilitating future research endeavors.

Results

Genome structure of the *P. lutea* cp genome

In this study, we assembled and annotated the complete cp genome, analyzed the sequencing depth, which reached an average of 3,012× sequencing depth (Fig. S1). Compared with that of other species [16, 17], this higher sequencing depth has enhanced the reliability of gene annotation and structural

analysis. The accuracy of the assembly and annotation of *P. lutea* cp genome was evaluated, respectively (Fig. S2; Fig. S3). The complete cp genome of *P. lutea* was 151,574 bp in length (Fig. 1), and it had a typical quadripartite structure with junction regions: a LSC region of 83,348 bp, an SSC region of 17,226 bp, and a pair of IR regions (IRa and IRb) of 25,500 bp each. The overall GC content of the complete *P. lutea* cp genome was 38.18%, and the corresponding values in the LSC, SSC, and IR regions were 36.21%, 32.36%, and 43.36%, respectively. The complete cp genome was found to contain 132 genes, which comprised 87 PCGs, 37 tRNA genes, and eight rRNA genes (Table 1). Nine PCGs, six tRNA genes, and four rRNA genes were duplicated in the IR regions. Nineteen genes contained two exons, and four genes (*clpP*, *ycf3*, and two *rps12*) contained three exons. Among 87 PCGs, 45 photosynthesis genes, 29 genes related to self-replication, six other genes and seven genes with unknown function were identified.

Table 1
Gene function statistics table of chloroplast genome in *P. lutea*.

Gene category	Gene function	Gene name
Photosynthesis gene	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>
	Subunits of photosystem II	<i>psbL, psbZ, psbM, psbN, psbA, psbB, psbC, psbD, psbE, psbF, psbT, psbH, psbl, psbJ, psbK</i>
	Subunits of NADH-dehydrogenase	<i>ndhG, ndhH, ndhI, ndhJ, ndhK, ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF</i>
	Subunits of cytochrome b/f complex	<i>petL, petN, petA, petB*, petD*, petG</i>
	Subunit for ATP synthase	<i>atpI, atpA, atpB, atpE, atpF*, atpH</i>
	Large subunit of rubisco	<i>rbcL</i>
Self-replication	Large subunit of ribosome	<i>rpl20, rpl22, rpl32, rpl23(2), rpl14, rpl33, rpl16*, rpl36, rpl2*(2)</i>
	Small subunit of ribosome	<i>rps11, rps14, rps15, rps16*, rps2, rps3, rps18, rps4, rps19, rps7(2), rps8, rps12**(2)</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	rRNA gene	<i>rrn5(2), rrn4.5(2), rrn16(2), rrn23(2)</i>
	tRNA gene	<i>trnR-UCU, trnE-UUC, trnT-GGU, trnS-GGA, trnV-GAC(2), trnR-ACG(2), trnL-UAA*, trnG-GCC, trnD-GUC, trnY-GUA, trnP-UGG, trnM-CAU, trnL-CAA(2), trnS-GCU, trnW-CCA, trnF-GAA, trnT-UGU, trnS-UGA, trnV-UAC*, trnG-UCC*, trnL-UAG, trnI-GAU*(2), trnH-GUG, trnH-CAU(2), trnQ-UUG, trnN-GUU(2), trnK-UUU*, trnA-UGC*(2), trnC-GCA</i>
Other genes	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP**</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of Acetyl-carboxylase	<i>accD</i>
	C-type cytochrome	<i>ccsA</i>

Gene category	Gene function	Gene name
	synthesis gene	
Unknown gene	Open reading frames (ORF,ycf)	<i>ycf1, ycf2(2), ycf3**, ycf4, ycf15(2)</i>
Notes: Gene *: Gene with one intron and two exons; Gene **: Gene with two introns and three exons; Gene (2): Number of copies of multi-copy genes;		

Simple sequence repeats analysis

SSRs analysis identified 47 SSRs in the *P. lutea* cp genome. They were mainly of three types: mononucleotides, dinucleotides, trinucleotides, tetranucleotide, pentanucleotide, and hexanucleotide types were not found. Among them, the mononucleotides repeat sequence (A/C/T/G) type was the most abundant, with 40, which were mainly dominated by A and T bases; there were six dinucleotide repeat sequences (AT/TA), and one trinucleotide repeat sequence (TCT) (Fig. 2). Out of 47 SSR loci, 37 were concentrated in the LSC region, eight were located in the SSC region, and only two were located in the IR regions (Table S1).

Codon preference analysis

A total of 24,537 codons in 87 PCGs of the cp genome of *P. lutea* were participated in translation protein expression. Codons with a RSCU over 1 are thought to be favored by amino acids. Overall, 31 codons had an RSCU > 1.0, with 29 codons ending in A or T and two codons ending in G or C. In addition, among the amino acid codes, leucine (Leu) had the highest encoding rate, with six synonymous codon codes (TTA, TTG, CTT, CTC, CTA, CTG), with a total of 2,629. The encoding rate of cysteine (Cys) was low, and there were two synonymous codons (TGT, TGC), totaling 279. methionine (Met) and tryptophan (Trp) had only one codon, while the rest of the amino acids had two to six codons (Fig. 3; Table S2). In addition, the ENC value of the PCGs was 45.83, and the GC1, GC2 and GC3 contents were 47.20%, 40.07% and 28.62%, respectively. Neutral-plot analysis showed that all genes were located above the diagonal line, indicating a weak correlation between GC12 (average of GC1 and GC2) and GC3; ENC-plot analysis showed that most of the genes were located below the standard curve and far away from the curve; and PR2-plot analysis showed that the genes were unevenly distributed in the four regions, and most of the genes were located in the lower right of the plot (Fig. 4). These results indicated that the codon preference of *P. lutea* was influenced more by natural selection [18–20].

IR/SC boundary comparison analysis

The sequence of the IR boundary regions may extend outward and expand inward, resulting in changes in the copy number of related genes or the generation of pseudogenes in the boundary regions, which is a common phenomenon in the evolution of the cp genome. In this study, we compared the boundary expansion and contraction of IR and SSC regions in eight species from six genus of the family

Acanthaceae (Fig. 5). The results revealed little variation in the size of the cp genomes across the eight species, with the exception of *S. cusia*, which exhibited a reduction of 5,094 to 7,536 bp compared to the other species. Similarly, the size of the IR regions of *S. cusia* showed a distinct shorter length. In six species (except *C. nutans* and *S. cusia*), *rps19* straddled the IRb/LSC (JLB) boundary, *ndhF* straddled the IRb/SSC (JSB), and *ycf1* straddled the IRa/SSC (JSA). *rpl2* was located within the IR boundary, and *trnH* was located near the LSC-IRa boundary junction. The changes in genes at the four boundaries of *P. lutea* were consistent with two species of *Dicliptera* genus, two species of *Justicia* genus and one species of *Peristrophe* genus. Compared with *C. nutans*, *rps19* at the JLB boundary expanded to the LSC region but contracted in the IRb direction, and *ndhF* showed obvious contraction in the IRb direction. The IR/SC borders were validated using Sanger sequencing [21], which was completely consistent with the sequencing results from the Illumina sequencing platform (Fig. S4).

Analysis of cp genome sequence variation

Visual analysis of the similarity among the eight cp genome sequences showed that there was little difference in the gene interval composition of the cp genomes of Acanthaceae plants, which was relatively consistent (Fig. 6). Among the four regions, the LSC region had the highest variability in changes, followed by SSC region, and the IR regions had the lowest variability and was the most conserved, which was consistent with the results of boundary analysis. From the perspective of the non-gene coding regions and gene coding regions, the degree of variation in the non-gene coding regions was relatively high, while the gene coding regions were relatively conserved. However, there was a significant difference in the degree of variation in the gene coding regions, such as those of *rps19*, *ycf1*, *ndhF*, *ndhA*, *clpP*, and *ycf3*. This indicated the existence of single nucleotide polymorphisms (SNPs) in the coding regions of genes such as *rps19*, *ndhF* and *ycf1*, which was further confirmed by gene sequence comparison (Fig. S5).

Substitution rates of PCGs

The Ka/Ks ratio in genetics is used to evaluate the existence of selection pressure on a certain PCG during evolution. A Ka/Ks value greater than 1 implies positive selection. A Ka/Ks value of 1 suggests neutral selection. When the Ka/Ks ratio is below 1, it signifies negative selection [22]. Selected 61 PCGs from *P. lutea* cp genome were compared with seven species from Acanthaceae for Ka/Ks calculation (Fig. 7; Table S3). To enhance the presentation of the data, only Ka/Ks values greater than 0.5 were shown. Overall, most genes of *P. lutea* cp genome experienced negative selection throughout evolution, as shown by the Ka/Ks values of 58 PCGs, being less than 1 when compared to other seven species. The high Ka/Ks values of *petD* from *P. lutea* compared to two *Dicliptera* species suggested that positive selection occurred during evolution. Furthermore, *petA* exhibited positive selection when comparing *P. lutea* with *P. japonica*, whereas *rpl20* showed positive selection when comparing *P. lutea* with *C. nutans*.

Phylogenetic analysis

To elucidate the evolutionary dynamics within the family Acanthaceae, particularly focusing on *P. lutea*, a comprehensive phylogenetic analysis was conducted, encompassing *P. lutea* and 34 other samples from 16 distinct genus, with two species assigned as outgroups for comparative context. This investigation was founded on the analysis of conserved PCGs (Fig. 8) and cp genomes (Fig. S6), facilitating a dual perspective on phylogenetic relationships. Overall, both phylogenetic trees exhibited the same evolutionary pattern. Namely, the species from the *Pseuderanthemum*, *Clinacanthus*, *Pachystachys*, *Justicia*, *Hypoestes*, *Peristrophe*, and *Dicliptera* exhibited well-defined clustering, indicating a clear evolutionary relationship among them. *P. lutea* and *C. nutans* were found to be especially closely related, as evidenced by a bootstrap support (BS) value of 100, which strongly supports the robustness of this phylogenetic relationship.

Discussion

The study of cp genome sequences gives extensive information for phylogenetic research of plants, and DNA barcoding utilizing cp markers allows for accurate identification of plant species [12]. In this study, the cp genome of *P. lutea* was sequenced, assembled and annotated by high-throughput sequencing technology, and its genome structure, SSR sites, codon preference and phylogeny were analyzed. The findings will provide valuable insights into the evolutionary relationships and genetic variability of *P. lutea*, which is essential for its preservation. Genomic data may provide references for the development of plans for habitat restoration and the preservation of genetic reservoirs.

The complete cp genome obtained had a total length of 151,574 bp, a total of 132 annotated genes, and a GC content of 38.18%. Compared with the reported homologous plants such as *C. nutans* (151,669 bp, 38.40%) [23], *D. peruviana* (150,811 bp, 38.00%) [24], *D. mucronata* (150,720 bp, 38.00%) [24], *J. flava* (150,888 bp, 38.20%) [25] and *P. japonica* (151,374 bp, 38.07%) [26], the Acanthaceae plants have highly similar genome sizes, structures, compositions, and GC contents, indicating that the plants of the Acanthaceae family exhibit good conservatism during evolution.

SSRs are widely distributed in most plants, mainly in the external and noncoding regions of genes, and are often used in species identification, genetic diversity analysis and molecular marker-assisted breeding [27]. In this study, a total of 47 SSR sites were found in the cp genome of *P. lutea*, of which mononucleotides were the most abundant (85.11%), followed by dinucleotides (12.76%). The repetitive units of SSRs were mainly composed of A and T base combinations, with fewer G/C repeats. This further indicated that SSRs in cp genomes exhibit significant AT preference, which may be related to the difficulty of AT and GC chain uncoupling. Combined with the results of cp genome sequence variation analysis, the degree of variation in the IR regions with higher GC content was significantly lower than that in the LSC and SSC regions with lower GC content. Therefore, it can be inferred that base preference may be positively correlated with the degree of sequence variation, and it also indicated that the structure of the cp genome of *P. lutea* was highly conserved. Consequently, the identified SSR loci hold great potential as they can offer a solid theoretical basis for the identification of *P. lutea*, phylogenetic analysis, and the development of molecular markers related to this particular plant species.

Codons serve as a crucial link among nucleic acids, proteins, and genetic material, thereby playing a significant role in the transmission of genetic information within organisms. Their preferred usage patterns offer dependable information for investigations into gene function, species evolution, and other related aspects [28, 29]. In the present study, Leu was identified as the most abundant amino acid, accounting for 10.71% within the cp genome of *P. lutea*. The cp genome of *P. lutea* demonstrated a preference for using codons ending with A/T. The occurrence of codon usage bias during the evolution of the cp genome is attributed to natural selection and mutations [30]. It is well-documented that an ENc value greater than 40 implies a weak codon preference [31, 32]. Given that the GC3 contents were less than 50%, it could be inferred that the codons in this genome tended to utilize A and T bases. This finding was in alignment with the results of the codon preference analysis conducted on *J. flava*, suggesting that the closer the phylogenetic relationship between species, the more similar their codon usage preferences are. This further corroborated the conclusions put forward by Parvathy et al [33]. Moreover, through Neutral-plot, ENC-plot, and PR2-plot analyses, it was determined that natural selection constitutes the primary factor influencing the codon usage bias observed in *P. lutea*.

The IR region of the cp genome is considered to be the most conserved region and plays an important role in maintaining the stability of the cp genome [34]. The contraction, expansion and deletion of IR boundaries can cause differences in cp genomes [35]. Comparisons between LSC-IR and SSC-IR boundaries, and the genome sequence variation in the eight species' complete cp genomes showed that the LSC/SSC regions had higher variability, while the IR regions had lower variability, but the whole genome was still relatively conserved. In addition, there was variation among genomes, variation was manifested as the location and number of base pairs in the borders of four genes, *rps19*, *ndhF*, *ycf1*, and *trnH*. This may be caused by the instability of LSC-IR and SSC-IR boundaries and the different degrees of expansion and contraction during the historical evolution of species. The *rps19* gene is a key component of the ribosome biogenesis process, which is of vital importance for plant protein synthesis. Any variation in *rps19* may have an impact on the overall growth rate of the plant. The *ycf1* is located on the plastid membrane and promotes the transmembrane transport of various molecules. Mutations in *ycf1* will alter the plant's response to environmental stress. The *ndhF* gene is associated with the NADH dehydrogenase-like complex, which is involved in the electron transport during photosynthesis. Changes in *ndhF* may affect the photosynthetic efficiency of the plant. These genetic variations may have enabled *P. lutea* to adapt to its native tropical habitats in South America. They may also have contributed to the successful cultivation of *P. lutea* in the southern regions of the Yangtze River Basin in China. Additionally, these regions with partial differences can provide molecular bases for the identification and phylogenetic analysis of different species in the family Acanthaceae.

Adaptive evolution has a profound implication on the study of structural and functional variation of genes, and Ka/Ks is an effective method to evaluate whether the adaptive evolution of PCGs has occurred [36]. Ks occurs more frequently than Ka in most genes of organisms, so Ka/Ks values are usually less than 1 [37]. In this study, we detected that the majority of genes in Acanthaceae species had Ka/Ks < 1, indicating that the cp genes of Acanthaceae species had been subjected to strong purifying selection during the long evolutionary process. The Ka/Ks of photosynthetic related genes such as *atpE*,

ndhE, *psbH* and *psbJ* were less than 1, indicating that they were subjected to strong purifying selection during the evolutionary process. While *petD*, *petA* and *rpl20* showed positive selection in other species, indicating that the aforementioned genes had a strong influence on the evolution trend of different species. Furthermore, PCGs of the cp genome encodes many critical proteins involved in photosynthesis and other metabolic processes, playing a role in the development of plant defense against pathogen ingress, stress tolerance, and ornamental traits [38, 39]. For example, the photosynthesis gene *psbA* encodes a critical and highly conserved component of the photosystem II reaction center, polypeptide D1, which is participated in the photosynthetic electron transport chain. Under conditions of high-light stress, the increased production of the *psbA* protein helps to protect the photosynthetic machinery from damage caused by excess light energy, thereby maintaining photosynthetic efficiency and ensuring the plant's survival and growth. Thus, the transcription and translation of *psbA* play an important role in high-light stress responses [40]. Ribosomal proteins are essential for cell survival, among which *rps15* and *rpl33* are important components of the ribosome, responsible for protein synthesis in the cell, and their absence is also causally responsible for the high chilling sensitivity in plants [41]. For example, tobacco plants lacking the ribosomal proteins *rps15* or *rpl33* exhibited heightened sensitivity to cold stress [42]. In this study, *psbA*, *rps15* and *rpl33* genes were found in cp genome of *P. lutea*, which may play a role in photosynthesis and resistance. The *psbA* gene can maintain highly efficient photosynthesis under the intense light conditions typical of tropical regions, ensuring a stable energy supply for the plants. Adequate energy serves as the cornerstone for the growth of *P. lutea* and also contributes to its continuous blooming. With stable energy input, *P. lutea* can sustain the metabolic activities necessary for flower production throughout the year. Tropical regions are characterized by frequent fluctuations in temperature, humidity, and light intensity. The *rps15* and *rpl33* genes facilitate rapid environmental adaptation of *P. lutea* by promoting efficient protein synthesis. When these two genes function properly, *P. lutea* can synthesize proteins in a timely manner to cope with environmental changes. This significantly enhances the plant's tolerance to various environmental stresses, thereby providing support for its year-round growth and blooming. Overall, the genetic variations in the cp genome of *P. lutea*, especially the presence and functionality of the *psbA*, *rps15*, and *rpl33* genes, may represent crucial factors underlying the unique growth and blooming characteristics of *P. lutea*.

The family Acanthaceae has many species and genera, wide distribution, outstanding morphological diversity, including shrubs, herbs, and even vines, and outstanding habitat diversity. For a long time, the family was considered difficult to study. Since 1789, when French botanist Antoine-Laurent de Jussieu (1789) [43] proposed the natural classification of plants according to the relative positions of stamens and ovaries, the family Acanthaceae was published. Many scholars, such as Nees (1832) [44], Bentham (1876) [45], Lindau (1895) [46], and Bremekamp (1965) [47] have proposed different classification systems according to different classification characteristics. In recent years, the study of Acanthaceae evolutionary biology has benefited greatly from the utilization of cp genomes. Gao et al. (2019) [1] determined the cp genomes of four *Echinacanthus* species and resolved the phylogenetic relationship within Acanthaceae, which exhibited that *Echinacanthus* was sister to *Strobilanthes*. Similarly, Huang et al. (2020) [24] demonstrated that *Justicia*, *Clinacanthus*, and *Dicliptera* belong to one branch, whereas

Echinacanthus and *Strobilanthes* belong to another branch. Our findings were consistent with the evolutionary connection among the five aforementioned taxa. Furthermore, our study identified the position of *P. lutea* in Acanthaceae for the first time. That is, *P. lutea* was closely related to members of Justiciinae, but it was most closely related to *C. nutans* of Diclipterinae, and they clustered into a small branch to form a sister relationship with a support value of 100%. McDade et al. (2000) [48] classified *Pachystachys* and *Clinacanthus* into the same subtribe. In this study, it was found that there was a close relationship between *P. lutea*, Justiciinae and Diclipterinae. As a monospecific genus, *Pachystachys* was suggested to be included in the Justicieae, but whether it can be divided into Justiciinae or Diclipterinae can be further discussed in future studies.

Conclusions

As a newly emerged species resulting from natural selection during the evolutionary process of the Acanthaceae family, *P. lutea* boasts a genome structure that is highly conserved and bears a remarkable similarity to those of other related species. Phylogenetic analysis has revealed that *P. lutea* shares a close kinship with *C. nutans*. Moreover, it can be grouped together with species from *Justicia*, *Hypoestes*, *Peristrophe*, and *Dicliptera* within the Justicieae tribe. The cp genome data obtained in this research are of significant importance for the conservation as well as the rational development of the germplasm resources of *Pachystachys*, thus playing a crucial role in safeguarding the biodiversity and facilitating the sustainable utilization of these valuable genetic materials within the Acanthaceae family. Subsequent research will focus on validating the functions of the adaptive genes identified in the cp genome of *P. lutea*. These genomic data can directly serve conservation and breeding efforts. By leveraging the identified genetic variations, it is promising to cultivate new varieties with outstanding ornamental traits, strong stress resistance, and excellent ecological adaptability, thus promoting the diverse applications of plants in the Acanthaceae family.

Materials and Methods

Genomic DNA extraction and sequencing

Fresh leaves of *P. lutea* were collected from Xinyang, Henan Province, China (the experimental base of Xinyang Agriculture and Forestry University: 114° 12' E, 32° 16' N, altitude: 102 m) (Fig. 9). The leaves were frozen in liquid nitrogen before DNA extraction. Total genomic DNA was extracted from leaves using the CTAB method [49] and was sent to Shanghai Origingene Biotechnology Co., Ltd. for DNA library construction. The DNA library was constructed according to the instructions of the TruSeq RNA Sample Prep Kit, with an average length of 500 bp. Then, the samples were sequenced by using the Illumina NovaSeq 6000 sequencing platform (Illumina, San Diego, CA). A total of 8.9 GB raw data were produced and deposited in the SRA database (Sequence Read Archive, <http://www.ncbi.nlm.nih.gov/Traces/sra>).

Genome assembly and annotation

The quality of the raw paired-end reads was assessed using FastQC v0.11.7 [50] software. To ensure the accuracy of subsequent biological information analysis, the sequencing data for adapter sequences, low-quality reads, sequences with a high N rate, and sequences with insufficient lengths included in the raw data reads were filtered to obtain high-quality clean reads. After the quality evaluation, the reads were assembled by using both Fast-plast v.1.2.8 (<https://github.com/mrmckain/Fast-Plast>) [51] and GetOrganelle v.1.7.0+ [52]. Single contigs containing the complete cp genome were generated. PGA and Geseq were used for gene prediction and annotation of the *P. lutea* cp genome [53, 54], with default parameters and percent identity cut-off for protein-coding genes (PCGs) and RNAs set at ≥ 60 and ≤ 85 , respectively. All sample annotation results were manually corrected. BWA (v.0.7.17-r1188) was used to process comparison data, SAMtools (v.1.9) was used to calculate coverage depth, and then visualized using ggplot2 in R [55]. After the annotation was completed, the sequence of the cp genome of *P. lutea* was deposited in the GenBank database with accession number OP546128. CPGview (<http://www.1kmpg.cn/cpgview/>) was used to generate a circular cp genome map [56].

Simple sequence repeats analysis

MISA v1.0 software [57] was used for simple sequence repeats (SSRs) analysis in the cp genome of *P. lutea*, and the number of repetitions was set to 10, 5, 5, 5, 5, and 5 from mononucleotide to hexanucleotide, respectively. The maximum cardinality between any two SSRs was set to 100 bp. MISA was also used to determine the specific location of SSRs in the cp genome of *P. lutea*.

Codon preference analysis

The PCGs sequence of cp genome in *P. lutea* were manually screened to remove duplicates and coding sequences less than 300 bp in length, and the eligible sequences were used for subsequent analysis. Using CodonW V1.4.4 software [58] and the online software CUSP (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>) [59] calculated the effective number of codons (ENC), the GC content of codon 1, 2 and 3 bases (GC1, GC2, GC3, respectively) and relative synonymous codon usage (RSCU), and Neutral-plot, ENC-plot, and PR2-plot analysis were performed.

Sequence variation map and variations of inverted repeat regions (IRs) sequences

The comparison between the genome of *P. lutea* and seven Acanthaceae species' cp genomic sequences was performed using the mVISTA program (<http://genome.lbl.gov/vista/mvista/submit.shtml>) to find interspecific variation [60]; the annotation of *P. lutea* was used as a reference in the Shuffle-LAGAN mode. Furthermore, comparisons between the borders of the IR (IRa and IRb), the small single-copy (SSC), and the large single-copy (LSC) regions were generated using the IRscope online program (<https://irscope.shinyapps.io/irapp/>) [61]. The IR/SC borders of the *P. lutea* were validated using Sanger sequencing [62].

Non-synonymous substitutions (Ka)/Synonymous substitutions (Ks) analysis

The Ka/Ks substitution rates of the PCGs in *P. lutea* cp genome were compared with seven related species from Acanthaceae. Sequence alignment was conducted using Mega 7.0 [63], while Ka/Ks calculations were performed using DnaSP v.6.12 [64].

Phylogenetic analysis

In this study, to determine the phylogenetic position of *P. lutea*, the complete cp genomes of 33 species were downloaded from NCBI. Among these, 31 from four subfamilies within the Acanthaceae family, and two (*Catalpa bungei* and *Catalpa fargesii*) from outside this family, serving as outgroup species. The Multiple sequence alignment of the 34 species including *P. lutea* was performed by using MAFFT v7.158b [65] software and Gblock software [66] to find the conserved sequences between different species. The best substitution model, GTR + G, was selected in the jModelTest v2.1.7 program [67]. The phylogenetic tree was constructed with the maximum-likelihood (ML) method using RAxML v8.2.12 software [68]. Bootstrap analysis was used to evaluate the support for individual clades with 1000 replicates.

Abbreviations

cp chloroplast

LSC large single-copy

SSC small single-copy

IRs inverted repeat regions

PCGs protein-coding genes

SSRs simple sequence repeats

ML maximum-likelihood method

ITS internal transcribed spacers

Enc effective number of codons

GC1, GC2, GC3 GC content of codon 1, 2 and 3 bases

RSCU relative synonymous codon usage

Ka Non-synonymous substitutions

Ks Synonymous substitutions

Leu leucine

Cys cysteine

Met methionine

Trp tryptophan

JLB IRb/LSC

JSB IRb/SSC

JSA IRa/SSC

JLA IRa/LSC

SNPs single nucleotide polymorphisms

BS bootstrap support

Declarations

Statement

Our experimental research and field studies on plants comply with relevant institutional, national, and international guidelines and legislation.

Author Contributions: C.D. designed and executed experiments, completed data analysis, and wrote the first draft of the paper. Y.D. and H.G. contributed to the experimental design and analysis. T.F. assisted in sample collection and species identification. J.Y. and Y.Z. were the project developer and leader, guiding the experimental design, data analysis, and paper writing and revision. The final text has been read and approved by all authors.

Funding: This work was supported by the Program for Innovative Research Team of Horticultural Plant Resources and Utilization in Xinyang Agriculture and Forestry University (XNKJTD-012), the Young Key Teachers Training Program in Xinyang Agriculture and Forestry University (2021), the Academic Core Teachers Program in Xinyang Agriculture and Forestry University (2022), and the Scientific Research Foundation for Young Teachers of Xinyang Agriculture and Forestry University (QN2023023).

Availability of data and materials

The original sequencing data have been submitted to the NCBI database and received GenBank accession number OP546128. The data used in this study are available in the public domain (<https://www.ncbi.nlm.nih.gov>). And the associated Bio-project, SRA, Bio-sample numbers are PRJNA884896, SRX17730640, and SAMN31059961, respectively.

Ethics approval and consent to participate

Pachystachys lutea were collected in July 2022 from Xinyang, Henan Province, China (the experimental base of Xinyang Agriculture and Forestry University: 114° 12' E, 32° 16' N, altitude: 102 m). The plant specimen was deposited at the Herbarium of the Horticultural Plant Biotechnology Laboratory, Xinyang Agriculture and Forestry University under voucher code PL20220716.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Gao CM, Deng YF, Wang, J. The complete chloroplast genomes of *Echinacanthus* species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. *Front. Plant Sci.* 2019; 9:1989. <https://doi.org/10.3389/fpls.2018.01989>
2. Manzitto-Tripp EA, Darbyshire I, Daniel TF, Kiel CA, McDade LA. Revised classification of Acanthaceae and worldwide dichotomous keys. *TAXON.* 2022; 71(1):103–153. <https://doi.org/10.1002/tax.12600>
3. Kiel CA, Daniel TF, McDade LA. Phylogenetics of new world ‘justicioids’ (Justicieae: Acanthaceae): major lineages, morphological patterns, and widespread Incongruence with classification. *Syst Bot.* 2018; 43(2):459–484. <https://doi.org/10.1600/036364418X697201>
4. McDade LA, Daniel TF, Kiel CA. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). *Am J Bot.* 2008; 95(9):1136–1152. <https://doi.org/10.3732/ajb.0800096>
5. Grant WF. A cytogenetic study in the Acanthaceae. *Brittonia.* 1955; 8(2):121–149. <https://doi.org/10.2307/2804856>
6. Das RJ, Pathak K. Use of indigenous plants in traditional health care systems by Mishing tribe of Dikhowmukh, Sivasagar district, Assam. *Int J Herb Med.* 2013; 1(3):50–57.
7. Kartika T. Pemanfaatan tanaman hias pekarangan berkehasiat obat di Kecamatan Tanjung Batu. *Sainmatika.* 2018; 15(1):48–55.
8. da Silva Ribeiro A, Polonio JC, Costa AT, dos Santos CM, Rhoden SA, Azevedo JL, Pamphile JA. Bioprospection of culturable endophytic fungi associated with the ornamental plant *Pachystachys lutea*. *Curr Microbiol.* 2018; 75:588–596. <https://doi.org/10.1007/s00284-017-1421-9>
9. da Silva Ribeiro A, Domingos Polli A, Oliveira A, dos Santos de Oliveira JA, Emmer A, Hamamura Alves L, Neves Pereira OC, Pamphile JA. Ornamental plant *Pachystachys lutea* as a source of promising endophytes for plant growth and phytoprotective activity. *Acta Scientiarum. Biological Sciences.* 2021; 43(1):e51737. <https://doi.org/10.4025/actascibiolsci.v43i1.51737>

10. Rosdewi M, Sada M, Fitriah F. Inventory and identification of natural dyes of ikat woven fabrics at Sanggar Bliran Sina Watublapi. *Jurnal Riset Ilmu Pendidikan*. 2023; 3:6–19.
11. Wicaksono A, Raihandhany R. Bioprospeksi bunga lolipop (*Pachystachys lutea* Nees) sebagai sumber alternatif untuk produksi gula: Bioprospecting of the lollipop plant (*Pachystachys lutea* Nees) as alternative source for sugar production. *Jurnal Sumberdaya Hayati*. 2023; 9(2):63–69. <https://doi.org/10.29244/jsdh.9.2.63-69>
12. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016; 17:134. <https://doi.org/10.1186/s13059-016-1004-2>
13. Li YT, Zhang J, Li LF, Gao LJ, Xu JT, Yang MS. Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis*- “wild plants with a tiny population”-and three other *Pyrus* species. *Int J Mol Sci*. 2018; 19(10):3262. <https://doi.org/10.3390/ijms19103262>
14. Raman G, Park S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbis* var. *longicalycinus*, from a comparative genomics perspective. *PLoS ONE*. 2015; 10(10):e0141329. <https://doi.org/10.1371/journal.pone.0141329>
15. Scotland RW, Vollesen K. Classification of Acanthaceae. *Kew Bul*. 2000; 55(3):513–589. <https://doi.org/10.2307/4118776>
16. Wu WW, Li JL, Liu Y, Jiang M, Lan MS, Liu C. Peculiarities of the inverted repeats in the complete chloroplast genome of *Strobilanthes bantonensis* Lindau. *Mitochondrial DNA B*. 2021; 6(4):1440–1447. <https://doi.org/10.1080/23802359.2021.1911699>
17. Kirankumar SI, Balaji R, Tanuja, Parani M. The complete chloroplast genome of *Ocimum basilicum* L. var. *basilicum* (Lamiaceae) and its phylogenetic analysis. *Mitochondrial DNA B*, 2023; 8(11):1169–1173. <https://doi.org/10.1080/23802359.2023.2275835>
18. Sueoka N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses. *J Mol Evol*. 2001; 53:469–476. <https://doi.org/10.1007/s002390010237>
19. Gao Y, Lu Y, Song Y, Jing L. Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus*. *BMC Genom Data*. 2022; 23:46. <https://doi.org/10.1186/s12863-022-01064-8>
20. He B, Dong H, Jiang C, Gao FL, Tao ST, Xu LA. Analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending. *Sci Rep*. 2016; 6:35927. <https://doi.org/10.1038/srep35927>
21. Chen HM, Shao JJ, Zhang H, Jiang M, Huang LF, Zhang Z, Yang D, He M, Ronaghi M, Luo X, Sun BT, Wu WW, Liu C. Sequencing and analysis of *Strobilanthes cusia* (Nees) Kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. *Front. Plant Sci*. 2018; 9:324. <https://doi.org/10.3389/fpls.2018.00324>
22. Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom Proteom Bioinf*. 2006; 4(4):259–263. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2)
23. Li MN, Pei HQ, Wang JH, Zhao KK, Zhu ZX, Wang HF. Complete plastome sequence of *Clinacanthus nutans* (Acanthaceae): a medicinal species in Southern China. *Mitochondrial DNA B*. 2019;

- 4(1):118–119. <https://doi.org/10.1080/23802359.2018.1532349>
24. Huang SN, Ge XJ, Cano A, Salazar BGM, Deng YF. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. *PeerJ*. 2020; 8:e8450. <https://doi.org/10.7717/peerj.8450>
25. Yaradua SS, Alzahrani DA, Albokhary EJ, Abba A, Bello A. Corrigendum to “Complete chloroplast genome sequence of *Justicia flava*: genome comparative analysis and phylogenetic relationships among Acanthaceae”. *BioMed Res Int*. 2020; 2020:2049759. <https://doi.org/10.1155/2020/2049759>
26. Chen JX, Wang L, Zhao YC, Qin MJ. The complete chloroplast genome of a Chinese medicinal plant, *Peristrophe japonica* (Thunb.) Bremek. (Lamiales: Acanthaceae) from Nanjing, China. *Mitochondrial DNA B*. 2021; 6(7):1888–1889. <https://doi.org/10.1080/23802359.2021.1934157>
27. Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, Yan A, Mueller LA. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res*. 2015; 43:D1036–D1041. <https://doi.org/10.1093/nar/gku1195>
28. Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *P Natl Acad Sci U S A*. 1995; 92(17):7759–7763. <https://doi.org/10.1073/pnas.92.17.7759>
29. Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N'Goran JAK, Lanaud C. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theor Appl Genet*. 2004; 108:1151–1161. <https://doi.org/10.1007/s00122-003-1533-4>
30. Li B, Lin FR, Huang P, Guo WY, Zheng YQ. Complete chloroplast genome sequence of *Decaïsnea insignis*: genome organization, genomic resources and comparative analysis. *Sci Rep*. 2017; 7:10073. <https://doi.org/10.1038/s41598-017-10409-8>
31. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*. 1981; 151(3): 389–409. [https://doi.org/10.1016/0022-2836\(81\)90003-6](https://doi.org/10.1016/0022-2836(81)90003-6)
32. Chen ZY, Liu Q, Xiao Y, Zhou GH, Yu PH, Bai J, Huang H, Gong YH. Correction to: Complete chloroplast genome sequence of *Camellia sinensis*: genome structure, adaptive evolution, and phylogenetic relationships. *J Appl Genet*. 2023; 64:601. <https://doi.org/10.1007/s13353-023-00769-5>
33. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep*. 2022; 49:539–565. <https://doi.org/10.1007/s11033-021-06749-4>
34. Fu JM, Liu HM, Hu JJ, Liang YQ, Liang JJ, Wuyun TN, Tan XF. Five complete chloroplast genome sequences from *Diospyros*: genome organization and comparative analysis. *PLoS One*. 2016; 11(7):e0159566. <https://doi.org/10.1371/journal.pone.0159566>

35. Wang WC, Chen SY, Zhang XZ. Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (Bambusoideae: Olyreae). *Molecules*. 2018; 23(7):1537. <https://doi.org/10.3390/molecules23071537>
36. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet*. 2003; 4:213–235. <https://doi.org/10.1146/annurev.genom.4.020303.162528>
37. Makiłowski W, Boguski MS. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *P Natl Acad Sci U S A*. 1998; 95(16):9407–9412. <https://doi.org/10.1073/pnas.95.16.9407>
38. Bobik K, Burch-Smith TM. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci*. 2015; 6:781. <https://doi.org/10.3389/fpls.2015.00781>
39. Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiol Plant*. 2020; 42:98. <https://doi.org/10.1007/s11738-020-03089-x>
40. Pfannschmidt T, Nilsson A, Tullberg A, Link G, Allen JF. Direct transcriptional control of the chloroplast genes *psbA* and *psaAB* adjusts photosynthesis to light energy distribution in plants. *IUBMB Life*. 1999; 48:271–276. <https://doi.org/10.1080/713803507>
41. Rogalski M, Schöttler MA, Thiele W, Schulze WX, Bock R. Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell*. 2008; 20(8): 2221–2237. <https://doi.org/10.1105/tpc.108.060392>
42. Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schöttler MA, Bock R. Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell*. 2011; 23(9):3137–3155. <https://doi.org/10.1105/tpc.111.088906>
43. Jussieu AL de. *Genera plantarum secundum ordines naturales disposita*. 2nd ed, Apud viduam Herissant et Theophilum Barrois: Paris, France, 1789; 307.
44. Nees Von Esenbeck CG. Acanthaceae. In *Plantae Asiaticae Rariores*; Wallich, N. Eds, Treuttel, Würtz, Richter, London. 1832; 3:77–112.
45. Bentham G. Acanthaceae. In *Genera Plantarum*; Bentham, G., Hooker, J.D. Eds, Reeve & Co.: London, 1876; v.2: pt.2:1066–1122.
46. Lindau Q. Acanthaceae. In: *Die Natürlichen Pflanzenfamilien*; Engler, A., Prantl, K, Verlag von Wilhelm Englemann: Leipzig, 1895; IV(3b):274–354.
47. Bremekamp CEB. Delimitation and subdivision of the Acanthaceae. *Bulletin of the Botanical Survey of India*. 1965; 7:21–30.
48. McDade LA, Masta SE, Moody ML, Waters E. Phylogenetic relationships among Acanthaceae: evidence from two genomes. *Syst Bot*. 2000; 25(1):106–121. <https://doi.org/10.2307/2666677>
49. Doyle J. DNA protocols for plants. In *Molecular Techniques in Taxonomy*, Nato ASI Subseries H: Cell Biology; 1st ed, Hewitt, G.M., Johnston, A.W.B., Young, J.P.W. Eds, Springer: Berlin, Heidelberg, Germany, 1991; 57:283–293. https://doi.org/10.1007/978-3-642-83962-7_18

50. Wingett SW, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]. F1000Research. 2018; 7:1338.
<https://doi.org/10.12688/f1000research.15931.2>
51. McKain MR, Wilson M. Fast-Plast: rapid de novo assembly and finishing for whole chloroplast genomes. <https://github.com/mrmckain/Fast-Plast> (2017)
52. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020; 21:241.
<https://doi.org/10.1186/s13059-020-02154-5>
53. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019; 15:50. <https://doi.org/10.1186/s13007-019-0435-7>
54. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017; 45(W1):W6–W11.
<https://doi.org/10.1093/nar/gkx391>
55. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
56. Liu SY, Ni Y, Li JL, Zhang XY, Yang HY, Chen HM, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 2023; 23(3):694–704.
<https://doi.org/10.1111/1755-0998.13729>
57. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003; 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>
58. Sun XY, Yang Q, Xia XH. An improved implementation of Effective Number of Codons (*N_c*). Mol Biol Evol. 2013; 30(1):191–196. <https://doi.org/10.1093/molbev/mss201>
59. Wang YH, Wei QY, Xue TY, He SX, Fang J, Zeng CL. Comparative and phylogenetic analysis of the complete chloroplast genomes of 10 *Artemisia selengensis* resources based on high-throughput sequencing. BMC Genomics. 2024; 25:561. <https://doi.org/10.1186/s12864-024-10455-3>
60. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000; 16(11):1046–1047. <https://doi.org/10.1093/bioinformatics/16.11.1046>
61. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018; 34(17):3030–3031.
<https://doi.org/10.1093/bioinformatics/bty220>
62. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. P Natl Acad Sci U S A. 1977; 74(12):5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
63. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33(7):1870–1874. <https://doi.org/10.1093/molbev/msw054>
64. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;

- 34(12):3299–3302. <https://doi.org/10.1093/molbev/msx248>
65. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019; 20(4):1160–1166. <https://doi.org/10.1093/bib/bbx108>
66. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000; 17(4):540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
67. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9:772. <https://doi.org/10.1038/nmeth.2109>
68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30(9):1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>

Figures

Pachystachys lutea

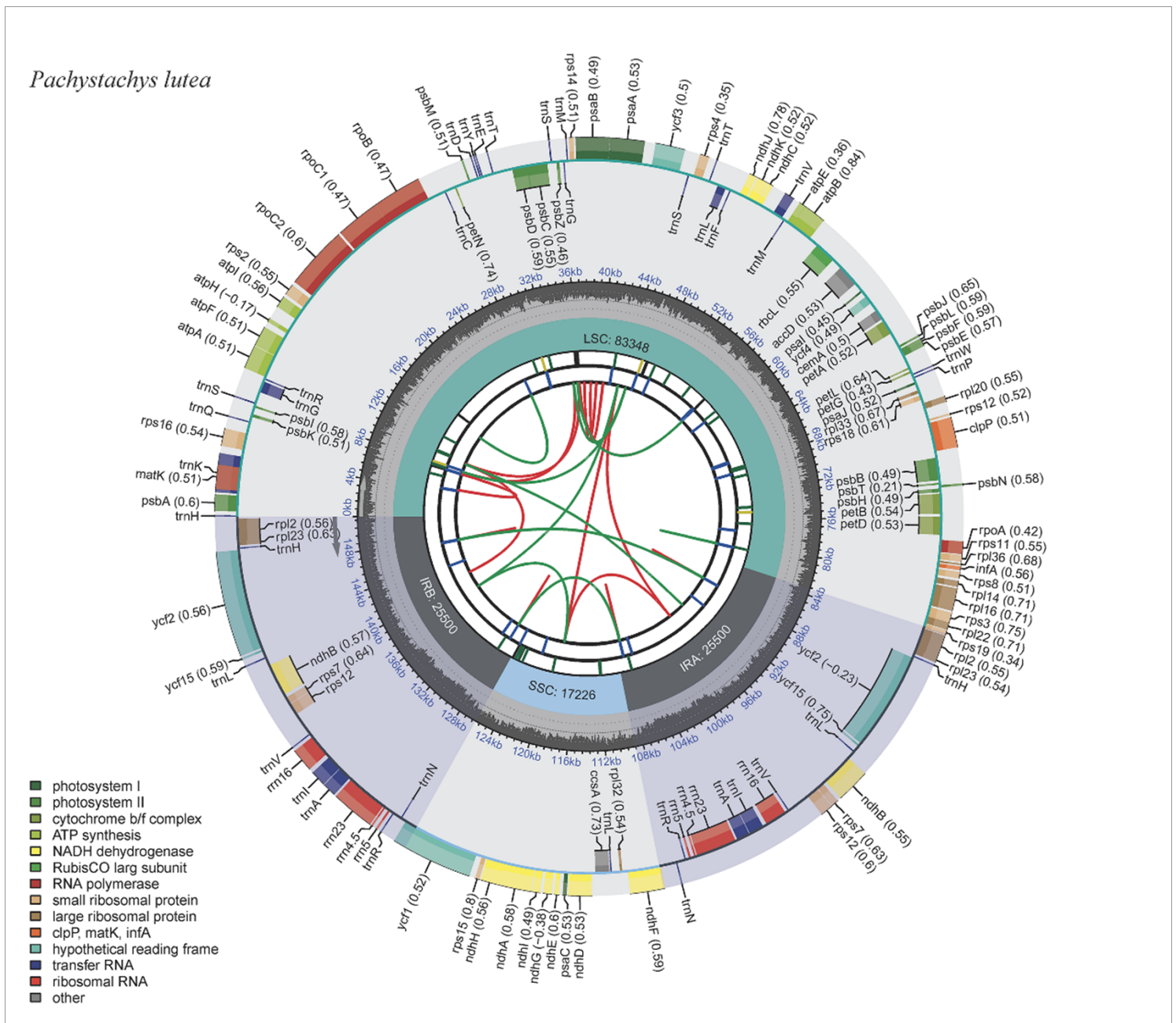


Figure 1

The chloroplast genome map of *P. lutea*. Genes outside the circle are transcribed clockwise, while genes inside the circle are transcribed counterclockwise. Different functional groups are represented by different colors. The darker and lighter gray in the inner indicated the GC and AT content, respectively.

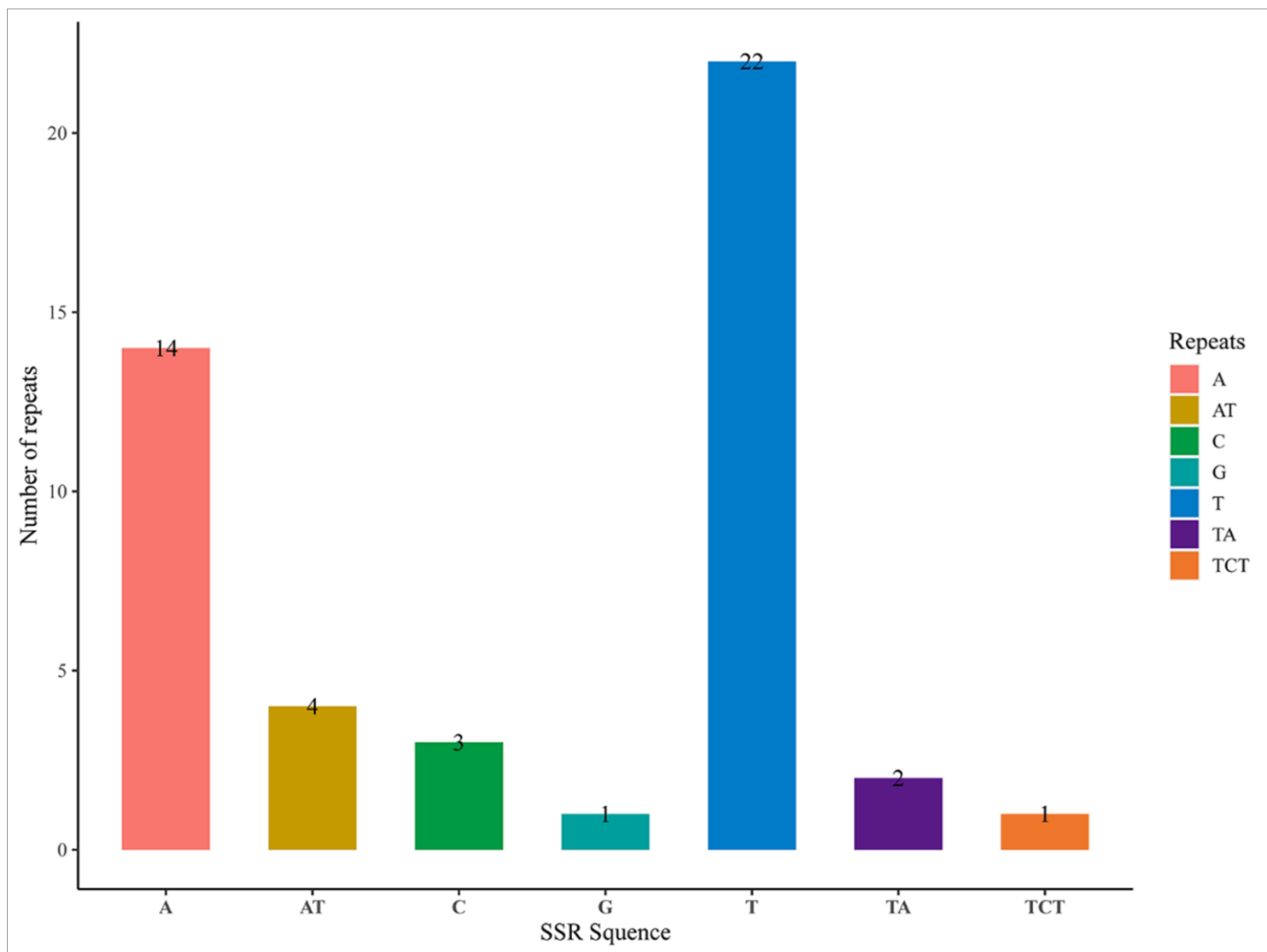


Figure 2

SSRs analysis of chloroplast genome in *P. lutea*. The x-axis represents the types of repetitive sequences in base pairs, while the y-axis indicates the number of SSRs found within each category.

(G3 + C3): the probability of G when the third base of codon is G or C. A3/ (A3 + T3): the probability of A when the third base of the codon is A or T.

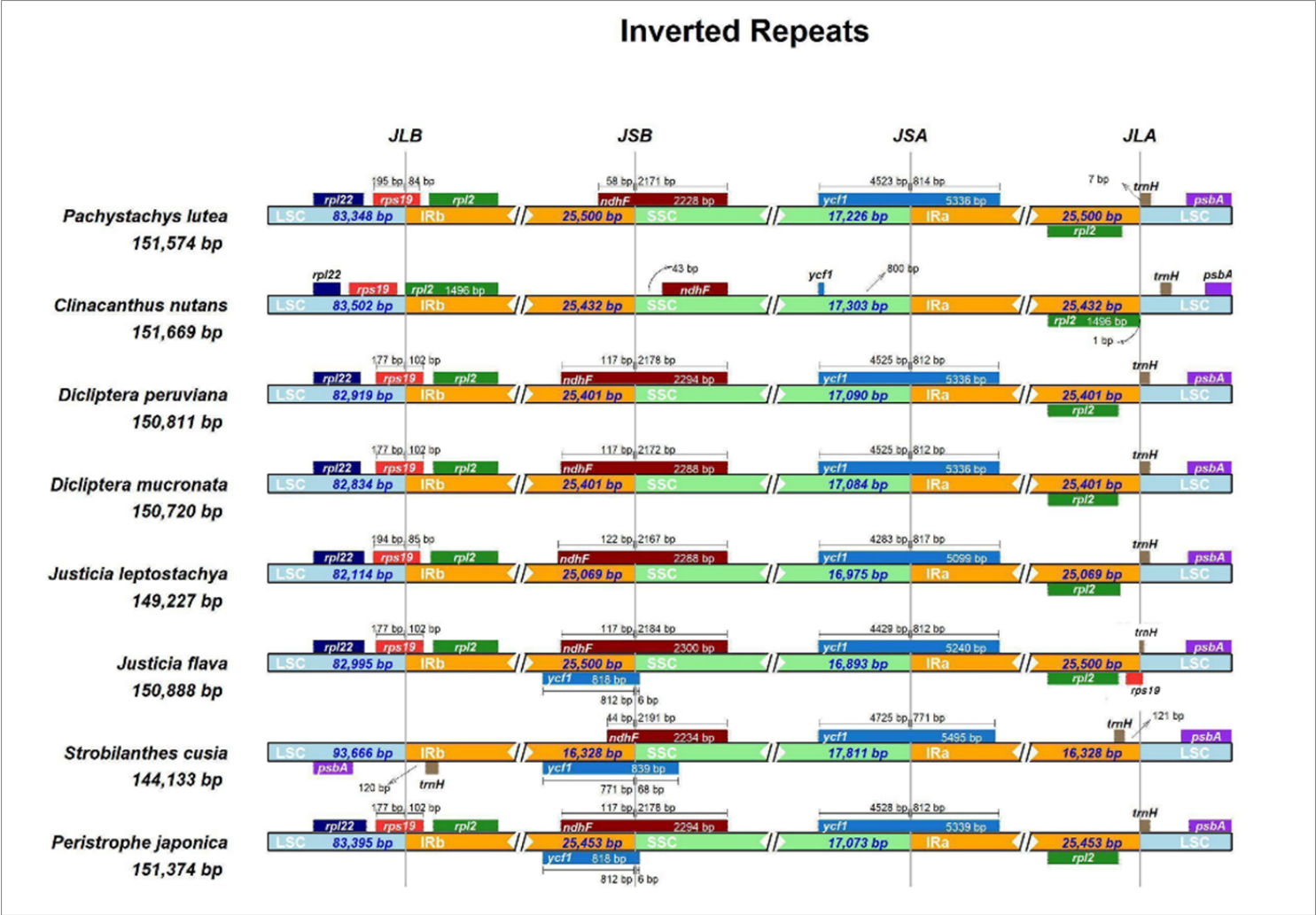


Figure 5

Comparison of the boundaries of LSC, IR and SSC among chloroplast genome of eight Acanthaceae species. The genes around the borders are shown above or below the main line. The JLB, JSB, JSA, and JLA represent junction sites of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC, respectively.

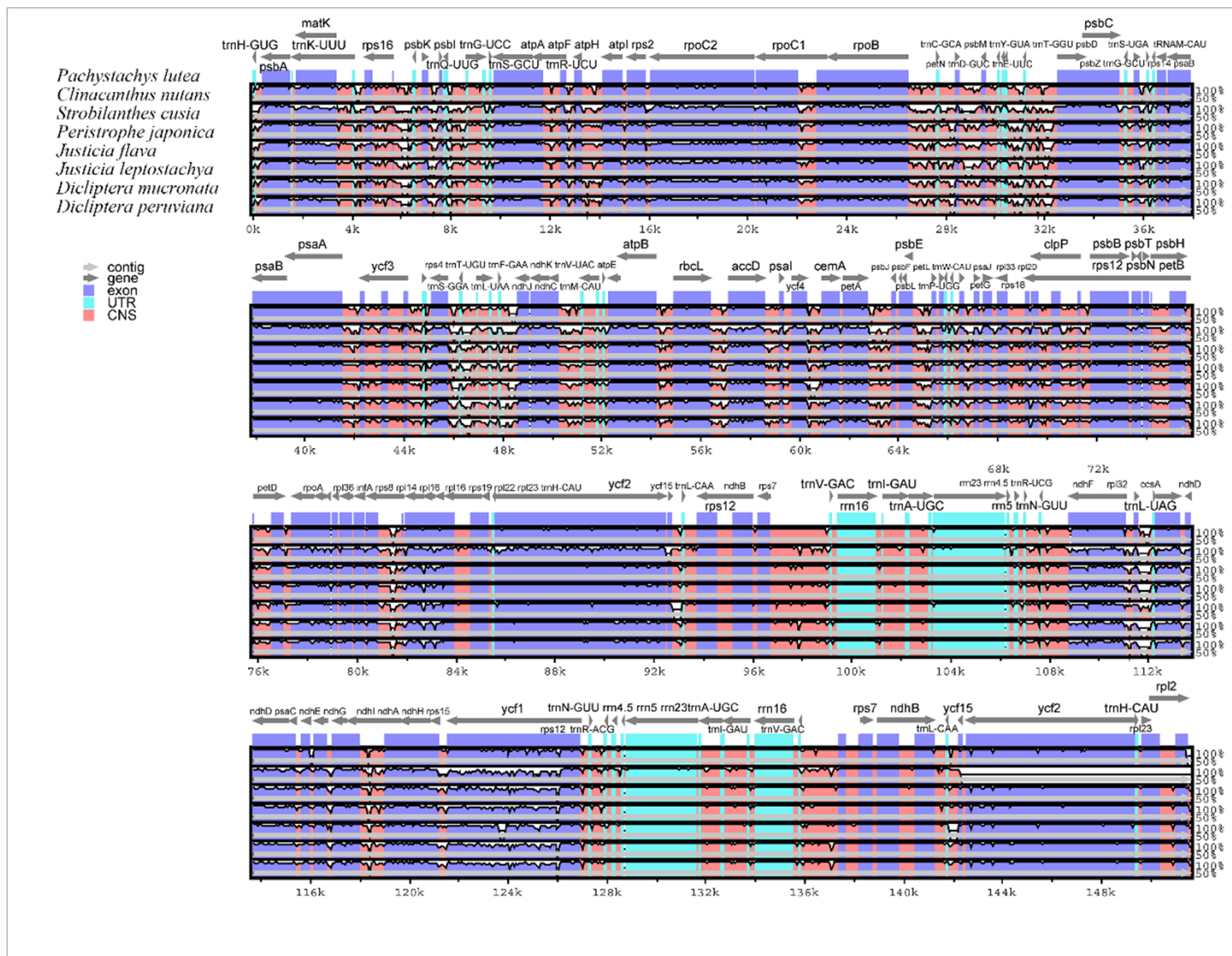


Figure 6

Global alignment analysis of chloroplast genome in eight Acanthaceae species. Annotated genes are displayed along the top. The horizontal axis indicates the coordinates in the chloroplast genome and the vertical axis indicates the percentage identity (between 50 and 100%). Genome regions are color-coded as exon, intron, and conserved non-coding sequences (CDS).

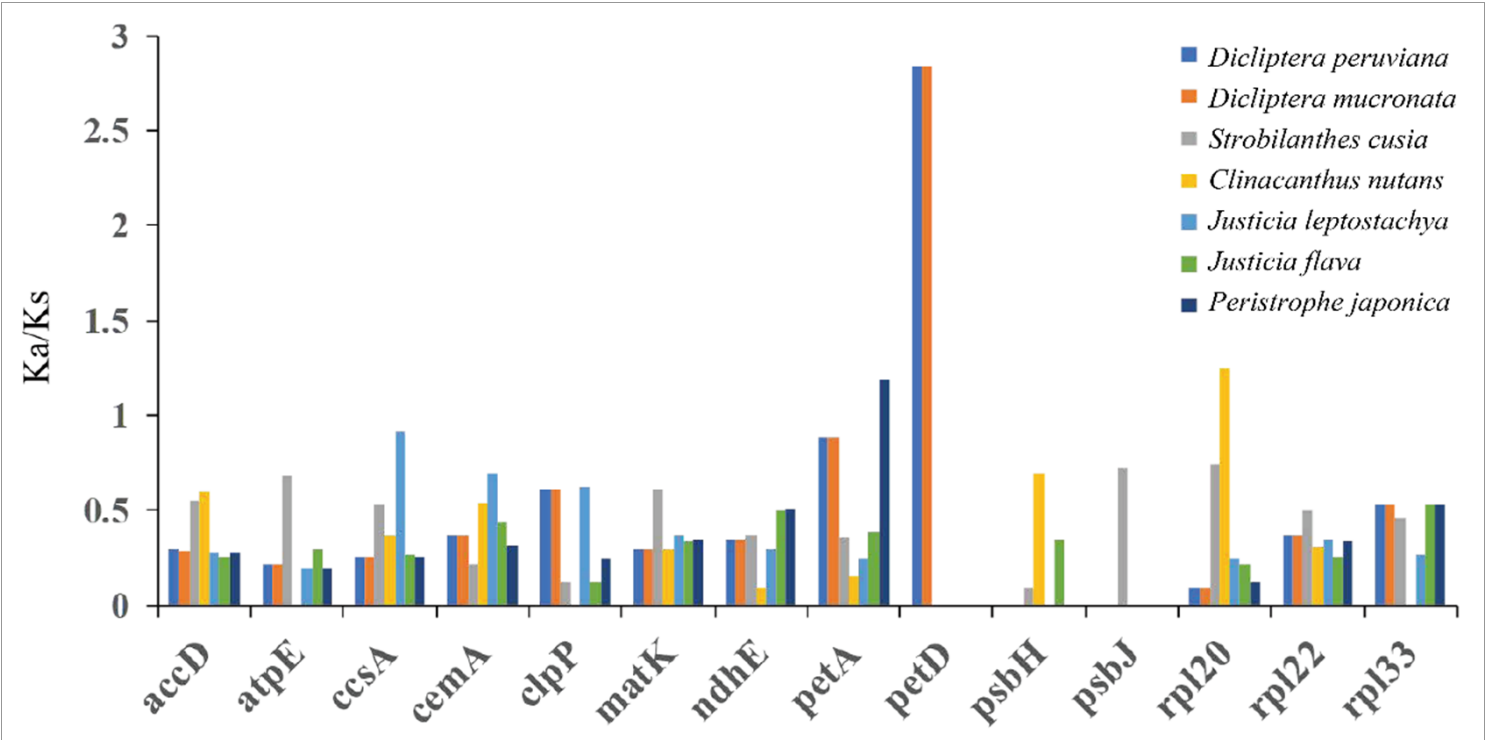


Figure 7

Ka/Ks values of selected PCGs of *P. lutea* with close species. The x-axis represents genes, and the order is alphabetical, while the y-axis represents the Ka/Ks value of each gene.

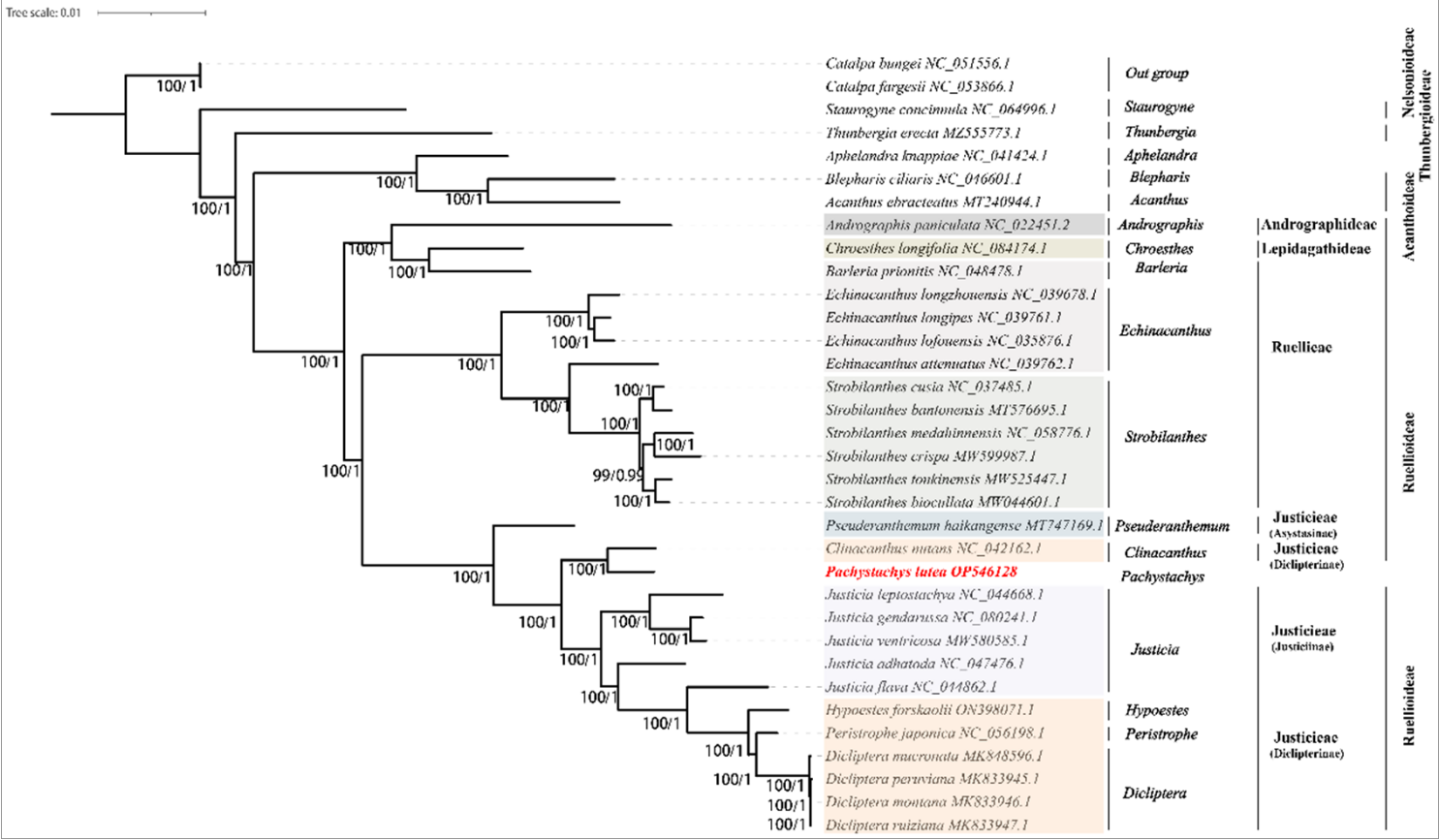


Figure 8

The phylogenetic tree of *P. lutea* was constructed by ML method based on the analysis of conserved PCGs of 34 species.

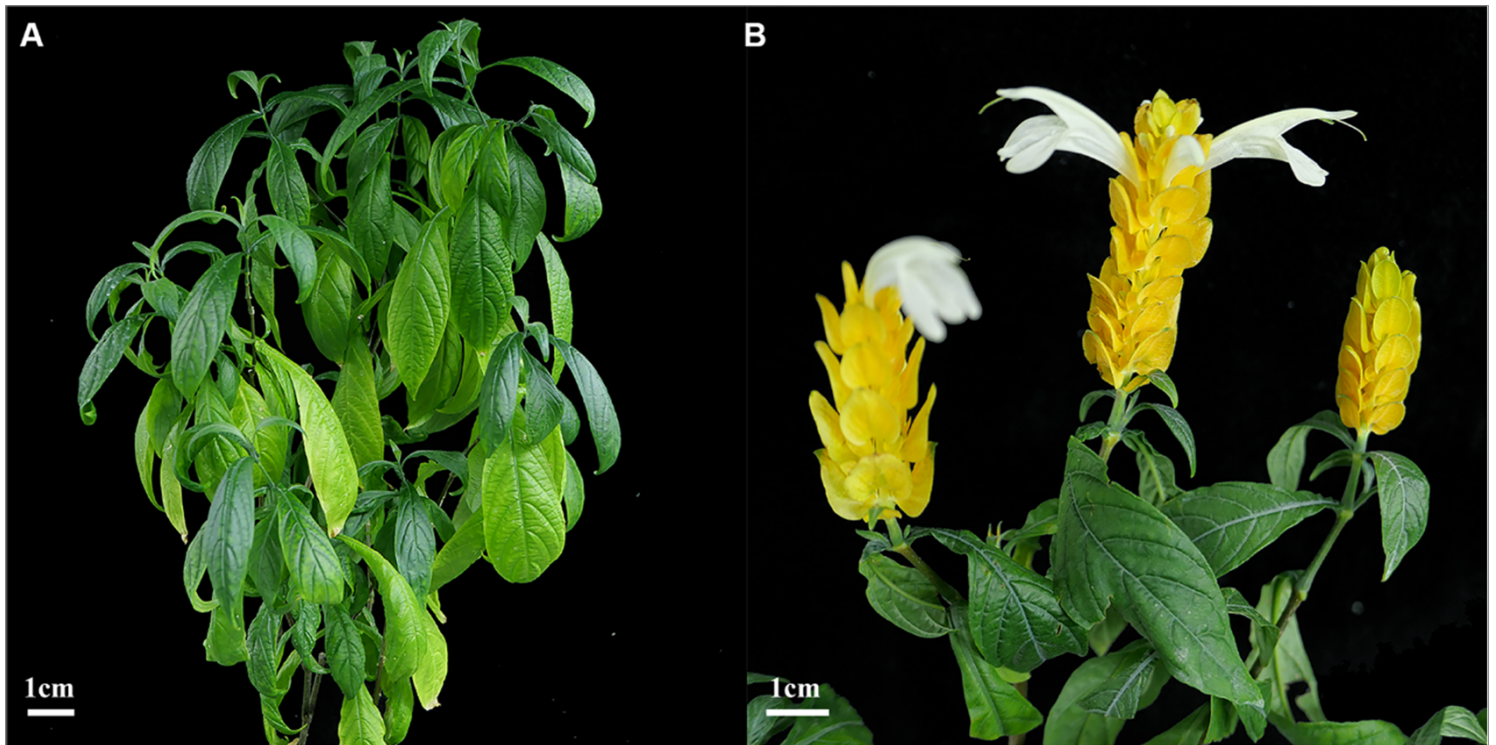


Figure 9

Plant morphological characteristics (A) and the floral organ morphology (B) of *P. lutea*.

Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [FigureS6.tif](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [SupplementaryInformationLegends.docx](#)